Please cite the Published Version

AN INTEGRATED COMPUTATIONAL AND EXPERIMENTAL APPROACH TO DESIGNING NOVEL NANODEVICES

Manchester Metropolitan University

N. A. SIMBANEGAVI

PhD 2014

AN INTEGRATED COMPUTATIONAL AND EXPERIMENTAL APPROACH TO DESIGNING NOVEL NANODEVICES

Nyevero Abigail Simbanegavi

A thesis submitted in partial fulfilment of the requirements of Manchester Metropolitan University for the degree of Doctor of Philosophy

2014

School of Science and the Environment

Division of Chemistry and Environmental Science

Manchester Metropolitan University

Abstract

Small molecules that can organise themselves through reversible bonds to form new molecular structures have great potential to form self-assembling systems. In this study, C_{60} has been functionalized with components capable of self-assembling *via* hydrogen bonds in order to form supramolecular structures. Functionalizing the cage not only enhances the solubility of C_{60} but also alters its electronic properties. In order to analyse the changes in electronic properties of C_{60} and those of the new derivatives and self-assembled structures, an integrated computational and experimental approach has been used.

DNA bases were among the components chosen for functionalizing C_{60} since these structures are the best example for self-assembly in nature. Experimental routes for novel C_{60} derivatives functionalised with thymine, adenine, cytosine and guanine have been explored using the Prato reaction. A number of challenges in synthesizing the aldehyde analogues of the DNA bases have been identified, and a number of different solutions have been tested. Samples of N-methyl-2-thymin-N-ylfulleropyrrolidine have been isolated and N-methyl-2-adenin-N-ylfulleropyrrolidine has also been characterised.

Zeolites, because of their ordered cages and channels, were selected as a template for the self-assembly of the fullerene derivatives to enable the formation of ordered arrays. The adsorption of the thymine-fullerene derivative (N-methyl-2-thymin-N-ylfulleropyrrolidine) into the openings of zeolite channels has been confirmed by powder X-ray diffraction and thermogravimetric analysis.

Two additional hydrogen-bonding fullerene derivatives have been modelled: N-propylthymine and diaminopyridine. Furthermore, methanofullerene derivatives functionalised with DNA bases have also been modelled. Monomer and dimer structures have been optimised using semi-empirical methods and density functional theory (DFT) using B3LYP/6-31G**. The HOMO, LUMO and HOMO-LUMO gap energies as well as the frontier HOMO and LUMO electron density distribution of the structures have been calculated to get an insight into the reactivity of the C_{60} derivatives. Results for the HOMO,

LUMO and band gap energies in the gas phase are comparable with previous studies. Solvent effects on the electronic properties of the monomers have been calculated. It is shown that the electronic properties are altered in the solvent environments, which may be important when choosing solvents to prepare devices. The dimers have been modelled in seesaw and dumb-bell conformations and the formation energies have been calculated to predict the preferred growth pattern of the molecular architectures.

In summary, it can be seen that the computational calculations can help experimentalists identify molecules with interesting properties and the challenges encountered developing effective synthesis pathways emphasise the need for even more integration of approaches to designing novel fullerene derivatives with tailored properties.

Contents

Section A
SYNTHETIC METHODOLOGY

Section B
COMPUTATIONAL

List of Figures

Chapter   8

Chapter   9

Chapter   10

List of Tables

Table 10.2   Figures showing the orientation of the methanofullerene ligands in the HIV-1

PR active site with calculated docking scores.

Glossary

A.U	Atomic units
Acc. Mass	Accurate Mass
AM1	Austin Model version 1
B3LYP	Becke-3 Lee Yang Par
BHJ	Bulk Heterojunction
BLY	Becke Lee Yang
BSSE	Basis-Set Superposition Error
COSMO	Conductor-like Screening Model
DAM-[60]	1,1-diadenine-1,9-methanofullerene
DCM-[60]	1,1-dicytosine-1,9-methanofullerene
DDM-[60]	1,1-di-diaminopyridine-1,9-methanofullerene
DEPT 135	Distortionless Enhancement by Polarization Transfer
DFT	Density Functional Theory
DGM-[60]	1,1-diguanine-1,9-methanofullerene
DNA	Deoxyribonucleic Acid
DNM-[60]	1,1-diN-propylthymine-1,9-methanofullerene
DTM-[60]	1,1-dithymine-1,9-methanofullerene
eV	Electron Volts
FT-IR	Fourier Transform Infrared Spectroscopy
HIV	Human Immunodeficiency Virus
HOMO	Highest Occupied Molecular Orbital
HRES	High Resolution
LCAO	Linear combination of Atomic Orbitals
LC-MS	Liquid Chromatography Mass Spectrometry
LUMO	Lowest Unoccupied Molecular Orbital
MM2	Molecular Mechanics 2

MM3	Molecular Mechanics 3
NMA-[60]	N-Methyl-2-adenin-N-ylfulleropyrrolidine
NMC-[60]	N-Methyl-2-cytosinefulleropyrrolidine
NMD-[60]	N-Methyl-2-diaminopyridinefulleropyrrolidine
NMG-[60]	N-Methyl-2-guaninefulleropyridine
NMN-[60]	N-Methyl-2-Npropylthyminefulleropyrrolidine
NMT-[60]	N-Methyl-2-thymin-N-ylfulleropyrrolidine
NMU-[60]	N-Methyl-2-uracilfulleropyrrolidine
NMR	Nuclear Magnetic Resonance
P3HT	Poly-(3-hexylthiophene)
$PC_{60}BM$	Phenyl-C_{61} Butyric Acid Methyl Ester
$PC_{70}BM$	Phenyl-C_{70} Butyric Acid Methyl Ester
PCE	Power conversion efficiency
PET	Photoinduced electron transfer
PM3	Parameterized Model Number 3
PMF	Potential of mean force
PPM	Parts per Million
PSC	Polymer Solar Cells
PW91	Perdew Wang exchange-correlation functional
RNA	Ribonucleic Acid
SEM	Scanning Electron Microscope
TEM	Transmission Electron Microscope
TGA	Thermogravimetric Analysis
TLC	Thin Layer Chromatography
UV-Vis	Ultra Violet Visible Spectroscopy
XRD	X-ray Diffraction

Acknowledgements

Firstly, I would like to thank my Director of Studies, Dr. Lindsey J. Munro for giving me this opportunity and for her expertise, patience and advice during the course of my studies. I also want to thank my second supervisor, Dr. Paul Birkett, for his expertise in the lab and motivation when needed, which have enable me to reach the end of my research.

I would also like to extend my gratitude to the NIB and Manchester Metropolitan University for providing me with the necessary funding to carry out my research.

Finally, I want to thank my parents, Piniel Simbanegavi and Shelyta Simbanegavi for their unfailing love and support throughout the course of my studies. It is to them that I dedicate this thesis.

Chapter 1
Introduction

1.1 C_{60}

The discovery[1] and isolation[2] of C_{60} (**1**) led to a lot of research activity, with researchers attempting to find the best application for this new type of carbon compound. Its truncated icosahedron shape and high symmetry make it an interesting compound to study. C_{60} comprises of 12 pentagons and 20 hexagons, with all 60 carbon atoms chemically equivalent (**Figure 1.1**). X-ray diffraction revealed C_{60} as a polyenic structure[3] consisting of two types of bonds: 6,6 junctions, which are double bonds and join two hexagons, and 5,6 junctions which are single bonds and join a hexagon and a pentagon.[4]
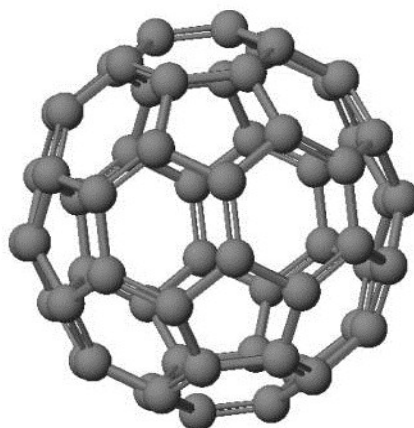
1

Figure 1.1 The structure of C_{60}.

The reactivity of C_{60} is typical of an electron-deficient olefin. It has low-lying triply-degenerate lowest unoccupied molecular orbitals (LUMO) and five-fold degenerate highest occupied molecular orbitals (HOMO).[5]

The low-lying LUMO orbitals allow C_{60} to accept up to six electrons in a reversible process, thus it readily reacts with nucleophiles.[6] In addition, C_{60} has enhanced reactivity which is

linked to its curvature, which leads to geometrical constraints on the planarity of the double bonds.[4] The main driving force for the reactivity of fullerenes is therefore the relief of the strain and the concomitant change in hybridisation from sp^2 to sp^3 on addition of a functional group to the C_{60}.[7]

In spite of its reactivity, C_{60} is only sparingly soluble in most organic solvents[6] and functionalization of the fullerene is considered the best way of increasing its solubility.[4]

1.1.1 Functionalization of C_{60}

In the quest to produce stable, characterizable C_{60} derivatives, several approaches have been developed. They include a range of now well-established cycloaddition reactions, including [2+2][8], [3+2][6] and [4+2][9] reactions which produce good yields of cyclobutanofullerenes (**2**), pyrrolidines (**3**) and cyclohexenofullerenes (**4**) all bonded to 6,6 junctions respectively.

2   **3**   **4**
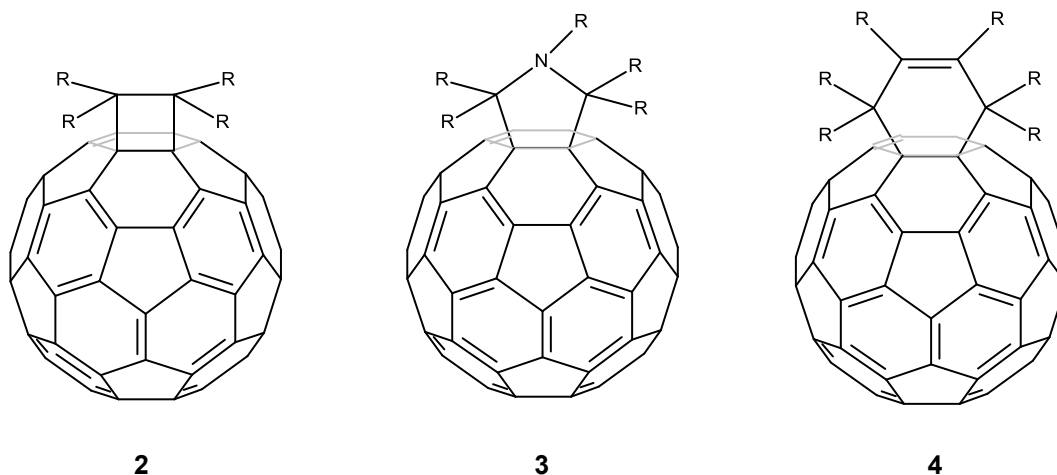
Figure 1.2 Products of [2+2] (cyclobutanofullerenes) (**2**), [3+2] (pyrrolidine) (**3**) and [4+2] (6-membered ring) (**4**) reactions respectively.

Other reactions such as the Bingel-Hirsch reaction,[10] which involves the nucleophilic addition of bromomalonates to C_{60}[11] or the addition of carbenes[12] from diazomethanes and tosylhydrazones to C_{60},[12, 13] leading to methanofullerenes are also popular. Methanofullerenes have a three-membered ring at the 6, 6 junction of fullerenes (**Figure 1.3** (**5**)).[4, 14]

5
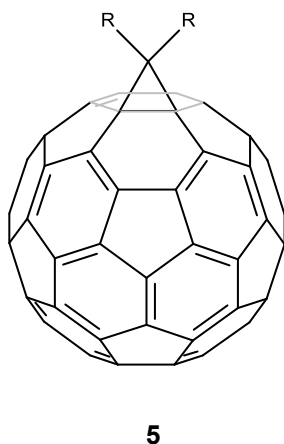
Figure 1.3 A three-membered ring (**5**) product from the Bingel-Hirsch reaction.

Of the four types of reactions mentioned, the [4+2] cycloaddition product (**4**) can contain up to six substituents. However, in the [4+2] cycloaddition reaction, the cycloadduct is not thermodynamically stable. The retro-reaction, which is strongly influenced by the electronic nature of the substituents on the diene, occurs readily. By contrast, the [3+2] cycloaddition product (**3**) not only has up to five different substituents,[4] but two substituents can be simultaneously inserted into the pyrrolidine ring resulting in a thermodynamically stable product.[15]

1.1.2 Fulleropyrrolidines

Four different approaches can be used to obtain fulleropyrrolidines:[16] (i) A two-step reaction which involves decarboxylation using N-methylglycine and a substituted aldehyde followed by 1,3 dipolar cycloaddition of the azomethine ylide (**Scheme 1.1**), (ii) the tautomerization route (**Scheme 1.2**) (iii) the reaction of an N-H fulleropyrrolidine with a suitable carboxylic acid chloride and acetic anhydride (**Scheme 1.3**) and lastly (iv) the thermal ring opening of an aziridine (**Scheme 1.4**).

The most commonly used functionalization procedure for C_{60} is the Prato reaction (**Scheme 1.1**).[17] This reaction forms stable fulleropyrrolidines (**3** and **9**) and is a powerful method that has steadily gained popularity due to its simple approach and readily available starting

materials (**Scheme 1.1**). Furthermore, the 1,3 cycloaddition reaction yields individual 6,6-closed isomers, where the electron density is higher.[6, 15]
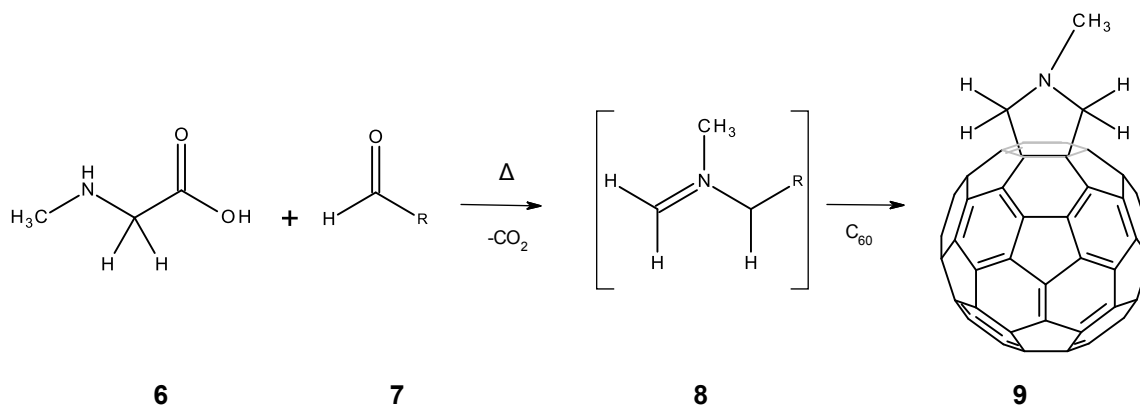
Scheme 1.1 Reaction scheme showing the decarboxylation of N-methylglycine (**6**) in the presence of paraformaldehyde (**7**) in refluxing toluene, and the subsequent formation of the N-methylfulleropyrrolidine derivative (**9**) *via* the azomethine ylide cycloaddition with C_{60} in the Prato method. (R = any group)

The intermediate azomethine ylide molecules (**8**) are planar reactive species that can be generated *in situ* in various ways. The most successful approach is the decarboxylation of immonium salts derived from the condensation of an α-amino acid with an aldehyde.[18] A typical example is the reaction of the α-amino acid, N-methylglycine (**6**) with paraformaldehyde (**7**) in refluxing toluene, which leads to a 41 % yield of **9** (R = H).[17]

Azomethine ylide addition to C_{60} changes the I_h symmetry of the C_{60} ring. Substituents present on the aldehyde **R** group (**7**) as well as the α-amino acid (CH_3) determine the symmetry of the fulleropyrrolidines (**9**). In instances where a symmetric aldehyde such as paraformaldehyde (**7 with R = H**) is used in combination with N-methylglycine (**6**), a symmetric product, N-methyl-2-fulleropyrrolidine is obtained (**9 with R = H**). The ^{13}C NMR spectrum displays fewer signals compared to those displayed for an unsymmetrical fulleropyrrolidine.[17] For example, the ^{13}C NMR spectrum for N-methyl-2-fulleropyrrolidine showed 19 signals which were attributed to the new C_{2v} symmetry of this product.[17] By contrast, an asymmetric derivative such as N-methyl-2-(4'-pyridyl)-3,4-fulleropyrrolidine (**10**)

shows a greater number of signals (40 in total) in its ^{13}C NMR spectrum.[16] The increased number of non-equivalent carbon atoms in the asymmetric fulleropyrrolidines results in challenging ^{13}C NMR spectroscopy characterisation.

10

Figure 1.4 Structure of N-methyl-2-(4'-pyridyl)-3,4-fulleropyrrolidine (**10**)

The Prato reaction tolerates a large variety of functional groups and is thus a versatile method of functionalizing C_{60}.[6, 16] In fact, the use of a different starting aldehyde or an N-functionalized glycine allows the synthesis of substituted fulleropyrrolidines and the availability of diverse aldehydes facilitates this process.[19]

The potential to form fulleropyrrolidine multi-adducts means that attention should be paid during synthesis of these derivatives.[6] Large excesses of reagents can result in up to nine pyrrolidine rings on the C_{60} cage.[6] Furthermore, the multi-adducts usually form complex isomer mixtures. For example, the bis-adduct can form up to eight isomers whilst the trisadduct forms up to forty-six isomers.[6] Therefore, reactions for mono-adduct products are often stopped when approximately 50-60 % of C_{60} has been consumed. At this point, only 10-20 % of bis-adducts will have been formed with approximately 40-50 % of the mono-adduct.[6] The mixture of adducts is normally purified by column chromatography.[17]

The three alternative approaches to synthesize fulleropyrrolidines all lead to more complex derivatives due to the presence of an additional functional group. Using the tautomerization

route, fulleropyrrolidines (fulleroprolines) (**13**) are formed when the ester of an amino acid (**11**) is refluxed together with an aldehyde (**12**) in the presence of C_{60}.[19] In these derivatives, the ester remains attached to the pyrrolidine ring as shown in **Scheme 1.2**.
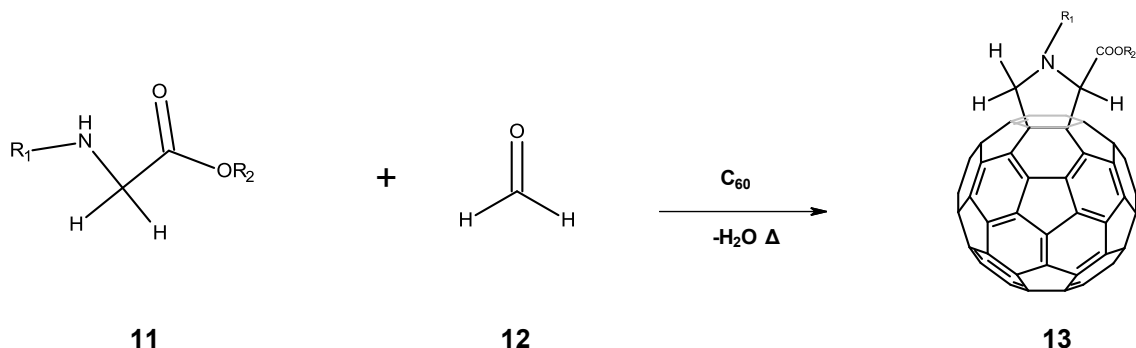
Scheme 1.2 The preparation of fulleropyrrolidines (fulleroprolines) *via* the tautomerization route.

Scheme 1.3 shows the formation initially of N-tritylfulleropyrrolidine (**15**). This fulleropyrrolidine can be synthesised by heating N-tritylazolodinones such as 3-triphenylmethyl-5-oxazolidinone (**14**) in toluene along with C_{60}, with a yield of 39 % after purification.[17, 19] Exposing (**15**) to triflouromethanesulphonic acid gives a precipitate which, after treatment with pyridine, generates the N-H fulleropyrrolidine (**16**). Reacting this product with a substituted carboxylic acid chloride and acetic anhydride produces the corresponding amide (**17**).
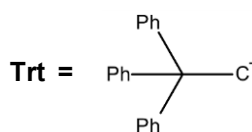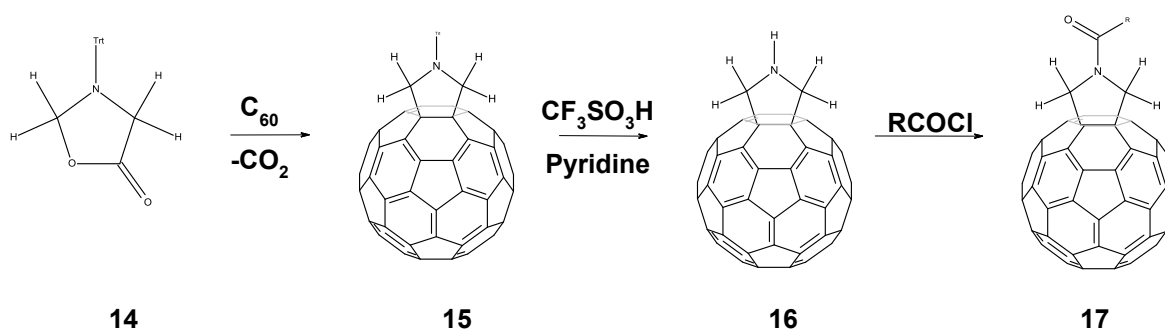
Scheme 1.3 Synthesis of an N-H fulleropyrrolidine and its reaction with a substituted carboxylic acid chloride and acetic anhydride.

The thermal ring opening of aziridines (**18**) to produce substituted fulleropyrrolidines (fullero-3,4-prolines) (**19**) was first described by Maggini *et al*. (**Scheme 1.4**).[20] To enable ring opening in this synthesis route, there must be at least one electron-withdrawing group on the three-membered ring (**18**). This reaction gives a yield of 40 % of the monoadduct (**19**) when refluxed in chlorobenzene.[17]

18   **19**
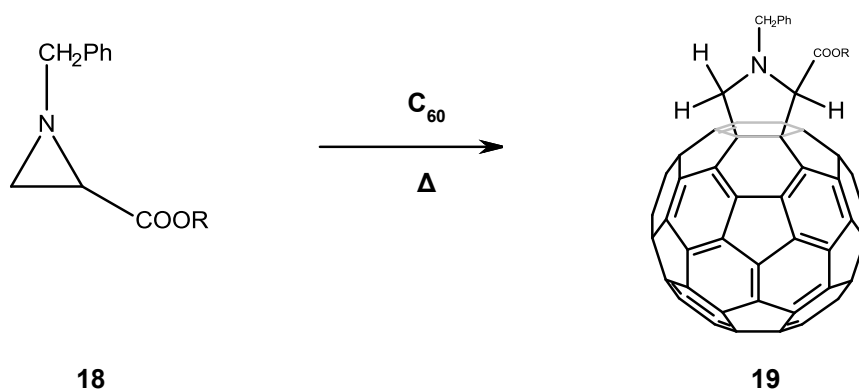
Scheme 1.4 A substituted fulleropyrrolidine (**19**) generated by the thermal ring opening of an aziridine (**18**).

Although all these methods are successful, the decarboxylation route (Prato reaction) has been the most widely used for obtaining functionalized fullerenes because the reaction leads to single monoadduct addition and because a wide variety of α-amino acids and aldehydes are commercially available.[6]

1.2 Applications of C_{60}

Target-directed synthesis of C_{60} derivatives with potential material and biological applications has received attention in recent years.[4, 21-24] The ability of C_{60} to accept six electrons and that of its derivatives to retain this ability enables C_{60} derivatives to function as electron acceptors with enhanced solubility in organic solvents. Furthermore, the derivatives also possess new properties due to the functionalization of C_{60}. This electron-accepting capability has been harnessed in the application of C_{60} derivatives as components of organic solar cells.[25-27] In addition, the hydrophobicity of C_{60} and its derivatives, along with the radius of the C_{60} cage,

was seen as an opportunity to explore the interactions of C_{60} derivatives with the HIV-1 PR protease.[28]

1.2.1 Material Applications

The drive towards greener sources of energy, which are low-cost and easy to make, has led to the intense research into organic photovoltaics worldwide.[29] Polymer solar cells (PSC) consist of a mixture of two materials, a donor and an acceptor, placed between two electrodes. In seeking better efficiency of the PSCs, bulk heterojunction solar cells were introduced, where the donor and acceptor materials were blended together in order to increase the contact area between the two materials.[30, 31]

1.2.1.1 C_{60} and Phenyl-C_{61}-Butyric Acid Methyl Ester

Phenyl-C_{61}-butyric acid methyl ester (PC$_{60}$BM) (**20**) is the most important electron acceptor synthesized to date.[32] With poly-(3-hexylthiophene) (P3HT) as the electron donor, polymer solar cells based on P3HT/PC$_{60}$BM have achieved power conversion efficiencies of 4-5 %.[25, 32]
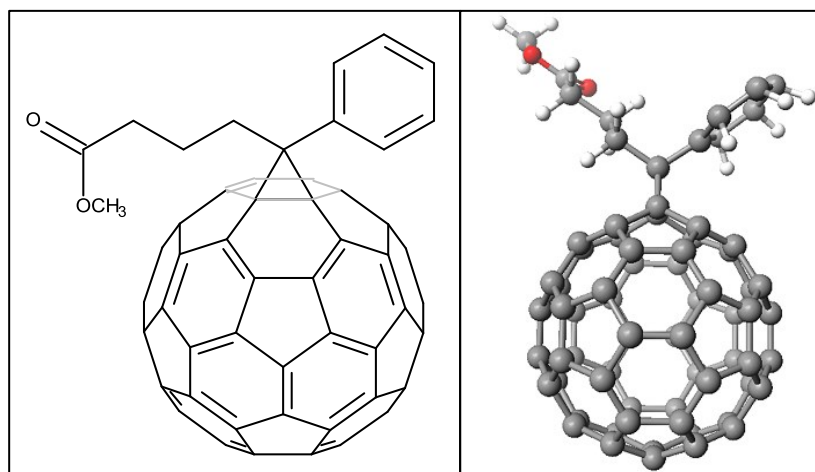
20

Figure 1.5 DFT B3LYP 6-31G** optimised structure of phenyl-C_{61}-butyric acid methyl ester (PC$_{60}$ BM) (**20**).

The LUMO energy for $PC_{60}BM$ has served as a benchmark for potential electron acceptors and for improving the performance of this type of solar cell. Experimentally obtained energy levels for $PC_{60}BM$ vary from publication to publication, ranging from -3.7 eV to -4.3 eV.[33-35] The calculated LUMO energy levels for $PC_{60}BM$ also vary from one publication to another: after full optimisation calculations using B3LYP/6-31G*, the LUMO energy level was -3.03 eV[36] and with B3LYP/6-311G**, the energy obtained was -3.05 eV.[37] These LUMO energies were higher than the calculated energy levels for pristine C_{60}, which were -3.14 eV after a full optimisation calculation with B3LYP/6-31G*[36] and -3.65 eV with B3LYP/6-311G**.[37] Making the LUMO energy of the electron donor less negative (higher) increases the efficiency of the polymer solar cell. The efficiency is also determined by the difference between the HOMO energy level of the electron donor and the LUMO energy level of the electron acceptor.[38] The commonly used donor/acceptor system is P3HT:PCBM and the experimentally determined HOMO energy level for the electron donor, P3HT is -4.90 eV.[25, 34]

Power conversion efficiencies (PCE) of 2.5 % were demonstrated by Shaheen *et al*. for the first time, with the poly[2-methoxy-5-(3′,7′-dimethyloctyloxy)-1,4-phenylenevinylene] (MDMO-PPV) system, which is an electron donor.[39] This report linked the morphology of the polymer-fullerene bulk heterojunction (BHJ) to its efficiency. It stated that the more contact there was between the donor and acceptor, the better the performance of the PSC.[39] The morphology of the device was analysed in chlorobenzene and toluene, and power conversion efficiencies of 2.5 % were observed for the chlorobenzene-based device whilst the toluene-based device had a power conversion efficiency of 0.9 %. This difference was attributed to the difference in solubility of their fullerene acceptor, phenyl-C_{61}-butyric acid methyl ester ($PC_{60}BM$), in chlorobenzene, which was double that in toluene.[39] Therefore, the morphologies of the devices were different: the chlorobenzene-based device showed more uniform mixing compared to the toluene-based device, which in turn, led to more contact between the two materials and thus better efficiency of the former device.[39]

1.2.2 Experimental Analysis of Potential Fullerene Electron Acceptors

The most commonly used fullerene derivative acceptor $PC_{60}BM$ (**20**) was first synthesized by Hummelen *et al*.[40] It was observed that due to the symmetry of C_{60}, the $PC_{60}BM$ compound had weak absorption in the visible region required in order for it to function as a photovoltaic material.[41] Therefore, to improve on this weak point, $PC_{70}BM$ was synthesised and indeed showed stronger absorption in the visible region from 400 to 700 nm. However, it is important to note that while $PC_{60}BM$ was a pure compound, $PC_{70}BM$ was a mixture of three isomers.[41] Although $PC_{70}BM$ has been used in some PSCs and showed a power conversion efficiency higher than $PC_{60}BM$ in corresponding devices, the cost of C_{70} is much higher than C_{60}. Therefore large-scale production of these devices would be costly.[41] It is because of this problem with C_{70} that scientists have continued to work on increasing the PCE of $PC_{60}BM$ as well as synthesizing new materials with the potential to function as electron acceptors.[25, 26]

The experimental analysis carried out by Wang *et al*. modified $PC_{60}BM$ structures (with a modified carbon chain ranging from 3 (**21**) to 7 carbons in length) and measured photophysical properties of the derivatives, including fluorescent quantum yields and the triplet quantum yields from fluorescence spectroscopy.[38] Compared to pristine C_{60}, it was found that the fluorescent quantum yields were enhanced whilst the triplet quantum yields were lowered.[38] These differences were attributed to the loss of symmetry of the $PC_{60}BM$-like derivatives compared to C_{60}, which results in the modification of the photophysical properties. The lowered triplet quantum yields meant that the derivatives were likely to make better electron acceptors in solar cells by preventing the triplet exciton formation that could lead to energy loses.[38] Increasing the carbon chain length induced minor changes between the derivatives which could be better understood by analysing the electron density distribution and LUMO energies calculated in the complementary computational calculations.[38]

21   **22**
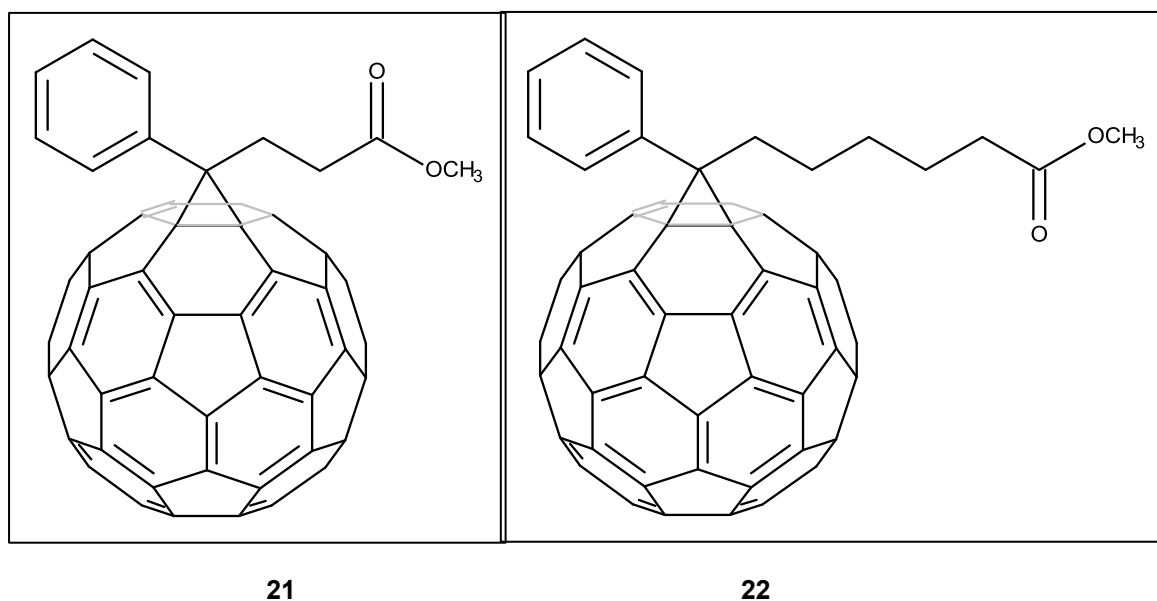
Figure 1.6 PC$_{60}$BM-like derivatives with a modified carbon chain with 3 (**21**) and 6 (**22**) carbons).

In the quest to find new electron acceptors with less negative LUMO energy levels, a new C$_{60}$ bisadduct derivative, indene-C$_{60}$ bis-adduct (ICBA) (**23**) was synthesised from C$_{60}$ and indene.[32] Analysis by cyclic voltammetry showed that ICBA had a less negative LUMO energy (-3.74 eV) than PC$_{60}$BM (-3.91 eV), leading to a difference between the LUMO energies of ICBA and PC$_{60}$BM of 0.17 eV.[32] These results also showed a superior LUMO to bisPC$_{60}$BM[42] which had a LUMO which was 0.10 eV (-3.81 eV) higher than PC$_{60}$BM. The PCE of the ICBA polymer solar cell with P3HT was 5.44 %, compared to 3.88 % of a PC$_{60}$BM device under the same conditions.[32] The optical properties of ICBA were analysed by UV-Vis, which showed strong absorption peaks in the visible region between 400-800 nm compared to PC$_{60}$BM. These results showed that ICBA could be a better electron acceptor compared to PC$_{60}$BM.[32]
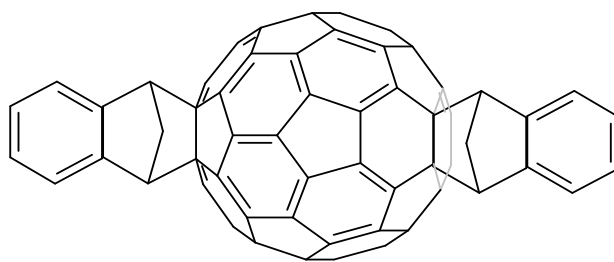
23

 Figure 1.7 Structure of ICBA (**23**).

One of the details that has been suggested to have an impact on the performance of solar cells is the solvent used in the fabrication of the device.[39, 43-45] The morphology of the BHJ active layer has been linked to the solubility of the fullerene acceptors in the solvent used. The fullerene derivatives with high solubility, which matched the solubility of the electron donor in the same solvent gave the highest power conversion efficiency. For a methanofullerene with 4 mg mL^{-1} solubility in chlorobenzene, the PCE was 0.02 %, whereas the methanofullerene with 80 mg mL^{-1} solubility in the same solvent had a PCE of 4.1 %.[44] Fullerenes that serve as electron acceptors in polymer solar cells should ideally have excellent solubility in common organic solvents so that the fullerenes can be dispersed.[26]

1.2.3 Solvated Properties of C_{60} and $PC_{60}BM$

Common solvents used in photovoltaic research are toluene, chlorobenzene (CB) and 1,2-dichlorobenzene (1,2-DCB).[26, 27, 44] Chlorobenzene and 1,2-dichlorobenzene have been the most-used because they are good solvents for both P3HT and $PC_{60}BM$.[46] Although fabrication of the active layer in solar cells occurs in solution, once the device has been assembled, it is followed by a drying process to remove the solvent.[27, 44] However, because the solvents used in the fabrication have aromatic rings, the van der Waals forces in these solvents mean it is possible that not all the solvent is removed. As a result, the solvent used, apart from influencing the morphology of the active layer, can also influence the electronic properties of this layer.

1.2.4 Computational Analysis of Potential Fullerene Electron Acceptors

In order to maximise the efficiency of BHJ solar cells, it is necessary to find the appropriate electron acceptor for each new donor that is used.[29] A lot of time is required to investigate the properties of all potential electron acceptors experimentally. Therefore, theoretical methods have been employed to assist experimentalists by offering a much better understanding of targeted molecules at a faster rate.[29, 47, 48] Computational data offers insight into the nature of a wider range of molecules than would be possible experimentally, whilst also designing better and possibly more efficient electron acceptors.[33]

Extensive research in this area therefore involved computational[29, 36-38, 48, 49] and experimental[25, 26, 45, 50, 51] work as scientists seek to improve the efficiency of these PSCs.

Quantum chemical methods were used to predict the acceptor strength of different C_{60} derivatives in order to understand the effects of substituents on the reduction potential of the derivative.[29] The reduction potential of fullerene derivatives was analysed because the power conversion efficiency of a solar cell is linked to the difference between the HOMO energy level of the electron donor and the LUMO energy level of the electron acceptor.[29] Therefore, the LUMO energy level of the electron acceptor can be raised (made less negative) to maximise the efficiency of the electron acceptor. However, for electron transfer from the donor to the acceptor to occur, the LUMO of the donor needs to be 0.3-0.5 eV higher than that of the acceptor (**Figure 1.7**, ΔE).[25, 42] The LUMO energies of $PC_{60}BM$ and a range of novel C_{60} derivatives was calculated.[29] The calculated LUMO energies were plotted against the first reduction potentials of the derivatives in order to establish a linear relationship. This linear relationship can be used to estimate the redox potentials of unknown C_{60} derivatives using their calculated LUMO energy levels.[29]
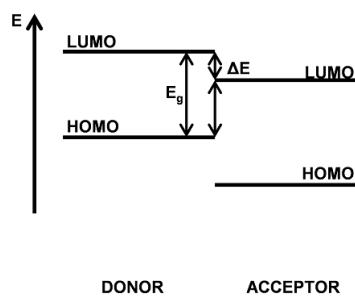
Figure 1.8 Energy level diagram for a donor-acceptor heterojunction. (E_g = band gap)

In a similar study, the LUMO levels of $PC_{60}BM$ and its modified derivatives, as well as those of other C_{60} derivatives, were calculated along with the electron affinity and electrophilicity.[37] The electron affinity of the C_{60} derivatives was calculated because of previous reports which stated that the efficiency of a fullerene in a solar cell depends on its electron affinity in order to effectively absorb electrons from the donor.[52] In addition, the electrophilicity was also calculated because it has been argued by the same group that the efficiency of fullerenes in solar cells also depends on its resistance against the back-transfer of electrons to the donor molecule.[52] After the calculations, linear relationships were observed for both the calculated electron affinity and electrophilicity versus the experimental open-circuit voltage. Since the studied fullerene derivatives had LUMO energy levels higher than $PC_{60}BM$, it was concluded that raising the LUMO levels of the electron acceptor results in an increase in the open-circuit voltage of the solar cell.[37]

In order to understand the structural characteristics of $PC_{60}BM$ and $PC_{70}BM$, the Gibbs free energies of $PC_{60}BM$ and $PC_{70}BM$, along with the enthalpies of formation and reaction have also been calculated.[48] Both computational and experimental results showed that the formation of $PC_{70}BM$ was more favoured over $PC_{60}BM$. The calculated and experimental standard enthalpies of reaction for $PC_{70}BM$ (-153.1 kJ mol^{-1}, 11.1 kJ mol^{-1}) were more exothermic than those for the $PC_{60}BM$ (-150.0 kJ mol^{-1}, 57.9 kJ mol^{-1}) respectively.[48]

$PC_{60}BM$ was also studied by Wang *et al.* in addition to a range of $PC_{60}BM$-like C_{60} derivatives.[38] The length of the butyl carbon chain (4 carbon atoms) in $PC_{60}BM$ was altered

from 3 (**21**) to 7 carbon atoms in order to investigate the effect of the carbon chain on the performance of the polymer solar cells (PSC). These effects were examined by calculating the HOMO and LUMO energies as well as the HOMO-LUMO gap and electron distribution on the C_{60} derivatives using PBEPBE/6-311G(d,p).[38] Since the performance of fullerene derivatives in PSCs is determined by the LUMO energy levels, the LUMO levels were examined and it was observed that the energies did not increase according to the increase in length of the carbon chain. Instead, an oscillating trend was observed in which $PC_{60}BM$ with a four-carbon chain, and a derivative structure with a six-carbon chain (+2 carbons) (**22**), had the highest calculated LUMO energy levels (-3.999 eV and -4.013 eV respectively). Consequently, these two derivatives had the higher photovoltaic performance of above 3.5 %, with P3HT as the donor.[53] The other derivatives (- 1 carbon, + 1 carbon and + 3 carbons) showed a lower efficiency of 3.0 % with calculated LUMO energies of -4.074 eV, -4.052 eV and -4.038 eV respectively. The oscillation of the LUMO energies was attributed to the electron-withdrawing effects of the $-COOCH_3$ group, it was observed that the HOMO electron densities of the derivatives were concentrated at the top of the C_{60} cage, where the functional group was attached.[53] The conformation of this $-COOCH_3$ group, which was *cis* to the C_{60} cage thus having the oxygen atom pointing towards the C_{60} cage (**22**), was also attributed to the difference in the LUMO energies, whereas the other three derivatives (- 1 carbon, + 1 carbon and + 3 carbons) had the $-COOCH_3$ group *trans* to the C_{60} cage (**21**). It is interesting to note that the experimentally measured LUMO energies, using cyclic voltammetry, were almost identical to each other (-3.9 eV) whereas the calculated values revealed the small differences in the LUMO energies. These could be within the error range of the calculations. Since the experimental results were almost identical to each other, it was stated that more accurate methods than cyclic voltammetry were required to measure these LUMO energies because of the small variations revealed by the electronic structure calculations. In addition, these results also show the importance of comparing computational and experimental results.[54]

From these studies, it is clear that computational studies have a significant role in the quest to optimise the performance of solar cells. The fast output of results from computational calculations means that they can be the starting point for experimental work in helping to eliminate unfavourable compounds for synthesis. Computational results can also be used alongside experimental results to get a better understanding of a reaction and to confirm results. For example, frontier orbital distribution gives a visual picture of the reactivity of a molecule, something that would not be possible experimentally.

1.3 Self-Assembling Systems

Two different modes of self-assembly will be considered as part of this project: (1) ordered arrays of C_{60} monomers on a zeolite and (2) supramolecular complexes formed from complementary pairs of C_{60} derivatives designed so that they can spontaneously bind to each other.

1.3.1 Organic and Inorganic Composites

The ability to produce ordered arrays of C_{60} fullerene derivatives with tailored properties would be a step towards developing nanodevices with specific applications. There are a number of different inorganic materials that have the potential to act as a template for this process. This field of research has promising applications in areas such as catalysts and sensors.[55] This relatively simple process allows researchers to tailor their products by choosing a suitable host matrix and organic molecule.[24]

C_{60}, because of its spherical shape and ability to reversibly receive up to six electrons, has been widely used as the organic component in the synthesis of organic-inorganic nanostructures. Surfaces commonly used as host matrices for C_{60} and its derivatives can be classed into two groups: non-porous structures including gold[56-59] and silicon[60-62] and porous materials which includes zeolites.[63-69] Metals like gold and silicon have served as substrates for self-assembled monolayers of either pristine C_{60} or its derivatives.

C_{60} tends to form aggregates in solution. Controlling these intermolecular interactions is a priority in the formation of fullerene-based supramolecular structures.[65] Therefore, using derivatives that do not aggregate or hosts such as zeolites, which facilitate ordered organisation helps to solve this problem.[65]

One of the aims of this work is to test the suitability of zeolites as a template for the self-assembly of fullerene derivatives. The zeolites should enable the C_{60} derivatives to assemble in ordered arrays, creating organic and inorganic composites.

1.3.2 C_{60} and Zeolites

Zeolites are ideal hosts for supramolecular organization, compared to the non-porous hosts like gold and silica, because they are made up of ordered cages and channels. They have high chemical and thermal stability[70] which allows for the testing of the effect of high temperatures on molecular adsorption.[63, 68, 71] In general, these structures are a class of well-defined crystalline aluminosilicates whose three-dimensional structures stem from an array of corner sharing $[SiO_4]^{4-}$ and $[AlO_4]^{5-}$ tetrahedra.[72] The pores sizes found in zeolites range from 7.4 Å to 32 Å.[67, 69] Thus, zeolites with a narrow enough channel opening allow the formation of ordered arrays of C_{60} on the surface, spaced far enough apart to minimise intermolecular interactions. The ball-like shape of C_{60} and the crystallographic diameter of 7.9 Å[67-69] is almost perfect for the zeolites channels as the fullerene can potentially settle at the channel opening, serving as a stop-cock, or penetrate inside the zeolites channels, if the channel is sufficiently large. Both possibilities and their effect on the photophysical properties of C_{60} have been studied by several groups.[63-71]

The majority of C_{60} studies has focused on the synthetic forms of the naturally occurring mineral faujasite, (microporous zeolites X and Y). The structure of zeolite Y (**24**) can be seen in **Figure 1.9**. A mesoporous aluminosilicate, MCM-41 has also been frequently used in the host-guest research with C_{60}.[67] The difference between these zeolites is the size of the channels: MCM-41 has a much larger channel diameter (32 Å) compared to the faujasites

(7.4 Å). The larger pore diameter enables easier adsorption of C_{60} into the pore of MCM-41.[67, 69] However, the smaller diameter of the faujasite pores means that the migration of C_{60} from the faujasite zeolite exterior into the channels does not take place at room temperature, instead C_{60} rests on the zeolite surface. Vapour-phase adsorptions at elevated temperatures have been documented with 650 °C identified as the optimum temperature at which this occurs.[63, 71] Despite vapour-phase adsorptions being a popular choice, other methods such as liquid-phase adsorption followed by hydrothermal techniques have also been used to adsorb C_{60} into the zeolite channels. Vapour-phase deposition does not lead to the penetration of C_{60} into the zeolite in all cases: VPI-5 is one known exception.[65]
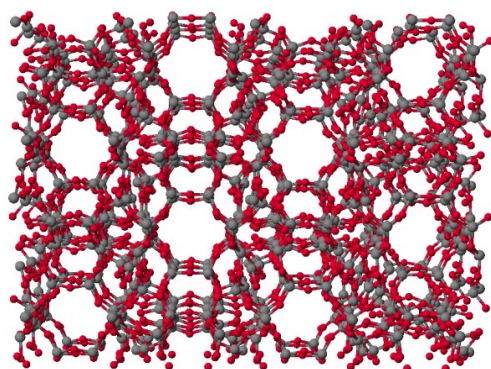
24

Figure 1.9 Modelled structure of a zeolite showing the open channels. (**24**)

Although the crystalline structure of the host zeolite does not change, that of the encapsulated C_{60} has been reported to alter after adsorption, if this occurs, the photophysical properties such as the fluorescence emission and UV-Vis of the fullerene are changed.[69] It is also interesting to note that a shift to a higher wavelength was reported for the confinement of C_{60} in zeolites with different channel sizes ((Na-Y (7.2 Å), UTD-1 (10 Å x 0.75 Å)) where the greatest shift was attributed to the zeolite with the smallest host cavities.[69] These studies demonstrate how the optical properties of C_{60} can be modified in different environments, and if fluorescence emission is being examined, how the host can be chosen as a function of the size of shifting required.[64, 69] These studies also demonstrate why zeolites stand out as a

host for organic structures. Their varying channel sizes and different structural composition offer a range of different environments for various applications.

1.3.3 Supramolecular Chemistry

Supramolecular chemistry has been one of the most studied objectives in nanotechnology in the last few decades.[73] With nature providing inspiration, scientists have come up with ways of making tailored compounds which are capable of functioning as building blocks that will self-assemble into structures ranging from simple to highly complex supramolecular materials.[74] The properties of the supramolecular structures are important for their potential applications and the excellent electron-accepting ability of C_{60} makes it an attractive building block.

A range of noncovalent interactions such as ion-ion, dipole-dipole, π-π stacking, van der waals and hydrogen bonds (H-bonds) are available for engineering multimolecular arrays.[75, 76] From these reversible interactions, hydrogen bonds have proved to be the most useful because of their specificity and high degree of directionality.[76] The H-bond, with binding energies between 4 and 120 kJ mol^{-1} has proved to be the most useful.[76] The presence of H-bonds in DNA, where stable structures are generated from multiple H-bonds and hydrophobic interactions, demonstrates the strength of these interactions.[76] Since Watson-Crick binding modes are the preferred interactions in DNA, replicating their stability in the construction of novel structures has been studied by several research groups.[77-82]

1.3.4 DNA Nucleobases

Double-stranded DNA is an excellent example of self-assembly in nature. DNA is made up of four base pairs: two purines, adenine (A) and guanine (G) and two pyrimidines, thymine (T) and cytosine (C). The presence of strong, non-covalent and reversible hydrogen bonds in the natural systems such as DNA has acted as a muse to the field of nanotechnology and supramolecular chemistry. By using a balanced combination of hydrogen bonds, van der Waals and other forces, DNA exhibits a rare, predictable control over its secondary

structure.[73] These unique inherent properties have inspired scientists to exploit DNA synthetically to build predictable and well-controlled structures. This field of science has led to the development of DNA nanotechnology.[83]

In DNA, the binding motifs are adenine-thymine (A-T) (**25**) and guanine-cytosine (G-C) (**26**). This is known as the Watson-Crick binding mode and is dominant in the constrained geometry of the double helix structure (**Figure 1.10**).[84]

A-T Watson-Crick (**25**)   G-C Watson-Crick (**26**)
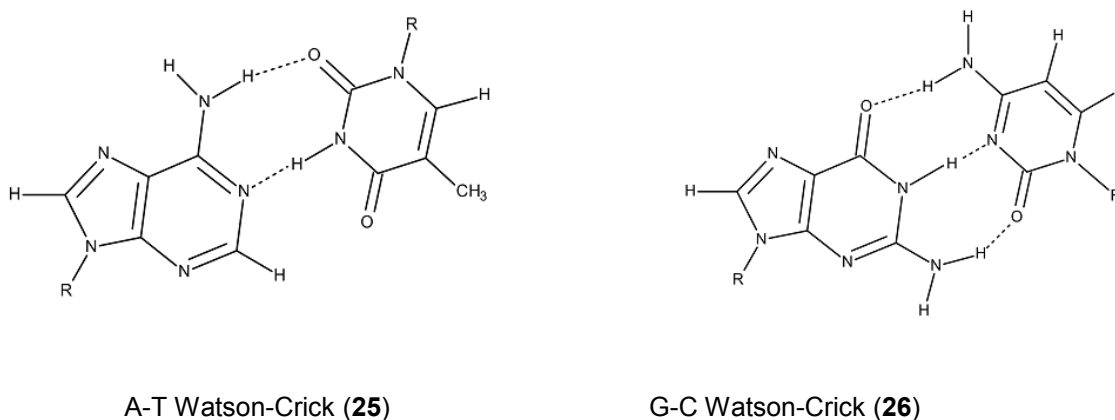
Figure 1.10 Watson-Crick binding modes of the purines and pyrimidines: adenine (A), thymine (T), guanine (G) and cytosine (C). A-T is bound by 2 H-bonds (**25**) and G-C by 3 H-bonds (**26**).

However, in the absence of the double helix feature of DNA, other binding modes are possible. These include Hoogsteen (**27**) and wobble (mismatched) (**28**) base pairs (**Figure 1.11**).[84] In contrast to the Watson-Crick A-T binding mode, which uses the C6 and N1 in adenine for binding, the Hoogsteen mode uses the C6 amino group as the donor and the N7 position of adenine as an H-bond acceptor.
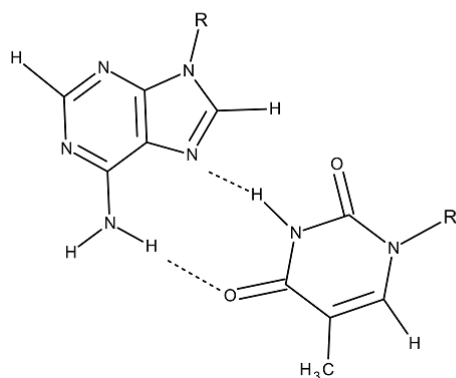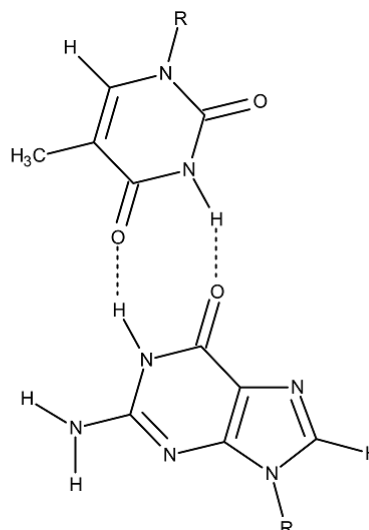
A-T Hoogsteen (**27**)   G-T Wobble (**28**)

Figure 1.11 Alternative binding modes of nucleotides, Hoogsteen (**27**) and wobble (**28**).

1.4 Aims and Objectives

Having identified fullerenes as molecules with interesting electronic properties, the focus is to develop fullerenes tailored for specific applications. As relatively good electron acceptors, it is also useful to test the performance of any fullerene in a range of other applications.

Aims:

To design a range of C_{60} derivatives with self-assembling properties and analyse the properties of the monomer and dimer structures in different environments using a combination of computational and experimental methods.

Objectives:

1. Computationally model and optimise novel C_{60} derivatives with H-bonding properties to obtain the conformations most likely to be formed experimentally as shown in Chapters 7- to 9.

2. Model and optimise dimer combinations of the monomers in two different orientations (dumbbell and seesaw) to explore the different self-assembly patterns which could occur experimentally.

3. Characterize the stability and possible applications of both the monomer and dimer lowest-energy conformations by calculating the binding energies, dipole moments and electronic properties of these lowest-energy conformers.

4. To synthesise and characterise the novel C_{60} derivatives.

5. To test the ability of C_{60} derivatives to adsorb on a zeolite surface.

6. In Chapters 7 and 10, the possible material and biological applications of the C_{60} derivatives are analysed. In particular, the photovoltaic applications and the abilities of the fullerene derivatives to act as inhibitors at the active site of the HIV-1 PR.

Chapter 2

Research Methods

2.1 Overview

Due to the time consuming nature of deriving synthesis pathways and computational calculations, these experiments were carried out in parallel. The computational calculations focused on probing the electronic properties of the monomers and predicting their ability to self-assemble into dimers. In Chapters 3-6, the synthesis pathways designed for each monomer are discussed, highlighting some of the challenges associated with synthesizing each derivative. Then in Chapter 7, the properties of the monomers based on computational calculations are analysed. In this chapter, the experimental methods used to analyse the products formed in each reaction step are outlined and the theoretical methods underpinning the computational calculations are discussed.

2.2 Computational Modelling

Computational modelling can be used to predict the structure, physical and chemical properties of new materials before they have been created in the laboratory.[85] This provides a cost effective alternative to carrying out experiments alone. It also speeds up the discovery process and allows one to search a broader number of candidate molecules and identify the most interesting species on which experimentalists should focus.[86]

One of the goals in computational chemistry is the design of molecules for specific applications.[87] As a result, this field of chemistry has become increasing popular in the last decade, not only as a starting point for research studies but also as a key component in understanding regioselectivity. In addition, reaction mechanisms, reactions outcomes and the properties of the product and stereochemistry of reactions can all be predicted. This is particularly useful when dealing with novel compounds, whose reactivity may be unknown or not clearly understood, or compounds with limited solubility like C_{60}.

Different levels of computational calculations are available. The accuracy of the level of theory depends on the parameters and assumptions that each method takes into account. The levels of accuracy required, as well as the size of the molecules and the properties to be analysed influence the level of theory used. Another influencing factor is the computational cost of the calculations. High-level calculations are computationally expensive and time-consuming, as a result, cheaper, low-level methods are often chosen as starting points.

2.3 Molecular Mechanics

The initial geometry of a molecule used in computational calculations is usually determined using molecular mechanics (in the form of a classical force field). This is the quickest approach to calculate the energy of each conformer. It also provides a good starting point for the equilibrium geometry, due to the large number of parameters included that are derived from experimental data.

Classical force fields are based on the assumption that all molecules are made up of similar interactions, even in different molecules. The explicit contribution of electrons is ignored. Instead, the energy is calculated based on nuclear positions. A modelled molecule is made up of atoms held together by bonds which are in constant motion. Classical force fields model atoms as spheres and bonds as springs. The atoms, therefore, interact through electrostatic and interatomic forces.[88, 89]

The force field energy can be expressed as a sum of individual energy terms:

$$E_{total} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{electro} + E_{cross} \qquad (2.1)$$

E_{str} = bond stretching energy

E_{bend} = angle bending energy

E_{tors} = dihedral twisting or torsion energy

E_{vdw} = van der Waals contributions

E_{electo} = electrostatic interactions

E_{cross} = coupling between the first three terms

The bond stretch and angle bend energies are both based on the simple harmonic oscillator model:

$$E_{str} = \sum \tfrac{1}{2} k_b (r - r_0)^2 \qquad\qquad (2.2)$$

$$E_{bend} = \sum \tfrac{1}{2} k_\theta (\theta - \theta_0)^2 \qquad\qquad (2.3)$$

where k_b, k_θ, r_0 and θ_0 are parameters specific to each bond type or angle derived from fitting to experimental data. The bond stretching energy (Equation 2.2) is the sum of the potential energy of each harmonically oscillating bond. The harmonic parameter, k_b, is the average of the bond force assigned for each type of bond, such as C-C, C-H or N-H based on IR spectra of a set of test molecules. Similarly, the average equilibrium bond length, r_0, for each bond length can be found from microwave spectra of the same set of molecules. The sum of the E_{bend} extends over all angles (Equation 2.3). The values of k_θ which controls the stiffness of the spring, and θ_0 which is the equilibrium angle, are specified for each type of angle, such as $\widehat{CCC}$ $\widehat{COH}$ and $\widehat{OCH}$.[88, 89]

Force fields usually differ in four aspects: the number of terms included, the function form of each energy term, the type of molecules used for fitting the parameters and the number of bond angles and dihedral types parameterised.[89] The parameters take into account the chemical environment of each atom; an sp^2 carbon atom is different from an sp^3 atom, a carbonyl carbon differs from an aromatic carbon. The set of molecules used to derive the experimental parameters determines the types of systems the force field can be used for since they are unlikely to be suitable for systems that have very different atom configurations. Thus MM2 and MM3 can be used to model small organic molecules such as those considered in this study.[89, 90]

2.4 Semi-Empirical Methods

Semi-empirical methods are normally used after force field calculations, and before *ab initio* calculations. This is because force field methods produce mostly qualitative results about the relative stability of conformers and *ab initio* calculations, though quantitative, are far more

time-consuming. Molecular mechanics calculations are also not able to provide thermodynamic information about the structures since they do not include electronic contributions explicitly, and thus cannot model the movement of electrons in bond formation and breaking processes.

Semi-empirical methods consider only the valence electrons explicitly and therefore have a considerably lower computational cost compared to *ab initio* methods. The calculation makes up for not including the core electrons of a molecule by introducing functions to model the combined repulsion due to the nuclei and core electrons or reducing the nuclear charge.[89]

AM1 and PM3 are examples of the commonly used semi-empirical methods. The two methods are similar as they were developed using the same set of data, based on carbon, hydrogen, nitrogen and oxygen.[91] The most notable difference between AM1 and PM3 is the way in which the parameters used for the energy calculations in the Hamiltonian were obtained: For PM3, the parameters were derived *via* an automated process whereas those for AM1 were derived from intuition and chemical knowledge. Consequently, different values are obtained for some of the parameters.[91]

The choice regarding which method to use is often determined by further applications of the resulting data. In this project, to calculate the solvation energies of the target molecules, the COSMO solvation model was employed as it can use the output from AM1 optimisation followed by a single-point density functional calculation.[92]

The structures and properties of C_{60} and its derivatives have been studied using a range of methods, including semi-empirical and *ab initio* approaches.[29, 93-95] Semi-empirical methods are usually paired with *ab initio* methods.[29] This is because of the simplicity of the Hamiltonian used in semi-empirical methods, whereas *ab initio* methods use the exact Hamiltonian.[96]

2.5 Density Functional Theory

Ab initio methods provide the most accurate calculations but due to the increase in the number of interactions included in the calculations, these are also the slowest. *Ab initio* calculations include all the electrons in a system in their calculations and are therefore the most favoured method when calculating electronic properties. Unlike force field and semi-empirical methods, *ab initio* calculations can be generated without the prior knowledge of experimental parameters.

2.5.1 Kohn-Sham Theorem

Hohenberg and Kohn[97] outlined the basis of density functional theory by showing that the energy of an electronic system can be written in terms of the electron-density function. In contrast to many *ab initio* methods such as Hartree-Fock theory, DFT only aims to calculate total electronic energy and electron density distribution rather than the full *N*-electron wavefunction.[91] This is less demanding computationally and can be used to look at larger systems. Unfortunately, the exchange-correlation energy functional is not known and instead approximate solutions are obtained using the Kohn-Sham (KS) orbitals.

In DFT, the energy functional is written as a sum of two terms:

$$E[\rho(\mathbf{r})] = \int V_{\mathrm{EXT}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + F[\rho(\mathbf{r})] \qquad (2.4)$$

where E is the energy of a unique functional of $\rho(\mathbf{r})$ and $V_{\mathrm{EXT}}(\mathbf{r})$ arises from the interaction of the electrons with an external potential. $F[\rho(\mathbf{r})]$ is the sum of the kinetic energy of the electrons and the contribution from inter-electronic interactions. However, the function $F[\rho(\mathbf{r})]$ is not known therefore, Kohn and Sham suggested that $F[\rho(\mathbf{r})]$ should be approximated as the sum of three terms. Thus, the energy functional in DFT is given by:

$$E[\rho(\mathbf{r})] = \int V_{\mathrm{EXT}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + E_{\mathrm{KE}}[\rho(\mathbf{r})] + E_{\mathrm{H}}[\rho(\mathbf{r})] + E_{\mathrm{XC}}[\rho(\mathbf{r})] \qquad (2.5)$$

$E_{\mathrm{KE}}[\rho(\mathbf{r})]$ is the kinetic energy, $E_{\mathrm{H}}[\rho(\mathbf{r})]$ is the electron-electron Coulombic energy and $E_{\mathrm{XC}}[\rho(\mathbf{r})]$ contains contributions from exchange and correlation.[91]

The $E_{\mathbf{XC}}[\rho(\mathbf{r})]$ functional is usually separated into an exchange energy (representing the energy when two electrons are exchanged between different orbitals) and a correlation energy functional representing the energy due to how the movement of one electron will affect the behaviour of all other electrons.[98]

There has been a debate as to the significance of the Kohn-Sham orbitals because Kohn-Sham orbitals are a set of non-interacting orbitals created to give the correct density, and therefore do not have the same physical meaning as those in Hartree-Fock (HF).[91] However, a number of papers have shown that the shape and symmetry of the Kohn-Sham orbitals are similar to those of HF methods, and the energy order of the occupied orbitals is consistent across a variety of methods.[99] Higher energy occupied orbitals are sometimes interchanged in energetic order.[99] Although the HOMO-LUMO energy gap calculated using a range of DFT functionals underestimates the experimental excitation energy (by 2 - 17 %), the trends seen for different molecules are consistent and thus should be considered as a useful predictor of the optical gap (rather than the fundamental gap which is the difference between the ionisation energy and electron affinity).[100]

2.5.2 Exchange-Correlation Functions

The main source of error in DFT comes from the approximate nature of the exchange-correlation energy functional. Various exchange-correlation functionals have been developed for use in DFT calculations. They include BLYP,[101] B3LYP[101] and PW91.[102] Their names indicate a pairing of an exchange functional and a correlation functional. B3LYP is a hybrid DFT calculation that uses Hartree-Fock corrections.[97] Hartree Fock is a method used to solve Schrodinger's equation for systems with many electrons. However, this method neglects electron correlation for electrons that are in opposite spin and this can lead to differences between the calculated results obtained and those from experimental data. Therefore, a method such as DFT which combines the Hartree Fock method with density-functional correlation and exchange contributions to the exchange-correlation functional is often used as an alternative.[36, 97]

Equation 2.6 shows how the Becke gradient-exchange (B88), the Lee-Yang-Par correlation functional (LYP) and HF contribution are mixed.

$$E_{XC}^{B3LYP} = (1 - a_0)E_X^{LDA} + a_0 E_X^{HF} + a_X \Delta E_X^{B88} + a_C E_C^{LYP} + (1 - a_C)E_C^{VWN} \qquad (2.6)$$

a_0, a_x and a_c are empirical coefficients obtained by least-squares fitting to experimental data and their values are $a_0 = 0.20$, $a_x = 0.72$ and $a_c = 0.81$.

2.5.3 Basis Sets

A basis set is a collection of functions which are combined linearly, which is known as the Linear Combination of Atomic Orbitals (LCAO), to form molecular orbitals.[89] Minimal basis sets are composed of the minimum number of basis functions required to represent all the electrons on an atom. These smaller basis sets are often used for large molecules. This is because smaller basis sets are less demanding computationally and therefore take less time.

An example of a commonly used basis set, particularly with C_{60} and its derivatives is the split valence double zeta basis set, 3-21G.[29, 103] The notation used describes the distribution of the Gaussian functions in a molecule; three Gaussian functions are used to describe the core orbitals and the valence electrons are described by a total of three Gaussians ,two Gaussians for the contracted part and one Gaussian for the diffuse part.[91]

The accuracy of a calculation can sometimes be improved by increasing the number of basis functions.[96] Basic split valence basis sets do not accurately represent the directional nature of orbital interactions in molecules. For example, the distribution about the 1s orbital of an isolated atom like hydrogen is symmetrical compared to its distorted distribution in a hydrogen molecule (H_2). The distortion is a result of the attraction of electrons towards other nuclei in the molecule and in order to represent this region, of high electron density, polarisation functions are introduced into the basis set.[91]

Polarisation functions can accommodate the directional nature of orbital interactions in molecules. They correspond to the addition of *p*-orbitals to hydrogen to add some *p*-type character to the electron distribution. This produces an *sp* hybrid and shifts the centre of the hybrid away from the atomic nucleus.[91] Similarly, *d*-orbitals added to the first and second-row elements facilitate the formation of more complex bonding. Polarisation functions are represented by an asterisk (*) next to a basis set name.[91] A polarised basis set in which six *d*-type functions are added for each atom other than H is written as 6-31G*. An additional polarisation function, 6-31G**, represents the addition of a set of three *p*-type polarisation functions for each H atom.[88] This particular polarised basis set is useful where hydrogen acts as a bridging atom. Alternatively, the polarised basis sets, 6-31G* and 6-31G** can be written as 6-31G(d) and 6-31G(d, p) respectively.[91]

2.6 Potential Energy Surface

Since molecular mechanics calculations using a force field are relatively quick, they can be used to calculate the energy of many different conformers of organic compounds. In order to identify the conformers a grid search is carried out. This search initially involves the identification of all the rotatable bonds in a molecule, which are then rotated through 360° using fixed increments of 72°. Every conformation that is derived is subject to energy minimisation in order to derive the associated minimum energy conformation. Once all of the possible conformations have been generated and minimised the search is stopped.[91]

The varying energies for the different geometric structures of the molecule are computed to produce an energy map.[91] This map, called the potential-energy surface, shows the minimum and maximum energies of a molecule. The minimum-energy points on the map are the most interesting because they indicate the most stable states of the molecule. On the potential-energy map, the lowest-energy point is referred to as the global minimum whilst the other lower points on the energy map are called local minima. The lowest-energy conformer is the conformation a molecule prefers to adopt at zero Kelvin. Experimentally, this is often the conformation adopted by the synthesised version of the molecule.[91] Once a set of low-

energy conformers have been obtained, further optimisation can be carried out, at higher levels of theory, prior to property calculations on the converged structures.

2.7 Optimisation

Geometry optimisation is a procedure that attempts to find the closest conformation of a molecule which is closest to the local minimum of the starting point.[98] The lowest-energy conformer from the semi-empirical potential energy surface is used as a starting point. This structure is first optimised using a semi-empirical method to find the global-energy minimum on that potential energy surface. To gain a more accurate result for the global minimum, the energy can be reoptimised using a single-optimisation calculation[93, 94, 104] or full geometry calculation at a higher level of theory.[33, 37]

Single-point calculations are popular for a structure like C_{60} because of its large size. This type of calculation finds the energy of a molecule at a single geometry using a DFT/*ab initio* approach and is therefore computationally cheaper and less time consuming.[93, 94, 104]

In comparison, a full geometry optimisation where the energy of all of the intermediate geometries of a molecule are calculated until the local minimum is found. At this point, the calculation converges to produce the final, minimum-energy conformer as well as the coordinates of the structure at this corresponding geometry. The calculation converges when the differences between the energies of the minimum-energy conformers becomes negligible by ($\Delta E < 10^{-5}$ or 10^{-6}).

Full geometry optimisation calculations yield more accurate results compared to single-point calculations. However, single-point calculations produce more accurate results compared to semi-empirical methods that produce the global-energy minimum. Finding a balance between the computational cost of a calculation and the accuracy of the results is important. The method used, single-point or full optimisation, or basis set used, play a vital role in both methods.

2.7.1 Basis Set Superposition Error

Calculating the binding energy of dimers by subtracting the optimised dimer energy from the energy of the isolated monomer energies results in a discrepancy known as basis set superposition error (BSSE).[91] This is a result of the incompleteness of the basis set used for the monomers: as the two monomers approach each other, the energy of the system falls because the basis functions on each molecule provide a better description of the electronic structure around the other atom and because of the favourable intermolecular interactions. Thus, this error leads to an overestimation of the true value of the binding energy.

The most widely used method of correcting the BSSE effects is the counterpoise correction method by Boys and Bernardi.[105] In this method, the energy of monomer A is calculated in the presence of ghost orbitals of monomer B and the same calculation is performed for B using ghost orbitals of A. The sum of these energies is then subtracted from the energy of the dimer as shown in Equation 2.7.

$$\Delta E = E_{AB}^{AB}(AB) - E_A^A(A) - E_B^B(B) - [E_A^{AB}(AB) - E_A^{AB}(A) + E_B^{AB}(AB) - E_B^{AB}(B)] \qquad (2.7)$$

2.8 Docking Algorithm

Docking and scoring are important during the drug discovery process because it predicts the most energetically favourable binding mode of a known active ligand and identifies new ligands using virtual screening.[106] Scigress, the programme used for the docking calculations, automated the docking of the ligands into the active sites by using a genetic algorithm with a fast and simplified potential of mean force (PMF).

PMF is a knowledge-based function.[107] This is because this function extracts structural information from experimentally determined protein-ligand complexes as stored in the Protein Data Bank (PDB) and uses the Boltzmann law to change the atom pair preferences into pair-wise potentials that are dependent on distance.[107]

Genetic algorithms are an optimization approach which was inspired by the evolution process where the best individuals have a greater chance of surviving. Genetic algorithms have several steps; (i) coding of variables – in this initial step, the genetic algorithm begins by creating random population chromosomes, which are positions, orientations and conformations of the ligand and then applying operators, such as crossover and mutation, produced by the combination of genetic material from more than 1 origin so as to preserve those which are the best fit; (ii) evaluation of the response – the best solutions are preserved by evaluating the PMF score of each chromosome and allocating reproductive opportunities so that the chromosomes with lower scores have more chances to multiply than those with poorer solutions; (iii) reproduction – the reproduction process continues for a specified number of generations, (1000 - 40 000).[108] Running for a greater number of generations increases the likelihood of finding the most energetically favourable binding mode and at least 40 000 generations are recommended to achieve this outcome.[108]

2.9 Solvation Calculations

Computational calculations often occur in the gas phase, since this is computationally less demanding.[38, 48] However, since most chemical reactions take place in solvents, it is important to model and predict the behaviour of the molecules in the liquid phase.[109] Since deposition of the donor-acceptor active layer can take place in solution, it is possible that the solvent may still be present after drying, and if present can affect the performance of the organic solar cells. The results from calculations in the solvent phase offer a good description of the properties and energetics of molecules in various solvents at costs comparable to gas phase calculations.[110] However, it is difficult to run calculations for long enough to mimic the behaviour of explicit solvent molecules in a real solution. Therefore, alternative approaches have been developed that treat the solvent implicitly as a continuum.[92]

2.10 Computational Studies of C_{60}

Semi-empirical methods are less time-consuming than density functional theory (DFT) methods. AM1 and PM3 are used as a starting point for geometry optimisation calculations of large molecules like C_{60} and can be followed by more accurate but slower calculations such as DFT. This combined approach is often used when calculating the electronic properties of C_{60}, such as the HOMO and LUMO energy levels, structure, reactivity and stability.[29, 93, 94]

Early calculations in 2000 on fullerenes were so computationally demanding that they relied on semi-empirical methods for energy calculations. AM1 was used for the full geometry optimisations of fulleropyrrolidines bearing the biologically active, derivatized, 1,4-dihydropyridine (**30**) as well as on monoimino-C_{60} (**29**). Single-point B3LYP/6-31G* and B3LYP/6-31G** energy calculations were then performed on the AM1-optimised structures for the 1,4-dihydropyridine and monoimino-C_{60} derivatives respectively. This was done in order to investigate the relative thermodynamic favourability of different routes for 1,3-dipolar cycloaddition reactions of the monoimino and 1,4-dihydropyridine appendages onto a C_{60} cage.[93, 94] AM1 was used for the full geometry optimisations of each of the reactants and products.

29   **30**
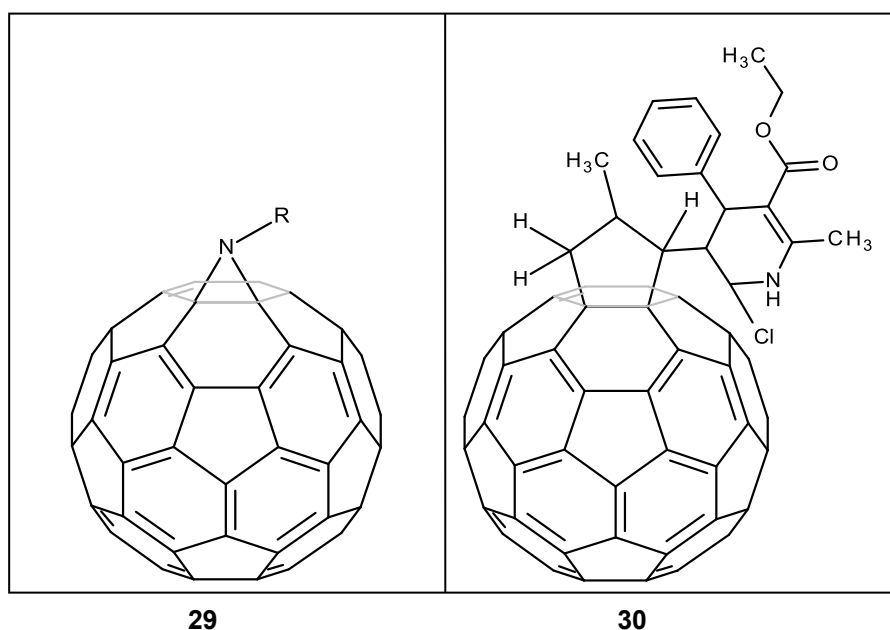
Figure 2.1 Structures of the monoimino-C_{60} (**29**) and 1,4-dihydropyridine-C_{60} (**30**) derivatives.

In 2008, PM3 was used for geometry optimisation prior to higher level single-point DFT calculations using B3LYP/3-21G* on C_{60}, [6,6]-phenyl-C_{61}-butyric acid methyl ester (PC$_{60}$BM) (**20**) and a range of PC$_{60}$BM derivatives. The calculations were carried out to find a relationship between the calculated LUMO energy levels and the efficiency of the organic solar cells.[29]

Where the electronic properties of C_{60} and its derivatives are required, larger polarised basis sets are required. Single-point calculations using B3LYP/6-31G* and 6-31G** were found to give sufficient accuracy when calculating the ionization potentials, electron affinity and total energy of C_{60}.[111] The formation energies of C_{60} derivatives *via* 1,3 dipolar cycloaddition have been studied using these basis sets.[93, 94] Reaction mechanisms of 1,3-dipolar cycloaddition to C_{60} have also been studied using these basis sets.

Despite the more representative data which can be calculated using 6-31G* and 6-31G** basis sets, they have not been used for full optimisation and property calculations on large structures like C_{60} and C_{60} derivatives. There have been very few publications at this level of theory because of the vast computational costs incurred.

SECTION A

SYNTHETIC METHODOLOGY

Chapter 3
Thymine

3.1 Introduction

Thymine (**32**) is a pyrimidine nucleobase which has been studied extensively in the biological[112, 113] and material science fields.[114, 115] It has been studied alongside adenine, its bonding base-pair, or on its own. Thymine binds to adenine *via* 2 H-bonds and can also bind to guanine, with the same number of bonds, in the Wobble binding mode.[73] So far, there has not been any published work regarding the functionalisation of thymine with C_{60}. The closest structure to thymine, whose interaction with C_{60} has been published, is uracil (**31**).[77, 81] Uracil is the un-methylated form of thymine, and is the complementary partner to adenine in RNA. It is not present in DNA.
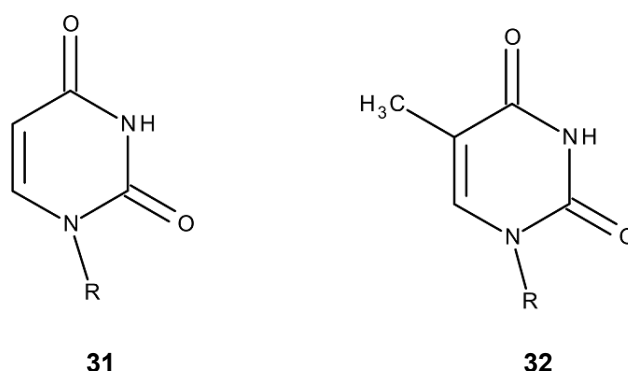
Figure 3.1 Diagrams showing the difference between uracil (**31**) and thymine (3**2**).

Uracil has been appended onto C_{60}, *via* the Prato reaction (**Scheme 3.1**),[77, 81] forming a N-methyl-2-uracilfulleropyrrolidine compound (**34**). In both studies, molecular modelling using DFT (B3LYP/3-21G* was used to complement the experimental work. A range of properties including H-bond distances were calculated in order to understand the base-pairing mechanism of N-methyl-2-uracilfulleropyrrolidine with adenine. This was important because

of the potential biological applications of fullerenes with receptor units that have high binding affinity.[77, 81]
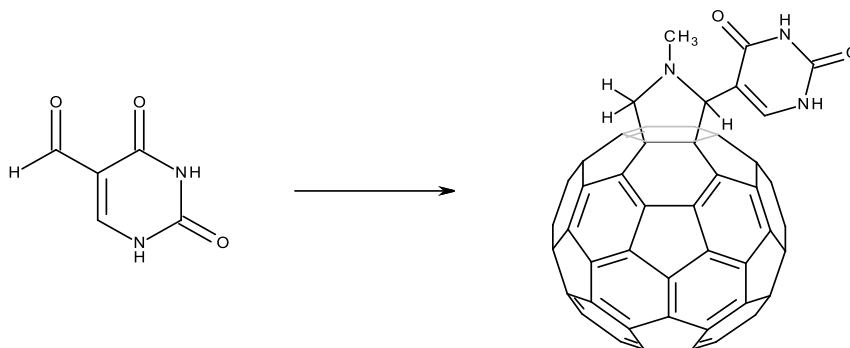
Scheme 3.1 Synthesis of N-methyl-2-uracilfulleropyrrolidine (**34**) using 5-formyl uracil (**33**) by the Prato reaction. Reagents: (i) C_{60}, N-methylglycine and toluene.

3.2 Experimental Results and Discussion

In this study, thymine has been used because the project is focused on DNA bases. The method used to synthesize the precursor to the intermediate (N-(2, 2-diethoxyethyl)thymine) (**35**) as well as the intermediate (2-(thymin-N-yl)ethanal) (**36**) was described by Ze-Qi *et al*. (**Scheme 3.2**).[116] The same method was used to synthesise the precursor and intermediate compounds for the adenine and cytosine base functionalised fullerene derivatives. A different method was used to synthesize the precursor and intermediate for the guanine base functionalised fullerene, which will be detailed in Chapter 6.

Synthesis of the novel N-methyl-2-thymin-N-ylfulleropyrrolidine (NMT-[60]) compound was a three-step process; the first was the *N*1-alkylation of thymine (**32**) to give the precursor, N-(2, 2-diethoxyethyl)thymine (**35**) as a thick yellow syrup which crystallised on standing.[116] The crystals were hydrolysed using dilute acid to produce 2-(thymin-N-yl)ethanal (**36**) as a white powder (**Scheme 3.2**). The final product was a result of the 1,3 dipolar cycloaddition of C_{60}

with 2-(thymin-N-yl)ethanal in toluene, using the Prato reaction (**Scheme 3.3**).[17] The final product was purified using column chromatography yielding the product as a brown powder with a 32 % yield.

3.2.1 Synthesis scheme of the precursor and the intermediate compounds to the cycloaddition reaction

32   **35**   **36**
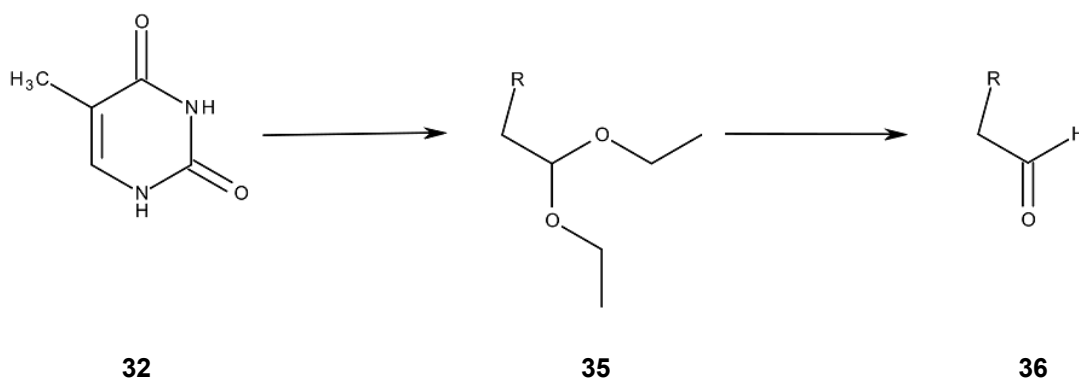
Scheme 3.2 Synthesis scheme of the precursor (**35**) and the intermediate (**36**) to the cycloaddition reaction. R = thymine, adenine, cytosine.

3.2.2 Synthesis Scheme of N-methyl-2-thymin-N-ylfulleropyrrolidine (NMT-[60])

36   **37**
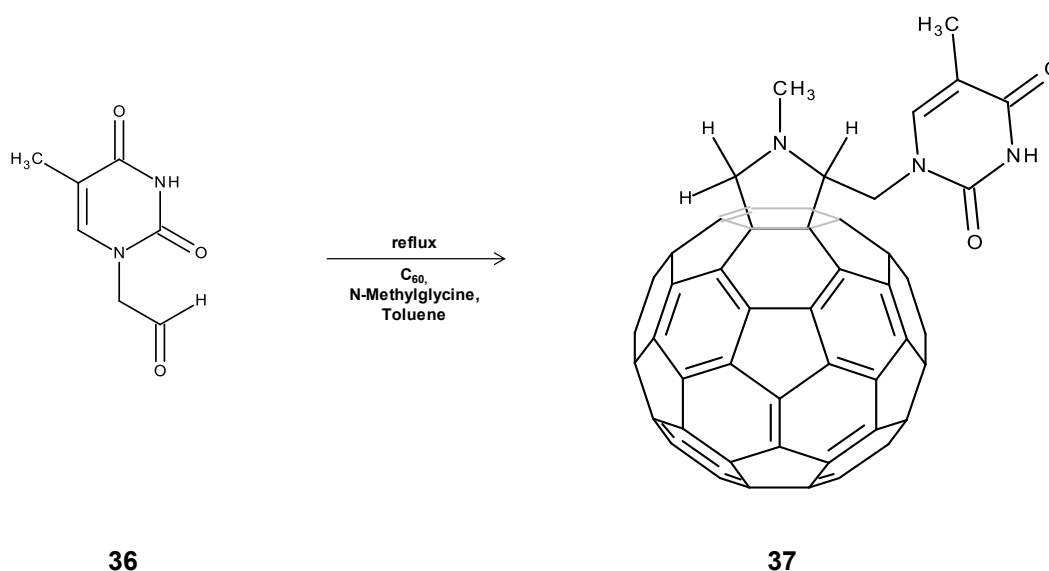
Scheme 3.3 Synthesis scheme of NMT-[60] (**37**) using the Prato reaction in toluene.

3.2.3 NMR and Mass Spectrometry of intermediates N-(2, 2-Diethoxyethyl)thymine and 2-(thymin-N-yl)ethanal

The *N*1-alkylated thymine product (**35**) was confirmed by the presence of a triplet signal at 1.07 ppm from the two ethyl groups, along with the pair of multiplets between 3.43-3.49 ppm and 3.61-3.69 ppm from the methylene groups (CH_2) in the ^1H NMR spectrum (**Figure 3.2** (**a**)), corresponding to literature data.[116] A single triplet, instead of two triplets, are observed for the two ethyl groups because they were in the same environment. The ^{13}C NMR spectrum had nine signals, where six were assigned to thymine and the ethoxy ethyl signals were present at 49.37, 62.47 and 99.32 ppm. DEPT 135 NMR spectroscopy confirmed the presence of the methylene groups by the inversion of the peaks at 49.37 ppm and 62.47 ppm in the spectrum. The signal at 62.47 ppm was assigned to the CH_2 atoms of the ethyl group because oxygen is more electronegative than nitrogen. Therefore, the adjacent carbon atom is de-shielded more than the carbon atom adjacent to nitrogen (NCH_2). Mass spectrometry analysis produced a molecular ion (M^+) at 243.0100 Da for (**35**) corresponding to the correct formula $C_{10}H_{18}N_4O_4$.

After hydrolysis, the corresponding ^1H NMR spectrum showed a carbonyl peak at 9.54 ppm coupled with the loss of the methylene multiplets between 3.43 to 3.69 ppm, from the ethoxy groups, and the presence of a new methylene peak at 4.58 ppm confirmed the formation of 2-(thymin-N-yl)ethanal (**Figure 3.2** (**b**)). The downfield shift of the methylene peak provides good evidence for the presence of the electron withdrawing carbonyl group. Mass spectrometry data provided a signal (M^+) for **36** at m/z = 169.0626 Da, corresponding to the required formula $C_7H_8N_2O_3$.
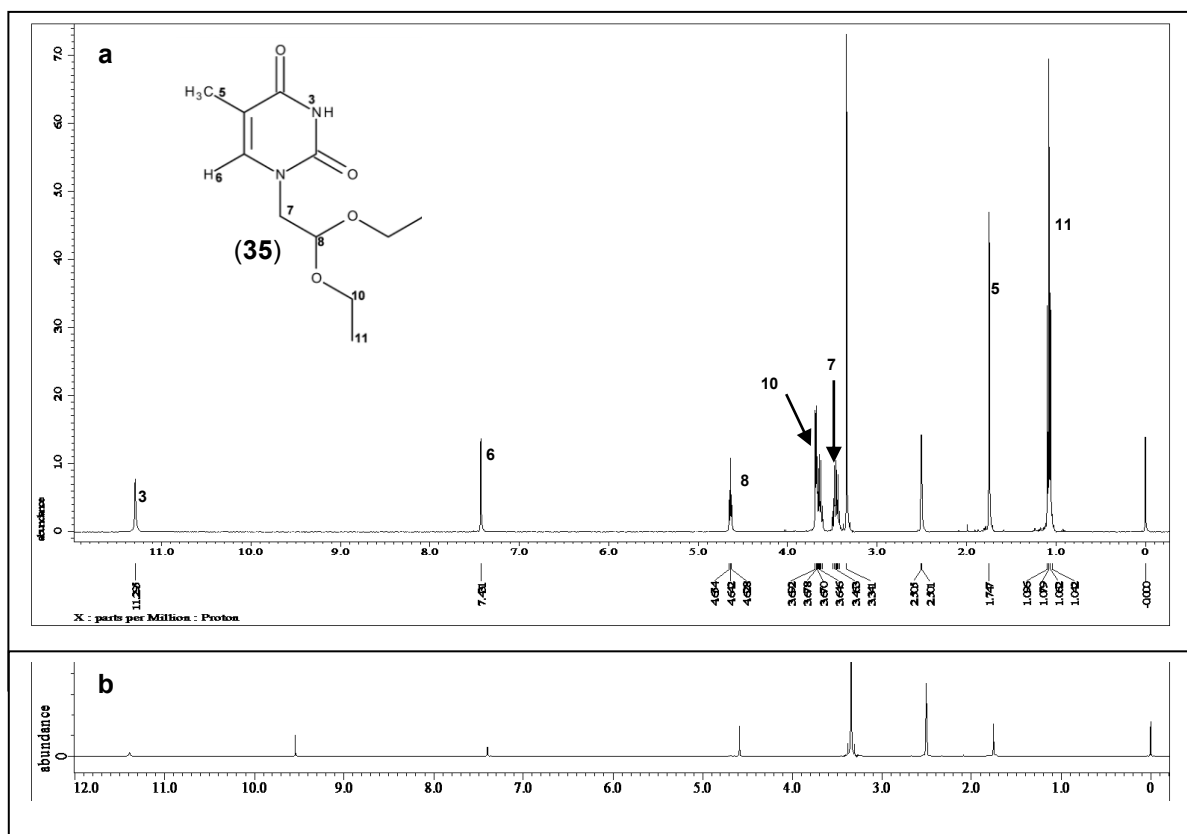
Figure 3.2 ^{1}H NMR spectrum of N-(2, 2-diethoxyethyl)thymine (**35**)(**a**) and 2-(thymin-N-yl)ethanal (**36**) (**b**) showing the transformation of the compound after hydrolysis with 1 M HCl.

3.2.4 NMR and Mass Spectrometry of NMT-[60]

Confirmation of fulleropyrrolidine formation was obtained from the ^{1}H NMR spectrum which had two doublets at 4.29 ppm (J = 10.07 Hz) and 4.84 ppm (J = 10.07 Hz) (**Figure 3.3**). These doublets were assigned to the two protons on the pyrrolidine ring, H[a] and H[b] respectively because the coupling constants were typical of ortho benzylic protons. It was observed that the signal for one proton (H[b]) was further downfield compared to the other (H[a]), even though they appear to be in the same environment. This was because the proton H[b] lies above a five-membered ring on the fullerene cage, whilst proton H[a] was above a six-membered ring. These environments were confirmed by the modelled structure of NMT-[60] (**Figure 3.3**). The ^{1}H NMR proton shifts correspond to published data which predicts two types of ring currents in the C_{60} cage.[18] The pentagons are thought to have strong

paramagnetic ring currents while the hexagons have mild diamagnetic currents.[117] Therefore, a proton located above a pentagon would be more de-shielded and would appear further downfield in an ^{1}H NMR spectrum compared to one located above a hexagon. The triplet at 4.55 ppm and the doublet at 4.69 ppm also support the formation of N-methyl-2-thymin-N-ylfulleropyrrolidine. These two signals were particularly important because they confirm the connection of thymine to the pyrrolidine ring on the C_{60} cage, as the latter signal (H[d]) was assigned to the methylene protons linking thymine and the five-membered ring and the former was a signal from the neighbouring proton (H[c]) on the pyrrolidine ring (**Figure 3.3**). Mass spectrometry analysis of NMT-[60] produced a molecular ion (M) signal 915.1000 Da corresponding to the formula $C_{69}H_{13}N_3O_2$ as required.
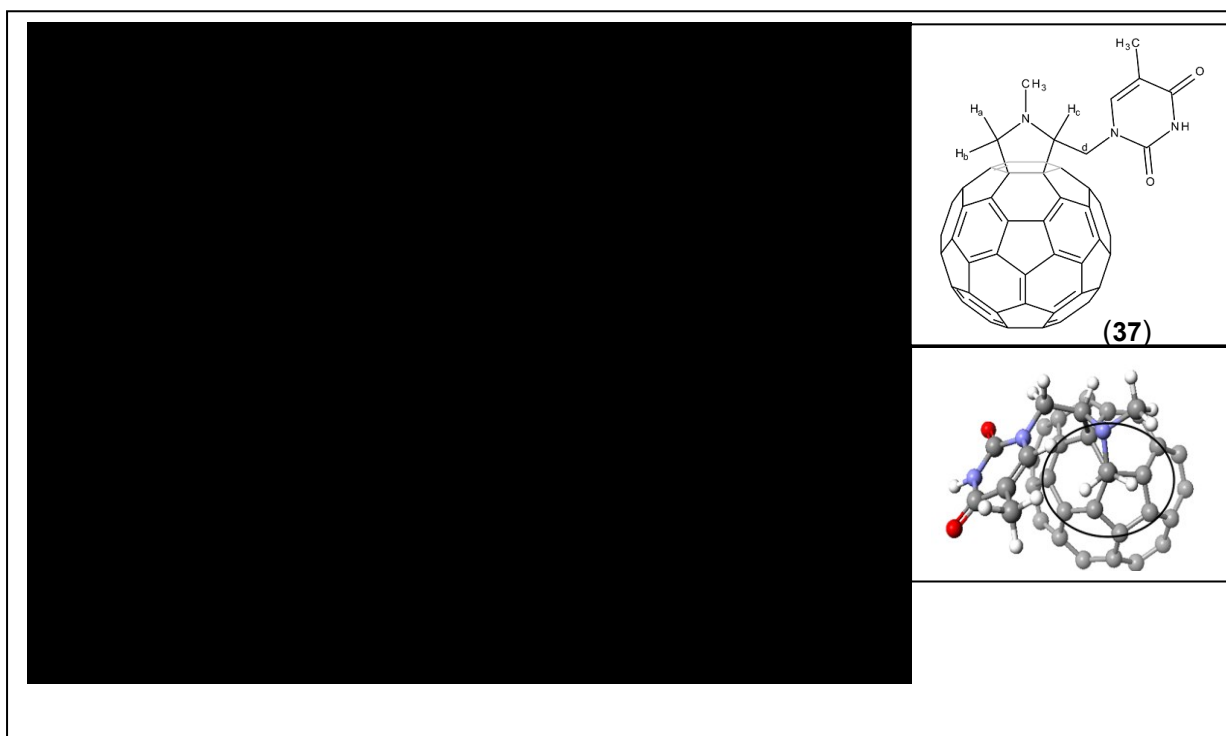

Figure 3.3 The pyrrolidine region of the ^{1}H NMR spectrum for NMT-[60] (**37**) showing the signals which confirm the formation on the derivative and the coupling constants. Insert: Structure of NMT-[60] showing the pyrrolidine H[b] and H[a] protons above the 5- and 6-membered rings on the C_{60} cage. Part of the cage has been deleted for clarity.

^{13}C NMR spectroscopy of NMT-[60] was not successful because of the insolubility of the fullerene compound. Several thousand runs were carried out in an attempt to obtain a spectrum but they were unsuccessful.

3.2.5 Infrared Spectroscopy of NMT-[60]

Successful cycloaddition of 2-(thymin-N-yl)ethanal to C_{60}, with a 32 % yield, was confirmed using infrared spectroscopy (**Figure 3.4**). Comparison of the spectrum for NMT-[60] with that for C_{60} shown in **Figure 3.4** (**b**) also shows the splitting of the fullerene signal at 574.84 cm^{-1} to smaller peaks, one observed at 593.60 cm^{-1} (**Figure 3.4** (**a**)). This change in the signal confirms the formation of a mono-adduct derivative.[6] The characteristic fullerene signals at 527.00 cm^{-1}, 1184.75 cm^{-1} and 1429.00 cm^{-1} are also present in the NMT-[60] spectrum, confirming the formation of the derivative (**Figure 3.4**). The strong IR peak at 1571.89 cm^{-1}, from the thymine carbonyl, confirmed the presence of thymine in the compound. In addition, the signals at 2922.77 cm^{-1} and 2852.03 cm^{-1} which represent the aromatic CH stretches were also present. The relatively low yield of the reaction is in line with published literature on fulleropyrrolidine derivatives.[17, 118]
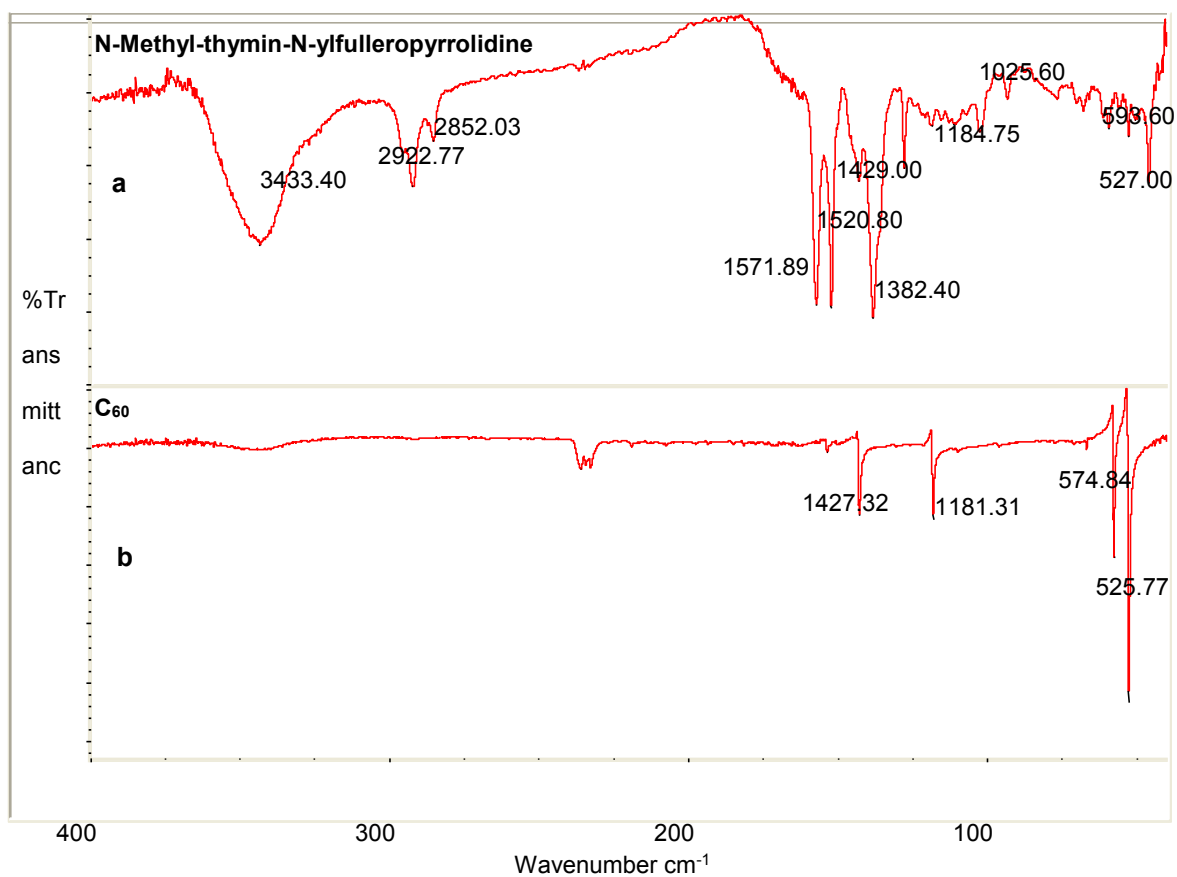
Figure 3.4 Comparison of the infrared spectra of NMT-[60] (**a**) and C$_{60}$ (**b**).

3.2.6 UV-Vis Spectroscopy of NMT-[60]

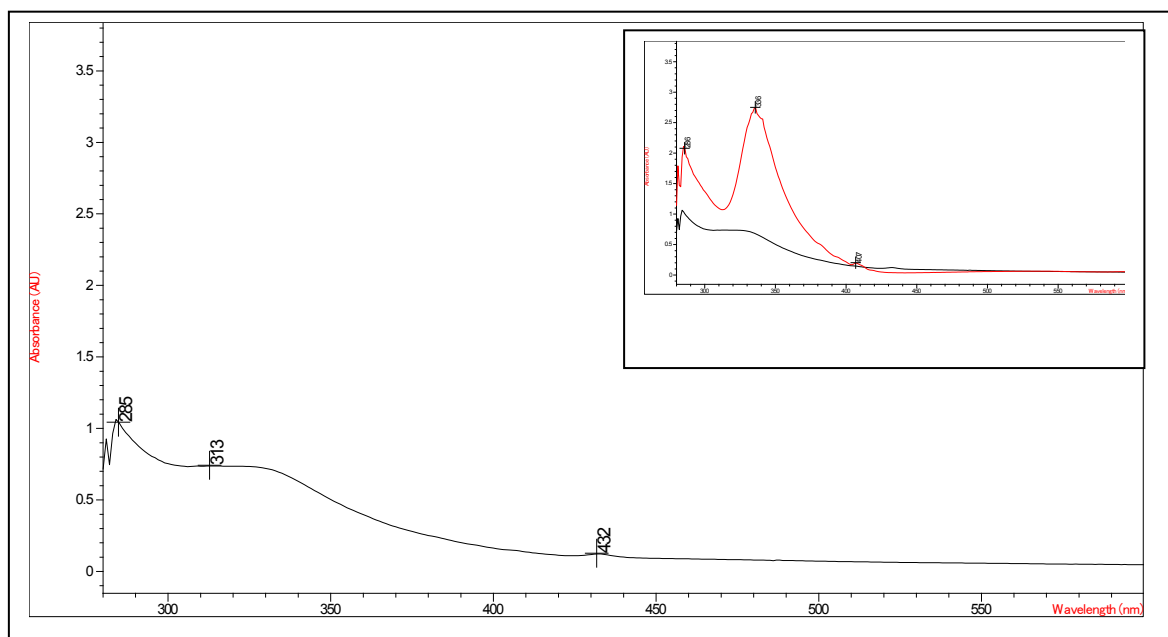The UV-Vis spectrophotometry of NMT-[60] produced peaks at 285 nm, 313 nm and 432 nm.

Figure 3.5 UV-Vis spectrum of NMT-[60] recorded in chloroform and the UV-Vis spectrum of NMT-[60] (black) and C$_{60}$ (red) (inset).

The small peak at 432 nm is characteristic of mono-adducts, and so confirms the formation of a mono-adduct.[17] **Figure 3.5** also shows the UV spectra of C_{60} and NMT-[60] overlaid and confirms that the derivative retains some of the electronic properties of C_{60}, such as the peaks at 285 and 313 nm.[17] This feature is one of the reasons why mono-adducts of fullerenes are attractive starting materials for a range of applications. A fullerene derivative is capable of exhibiting the electron accepting properties of pristine C_{60}, whilst possessing the new properties of the new derivative, in this case NMT-[60].

3.2.7 Thermogravimetric Analysis of NMT-[60] + zeolite Y

To test the use of zeolites as a template for self-assembly, the initial step was to adsorb C_{60} and the NMT-[60] monomer as test compounds onto the surface of zeolite Y which has narrow channels (7.4 Å in diameter). To assess whether NMT-[60] was adsorbed onto zeolite Y, thermogravimetric analysis (TGA) and X-ray diffraction (XRD) were used (**Table 3.1**).

TGA results of pristine C_{60} showed that the compound started decomposing from around 300 °C to 550 °C, with major decomposition between 350 °C and 500 °C.[119] C_{60} + zeolite Y showed a decomposition peak from around 500 °C to 550 °C. NMT-[60] adsorbed on the zeolite exhibited a major decomposition curve, though not as pronounced as that of C_{60}, from around 450 °C to 550 °C.

Structure	Temperature Range °C From	To	Weight loss of sample (mmoles)
C_{60}	300	550	
C_{60} + zeolite	500	550	0.14
NMT-[60] + zeolite	450	550	0.26
Zeolite	No decomposition observed		

Table 3.1 Thermogravimetric analysis and decomposition temperature ranges for C_{60} and the C_{60}/NMT-[60] + zeolite composites.

After refluxing NMT-[60] with the white zeolite, the resulting composite was the characteristic brown colour of fullerene derivatives. This indicated the presence of NMT-[60] adsorbed onto the zeolite framework. Soxhlet extraction, with toluene as the refluxing solvent, was used to wash the unadsorbed NMT-[60].

The number of moles of weight loss observed for NMT-[60] with zeolite Y was higher than that observed for pristine C_{60} with zeolite Y (**Table 3.1**). These results signal the synthesis of a novel composite as all previous studies have focused on pristine C_{60} and its association with a range of zeolites.[63, 65, 67, 68, 71, 120]

3.2.8 X-ray Diffraction

The X-ray diffraction spectrum of NMT-[60]/zeolite composite shows a strong C_{60} peak at around 17.2 2θ. This fullerene peak has been well documented as the peak between 15 - 20 2θ. There were no thymine peaks observed in the spectrum therefore this data did not confirm the presence of NMT-[60] in the zeolite. However, it is possible that the thymine signals were shielded by the strong zeolite signals. Pristine C_{60} with the zeolite, did not exhibit any fullerene peaks.[121, 122] From this data coupled with the weight loss observed in the TGA results for C_{60} + zeolite, it is possible that only a small amount of pristine C_{60} was adsorbed on the zeolite and the C_{60} peak was suppressed by the strong zeolite signals. Unfortunately, these characterisation methods and results do not give an indication of the position or orientation of NMT-[60]. However, it is assumed that the derivative rests at the pore openings because the C_{60} cage is larger than the zeolite pores. Further analysis such as high resolution scanning electron microscopy (SEM) or transmission electron microscopy (TEM) would be required to confirm the orientation of the fullerene derivative.

3.3 Conclusion

The successful synthesis of a novel C_{60} derivative with the potential to self-assemble when coupled with an adenine-functionalised C_{60} derivative has been described. UV-Vis absorption spectroscopy and IR spectroscopy confirmed the formation of a mono-adduct with the characteristic fullerene mono-adduct peak at 432 nm and the splitting of the signal at 574.84 cm^{-1} in C_{60} to 593.60 cm^{-1} and other small signals in NMT-[60]. The newly synthesized of N-methyl-2-thymin-N-yl-fulleropyrrolidine has been successfully adsorbed onto zeolite Y. Adsorption was confirmed by thermogravimetric analysis.

3.4 Experimental Section

N-(2, 2-Diethoxyethyl)thymine (35). Dried thymine (4.19 g, 33.2 mmol) and K_2CO_3 (4.60 g, 33.3 mmoles) were mixed with bromoacetaldehyde diethyl acetal (9.90 g, 50.2 mmol) in DMF (38 cm^3) and refluxed at 130 °C for 10 hours. The solids were filtered whilst still hot and washed with chloroform. The organic phase was concentrated by evaporation and triturated with ether: petroleum ether (1:2, 10 cm^3) to give the required compound, N-(2, 2-diethoxyethyl)thymine (3.62 g, 44 %) as a thick syrup which crystallised on standing. m.p 100-104 °C. **FT-IR** (KBr) v_{max}/cm^{-1} 3200 cm^{-1}, 2900 cm^{-1}, 1676-1709 **¹H NMR** (DMSO) δ 1.06 (t, 6, CH_3 J = 13.4 Hz), 1.74 (s, 3, CH_3 of thymine), 3.43-3.49 (m, 2, CH_2 J = 16.4 Hz), 3.61-3.69 (m, 4, CH_2 J = 25.6 Hz), 4.64 (t, 1, CH of acetal J = 5.5 Hz), 7.43 (s, 1, H_6), 11.29 (s, 1, NH); **¹³C NMR** (DMSO) δ 11.73 (thymine), 15.12 (CH_3 of Et), 49.37 (NCH_2), 62.47 (CH_2 of Et), 99.32 (CH of acetal), 107.79, 142.25, 150.91 and 164.14 (thymine). **UV-VIS** λ_{max} ($CHCl_3$) 424, 432 nm. **Acc. mass** (HRES⁺) M⁺ calc. m/z = 242.2739 Da, measured m/z = 243.0134 Da.

2-(Thymin-*N*-yl)ethanal (36). N-(2, 2-Diethoxyethyl)thymine (0.37 g,1.53 mmol) was stirred in 1 M HCl (6.5 cm^3) at 100 °C for 45 minutes. The clear solution was evaporated under vacuum to give a white solid. The white solid was dried at 100 °C for 24 hours to give the required product, 2-(thymin-*N*-yl)ethanal. m.p 185-188 °C. **FT-IR** v_{max}/cm^{-1} 2892 cm^{-1},1652-

1701 cm^{-1}. ^{1}**H NMR** (DMSO) δ 1.75 (s, 3, CH$_3$), 4.58 (s, 2, CH$_2$), 7.40 (s, 1, H$_6$), 9.54 (s, 1, CH=O) and 11.39 (s, 1, NH); ^{13}**C NMR** (DMSO) δ 12.44 (thymine), 57.04 (CH$_2$), 109.00, 142.20, 151.54, 164.87 (thymine), 198.10 (CH=O). **UV-VIS** λ$_{max}$ (MeOH) 359 nm. **Acc. mass** (HRES$^+$) M$^-$ calc. m/z = 168.0530 Da, measured m/z = 167.0461 Da.

N-methyl-2-thymin-N-ylfulleropyrrolidine (**37**). C$_{60}$ (0.20 g, 0.28 mmol)**,** 2-(thymin-*N*-yl)ethanal (0.15 g, 0.96 mmol) and N-methylglycine (0.062 g, 0.70 mmol) were stirred in toluene (180 cm^3) and refluxed at 60°C for 12 hours. After cooling, the reaction mixture was purified by column chromatography with toluene then toluene/ethyl acetate [8:2] to give the product NMT-[60] (0.0642 g, 32 % yield). **FT-IR** v_{max}/cm^{-1} (KBr) 2927 cm^{-1} 2852 cm^{-1}, 1571 cm^{-1}, 1429 cm^{-1} 1184 cm^{-1} 593 cm^{-1} 527 cm^{-1}. ^{1}**H NMR** (CDCl$_3$:CS$_2$ [1:2]) 1.93 (s, 3, CH$_3$, thymine), 3.04 (s, 3, NCH$_3$), 4.29 (d, 1, H *J* = 10.07 Hz), 4.55 (t, 1, H *J* = 11.45 Hz), 4.70 (d, 2, CH$_2$ *J* = 5.95 Hz), 4.84 (d, 1, H *J* = 10.07 Hz), 8.43 (s, 1, H) **UV-VIS** λ$_{max}$ (chloroform) 284, 432 nm. **Acc. mass** (HRES$^+$) M calc. m/z = 915.8604 Da measured m/z = 915.1000 Da.

Zeolite Y + C$_{60}$. C$_{60}$ (0.01 g, 0.01 mmol) and zeolite Y (0.70 g) were refluxed in toluene (20 cm^3) for 4 days. The reaction mixture was cooled to room temperature and purified by soxhlet extraction. The composite was collected and dried, first at room temperature, then at 200 °C for 48 hours.

Zeolite Y + N-methyl-2-thymin-N-ylfulleropyrrolidine. C$_{60}$ (0.01 g, 0.01 mmol) and zeolite Y (0.70 g) were refluxed in toluene (20 cm^3) for 4 days. The reaction mixture was cooled to room temperature and purified by soxhlet extraction. The composite was collected and dried, first at room temperature, then at 200 °C for 48 hours.

Chapter 4
Adenine

4.1 Introduction

Adenine is one of the four bases in DNA and is involved in base-pairing with thymine.[123] It is made up of a six-membered ring attached to a five-membered ring and is thus a member of the purine family.[124] Adenine binds to thymine *via* 2 H-bonds and can take part in the Hoogsteen binding mode with thymine, with the same number of H-bonds.[73]

To date, there has only been a very limited amount of research into the interactions of adenine with C_{60}. In 2009, a fullerene-modified electrode was successfully used to analyse adenine in human blood plasma and urine samples.[124] C_{60} was cast onto the glassy carbon electrode in a solution *via* a microsyringe then dried in hot air to produce microcrystals. One of the key reasons why C_{60} was chosen as an electrode-surface modifier is that it exhibited good electrochemical stability.[124]

Krainara *et al.*'s DFT study probed the stability and electronic structure of adenine and its unusual tautomer.[123] In addition, Boron-Nitrogen (BN)-doped fullerene ($C_{58}BN$) was used as a sensor to discriminate between the two Cu (II) complexes of adenine-Cu (II) and its tautomer-Cu (II).[123]

In this chapter, the possibility of attaching an adenine group to C_{60} will be explored. Since adenine is the base pair for thymine, we are interested in the self-assembling abilities of the DNA-based fullerene derivatives.

4.2 Experimental Results and Discussion

4.2.1 Synthesis of N-(2, 2-diethoxyethyl)adenine and 2-(adenin-N-yl)ethanal

Synthesis of N-(2, 2-diethoxyethyl)adenine and 2-(adenin-N-yl)ethanal, the precursor and intermediate to the cycloaddition of NMA-[60], followed **Scheme 3.2**.[116] N-(2, 2-diethoxyethyl)adenine was produced as a white solid after purification using column chromatography.[116] N-(2, 2-diethoxyethyl)adenine was then hydrolysed in dilute acid and after evaporation under vacuum, the product was washed with 50 % ethanol, ether and dichloromethane, producing 2-(adenin-N-yl)ethanal as an impure yellow solid. In order to purify 2-(adenin-N-yl)ethanal, the solubility of the compound was tested in a range of solvents and solvents mixtures including chloroform, ethyl acetate, methanol/dichloromethane and ethyl acetate/hexane. However, 2-(adenin-N-yl)ethanal was insoluble in these solvents and protection with BOC was explored to increase its solubility and enable the production of the pure aldehyde.

4.2.2 NMR and Mass Spectrometry of N-(2, 2-diethoxyethyl)adenine

Successful alkylation of adenine to produce N-(2, 2-diethoxyethyl)adenine (**38**) was initially confirmed by ^{1}H, ^{13}C and DEPT 135 NMR spectroscopy. ^{1}H NMR spectroscopy exhibited the single triplet signal from the two methyl groups at 1.02 ppm as well as a doublet from the methylene protons on the acetal at 4.21 ppm. Crucially, two singlets at 8.05 ppm and 8.14 ppm were also present from the H8 and H2 respectively. These signals confirmed that alkylation was successful on adenine. Similarly, the ^{13}C NMR spectrum had five signals from adenine as well as the expected four from the acetal group.

Two signals at 47.5 ppm and 62.7 ppm were inverted in the DEPT 135 NMR, signalling their *sp³* hybridization and the presence of two protons. These signals were from the methylene group next to the N9 on the acetal (NCH$_2$) and the two methylene groups next to the oxygen atoms also on the acetal group (-OCH$_2$CH$_3$), respectively. It was clear that both methylene groups experience de-shielding effects from the neighbouring nitrogen and oxygen atoms.

Since oxygen is more electronegative than nitrogen, the methylene protons next to oxygen are further upfield than those next to N. Mass spectrometry produced a molecular ion at 252.1469 Da, corresponding to the required formula $C_{11}H_{18}N_5O_2$ for (**38**).
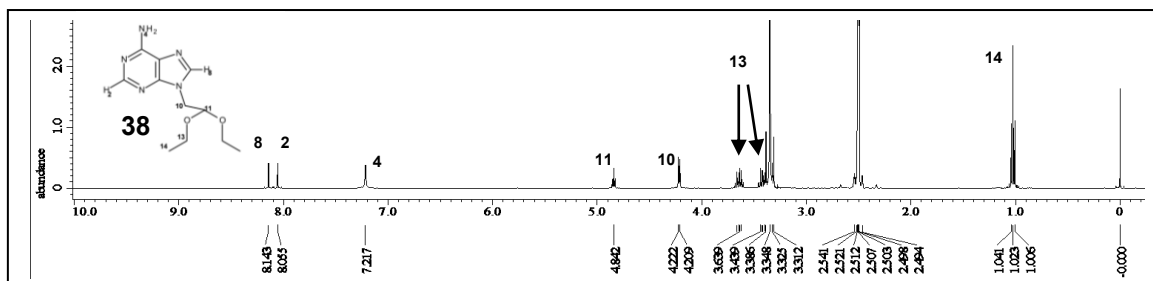

Figure 4.1 ^{1}H NMR spectrum of N-(2, 2-diethoxyethyl)adenine (**38**) showing the assignments of the signals.

4.2.3 NMR and Mass Spectrometry of 2-(adenin-N-yl)ethanal

The ^{1}H NMR spectrum of the product from the hydrolysis of N-(2, 2-diethoxyethyl)adenine (**38**) showed the resulting compound as a mixture of two compounds (**39a**) and (**39b**). **Figure 4.2** showed a signal at 9.70 ppm for the aldehyde proton and this corresponded to two of the four singlet signals in the aromatic region, at 8.37 ppm and 8.47 ppm for the H2 and H8 protons on adenine. Furthermore, there was another singlet at 5.37 ppm, which was assigned to the deshielded methylene protons. This combination of signals was assigned to 2-(adenin-N-yl)ethanal compound (**39a**).

The additional signals in the spectrum were two singlets at 8.42 ppm and 8.50 ppm for the H2 and H8 adenine protons as well as a doublet at 4.16 ppm and a triplet signal at 5.11 ppm. The doublet was assigned to the methylene signals and the proton, which was deshielded by the OH groups, was observed as the triplet. These signals matched 2-adenine diol (**39b**)

Mass spectrometry of this mixture showed a molecular ion (m/z = 196.0841 Da) which corresponded to 2-adenine diol (**39b**) with the formula $C_7H_9N_5O_2$ and another ion (m/z =

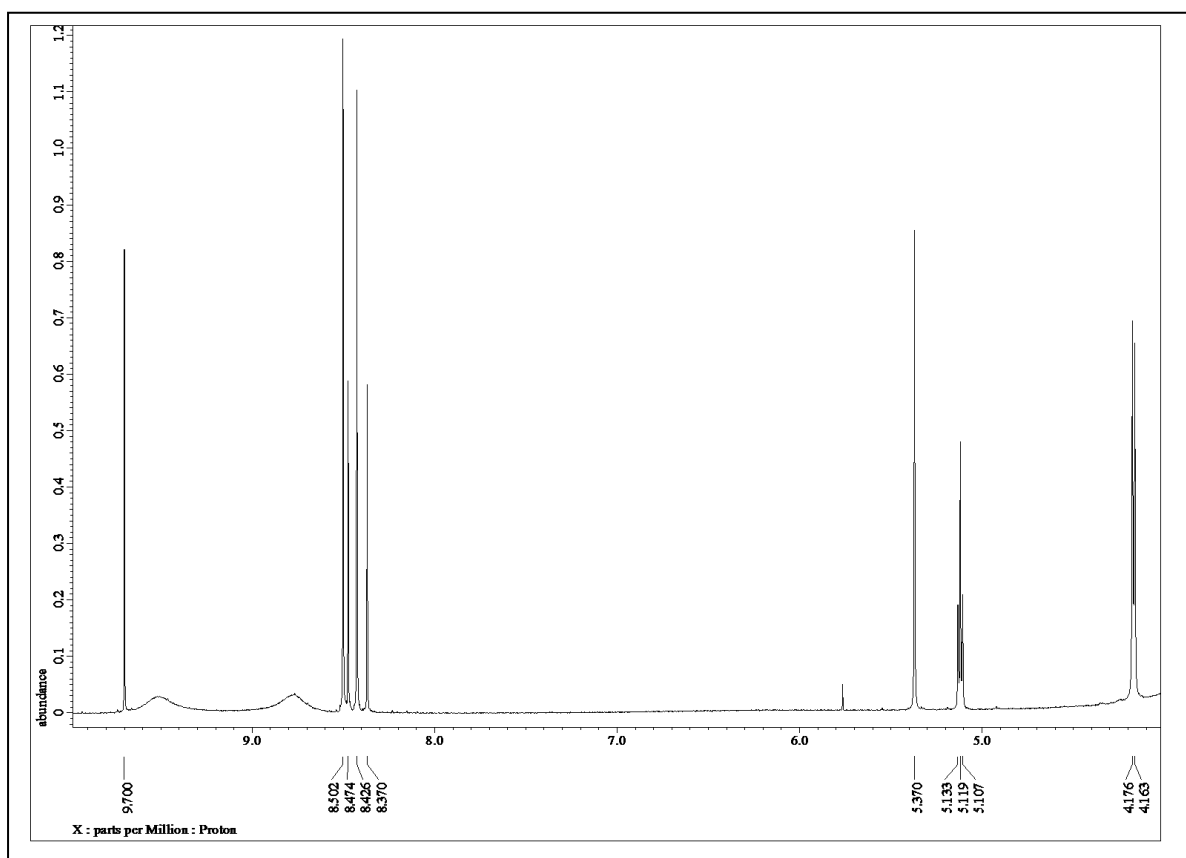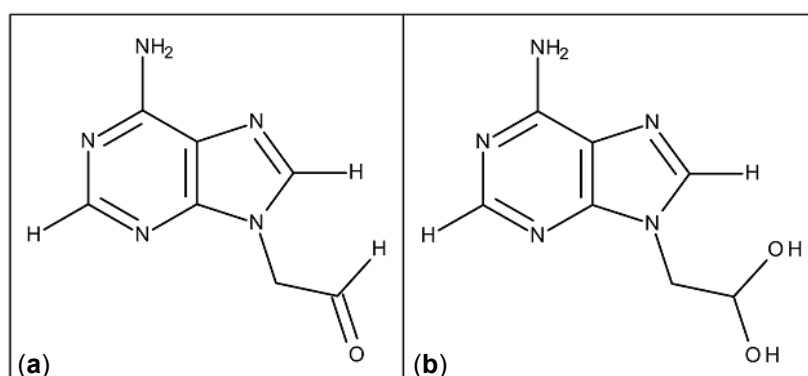178.0723 Da), which corresponded to the required formula $C_7H_7N_5O$ for 2-(adenin-N-yl)ethanal (**39a**).

Figure 4.2 Part of the ^{1}H NMR spectrum of the mixture of 2-(adenin-N-yl)ethanal and 2-adenine diol (**39**).

(**39**)

 Figure 4.3 Structures of the compounds in the mixture (a) 2-(adenin-N-yl)ethanal and (b) 2-adenine diol

### 4.2.4 Increasing the solubility of adenine analogues

After synthesis of N-(2, 2-diethoxyethyl)adenine (**38**), hydrolysis of this compound to make 2-(adenin-N-yl)ethanal (**39**) resulted in an impure compound which was insoluble in toluene, and 1,2-dichlorobenzene and could therefore not be purified. Since it was possible that covalent bond formation between the 6-amino group of adenine and the carbonyl of the N9 substituent were the cause of the insolubility, it was decided to protect the 6-amino group with di-*tert*-butyl dicarbonate (**Scheme 4.1**).

BOC was selected because it not only increases solubility, but also stops covalent bonding between the 6-amino group of adenine and the carbonyl of the N9 substituent. Other alternative protection methods such as benzoylation[125, 126] were considered but BOC-protection was thought to be the most useful here because of the feasibility of removing the protecting group without any side-reactions on the fullerene cage. In addition, C_{60} is known to be stable towards acids and BOC is acid labile.[127]

Not only is it acid-labile, but BOC can also be removed under neutral conditions.[125, 128, 129] The BOC group is stable to hydrogenation, and often produces high yields of BOC-protected products of approximately 80-99 %.[128, 129] BOC-protection has been carried out in solution[125] and under solventless conditions.[130] Although the solventless protocol produced relatively high yields, between 70-99 % for the majority of the reactions, the reactions in solution were chosen because this route is more established in the literature.

BOC-protected products have excellent solubility. They have decreased self-aggregation due to a lack of H-bond donors and they often do not require elaborate purification steps.[125, 131] It is often possible to remove the BOC groups from a compound selectively. For example, converting tris-*N*-BOC-adenine to bis-*N*-BOC-adenine and subsequently mono-*N*-BOC-adenine, without the use of an acid, under neutral conditions is possible and gives good yields.[128]

The route to protect adenine (**40**) is given in **Scheme 4.1**.
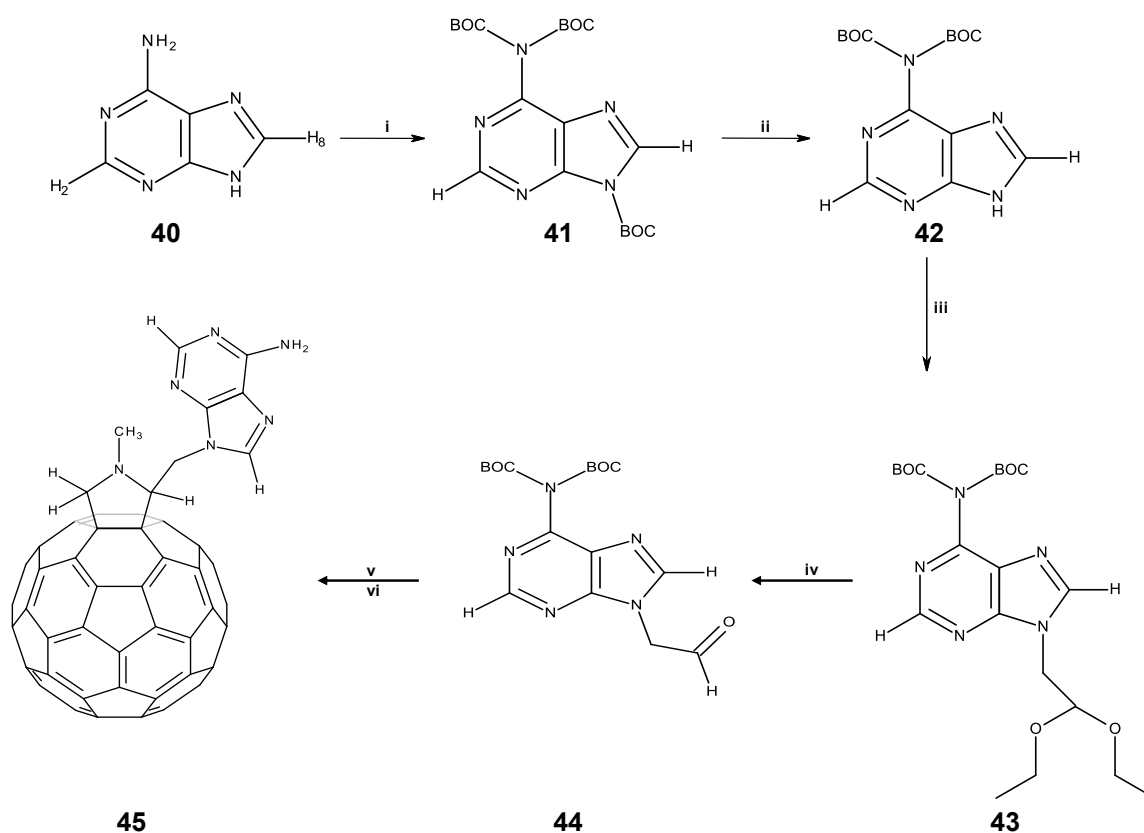
Scheme 4.1 Proposed synthesis scheme for N-methyl-2-adenin-N-ylfulleropyrrolidine (**45**) *via* tris-boc-adenine (**41**). Reagents: (i) BOC, DMAP, THF; (ii) NaHCO$_3$, MeOH, H$_2$O, reflux; (iii) K$_2$CO$_3$, bromoacetaldehyde diethyl acetal, DMF (iv) 1 M HCl (v) C$_{60}$, N-methylglycine, 1,2-dichlorobenzene (vi) TFA.

Since N-(2, 2-diethoxyethyl)adenine had already been successfully synthesized, steps i-iii (**Scheme 4.1**) were skipped and the intermediate (**38**) was protected with di-*tert*-butyl dicarbonate, producing (**43**).

4.2.5 Protection of N-(2, 2-diethoxyethyl)adenine with BOC - Method A (i)

N-(2, 2-diethoxyethyl)adenine (**38**) was mixed with THF, DMAP and BOC under nitrogen and stirred at 60 ˚C, following the literature method for protecting adenine,[129] yielding bis-BOC-N-(2, 2-diethoxyethyl)adenine as a white solid. The successful protection of N-(2, 2-

diethoxyethyl)adenine (**Scheme 4.2** (**38**)) was evident in the ^{1}H NMR spectrum, where a strong BOC peak was present at 1.46 ppm. The remainder of the signals were the same as those for N-(2, 2-diethoxyethyl)adenine. ^{13}C NMR spectroscopy of the compound showed the signal for BOC at 27.7 ppm as well as the signals N-(2, 2-diethoxyethyl)adenine as outlined in Section 4.2.2. High-resolution mass spectrometry analysis produced the molecular ion as a monomer (M$^+$) at 452.2500 Da, corresponding to the correct formula of (**44**) $C_{21}H_{33}N_5O_6$.

38   **44**
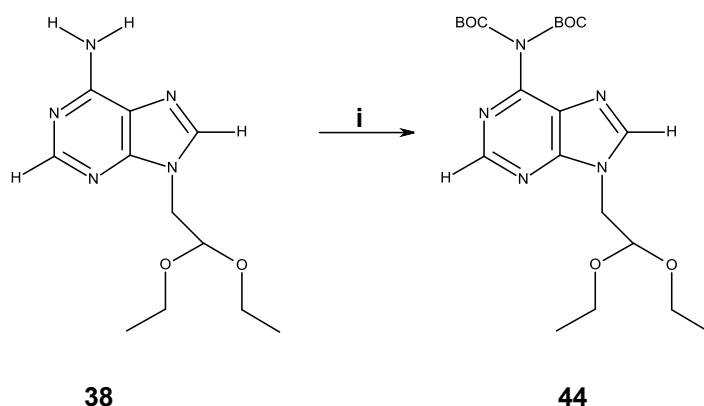
Scheme 4.2 Reaction scheme for the protection of N-(2, 2-diethoxyethyl)adenine (**38**) with BOC. (i) = di-tert butyl dicarbonate, DMAP and THF under N_2.

4.2.6 Synthesis of bis-BOC-2-(adenin-N-yl)ethanal - Method A (ii)

In order to hydrolyze bis-BOC-N-(2, 2-diethoxyethyl)adenine (**44**), the white solid was stirred in 1 M HCl as described by Xu *et al.* for N-(2, 2-diethoxyethyl)adenine.[116] However, upon characterisation, ^{1}H NMR spectroscopy showed that the BOC group had been removed by the acid, therefore alternative methods were used to synthesize BOC-protected 2-(adenin-N-yl)ethanal.

4.2.7 Protection of the 2-(adenin-N-yl)ethanal/2-adenine diol mixture with BOC - Method B

After the failure of **Method A** (**ii**) a second method, **Method B** (**Scheme 4.3**) was used. As a test reaction, the mixture of 2-(adenin-N-yl)ethanal/2-adenine diol (**39**) was protected with BOC under the same conditions as those used for adenine protection.[129]

39   **46**
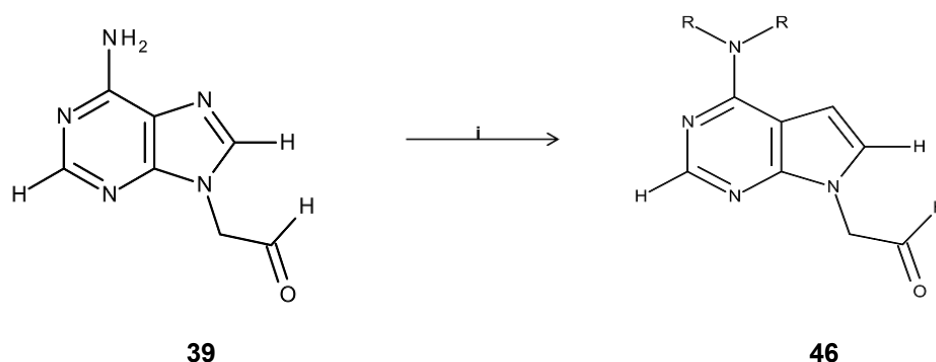
Scheme 4.3 Synthesis scheme for the protection of the 2-(adenin-N-yl)ethanal/2-adenine diol mixture with BOC. R = BOC or H. (i) = di-tert butyl dicarbonate, DMAP and THF under N_2.

4.2.8 Cycloaddition of the BOC-protected 2-(adenin-N-yl)ethanal/2-adenine diol mixture

The product from the BOC-protection of the 2-(adenin-N-yl)ethanal/2-adenine diol mixture was refluxed with N-methylglycine and C_{60} in 1,2-dichlorobenzene. The reaction was monitored by TLC and stopped when the change from the characteristic purple colour of C_{60} to the brown colour of fullerene derivatives was observed.

4.2.9 NMR spectroscopy of the cycloaddition product

The ^1H NMR spectrum of the cycloaddition product of C_{60} with the BOC-protected 2-(adenine-N-yl)ethanal is shown in **Figure 4.4** showing the peaks in the pyrrolidine region and the H2 and H8 proton peaks in the aromatic region at 8.33 ppm and 8.39 ppm. An expanded section of the pyrrolidine region, showing just three of the four expected peaks, two doublets and a triplet at 4.30 ppm, 4.94 ppm and 4.72 ppm is shown in **Figure 4.5**. These three peaks

match the peaks observed in **Figure 3.3** for the NMT-[60] monoadduct whilst the missing doublet peak was from the one of the two protons adjacent to each other on the pyrrolidine ring. The same coupling constants (J = 9.62 Hz) observed on the more deshielded doublet (4.94 ppm) and the triplet show that these two peaks are from protons which are adjacent to each other, which are the single proton on the pyrrolidine ring and methylene protons on the carbon connecting the pyrrolidine ring and the adenine entity. In addition, the coupling constant on the remaining doublet (J = 10.07 Hz) matched the coupling constant observed for the same proton in the NMT-[60] monoadduct. This data shows the possible presence of the BOC-protected NMA-[60] monoadduct in the product mixture.

Figure 4.4 A section of the ^{1}H NMR spectrum of the cycloaddition product of the BOC-protected 2-(adenin-N-yl)ethanal/2-adenine diol mixture with C_{60} showing all the peaks in the pyrrolidine region and the signals from the H2 and H8 adenine protons.

Figure 4.5 A section of the ^{1}H NMR spectrum of the cycloaddition product of BOC-protected 2-(adenin-N-yl)ethanal/2-adenine diol mixture with C$_{60}$ focusing on the major peaks in the pyrrolidine region and the coupling constants.

4.3 Conclusion

This work has demonstrated the possibilities of synthesizing NMA-[60]. The key step of synthesizing a soluble adenine aldehyde by hydrolysing the intermediate, N-(2, 2-diethoxyethyl)adenine, proved to be a challenge. However, a potential route has been identified. Once this step has been refined, synthesis of the pure NMA-[60] derivative will be possible.

4.4 Experimental Section

N-(2, 2-diethoxyethyl)adenine (**38**). Adenine (13.0 g, 96.3 mmol), K$_2$CO$_3$ (14.6 g, 105.6 mmol), bromoacetaldehyde diethyl acetal (17.4 cm^3, 115.6 mmol) and DMF (160 cm^3) were refluxed at 140 ˚C for 23 hours. The mixture was filtered whilst hot and washed with DMF (50 cm^3). The filtrate was concentrated to approximately 20 cm^3 then cooled to 0 ˚C. The

precipitated product was collected by filtration, dried *in vacuo* and purified using column chromatography in methanol: dichloromethane (1:9) to give N-(2, 2-diethoxyethyl)adenine (7.21 g, 30 %). Recrystallization from methanol: dichloromethane (1:9) furnished clear crystals. m.p 222-224 °C. **FT-IR** v_{max}/cm^{-1} 3276, 3112, 2979, 2877, 1673, 1602, 1576, 1474, 1416, 1384, 1372, 1326, 1305, 1249, 1212, 1131, 1111, 1057, 1025, 933, 904, 867, 799, 779, 721 and 689. **^{1}H NMR** (DMSO) δ 1.02 (t, 6, CH$_3$ of Et, *J* = 14.2 Hz), 3.31-3.38 (m, 2, CH$_2$), 3.63 (m, 2, CH$_2$ *J* = 16.9 Hz), 4.21 (d, 2, CH$_2$ *J* = 5.5 Hz), 4.83 (t, 1, CH), 8.05 (s, 1, H2) and 8.14 (s, 1, H8) ppm. **^{13}C NMR** (DMSO) δ 15.02 (CH$_3$ of Et), 42.90 (NCH$_2$), 62.13 (CH$_2$ of Et), 99.34 (CH of acetal), 118.27, 141.18, 149.27, 152.39, 155.83 (adenine) ppm. **UV-VIS** λ_{max} (MeOH) 355, 388, 398, 392, 405 nm. **Acc. mass** (HRES$^+$) M calc m/z = 252.2929, measured m/z = 252.1469 corresponding to $C_{11}H_{18}N_5O_2$ for (**38**).

2-(Adenin-N-yl)ethanal/2-adenine diol mixture (**39**). N-(2, 2-diethoxyethyl)adenine (3.0 g, 12.0 mmol) was stirred in 1 M HCl (40 cm^3) at 100 °C for 1 hour whereupon it was evaporated to dryness. The solid residue was washed with 50 % ethanol (3 cm^3), CH$_2$Cl$_2$ (50 cm^3) and ether (50 cm^3) to give a yellow solid which was dried at 100 °C for 3 days (2.9 g, 97 %). m.p > 300 °C. **FT-IR** v_{max}/cm^{-1} 2988.37, 2359.41, 1684.85, 1053.98. **^{1}H NMR** (DMSO) δ 4.16 (d, 2, CH$_2$), 5.11 (t, 1, CH-(OH), 5.37 (s, 1, CH=O), 8.36 (s, 1, H2), 8.46 (s, 1, H8), 8.41 (s, 1, H2), 8.49 (s, 1, H8) 9.69 (s, 1, C=O) ppm. **^{13}C NMR** (DMSO) δ 50.21, 53.39, 87.92, 118.14, 118.28, 144.81, 145.12, 145.48, 146.17, 149.30, 150.37, 151.10 196.44 (C=O). **UV-VIS** λ_{max} (MeOH) 409, 421 nm. **Acc. mass** (HRES$^+$) M calc m/z =178.1713, measured m/z = 178.0723 corresponding to $C_7H_8N_5O$ for (**39a**) and 196.0841 Da corresponding to $C_7H_9N_5O_2$ for (**39b**)

Tris-BOC-adenine (**41**). To a 100 ml Ar flushed flask, adenine (0.501 g, 3.7 mmol) and DMAP (0.04 g, 0.37 mmol) were added to dry THF (25 cm^3). Boc$_2$O (3.23 g, 14.8 mmol) was added to the flask and the suspension was stirred overnight at room temperature, under Ar atmosphere. Once TLC indicated the complete conversion of the starting material after 18 hours, the excess amount of THF was removed and the residual yellow oil was purified by

flash chromatography on silica gel (hexane/ethyl acetate, 7:3) to give tris-BOC-adenine (**39**) (1.49 g, 92 %) as a pale-yellow oil. R_f = 0.43 (hexane/ethyl acetate,7:3) **FT-IR** v_{max}/cm^{-1} 1119.31, 1763.53, 2981.76. **^{1}H NMR** (CDCl$_3$): δ 1.43 (s, 18, CH$_3$), 1.72 (s, 9, CH$_3$), 8.52 (s, 1, H$_2$), 9.02 (s, 1, H$_8$) ppm. **^{13}C NMR** (CDCl$_3$): δ 28.03, 84.09, 87.60, 129.69, 143.29, 145.72, 150.13, 151.30, 152.56, 154.20 ppm. **UV-VIS λmax** (CHCl$_3$) 361 nm. **Acc. Mass** (HRES$^+$) M calc. m/z = 435.4790 Da, measured m/z (M+23) = 458.2009 Da.

Bis-BOC-adenine (**42**). Tris-BOC-Adenine (0.301 g, 0.69 mmol) was stirred in MeOH (10 cm^3) and saturated NaHCO$_3$ (5 cm^3) at 50 °C 0 for 1 hr. Methanol was then evaporated, followed by the addition of water (6 cm^3) to the suspension. The aqueous layer was extracted with CH$_2$Cl$_2$ (3 x 12 cm^3) and the organic layer was dried with MgSO$_4$, filtered and evaporated to give a white solid. The solid was purified by flash chromatography on silica gel (ethyl acetate, DCM/MeOH [9:1]) to give bis-BOC-adenine (**42**) (0.0539 g, 23 %) as a white solid. R_f = 0.40 (DCM/MeOH) **FT-IR** v_{max}/cm^{-1} 3336.03, 2979.72, 2359.28, 1248.89, 1149.54 cm^{-1}. **^{1}H NMR** (CDCl$_3$): δ 1.60 (s, 18, H), 8.29 (s, 1, H), 8.76 (s, 1, H) ppm. **^{13}C NMR** (CDCl$_3$): δ 28.16, 83.79, 112.09, 143.84, 144.33, 152.32, 153.08 ppm. **UV-VIS** $λ_{max}$ (CHCl$_3$) 277, 286, 361 nm. **Acc. Mass** (HRES$^+$) M calc. m/z = 335.3627 Da, measured m/z (M$^-$) = 334.1520 Da corresponding to (**42**).

Bis-BOC-N-(2, 2-diethoxyethyl)adenine (**43**) **– Method A**. N-(2, 2-diethoxyethyl)adenine (0.355 g. 1.41 mmol), DMAP (0.017 g, 0.14 mmol) and Boc$_2$O (1.23 g, 5.66 mmol) were stirred in dry THF (10 cm^3), under N$_2$, at 60 °C for 6 hours. When TLC indicated the complete conversion of N-(2, 2-diethoxyethyl)adenine to the product after 6 hours, the excess THF was removed and the crude product was purified by flash chromatography on silica gel (DCM/MeOH [9.9:0.1]). The solvent was removed under vacuum to give a white solid (**43**) (0.23 g, 36 %) **FT-IR** v_{max}/cm^{-1} 2978, 2932, 1739, 1600, 1577, 1506, 1480, 1449, 1393, 1369, 1336, 1275, 1249, 1221, 1111, 1058, 1036, 937, 880, 850, 808, 793, 779, 769, 753, 719, 710 and 669. **^{1}H NMR** (CDCl$_3$) δ 1.16 (t, 6, CH$_3$ of Et, J = 34.3 Hz), 1.44 (s, 18, CH$_3$), 3.47-3.52 (m, 2, CH$_2$, J = 30.2 Hz), 3.73-3.77 (m, 2, CH$_2$, J = 30.6 Hz) 4.37 (d, 2,

CH$_2$, J = 5.5 Hz), 4.73 (t, 1, CH. J = 10.5 Hz), 8.17 (s, 1, H2), 8.86 (s, 1, H8) ppm. **^{13}C NMR** (CDCl$_3$) δ 15.13 (CH$_3$ of Et), 27.70 (CH$_3$), 46.33 (N-CH$_2$) 63.89 (CH$_2$) 83.61 (CH), 100.01, 145.82, 150.14, 150.37, 151.94 (adenine) ppm. **UV-VIS** λ_{max} (CHCl$_3$) 358, 426 nm. **Acc. mass** (HRES) M$^+$ calc m/z = 451.514 Da, measured m/z = 452.2500 Da corresponding to the required product.

Cycloaddition of bis-BOC-2-(adenin-N-yl)ethanal mixture. C$_{60}$ (0.025 g, 0.035 mmol), bis-BOC-2-(adenin-N-yl)ethanal (0.0062 g, 0.17 mmol) and N-methylglycine (0.078 g, 0.07 mmol) were stirred in 1,2-DCB (30 cm^3) at 100 °C. **FT-IR** v_{max}/cm^{-1} 2362, 2341, 1655, 1384, 1101, 669, 576. **^{1}H NMR** (CDCl$_3$): δ, 2.99 (s, 3, CH$_3$) (pyrrolidine region) 4.30 (d, 2, H, J = 10.07 Hz), 4.72 (t, 1, H, J = 9.62 Hz), 4.94 (d, 2, CH$_2$, J = 9.62 Hz), 5.13, 5.17, 5.31, 5.38, 5.42, 5.33, 8.33 (s, 1, H2), 8.39 (s, 1, H8) ppm. **UV-VIS** λ_{max} (chloroform) 259, 330, 430 nm.

Chapter 5
Cytosine

5.1 Introduction

Cytosine (**47**) is the second pyrimidine present in DNA. In comparison to thymine, cytosine has an amine group but does not have a methyl group as shown (**Figure 5.1**).

(47) **(48)**
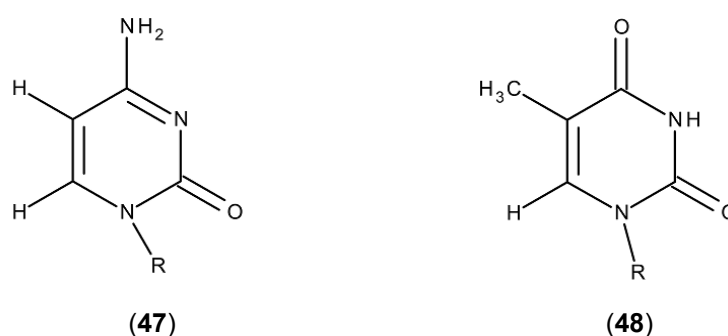
 Figure 5.1 Comparison of pyrimidines, cytosine (**47**) and thymine (**48**).

5.2 Experimental Results and Discussion

5.2.1 Synthesis Scheme for Starting Materials

The precursor to the aldehyde required for the cycloaddition of N-methyl-2-cytosin-N-ylfulleropyrrolidine (NMC-[60]) was synthesized following **Scheme 3.2**. N-(2,2-diethoxyethyl)cytosine (**49**) was produced as a white solid after purification by column chromatography with dichloromethane/methanol as the solvent mixture. N-(2,2-diethoxyethyl)cytosine was then hydrolysed with 1M HCl for an hour. The excess acid was removed under vacuum, producing a thick yellow liquid which crystallised on standing. Characterisation with ^{1}H NMR spectroscopy revealed that the compound was impure, and to purify the compound, it was washed and extracted with water and dichloromethane. However, when characterisation of the extracted compound did not reveal the required 2-2-(cytosin-N-yl)ethanal, the same protection route with di-*tert*-butyl dicarbonate used for adenine was considered.

5.2.2 NMR and Mass Spectrometry of N-(2,2-diethoxyethyl)cytosine

Successful synthesis of N-(2,2-diethoxyethyl)cytosine (**49**) was confirmed by ^{1}H, ^{13}C and DEPT 135 NMR spectroscopy. The methyl triplet at 1.04 ppm and the pair of multiplets from 3.34 - 3.42 ppm and 3.60 - 3.66 ppm from the methylene protons confirmed the alkylation of cytosine. More importantly, the pair of doublets from the H5 and H6 cytosine protons were present at 5.61 ppm and 7.45 ppm respectively and had coupling constants of 7.3 Hz (**Figure 5.2**). H6 is adjacent to nitrogen and is therefore more downfield than H5 because it is adjacent to carbon, which is less electronegative than nitrogen. The ^{13}C NMR spectrum exhibited eight signals from the compound and an inverted sp^3 signal at 63.1 ppm in the DEPT 135 NMR spectrum. Confirming formation of N-(2,2-diethoxyethyl)cytosine, mass spectrometry produced a molecular ion (M$^+$) at 228.1346 Da corresponding to the required formula $C_{10}H_{17}N_3O_3$ for (**49**).
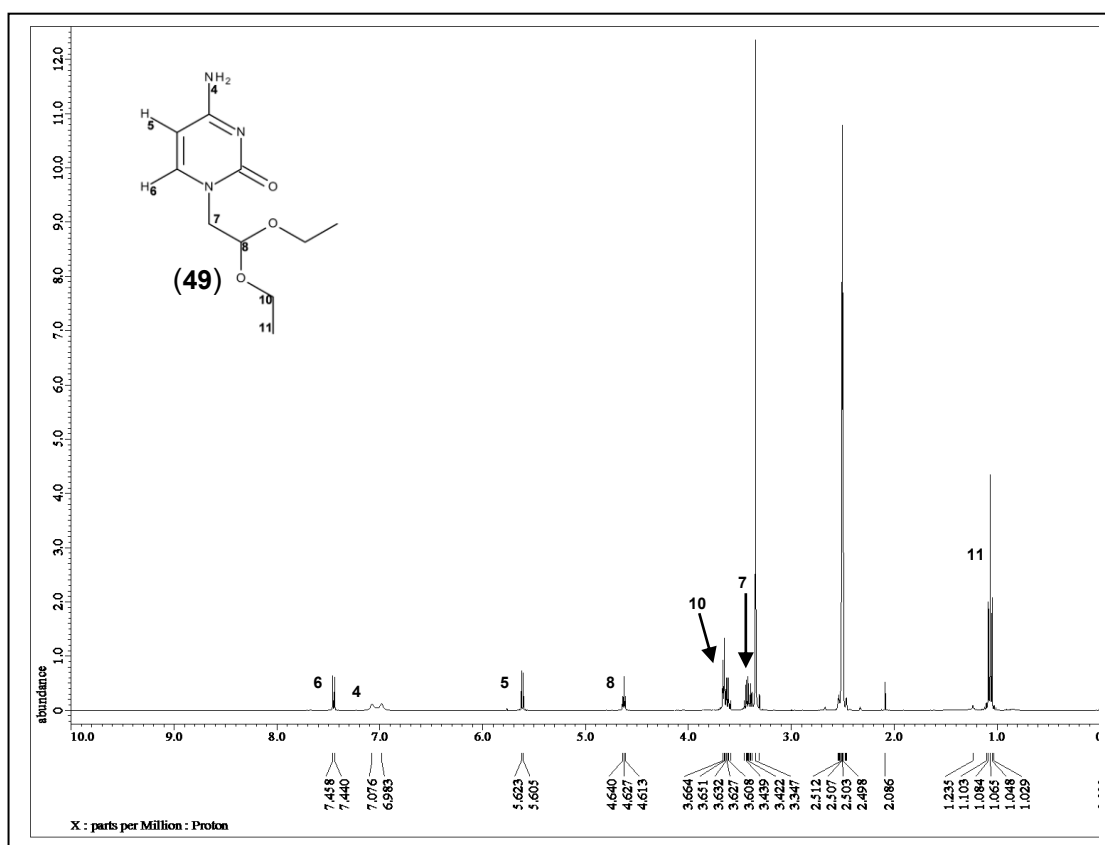
Figure 5.2 ^{1}H NMR spectrum of N-(2,2-diethoxyethyl)cytosine (**49**).

5.2.3 NMR and Mass Spectrometry of 2-(cytosin-N-yl)ethanal

Hydrolysis of N-(2,2-diethoxyethyl)cytosine (**49**) with 1 M HCl yielded an impure compound that contained 2-(cytosin-N-yl)ethanal. The ^{1}H NMR spectrum had signals from the carbonyl methylene signal at 4.79 ppm as well as the de-shielded aldehyde proton at 9.56 ppm. The signals from the H5 cytosine proton was present as a doublet at 6.17 ppm (J = 7.7 Hz) while the H6 signal was observed as 7.88 ppm as a split doublet (J = 12.8 Hz). The ^{1}H NMR spectrum also displayed a range of minor peaks, including broad singlets on either side of the aldehyde singlet. These signals are characteristic of quaternary ammonium salts. This suggested that the compound formed was possibly a mixture of the aldehyde and the ammonium salt of cytosine. The ^{13}C NMR spectrum had signals at 197.03, 160.35, 160.17, 150.61 93.98 and 58.01 ppm for the aldehyde carbon, the C-NH$_2$ carbon atom, CH6, CH5, the C=O carbon atom and the methylene carbon respectively. It also exhibited six other signals, which confirmed the compound as a mixture of the salt and the required compound.

In order to isolate 2-(cytosin-N-yl)ethanal from the ammonium salt, this mixture was extracted with a mixture of dichloromethane and water. It was expected that the ammonium salt would dissolve in the water layer, whilst 2-(cytosin-N-yl)ethanal would be dissolved in the organic layer. After separation of the organic and aqueous layers, TLC of each layer revealed that the compound was present in the aqueous layer. The water layer was then evaporated, leaving a yellow solid which was further dried under vacuum over phosphorous pentoxide.

5.2.4 Analysis of the extracted compound

The ^{1}H NMR spectrum of the washed compound had three doublet peaks at 3.68 (J = 5.04 Hz), 5.88 (J = 7.33 Hz) and 7.48 (J = 7.33 Hz) ppm as well as a triplet at 5.07 (J = 10.5 Hz) ppm. The doublet peak with the same coupling constants were assigned to the H5 and H6 protons on the cytosine structure. The doublet peak which was shifted highfield (3.68 ppm) was assigned to the methylene protons and the triplet was assigned to the single proton on the carbon atom adjacent to the methylene protons. In the ^{13}C NMR spectrum, the most de-

shielded peak was the C-NH$_2$ signal at 162.81 ppm. This eliminated the presence of a carbonyl (CH=O) carbon in the compound which had been present at 197.03 ppm before the compound was washed. The methylene peak was present at 54.25 ppm and was confirmed in the DEPT 135 spectrum by its inversion. Mass spectrometry of this compound gave a molecular ion (M$^+$) at 172.0708 Da. This data indicated that this compound was a diol derivative of cytosine (**50**). This data corresponded to literature data by Heemstra and Lin on the same compound.[132] However, because the solubility of this compound was limited to water, this compound was also a salt (**50**).

50
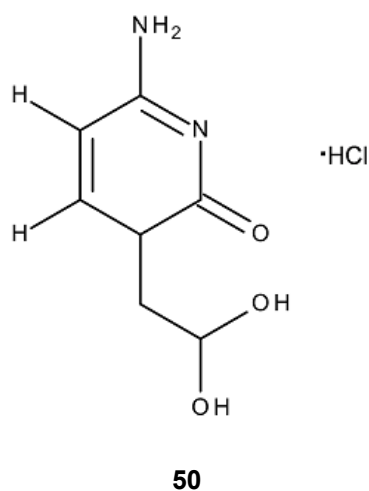
Figure 5.3. Structure of 2-cytosine-diol (**50**).

5.2.5 Protection of cytosine with di-*tert*-butyl dicarbonate

The proposed reaction scheme for protecting cytosine and the subsequent synthesis of NMC-[60] (**Scheme 5.1**) was similar to **Scheme 4.2** proposed for the synthesis of NMA-[60]. In a similar manner to adenine, cytosine was protected with di-*tert*-butyl dicarbonate in order to prevent intramolecular interactions and to increase its solubility. Tris-BOC-cytosine (**51**) was then alkylated (**52**) to facilitate hydrolysis leading to the formation of bis-BOC-2-(cytosine-N-yl)ethanal (**53**). The aldehyde is one of the reactants required for the Prato reaction, along with C$_{60}$ and N-methylglycine.

5.2.6 Formation of tris-BOC-cytosine

After carrying out step (i) (**Scheme 5.1**), the product was analysed and a strong signal characteristic of the BOC group was present at 1.56 ppm in the ^{1}H NMR spectrum. ^{13}C and DEPT 135 NMR spectroscopy confirmed the successful synthesis of tris-BOC-cytosine **Scheme 5.1** (**51**). Mass spectrometry analysis gave the molecular ion (M + 23) at 4.1895 ppm, corresponding to the correct formula $C_{19}H_{29}N_3O_7$ (**51**).
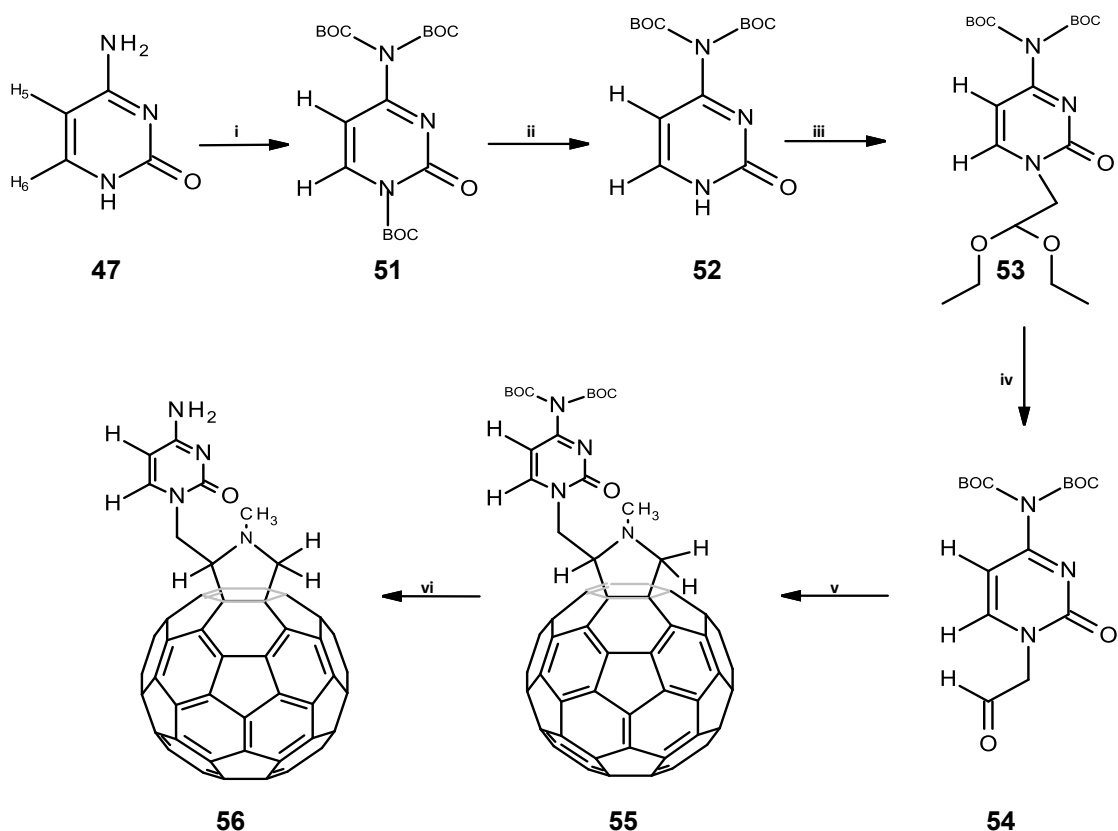
Scheme 5.1 Proposed reaction scheme for the synthesis of N-methyl-2-cytosinefulleropyrroline NMC-[60] (**56**) *via* the BOC-protection route. Reagents: (i) Boc$_2$O, DMAP, THF (ii) MeOH, NaHCO$_3$ (iii) K$_2$CO$_3$, BrCH$_2$CH(OC$_2$H$_5$)$_2$, NEt$_3$, DMF, reflux; (iv) 1 M HCl, reflux, (v) C$_{60}$, N-methylglycine, 1,2-DCB, reflux; (vi) TFA.

5.2.7 Subsequent Synthesis Problems

This synthesis pathway (**Scheme 5.1**) was not carried further than the synthesis of tris-BOC-cytosine (**51**) (step i). This was because of the negative results obtained from the test

reaction of bis-BOC-N-(2, 2-diethoxyethyl)adenine with 1 M HCl: bis-BOC-N-(2, 2-diethoxyethyl)adenine (**Method A** (**ii**)) was de-protected after gently refluxing the compound in 1 M HCl. Since hydrolysis of BOC-protected cytosine was also part of the planned synthesis route (**Scheme 5.2, step** (**iv**)) it was abandoned after it failed for bis-BOC-N-(2, 2-diethoxyethyl)adenine. Additionally, **Method B,** which involved protection of the aldehyde, could not be used since it required the isolation of 2-(cytosine-N-yl)ethanal which was not possible.

5.3 Conclusion

The successful synthesis of the alkylated cytosine has been shown. The presence of the 2-(cytosine-N-yl)ethanal signals after hydrolysis showed that it is possible to make this compound. The synthesis route requires optimisation in order to successfully synthesize pure 2-(cytosine-N-yl)ethanal. The refinement of the corresponding adenine product will also provide a useful template.

5.4 Experimental Section

N-(2,2-diethoxyethyl)cytosine (**49**). A mixture of freshly dried cytosine (1.0 g, 9 mmol) and K_2CO_3 was refluxed in bromoacetaldehyde diethyl acetal (1.77g, 9 mmol) and DMF (10 cm^3) at 130 °C for 10 hours. The solids were filtered whilst hot and washed with CHCl$_3$ (10 cm^3). The filtrate was dried *in vacuo* then triturated with ether:petroleum ether (1:2, 2.5 cm^3). The crude product was purified by column chromatography (DCM/MeOH [8.5:1.5]) to give the required product, N-(2,2-diethoxyethyl)cytosine (0.16 g, 16 %). R_f = 0.27. m.p = 234-238 °C. **FT-IR** v_{max}/cm^{-1} 1056 cm^{-1} 1613 cm^{-1}, 1659 cm^{-1}, 2896 cm^{-1}, 2929 cm^{-1}, 2973 cm^{-1}. **^{1}H NMR** (DMSO) δ 1.07 (t, 6, H, J = 29.7 Hz); 3.34-3.42 (m, 2, CH$_2$, J = 30.6 Hz); 3.60-3.66 (m, 4, CH$_2$, J = 29.3 Hz); 4.62 (t, 1H, J = 10.9 Hz); 5.62 (d, 1H, J = 7.3 Hz); 7.03 (d, 2H, J = 37.0 Hz); 7.44 (d, 1H, J = 7.3 Hz) ppm. **^{13}C NMR** (DMSO) δ 15.7, 52.0, 63.1, 93.3, 100.1, 147.5, 156.4, 166.5 ppm. **UV-VIS** λ_{max} (EtOH) 318, 327 nm. **Acc. Mass**. (HRES) M$^+$ calc m/z = 227.2630 Da measured m/z = 228.1346 Da

2-(Cytosin-N-yl)ethanal mixture. N-(2,2-diethoxyethyl)cytosine (0.51 g, 2.2 mmol) was gently refluxed in 1 M HCl (20 cm^3) for 1 hour. The excess HCl was removed under vacuum, producing a white solid. **FT-IR** v_{max}/cm^{-1} 2988.13, 2359.38, 1668.67, 1051.39. **^{1}H NMR** (DMSO) δ 4.79 (s, 1, H); 6.17 (d, 1, H, J = 7.7 Hz); 7.88 (d, 1, H, J = 12.8 Hz),9.56 (s, 1, CH=O) **^{13}C NMR** (DMSO) δ 54.80, 58.01 (CH$_2$) 87.07, 93.07, 93.98 (CH5) 148.12, 148.17, 150.61 (CH6), 151.57 (C=O), 160.17, 160.35 (C-NH$_2$) ppm. **UV-VIS** λ_{max} (CHCl$_3$) 354 nm. **Acc. Mass**. M$^+$ calc m/z = 171.1555 Da measured, m/z = 172.0708 Da.

2-Cytosine diol (**50**). 2-(Cytosin-N-yl)ethanal mixture (0.50 g) was washed with water (10 cm^3), and dichloromethane (15 cm^3). The organic layer was dried with magnesium sulphate and removed under vacuum, yielding a white solid (0.45 g). **^{1}H NMR** (D$_2$O) δ 3.68 (d, 2, CH$_2$, J = 5.04 Hz); 5.07 (t, 1, H, J = 10.5 Hz); 5.88 (d, 1, H5, J = 7.33 Hz); 7.48 (d, 1, H6, J = 7.33 Hz) **^{13}C NMR** (D$_2$O) δ 38.70, 54.25, 87.09, 94.88, 149.10, 162.81 ppm. **UV-VIS** λ_{max} **Acc. Mass.** (HRES) M$^+$ calc m/z = 153.1644 Da measured, m/z = 154.0605 Da.

Tris-BOC-cytosine (**51**). Cytosine (1.025 g, 9.1.mmol) was added to a flask flushed with nitrogen. DMAP (0.12 g, 0.98 mmol), Boc$_2$O (7.86 g, 36.0 mmol) and dry THF (25 cm^3) were added to the flask and stirred overnight at room temperature under nitrogen. Once all the starting material had reacted, the excess THF was removed under vacuum. The crude product was dissolved in ethyl acetate (300 cm^3) and washed with 1 M HCl (1 x 15 cm^3) and brine (2 x 15 cm^3) then dried in MgSO$_4$ and concentrated *in vacuo*. The yellow oil was purified using flash chromatography affording tris-BOC-cytosine (**51**) as a clear oil. R_f =0.34 (hexane/ethyl acetate [8:2]). **FT-IR** v_{max}/cm^{-1} 2979.10, 2359.54, 1784.84, 1243.08, 1114.01. **^{1}H NMR** (CDCl$_3$) δ 1.56 (s, 27, H), 7.09 (d, 1, H5, J = 7.7 Hz), 7.97(d, 1, H6, J = 8.24 Hz) ppm. **^{13}C NMR** (CDCl$_3$) δ 27.91, 85.48, 96.99, 143.57, 149.23, 150.32, 151.38, 162.45 ppm. **UV-VIS** λ_{max} 257, 297, 359 nm. **Acc. Mass**. (HRES) calc m/z = 411.4532 Da, measured m/z (M+23) = 434.1895 Da.

Chapter 6
Guanine

6.1 Introduction

Guanine (**57**) is the second purine found in DNA and just like adenine, consists of a six-membered ring fused with a five-membered ring. However, guanine forms three H-bonds with cytosine when binding using the Watson-Crick mode, compared to the two bonds adenine forms with thymine.[133]

Guanine is notoriously insoluble in organic solvents, which makes it challenging to synthesize a fullerene derivative that incorporates this group.[128] To overcome the low solubility of the purine, other derivatives of guanine such as N^2-isobutyrylguanine[134] (**58**) and 2-amino-6-chloropurine[129] (**59**) have been used by Liu *et al*. and Thomson *et al.* respectively (**Figure 6.1**). Protecting routes with di-*tert* butyl dicarbonate have also been used by Dey and Garner.[125]

57 **58** **59**
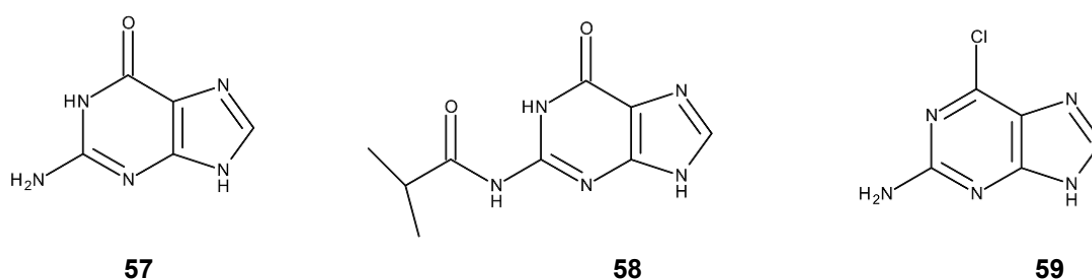
Figure 6.1 Guanine (**57**) and some of the used analogues, N^2-isobutyrylguanine (**58**) and 2-amino-6-chloro-purine (**59**).

6.2 Experimental Results and Discussion

The synthesis route for the precursor to the aldehyde and the aldehyde for the cycloaddition of N-methyl-2-guanin-N-ylfulleropyrrolidine (NMG-[60]) (**65**) did not follow the same route as described for the thymine, adenine and cytosine bases (**Scheme 3.2**). An alternative route

that used the more soluble 2-amino-6-chloropurine (**60**) as the starting material was used. This route, which included the allylation (step i), followed by hydrolysis (step ii) to produce **61** and **62** respectively, was chosen because it enabled the formation of guaninealdehyde (**63**) *via* oxidation with osmium tetroxide. Once **63** was produced, protecting it would enhance its solubility and enable cycloaddition to form NMG-[60] (**65**).

6.2.1 Synthesis Scheme for N-Methyl-2-guanin-N-ylfulleropyrrolidine (NMG-[60])
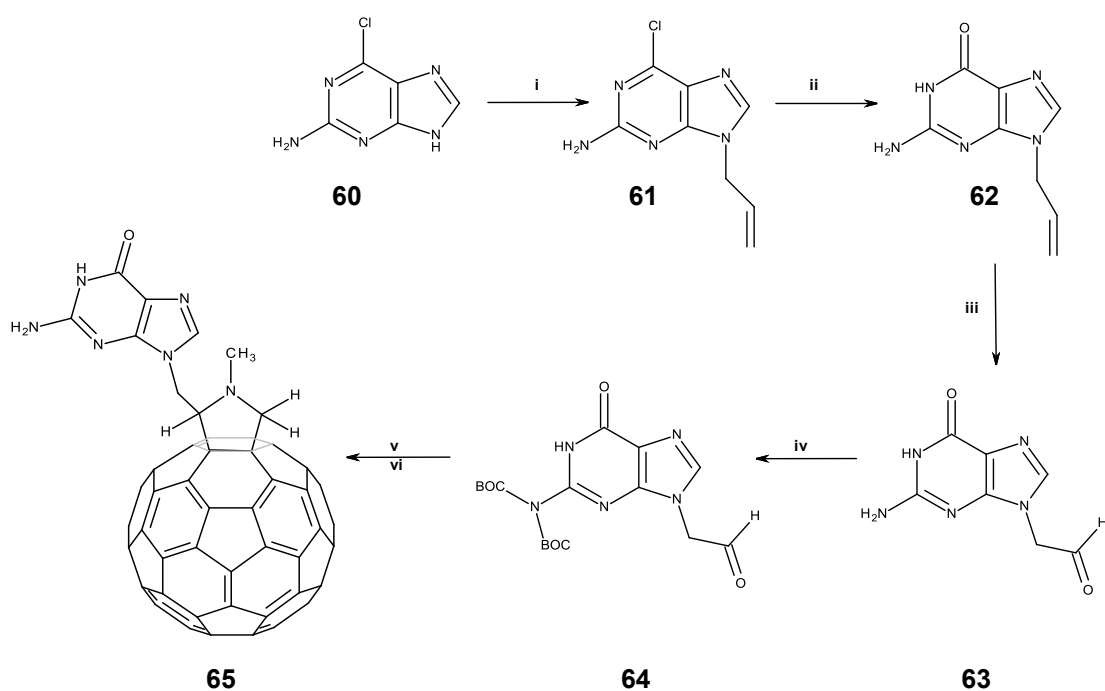
Scheme 6.1 Proposed reaction scheme for the synthesis of NMG-[60] (**65**) *via* the N9-allylguanine (**62**) route. Reactants: (i) K_2CO_3, allylbromide, DMF, 50 °C; (ii) 1 M HCl, NaOH; (iii) 1,4-dioxane, $NaIO_4$, OsO_4, water; (iv) Boc_2O, THF, DMAP; (v) C_{60}, N-methylglycine, 1,2-DCB (vi) TFA.

6.2.2 NMR and Mass Spectrometry of 2-Amino-6-chloro-9-allylpurine

^{1}H NMR spectroscopy confirmed the formation of 2-amino-6-chloro-9-allylpurine (**61**) and matched the results found in literature (**Figure 6.2**).[135] In particular, the multiplet signals at 5.98 ppm which arose from the coupling of H11 with H10b, H12a and H12b protons. H10a and H10b appeared as a doublet on the ^{1}H NMR spectrum because they were in the same

environment. In contrast, H12a and H12b appeared as a pair of doublets at 5.19 and 5.31 ppm respectively since H12b is *cis* to the molecule and is therefore more downfield. ^{13}C NMR spectroscopy also confirmed the presence of the allyl group with peaks at 45.64 ppm, 119.24 ppm and 131.28 ppm, which belong to the allyl group. High-resolution mass spectrometry analysis showed a molecular ion at 210.0544 Da which corresponds to the required formula $C_8H_8N_5Cl$ (**61**).
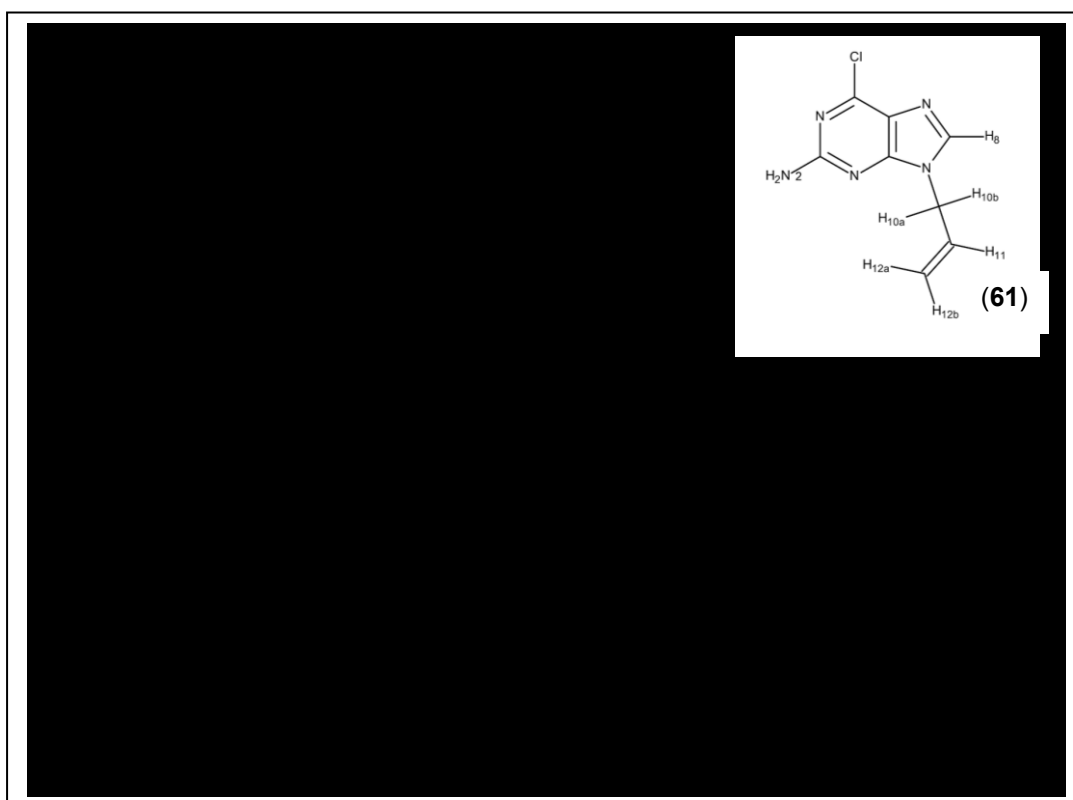
Figure 6.2 ^{1}H NMR spectrum of 2-amino-6-chloro-9-allylpurine (**61**) showing the allyl group protons.

6.2.3 NMR and Mass Spectrometry of N9-allylguanine

The ^{1}H NMR (**Figure 6.3**) and ^{13}C NMR spectroscopy data for N9-allylguanine (**62**) were similar to those observed for 2-amino-6-chloro-N9-allylpurine (**Figure 6.2**). Therefore, successful hydrolysis was confirmed by mass spectrometry and infrared spectroscopy (**Figure 6.4**). High-resolution mass spectrometry produced a molecular ion (M$^+$) at 192.0883 Da, corresponding to the required formula $C_8H_9N_5O$ (**62**).
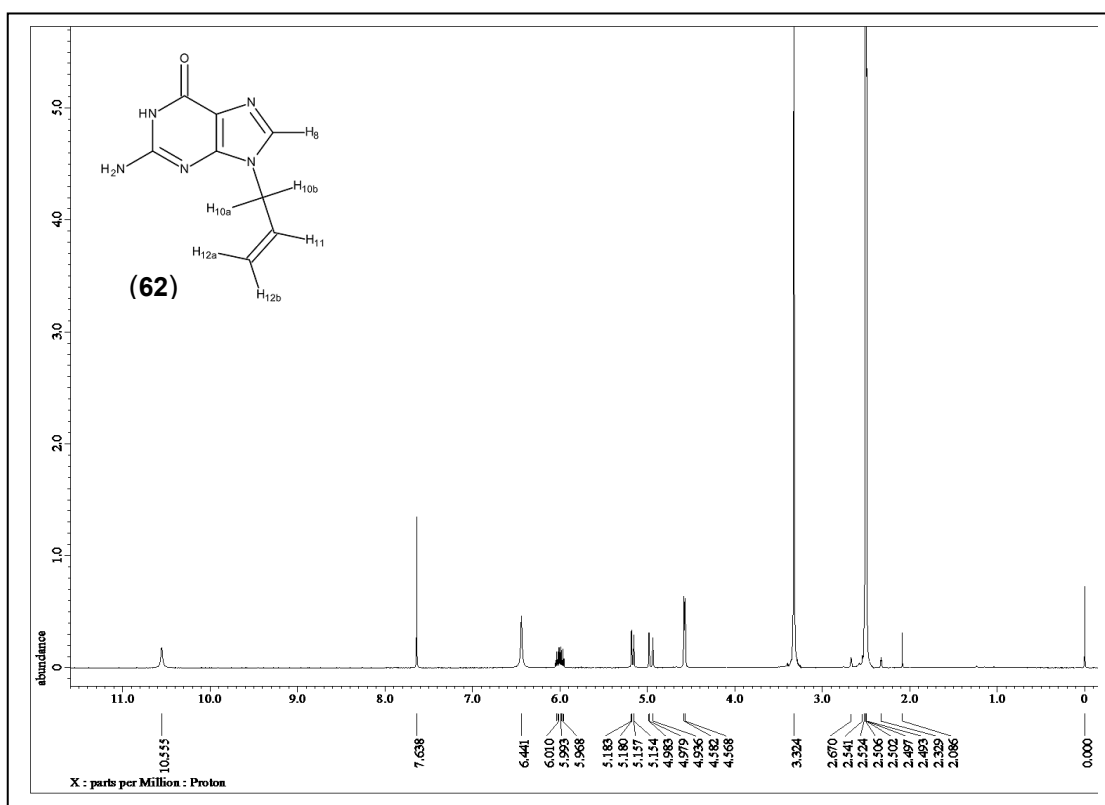
Figure 6.3 ^{1}H NMR spectrum of N9-allylguanine (**62**) showing the differences with 2-amino-6-chloro-N9-allylpurine (**61**).

The similarities in the ^{1}H and ^{13}C NMR were because both molecule structures were similar, apart from the change of the chlorine atom in 2-amino-6-chloro-N9-allylpurine to an oxygen atom in N9-allylguanine *via* hydrolysis with 1 M HCl.[135] The addition of a carbonyl signal was more evident in the infrared spectrum for N9-allylguanine (**Figure 6.4 (a)**).

Comparison of the IR spectra of 2-amino-6-chloro-N9-allylpurine with that of N9-allylguanine showed differences in the two compounds (**Figure 6.4 (a)** and (**b**)). The IR spectrum for 2-amino-6-chloro-N9-allylpurine (**b**) shows a sharp signal at 905.24 cm^{-1} due to the C-C stretch which was not present in the N9-allylguanine spectrum. Another important difference between the IR spectra was the signal at 1699.67 cm^{-1}, due to the C=O stretch, in the N9-allylguanine spectrum (**Figure 6.4 (a)**).
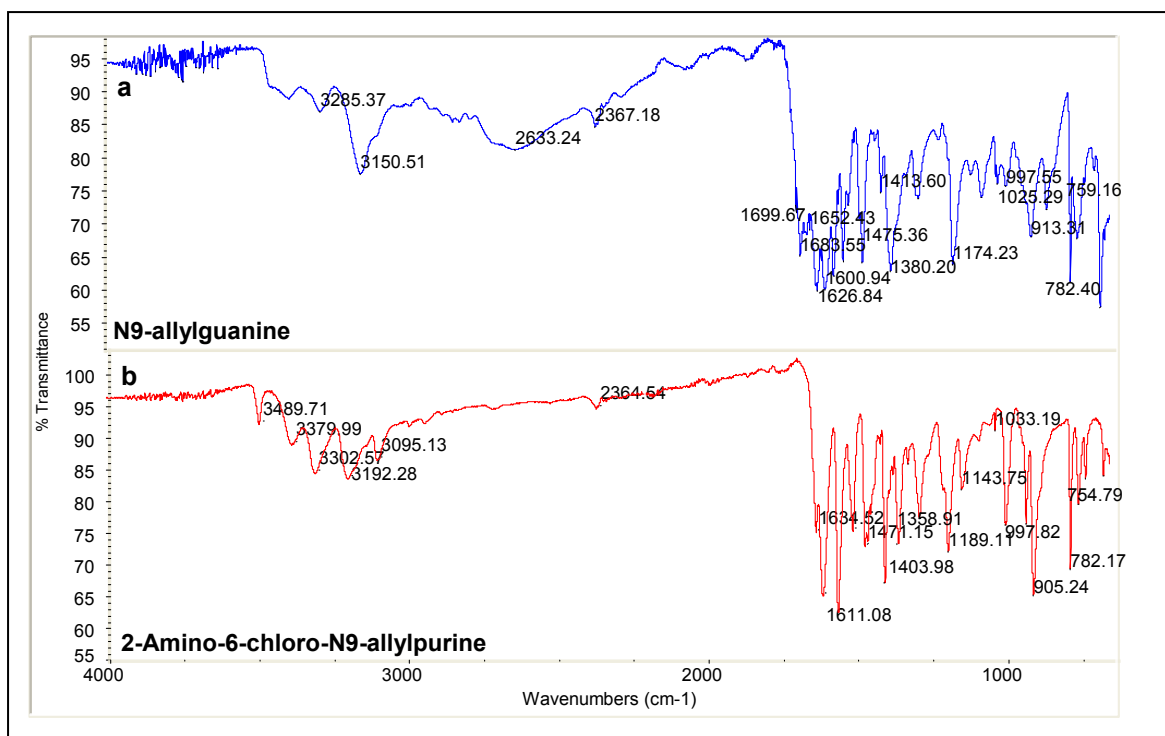
Figure 6.4 Infrared spectra of N9-allylguanine (**a**) and 2-amino-6-chloro-N9-allylpurine (**b**) showing the differences between the two compounds.

6.2.4 Synthesis of N9-Guaninealdehyde

In order to oxidise N9-allylguanine (**62**) to get N9-guaninealdehyde (**63**), osmium tetroxide was used (**Scheme 6.1**, step (iii)). This oxidising agent was used because the Glaxo group showed that 7-Methyloxy-1-(2-propen-1-yl)-1,5-naphthyridin-2(1H)-one (**66**) can be oxidised to form [7-methyloxy-1-(2-propen-1-yl)-1,5-naphthyridin-1(2H)-yl]acetaldehyde (**67**) using osmium tetroxide.[136] The similarities in their starting material to the guanine entity led us to use this method.

Test oxidation of N9-allylguanine (**62**) with osmium tetroxide failed. ^{1}H NMR spectroscopy of the product revealed that the starting material, 2-amino-6-chloro-9-allylpurine was recovered.

<div align="center">
66   **67**
</div>
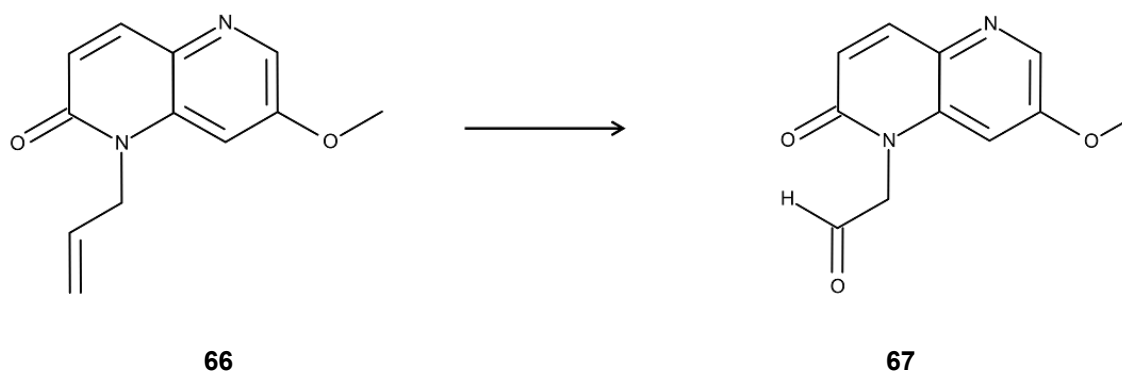
Scheme 6.2. Oxidation of 7-methyloxy-1-(2-propen-1-yl)-1,5-naphthyridin-2(1H)-one (**66**) to [7-methyloxy-1-(2-propen-1-yl)-1,5-naphthyridin-1(2H)-yl]acetaldehyde (**67**) using osmium tetroxide.

The synthesis of NMG-[60] (**Scheme 6.1(65)**), like its binding partner NMC-[60], is currently a step away from producing the required product. Once the oxidation of N9-allylguanine has been refined to produce aldehyde, it is likely that synthesis of the final product will be successful.

6.3 Conclusion

A new route for synthesizing 9-guaninealdehyde was developed and tested. Oxidation of 2-amino-6-chloro-9-allylpurine was unsuccessful at this time. However, once this step has been refined, synthesis of the required C_{60} derivative will be possible.

6.4 Experimental

2-Amino-6-chloro-N9-allylpurine (**61**). To a solution of 2-amino-6-chloropurine (0.50 g, 29.5 mmol) in DMF (15 cm^3), K$_2$CO$_3$ (12.1 g, 88 mmol) was added and stirred for 10 minutes at room temperature. Allyl bromide (2 cm^3, 23.1 mmol) was then added in small portions and the mixture was stirred for a further 15 minutes at room temperature then refluxed at 50 °C for 16 hours. The reaction mixture was cooled to room temperature and concentrated *in vacuo*. It was then partitioned between ethyl acetate (325 cm^3) and water (100 cm^3). The organic phase was washed with brine (2 x 100 cm^3), dried in MgSO$_4$ and concentrated *in vacuo*. Flash chromatography (ethyl acetate/hexane [1:1]) afforded the required compound, 2-amino-6-chloro-N9-allylpurine (**61**) as a white solid (1.9 g, 50 %). m.p 148-150 °C; R_f = 0.1.

FT-IR v_{max}/cm^{-1} 1634.52, 1189.11, 905.24 (C-C), 782.17. **^{1}H NMR** (CDCl$_3$) δ 4.69 (d, 2, CH$_2$, J = 6.8 Hz), 5.10 (br, 2, CH$_2$) 5.19 (d, 1, CH, J = 16.9 Hz) 5.31 (d, 1, CH, J = 10.5 Hz), 5.98 (ddt, 2, CH$_2$, J = 27.4 Hz), 7.76 (s, 1, CH) ppm; **^{13}C NMR** (CDCl$_3$) δ 45.64, 119.24, 125.16, 131.28, 142.13, 151.34, 153.61, 159.05 ppm. **UV-VIS** λ_{max} 318, 355, 366 nm. **Acc. Mass** (HRES) M$^+$ calculated m/z = 209.6375 Da measured m/z = 210.0544 Da.

N9-Allylguanine (**62**). A solution of 2-amino-6-chloro-N9-allylpurine (1.9 g, 9.9 mmol) was refluxed in 1 M HCl for 2 hours. After cooling to room temperature, the pH was adjusted with solid NaOH and cooled in an ice bath for 30 minutes. The precipitate was collected by filtration and dried *in vacuo* to give N9-allylguanine (**62**) (1.75 g). R_f = 0.77. m. p 163 -166 °C. **FT-IR** v_{max}/cm^{-1} 1699.67, 1174.23, 782.40. **^{1}H NMR** (DMSO) δ 4.57 (d, 2, H, J = 5.50 Hz), 4.98 (d, 1H, J = 18.7 Hz), 5.18 (d, 1H, J = 11.9 Hz), 5.98 (ddt, 1, H, J = 38.0 Hz), 6.44 (br, 2, H), 10.56 (br, 1, H) ppm. **^{13}C NMR** (DMSO) δ 45.11, 117.42, 122.9, 134.13, 137.76, 149.1, 154.15, 157.36 ppm. **UV-VIS** λ_{max} 355, 359 nm. **Acc. Mass** (HRES) M$^+$ calculated m/z =191.1934 Da measured m/z = 192.0883 Da.

9-Guaninealdehyde (**63**). To a solution of N9-allylguanine (0.1 g) in 1,4-dioxane (1.6 cm^3) and H$_2$O (0.8 cm^3), sodium periodate (0.2 g, 0.94 mmol) and osmium tetroxide (solution) (0.07 g, 0.28 mmol) were added consecutively. The mixture was stirred at room temperature for 24 hours. An additional amount of sodium periodate (0.02 g) and osmium tetroxide (0.007 g) was added to the mixture and stirred for another 72 hours. The mixture was filtered and the filtrate extracted with CH$_2$Cl$_2$/MeOH (9:1) (10 and 5 cm^3). The organic phases were combined and dried over MgSO$_4$ then filtered and evaporated to dryness under vacuum to give a white solid.

SECTION B

COMPUTATIONAL

Chapter 7
Computational Results

7.1 Introduction

In addition to the four DNA base-functionalised C_{60} derivatives mentioned in chapters 3 to 6, two novel fullerene derivatives were modelled: N-methyl-2-N-propylthyminefulleropyrrolidine (NMN-[60]) and N-methyl-2-diaminopyridinefulleropyrrolidine (NMD-[60]) (**Figure 7.1**). N-propylthymine and diaminopyridine were selected due to the potential for hydrogen bonding *via* the functional groups attached to C_{60}. Just like the NMG-[60]-NMC-[60] bonding pair, NMN-[60] and NMD-[60] can self-assemble *via* three H-bonds.

7.2 Computational Method

The global minimum structures for NMT-[60] (**68**), NMA-[60] (**69**), NMC-[60] (**70**), NMG-[60] (**71**), NMN-[60] (**72**) and NMD-[60] (**73**) were found after a full DFT optimisation calculation using B3LYP/6-31G**. Results from the optimisation calculation were analysed in order to determine the electronic properties such as the HOMO and LUMO energies as well as the electron distribution.

7.3 Computational Results and Discussion

7.3.1 Global Minimum Conformer

The DFT-optimised global minimum conformers of NMT-[60], NMA-[60], NMC-[60], NMG-[60], NMN-[60] and NMD-[60] are shown in **Figure 7.1**. The two pyrimidine-functionalised derivatives, NMT-[60] (**68**) and NMC-[60] (**70**) had similar conformations showing the thymine and cytosine entities pointing away from the C_{60} cage. Similarly, the purine-functionalised derivatives, NMA-[60] (**69**) and NMG-[60] (**71**), also had similar conformations where the adenine and guanine entities were closer to the C_{60} cage compared to the thymine and cytosine entities in structures (**68**) and (**70**) (**Figure 7.1**). NMD-[60] (**73**) also

exhibited a similar conformation to the purine-functionalised derivatives. By contrast, the thymine entity in the NMN-[60] derivative (**72**) was further from the C_{60} cage than observed in the other derivatives because of the additional methylene groups on the propyl chain. However, despite the distance from the cage, the thymine entity appeared to be at a similar orientation to that observed for the same moiety on the NMT-[60] derivative (**68**).

	1D model	B3LYP/6-31G** optimised model
NMT-[60] (**68**)		
NMA-[60] (**69**)		
NMC-[60] (**70**)		

NMG-[60] **(71)**		
NMN-[60] **(72)**		
NMD-[60] **(73)**		
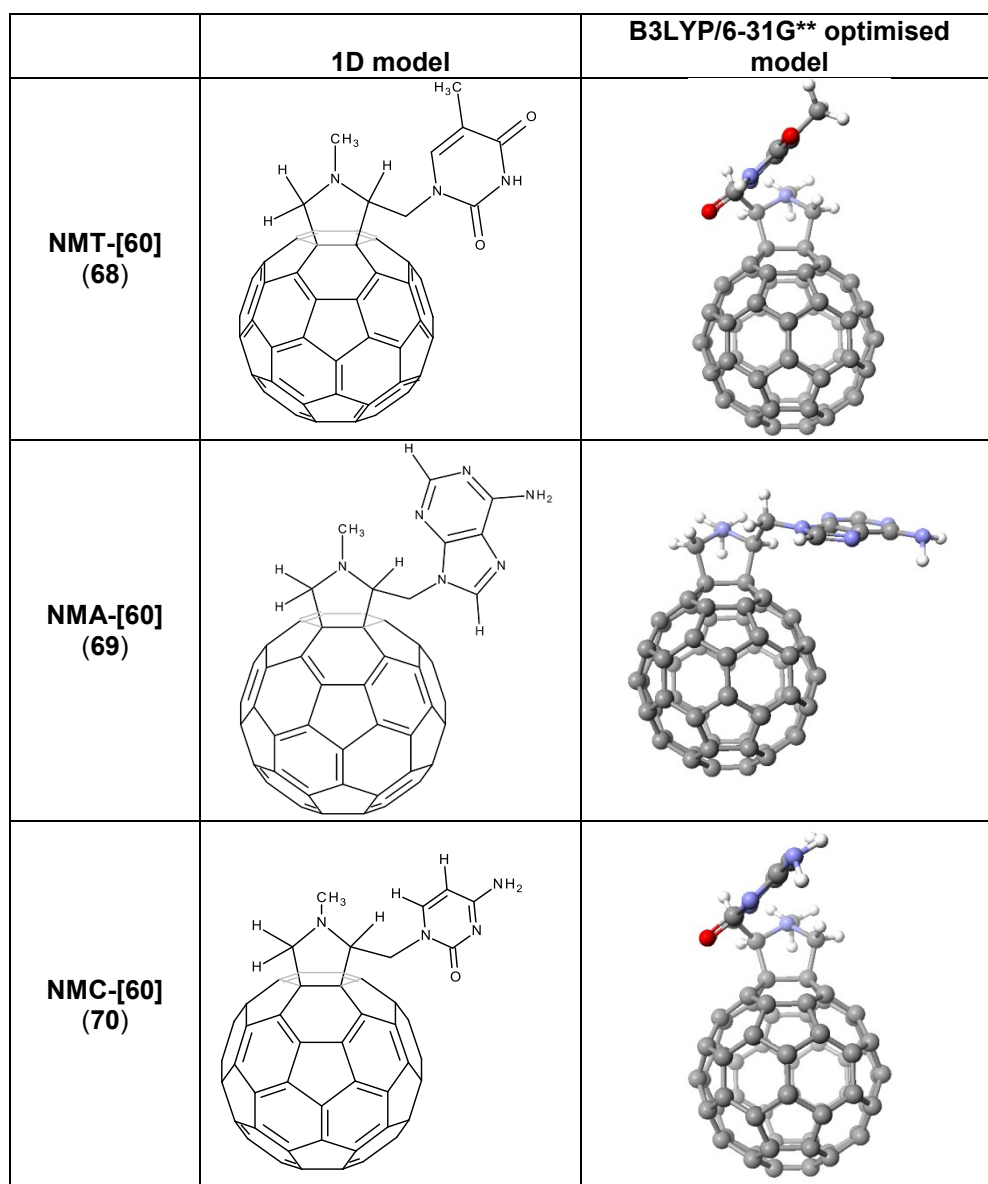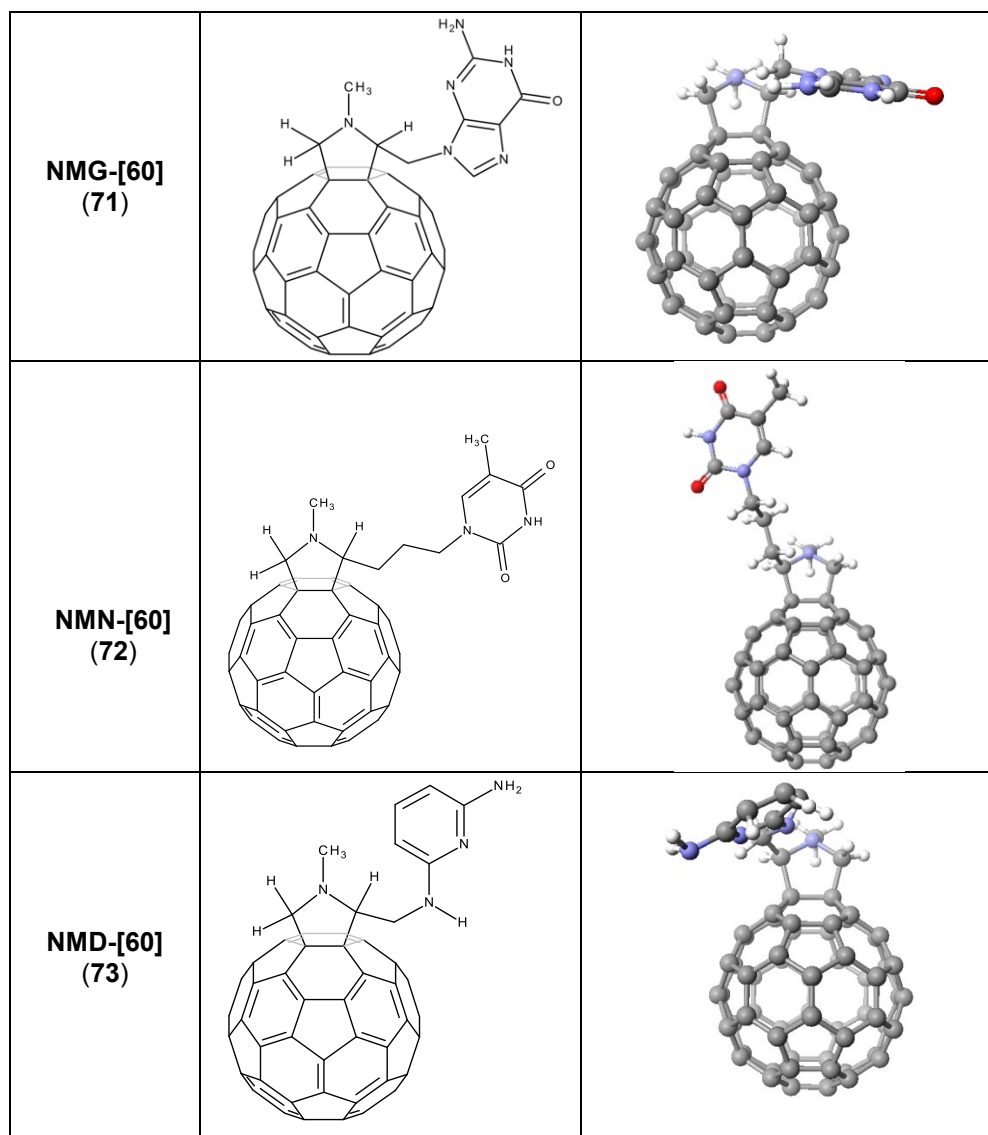
Figure 7.1 Lowest-energy conformers of the six modelled structures and the 1-dimensional structures showing the point of attachment on the pyrrolidine ring.

7.4 Electronic Properties

The electronic properties of the six monomers were calculated in the gas phase and three organic solvents, 1,2-dichlorobenzene (1,2-DCB), chlorobenzene (CB) and toluene. These solvents were selected because they are normally used for solar cell fabrication since C_{60} derivatives are more soluble in these solvents. The properties of the monomers in water were also calculated as this is applicable to biological systems.

7.4.1 Gas Phase

The HOMO, LUMO and HOMO-LUMO gap of NMD-[60], NMG-[60], NMC-[60], NMA-[60], NMT-[60] and NMN-[60] in the gas phase were compared in **Table 7.1**. In order to design and synthesize new soluble fullerene derivatives with improved efficiencies, it is necessary first to analyse the properties of C_{60} and $PC_{60}BM$.[38, 48] Once the properties of $PC_{60}BM$ have been analysed, they can be compared to the properties of the new derivatives, whose function can then be determined. Therefore, the calculated HOMO, LUMO and HOMO-LUMO gap for C_{60} and $PC_{60}BM$ are also shown (**Table 7.1**). Comparison of the LUMO energies of the six derivatives with $PC_{60}BM$ shows that the LUMO energies were within 0.09 eV of the LUMO energy for PCBM.

	HOMO (eV)	LUMO (eV)	HOMO – LUMO gap (eV)	HOMO – LUMO gap (nm)
C_{60}	-5.89	-3.13	2.76	449
NMD-[60]	-5.06	-2.90	2.16	574
NMG-[60]	-5.56	-3.08	2.48	499
NMC-[60]	-5.42	-2.92	2.53	490
NMA-[60]	-5.59	-3.05	2.54	487
NMT-[60]	-5.60	-3.06	2.54	488
NMN-[60]	-5.63	-3.08	2.55	486
$PC_{60}BM$	-5.57	-2.99	2.58	492

Table 7.1 B3LYP/6-31G** calculated energy of the frontier HOMO and LUMO orbitals, the HOMO-LUMO gap for C_{60}, six modelled fullerene derivatives and $PC_{60}BM$ in the gas phase. The derivatives are ranked according to increasing HOMO-LUMO gap energy.

Since calculated results quoted in published data are usually from gas phase calculations[29, 37, 38] the differences in the LUMO energies show that there is a potential for the six derivatives to function as electron acceptors in organic solar cells. It has been documented that with P3HT as the donor, making the LUMO energy of the acceptor less negative is a dominant way of increasing the efficiency of the solar cell.[29, 35, 38, 42] The results for the novel C_{60} derivatives in **Table 7.1** show the calculated LUMO energies of NMD-[60] and NMC-[60] are slightly higher (less negative) than the value shown for $PC_{60}BM$. This means that these derivatives are likely to show better power conversion efficiency compared to $PC_{60}BM$, thus being better electron acceptors.

The calculated LUMO energy levels shown for C_{60} and $PC_{60}BM$ (**Table 7.1**) were consistent with published computational data.[36, 37] For C_{60}, the LUMO energy from a B3LYP/6-31G* full optimisation was only 0.01 eV (0.3 %) lower (-3.14 eV) than the result shown in the gas phase in **Table 7.1**.[36] A bigger difference, 17 % (-3.75 eV) was observed for the experimentally calculated LUMO energy for the fullerene compared to the calculated energy in **Table 7.1**.[137] However, an equally large difference (16 %) is also observed between the calculated literature energy (-3.14 eV) and the published experimental LUMO energies (-3.75 eV).Therefore the values shown in **Table 7.1** for C_{60} are consistent with literature values.[36, 137]

The calculated LUMO energy for $PC_{60}BM$ using B3LYP/6-31G* had a 1.3 % (-3.03 eV) difference to the B3LYP/6-31G** results for the same structure shown in **Table 7.1**.[36] Comparison with results from B3LYP/6-311G** calculations showed a 2 % (-3.05 eV) difference in the LUMO energies to the results in **Table 7.1**.[37] The difference between experimental LUMO energy and the calculated energy in **Table 7.1** was 18 %.[137] However, other experimental LUMO energies for $PC_{60}BM$ vary between -3.7 eV and -4.3 eV (19 % and 30 %).[33-35] In addition, the differences between the published LUMO energies and the published calculated energies are also equally high, confirming the consistency of the energies shown for PCBM in **Table 7.1** with literature values.

7.4.1.1 HOMO-LUMO Gap

The HOMO-LUMO gap energies for NMC-[60], NMA-[60], NMT-[60] and NMN-[60] were within 0.05 eV of the HOMO-LUMO gap for $PC_{60}BM$ (**Table 7.1**). The exceptions were NMD-[60] and NMG-[60] which was 0.42 eV and 0.10 eV lower than $PC_{60}BM$. Comparison of the HOMO-LUMO gap for C_{60} and the four derivatives showed that these energies were consistent with the change observed in the HOMO-LUMO gap for $PC_{60}BM$, showing that these modelled structures not only behave in a similar manner to $PC_{60}BM$ but also possess similar properties to C_{60}. The changes in the HOMO-LUMO gap energies also confirm that functionalising the C_{60} cage alters the electronic properties of C_{60}.[17]

7.4.2 1,2-Dichlorobenzene

In this study, the effect of three key solvents on the fullerene acceptors has been calculated and the results will now be compared (**Table 7.2**, **7.3** and **7.4**).

In the presence of 1,2-dichlorobenzene, there was an increase in the HOMO energies and a decrease in the LUMO energies of C_{60}, the six derivatives and $PC_{60}BM$ (**Table 7.2**) compared to the energies in the gas phase (**Table 7.1**). In addition, the HOMO-LUMO gap energies were also significantly lower than the energies observed in the gas phase. This showed that the presence of 1,2-DCB affects the electronic properties of C_{60} as well as the six derivatives and $PC_{60}BM$, which would affect the behaviour of these electron acceptors in organic solar cells. Since the LUMO energies were lower (more negative) in the presence of 1,2-DCB compared to the LUMO energies in the gas phase, the efficiency of the solar cell would also decrease if residual 1,2-DCB was left during the fabrication of the solar cell.

	HOMO (eV)	LUMO (eV)	HOMO – LUMO gap (eV)	HOMO – LUMO gap (nm)
C_{60}	-5.75	-4.09	1.66	746
NMD-[60]	-4.74	-3.96	0.78	1589
NMN-[60]	-5.35	-4.01	1.34	925
NMG-[60]	-5.38	-4.01	1.36	911
NMA-[60]	-5.45	-4.02	1.43	869
NMC-[60]	-5.41	-3.98	1.43	867
NMT-[60]	-5.47	-4.01	1.45	850
$PC_{60}BM$	-5.51	-4.02	1.49	832

Table 7.2 B3LYP/6-31G** calculated energy of the frontier HOMO and LUMO orbitals, the HOMO-LUMO gap for C_{60} and the six modelled fullerene derivatives in 1,2-dichlorobenzene (1,2-DCB). The derivatives are ranked according to increasing HOMO-LUMO gap energy.

7.4.2.1 Chlorobenzene

In a similar manner to the calculations in the presence of 1,2-DCB, the calculations in the presence of chlorobenzene (CB) also showed an increase in the HOMO energies, a decrease in the LUMO energies as well as lower HOMO-LUMO gap energies (**Table 7.3**).

	HOMO (eV)	LUMO (eV)	HOMO – LUMO gap	
			(eV)	(nm)
C_{60}	-5.78	-4.11	1.66	746
NMD-[60]	-4.75	-3.97	0.78	1589
NMG-[60]	-5.38	-4.03	1.35	918
NMN-[60]	-5.41	-4.03	1.38	898
NMA-[60]	-5.48	-4.04	1.44	861
NMC-[60]	-5.44	-3.99	1.45	855
NMT-[60]	-5.50	-4.03	1.47	843
$PC_{60}BM$	-5.52	-4.04	1.48	837

Table 7.3 B3LYP/6-31G** calculated energy of the frontier HOMO and LUMO orbitals, the HOMO-LUMO gap for C_{60} and the six modelled fullerene derivatives in chlorobenzene (CB). The derivatives are ranked according to increasing HOMO-LUMO gap energy.

7.4.2.2 Toluene

Likewise, results from calculations with toluene as the solvent also produced similar results to the results in **Tables 7.2** and **7.3** where 1,2-dichlorobenzene and chlorobenzene were solvents.

	HOMO (eV)	LUMO (eV)	HOMO – LUMO gap (eV)	(nm)
C_{60}	-5.75	-4.09	1.66	746
NMD-[60]	-4.78	-3.98	0.80	1549
NMG-[60]	-5.42	-4.08	1.34	921
NMN-[60]	-5.52	-4.09	1.43	867
NMA-[60]	-5.53	-4.06	1.46	848
NMC-[60]	-5.46	-3.99	1.47	843
NMT-[60]	-5.56	-4.06	1.49	832
$PC_{60}BM$	-5.56	-4.07	1.49	832

Table 7.4 B3LYP/6-31G** calculated energy of the frontier HOMO and LUMO orbitals, the HOMO-LUMO gap for C_{60} and the six modelled fullerene derivatives in toluene. The derivatives are ranked according to the increasing HOMO-LUMO gap energy.

7.4.2.3 Water

	HOMO (eV)	LUMO (eV)	HOMO – LUMO gap (eV)	HOMO – LUMO gap (nm)
C_{60}	-5.71	-4.05	1.66	746
NMD-[60]	-4.72	-3.95	0.77	1610
NMN-[60]	-5.28	-3.97	1.31	946
NMC-[60]	-5.36	-3.98	1.38	898
NMG-[60]	-5.37	-3.98	1.39	892
NMA-[60]	-5.41	-3.99	1.41	879
NMT-[60]	-5.40	-3.99	1.41	879
$PC_{60}BM$	-5.49	-4.00	1.49	832

Table 7.5 B3LYP/6-31G** calculated energy of the frontier HOMO and LUMO orbitals, the HOMO-LUMO gap for C_{60} and the six modelled fullerene derivatives in water. The derivatives are ranked according to the increasing HOMO-LUMO gap energy.

Comparison of the calculated LUMO energies in the presence of the three organic solvents showed that calculations in the presence of toluene (**Table 7.5**) had the biggest change from the gas phase, and the smallest change was observed when 1,2-dichlorobenzene was the solvent (**Table 7.1**). The LUMO energies for the novel C_{60} derivatives were consistent in each solvent (0.01 to 0.07 eV). Based on these results, NMD-[60] and NMC-[60] have the highest LUMO energy in all environments and are thus most likely to show as good or better

power conversion efficiency compared to $PC_{60}BM$. Furthermore, the changes in the HOMO, LUMO and HOMO-LUMO gap energies of $PC_{60}BM$ were also consistent with the changes in the six modelled derivatives.

7.4.3 Frontier Orbital Distribution

The frontier HOMO and LUMO orbital distribution of the derivatives in the gas phase and in solution were compared for each derivative in order to clearly observe the effects of the different solvents on each derivative.

7.4.3.1 N-methyl-2-thymin-N-ylfulleropyrrolidine (NMT-[60])

The frontier HOMO and LUMO orbital distributions were modified upon functionalisation of C_{60} to NMT-[60] (**Figure 7.2**). Both HOMO and LUMO orbitals were no longer five-/three-fold degenerate in energy respectively, and so the orbital distribution on NMA-[60] was not as dense as it was on pristine C_{60}. Comparison of the frontier LUMO orbitals in the gas phase and the different solvent environments on the NMT-[60] derivative showed the same distribution pattern which was localised on the C_{60} cage (**Figure 7.2**). By contrast, the frontier HOMO orbitals showed similarities in the distribution of the orbitals calculated in the presence of 1,2-DCB, CB and toluene as well as the gas phase. In these environments, the HOMO electron distribution was concentrated on the C_{60} cage. In addition, 1,2-DCB as the solvent, the HOMO electron density was also located on the five-membered pyrrolidine ring as well as the C_{60} cage. With water as the solvent, the majority of the frontier HOMO orbitals were located on the five-membered pyrrolidine ring, with some orbitals present on the thymine structure. The similarities of the electron density of the HOMO frontier orbitals with distribution observed when water, CB and toluene were solvents indicates that 1,2-DCB has the combined effect of water, CB and toluene on NMT-[60] (**Figure 7.2**).

There were no significant contributions on the C_{60} cage, apart from the area around the pyrrolidine ring and opposite the position of the pyrrolidine ring, on the C_{60} cage.
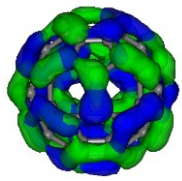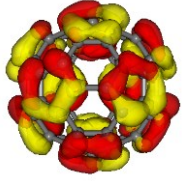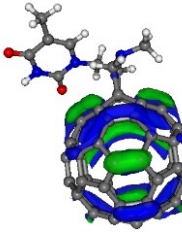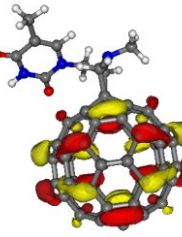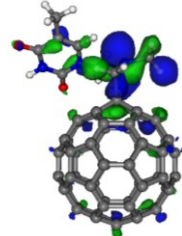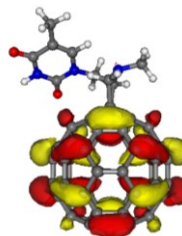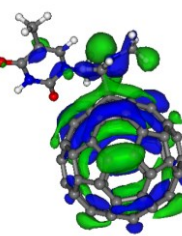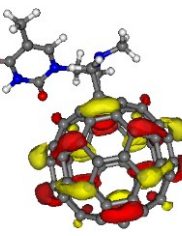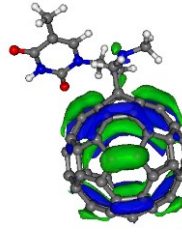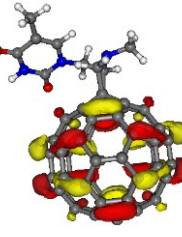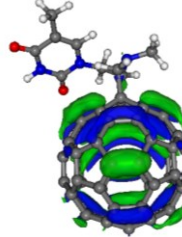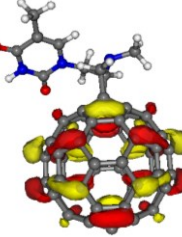
	HOMO	**LUMO**
C₆₀		
NMT-[60] (Gas phase)		
NMT-[60] (Water)		
NMT-[60] (1,2-DCB)		
NMT-[60] (CB)		
NMT-[60] (Toluene)		

Figure 7.2 B3LYP/6-31G** calculated frontier HOMO and LUMO orbitals for C_{60} and NMT-[60] in the gas phase, water, dichlorobenzene (DCB), chlorobenzene (CB) and toluene. Isosurface = 1.50 au.

The location of the LUMO electron density shows that nucleophilic attack will occur on the C_{60} cage of the derivative. This result is not unexpected, as C_{60} is known to possess triply degenerate lowest unoccupied molecular orbitals, which facilitate its behaviour as an excellent electron acceptor.[6] Thus, the fullerene retains some of its basic characteristics as can be observed in **Figure 7.2**, from the C_{60} cage to NMT-[60] in the gas phase.

The difference in the electron distribution is interesting as it gives an insight into the possible behaviour of NMT-[60] in the different solvents. The near-absence of the HOMO on the C_{60} cage illustrates the near impossible task of oxidising the C_{60} cage in water. More remarkable is the delocalization of the HOMO in 1,2-DCB compared to the localization of the orbitals on the C_{60} cage in the other organic solvents, CB and toluene. Additionally, the presence of the HOMO on the C_{60} cage as well as the pyrrolidine ring and thymine entity give an indication of the oxidation sites in the different solvents; where the more polar solvents enable oxidation on the pyrrolidine and thymine structures.

7.4.3.2 N-methyl-2-adenin-N-ylfulleropyrrolidine (NMA-[60])

As observed for NMT-[60], the frontier LUMO orbitals were localized on the C_{60} cage in the gas phase and solvent environments, indicating that reduction would occur on the C_{60} cage in this derivative. The frontier HOMO orbital distribution with water, 1,2-DCB and CB as solvents was very similar (**Figure 7.3**). The HOMO was delocalized on the five-membered pyrrolidine ring and the C_{60} cage. In all three solvents, the HOMO was delocalized on both the pyrrolidine ring and sparsely around the C_{60} cage. A few orbitals were scattered on the adenine entity whilst there was a larger concentration of orbitals on the five-membered ring. There was also a small concentration of orbitals on the C_{60} cage opposite the pyrrolidine ring.
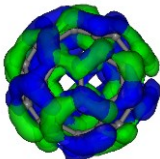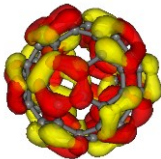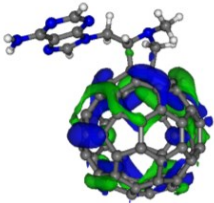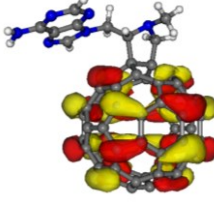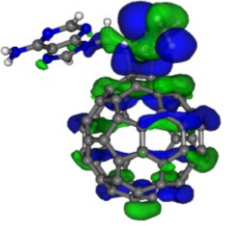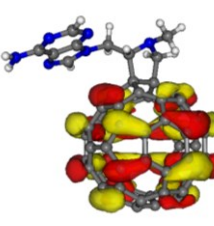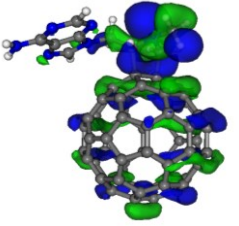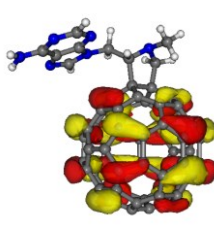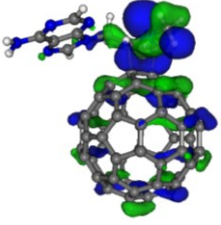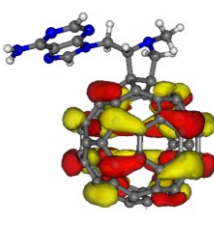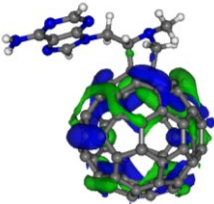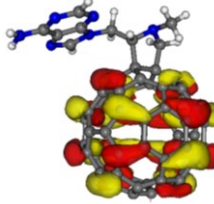
	HOMO	LUMO
C_{60}		
NMA-[60] (Gas phase)		
NMA-[60] (Water)		
NMA-[60] (1,2-DCB)		
NMA-[60] (CB)		
NMA-[60] (Toluene)		

Figure 7.3 B3LYP/6-31G** calculated HOMO and LUMO orbitals for C_{60} and NMA-[60] in the gas phase, water, 1,2-dichlorobenzene (1,2-DCB), chlorobenzene (CB) and toluene. Isosurface = 1.50 au.

The frontier HOMO distribution patterns offer insight into the orientation of the derivative on the zeolite surface or inside the zeolite channels. Since the HOMO orbitals were concentrated on the pyrrolidine ring, NMA-[60] has the potential to adsorb in the channels with the pyrrolidine ring pointing towards the channels if they are wide enough. However, since there is also a small concentration of orbitals on the C_{60} cage opposite the pyrrolidine ring, there is also the possibility that NMA-[60] may orientate with the C_{60} ball sitting on the opening to the zeolite channels with the adenine entity pointing outwards, facilitating self-assembly.

7.4.3.2.1 Electronic Properties and Distribution of bis-BOC-NMA-[60]

In order to increase the solubility of adenine and subsequently NMA-[60], the starting materials of the reaction required BOC-protection. To assess any effects of attaching the BOC group on NMA-[60], bis-BOC-NMA-[60] was modelled and analysed. The calculated properties versus those of NMA-[60] in the gas phase are shown in **Table 7.6** and **Figure 7.4**.

There were obvious differences in conformational structure between NMA-[60] and bis-BOC-NMA-[60]: these 6-31G** optimised structures showed the adenine entity on bis-BOC-NMA-[60] approximately perpendicular to the fullerene cage, whereas for NMA-[60], it was almost at right angles to the cage (**Figure 7.4**).

	HOMO (eV)	LUMO (eV)	HOMO – LUMO gap (eV)	(nm)
NMA-[60] (gas phase)	-5.59	-3.05	2.54	487
Bis-Boc-NMA-[60] (gas phase)	-5.76	-3.21	2.55	486

Table 7.6 B3LYP/6-31G** calculated energy of the frontier HOMO and LUMO orbitals and the HOMO-LUMO gap of the NMA-[60] and bis-boc-NMA-[60] the gas phase.
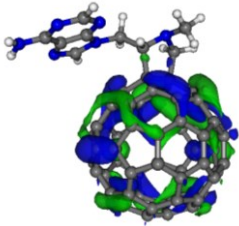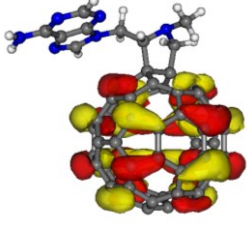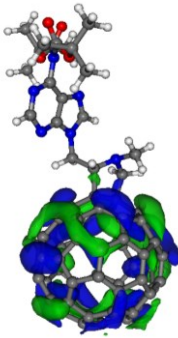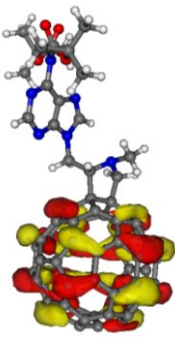
	HOMO	LUMO
NMA-[60] **(gas phase)**		
Bis-Boc-NMA-[60] **(gas phase)**		

Figure 7.4 B3LYP/6-31G** calculated HOMO and LUMO orbitals for bis-boc-NMT-[60] and bis-boc-NMA-[60] in the gas phase. Isosurface = 1.50 au.

The electron distribution in the HOMO and LUMO of bis-BOC-NMA-[60] were both concentrated on the fullerene cage as observed in NMA-[60] (**Figure 7.3**). This indicates that the BOC group had little influence on the electronic properties.

7.4.3.3 N-methyl-2-cytosin-N-ylfulleropyrrolidine-(NMC-[60])

There was a clear change in the frontier HOMO and LUMO orbital distribution from C_{60} to NMC-[60] (**Figure 7.5**). In the gas phase, the frontier HOMO orbitals were localized on the C_{60} cage. This distribution pattern was also observed when the HOMO distribution was plotted with toluene as the solvent, indicating the area most likely to undergo electrophilic attack. However, in CB, the electron density distribution pattern was on both the C_{60} cage and the pyrrolidine ring. The electron distribution calculated in water and 1,2-DCB also displayed a delocalised distribution which was concentrated on the pyrrolidine ring with a few orbitals on the C_{60} cage.
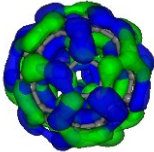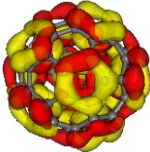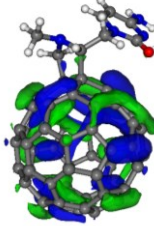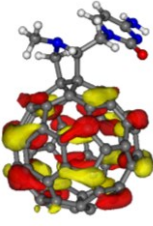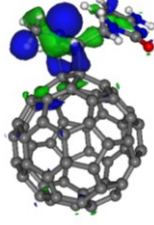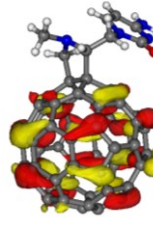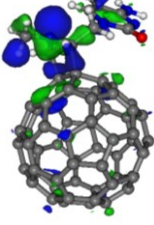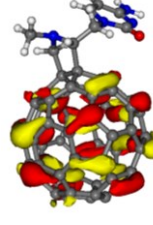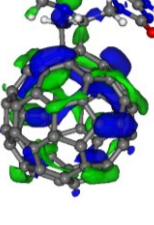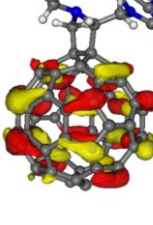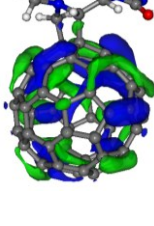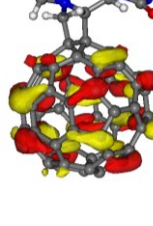
	HOMO	LUMO
C$_{60}$ **(Gas phase)**		
NMC-[60] **(Gas phase)**		
NMC-[60] **(Water)**		
NMC-[60] **(1,2-DCB)**		
NMC-[60] **(CB)**		
NMC-[60] **(Toluene)**		

Figure 7.5 B3LYP/6-31G** calculated HOMO and LUMO orbitals for C$_{60}$ and NMC-[60] in the gas phase, water, 1,2-dichlorobenzene (1,2-DCB), chlorobenzene (CB) and toluene. Isosurface = 1.50 au.

The localized HOMO electron distribution on the pyrrolidine ring and the area surrounding the ring on the C_{60} cage suggests that C_{60} ring protruding out of the zeolite channels towards the positively charged ion channels of the zeolite, acting as a stopcock. This orientation would not facilitate self-assembly. CB or toluene would be more appropriate solvents to use for dissolving NMC-[60] in order to enable self-assembly to occur. Since with these solvents NMC-[60] would orientate with the C_{60} ring in the channel opening and the cytosine entity pointing outwards.

7.4.3.4 N-methyl-2-guanin-N-ylfulleropyrrolidine

Figure 7.6 shows the electron density distribution of the frontier HOMO and LUMO orbitals on the NMG-[60] derivative. Since the HOMO orbitals were localised on the guanine entity in all the solvents environments and the gas phase, oxidation reactions would occur on this region of the derivative. By contrast, the frontier LUMO orbitals were entirely localized on the C_{60} cage, as observed in the aforementioned derivatives (**Figure 7.2**, **7.3**, and **7.5**). These figures confirm that C_{60} is an excellent fullerene acceptor and after functionalization some of these properties are retained.[17]
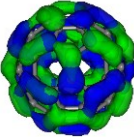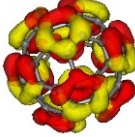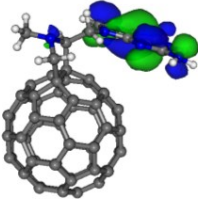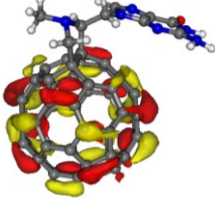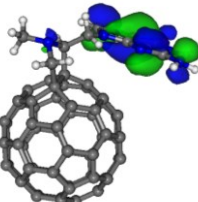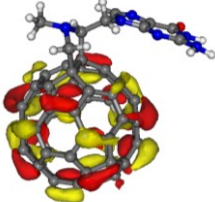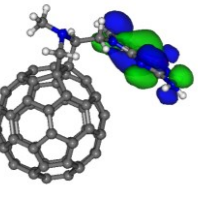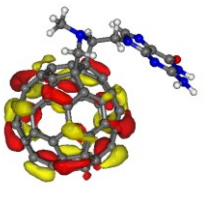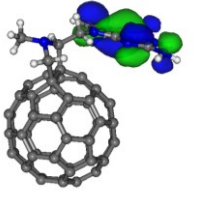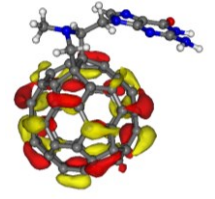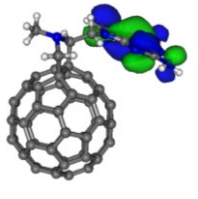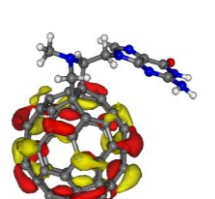
	HOMO	LUMO
C_{60} ring		
NMG-[60] (gas phase)		
NMG-[60] (water)		
NMG-[60] (1,2-DCB)		
NMG-[60] (CB)		
NMG-[60] (Toluene)		

Figure 7.6 B3LYP/6-31G** calculated HOMO and LUMO orbitals for C_{60} and NMG-[60] in the gas phase, water, 1,2-dichlorobenzene (1,2-DCB), chlorobenzene (CB) and toluene. Isosurface = 1.50 au.

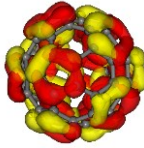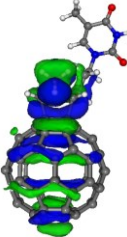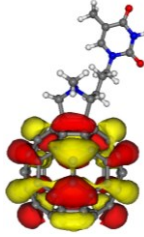## 7.4.3.5 N-methyl-2-N-propylthymin-N-ylfulleropyrrolidine
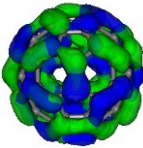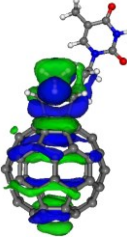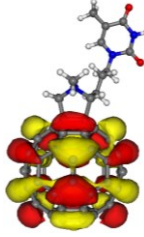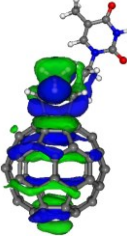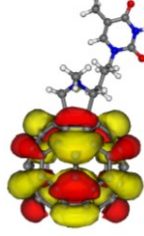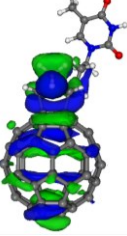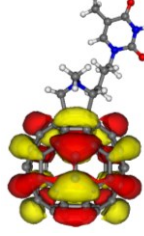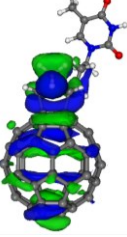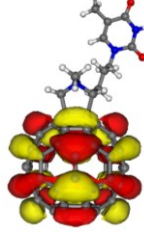
	HOMO	LUMO
C$_{60}$ (gas phase)		
NMN-[60] (gas phase)		
NMN-[60] (water)		
NMN-[60] (1,2-DCB)		
NMN-[60] (CB)		
NMN-[60] (Toluene)		

Figure 7.7 B3LYP/6-31G** calculated HOMO and LUMO orbitals for C$_{60}$ and NMN-[60] in the gas phase, water, 1,2-dichlorobenzene (1,2-DCB), chlorobenzene (CB) and toluene. Isosurface = 1.50 au.

The frontier orbital distribution of NMN-[60] in the gas phase, water, 1,2-DCB, CB and toluene is shown in **Figure 7.7**. The electron distribution of the HOMO orbitals distribution in all the solvents as well as the gas phase as delocalized on the five-membered pyrrolidine ring as well as the C_{60} cage. The orbitals were concentrated on the pyrrolidine ring as well as on the C_{60} cage, close to the five-membered ring as well as opposite side of the five-membered ring. There was no HOMO distribution on the N-propylthymine entity. The delocalisation of the electron density in the HOMO around both the C_{60} cage and the pyrrolidine ring makes it hard to predict how these molecules will orientate at the zeolite channel openings. Due to the lack of electron density on the functional group, it is potentially likely that the C_{60} cages will sit in the zeolite opening.

By contrast, the LUMO orbitals were localized on the C_{60} cage, as observed in all the DNA base-functionalized derivatives. Just like in the frontier HOMO orbital distribution, the LUMO distribution was less concentrated on the C_{60} cage because of the loss of degeneracy after functionalization. However, the presence of frontier LUMO orbitals on the C_{60} cage mean that NMN-[60] retains some of the electron accepting properties of pristine C_{60}.

Despite the similarities of NMT-[60] and NMN-[60] derivatives with the thymine entity attached, there were no significant similarities observed in the electron distribution. The only similarities observed were for the HOMO electronic distribution for NMT-[60] in water and 1,2-DCB where the orbitals were also delocalised on the five-membered pyrrolidine ring as well as the C_{60} cage, as observed in the NMN-[60] derivative in all the environments (**Figure 7.2** and **Figure 7.7** respectively).

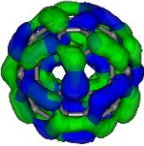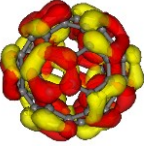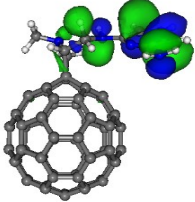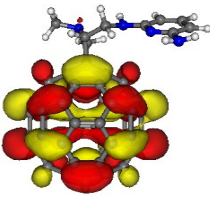## 7.4.3.6 N-methyl-2-diaminopyridinefulleropyrrolidine (NMD-[60])
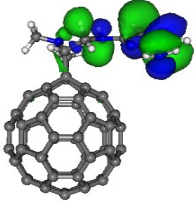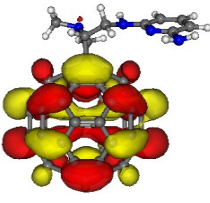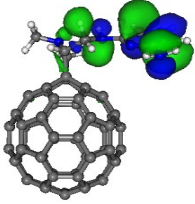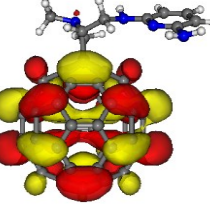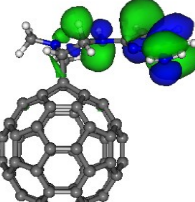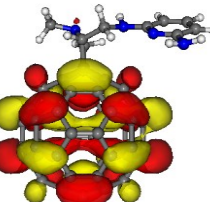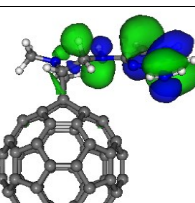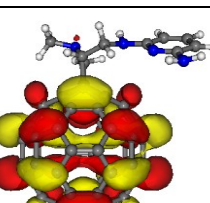
	HOMO	LUMO
C_{60} (gas phase)		
NMD-[60] (gas phase)		
NMD-[60] (water)		
NMD-[60] (1,2-DCB)		
NMD-[60] (CB)		
NMD-[60] (Toluene)		

Figure 7.8 B3LYP/6-31G** calculated HOMO and LUMO orbitals for C_{60} and NMD-[60] in the gas phase, water, 1,2-dichlorobenzene (1,2-DCB), chlorobenzene (CB) and toluene. Isosurface = 1.50 au.

The electron distribution of the frontier HOMO orbitals on NMD-[60] was delocalized on the five-membered ring and the 2,6-diaminopyridine entity (**Figure 7.8**). A similar distribution pattern was observed for NMG-[60] (**Figure 7.6**), where the frontier HOMO distribution was localized on the guanine entity. For NMD-[60], this distribution pattern was observed in the gas phase and in all the solvent environments. As observed in NMG-[60], the location of the frontier HOMO orbitals on the 2,6-diaminopyridine entity and the pyrrolidine ring indicates that when adsorbed on a zeolite surface, this derivative will likely have the 2,6-diaminopyridine entity inside the channels. This conformation would not enable self-assembly, but if the H-bonding partner, NMN-[60] was adsorbed first, the H-bonds might direct the conformation of the dimer. In comparison, the frontier LUMO orbitals were localized on the C_{60} cage as observed in the DNA base-functionalized derivative and NMN-[60] (**Figure 7.6**).

7.5 Conclusion

The HOMO-LUMO gap for the modelled derivatives in the gas phase was lower than the gap for C_{60} in the gas phase. The gap further decreased in all the solvents due to the increase in the HOMO energies. The electron density distribution in the HOMO indicates that the oxidation site of each derivative will be influenced by the solvent chosen. In contrast, the electron distribution in the LUMO was consistently concentrated on the C_{60} cage for all the derivatives in all environments. The electron-accepting ability of all six fullerene derivatives could be further explored experimentally.

Comparison of the calculated LUMO energy for the monomers in the gas phase has two derivatives (NMD-[60] and NMC-[60]) with energies higher than that calculated for $PC_{60}BM$. From this data, these two derivatives could potentially function as improved electron acceptors in organic solar cells. The other 4 derivatives having calculated LUMO energies that are only slightly lower compared to $PC_{60}BM$, the differences are small, (0.06 (2 %) - 0.08 eV (2.6 %)). Therefore, these derivatives could also potentially be used as electron acceptors in organic solar cells.

Due to van der Waals attractions between the solvents used in the organic solar cell fabrication and the electron donor and acceptor, it is not possible to remove all the solvent molecules during the drying process. The decrease in the LUMO energies for the derivatives including $PC_{60}BM$ shows that the presence of the solvents would lower the efficiency of the organic solar cells

It is predicted that the orientation of the monomers on the zeolite surface will be influenced by the HOMO electron distribution. Therefore, once the experimental methods have been refined, it would also be interesting to analyse the orientation of the monomers and self-assembled dimers on the zeolite surface using SEM or TEM.

Chapter 8

Dimers

8.1 Introduction

Fullerene derivatives are attractive building blocks for supramolecular chemistry because of their electronic properties.[138] They retain the main properties of C_{60}, namely the ability to accept up to six electrons upon reduction, and unlike pristine C_{60}, became more soluble in organic solvents. Furthermore, the multiple attachment points on C_{60} enable easy chemical functionalization, which allows the incorporation of most organic functional groups.[139] Although fullerene dimers exhibit poor solubility in common organic solvents, they are of interest because, in comparison to a mono fullerene, the molar absorptivity is expected to improve because of the presence of two C_{60} cages. As a result, the dimer can act as a multi-electron acceptor system with potential function in photovoltaic applications.[139]

There are no previous studies into the self-assembly of C_{60} derivatives functionalised with DNA bases, however, some fullerenes have been functionalized with uracil.[77, 81, 140, 141] In their study, Marczak *et al*. synthesized N-methyl-2-uracilfulleropyrrolidine,[77] in order to analyse and understand the base-pairing mechanisms between this C_{60} derivative and adenine. Computational and experimental studies were performed on the modelled structures. N-methyl-2-uracilfulleropyrrolidine was paired with adenine and computational analysis was carried out to determine the geometry of the base pair.[77] The uracil and adenine rings were positioned in a plane without steric hindrance from the bulky C_{60} ring. From the H-bond distances, it was found that the pair formed a stable Watson-Crick complex based on the lengths of the H-bonds.[77]

The same uracil C_{60} derivative was used by D'Souza *et al*.[81] to form a donor-acceptor conjugate with 2-aminopurine (2-AP), *via* Watson-Crick H-bonds (**74**), in order to experimentally analyse its photoinduced electron transfer abilities. 2-AP was chosen for this

experiment because it is a fluorescent isomer of adenine.[81] In this study, it was found that electron transfer occurred in the conjugate by the presence of the electron transfer products, 2-AP$^{\bullet+}$ and N-methyl-2-uracilfulleropyrrolidine$^{\bullet-}$.[81]
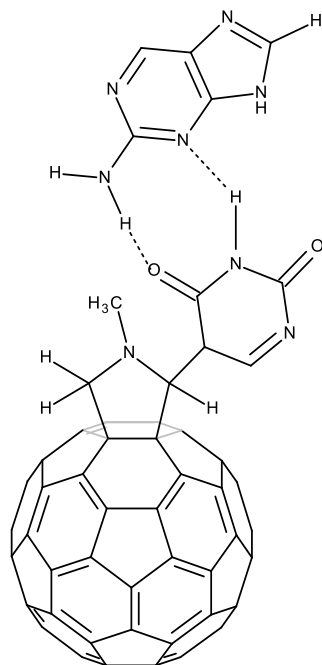
(**74**)

Figure 8.1 A donor-acceptor conjugate with 2-aminopurine and N-methyl-2-uracilfulleropyrrolidine (**74**)

In another study, an H-bonding fullerene also bearing a 2,6-diacylaminopyridine (DAP-C$_{60}$) was synthesized.[140] In addition, a uracil-bearing U-PPV structure was prepared for self-assembly.[142] In this self-assembled system, the uracil-bearing U-PPV acted as the electron donor while DAP-C$_{60}$ was the electron acceptor. The electron transfer possibilities of this supramolecule encouraged the authors to look into the study of photoinduced charge transfer within this system.[140]

The same derivative (DAP-C$_{60}$) was synthesized and mixed with a compound bearing a uracil moiety.[140] A donor-acceptor system was created *via* self-assembly, with DAP-C$_{60}$ as the electron acceptor. Evidence of H-bonding was observed from the downward shift of the amidic protons on DAP-C$_{60}$ in the ^{1}H NMR spectrum. These protons were said to shift

downfield because of the decrease in their electron densities, which results in the proton signals being shifted to lower magnetic fields.[140]

In a similar study to that carried out by D'Souza *et al.*, the photoinduced electron transfer of a porphyrin-fullerene dyad with Watson-Crick bonding interactions was analysed.[80] The porphyrin was appended with cytidine while the C_{60} derivative was appended with guanosine. To enhance their solubility, the protecting group, tert-butyldimethylsilyl (TBDMS) was used. After self-assembly of the two structures, spectroscopy methods confirmed the formation of the porphyrin radical cation and the fullerene radical anion which indicated electron transfer.[80]

Studies involving the self-assembly C_{60} derivatives have not been confined to Watson-Crick base pairing. After their first works of synthesizing an H-bonding fullerene with a single coupling unit capable of forming a dimer *via* four H-bonds (**75**),[143] Sanchez, Hummelen and Rispens reported the first example of an H-bonded fullerene array with two coupling units, thus forming eight H-bonds (**76**).[144]
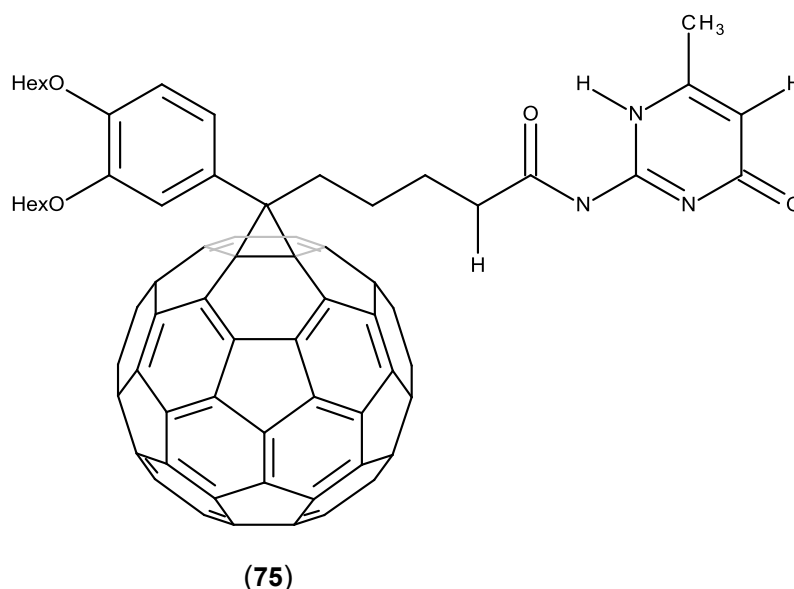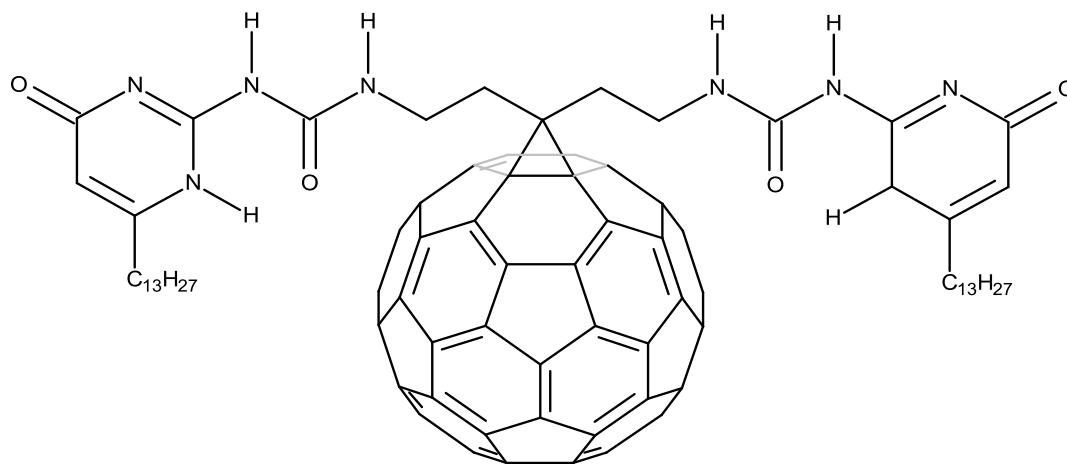

(**75**)

Figure 8.2 Structure of an H-bonding fullerene with a single coupling unit capable of forming a dimer *via* four H-bonds (**75**).

(76)

Figure 8.3 Structure of an H-bonding fullerene array with two coupling units capable of forming a dimer *via* eight H-bonds (**76**).

Whilst, for the DNA-based derivatives, the focus of this study was on the Watson-Crick binding modes, the alternative binding modes, Hoogsteen, Wobble and the self-dimer, were also modelled. Their electronic properties were compared to those obtained for the derivatives with Watson-Crick binding modes as well as the NMD-[60]-NMN-[60] dimer.

	AT	GC
Watson-Crick		
Hoogsteen	**Wobble**	**GG Self-Dimer**

Table 8.1 Pictures of the Watson-Crick, Hoogsteen and wobble binding modes as well as the guanine self-dimerization.

8.2 Computational Method

The electronic properties for the NMA-[60]-NMT-[60] (**77**), NMG-[60]-NMC-[60] (**78**) and NMN-[60]-NMD-[60] (**79**) dimers in the dumbbell and seesaw orientations were obtained using DFT (B3LYP/6-31G**). In the dumbbell orientation, the C_{60} cages were modelled on the same plane whereas in the dumbbell orientation, the C_{60} cages were on opposite planes. These two orientations were modelled in order to assess the preferred resting orientation of the dimers on a surface like that of the zeolites. The dumbbell orientation might be expected to rest on the zeolite surface with the two C_{60} cages on the surface. Depending on the spacing of the zeolite channels, both spheres could be at channel openings or one sphere at the opening and one on the planar surface of the zeolite.

8.3 Computational Results and Discussion

8.3.1 Global Minimum Conformers

The B3LYP/6-31G** optimised global minimum conformers for the dumbbell orientations of the NMA-[60]-NMT-[60], NMG-[60]-NMC-[60] and NMN-[60]-NMD-[60] dimers are shown in **Figure 8.4**. In the optimised structure for (**77**) the C_{60} cage for NMT-[60] appeared to rest at an angle and the pyrrolidine ring was not aligned with the pyrrolidine ring on the NMA-[60] derivative. H-bonding DNA bases appear almost perpendicular to the C_{60} cages. In a similar manner, the pyrrolidine rings in the NMG-[60]-NMC-[60] dimer (**78**) were not aligned because the NMC-[60] monomer rested on the side, in relation to the pyrrolidine ring. The adopted global minimum orientation meant that on adsorption on the zeolite surface, the cytosine-based derivative would rest on its side, with the pyrrolidine ring pointing to the side whilst the guanine-based derivative would rest with the pyrrolidine ring pointing upwards. The pyrrolidine rings for the global minimum conformer for the NMN-[60]-NMD-[60] dimer (**79**) in the dumbbell orientation were also not aligned.
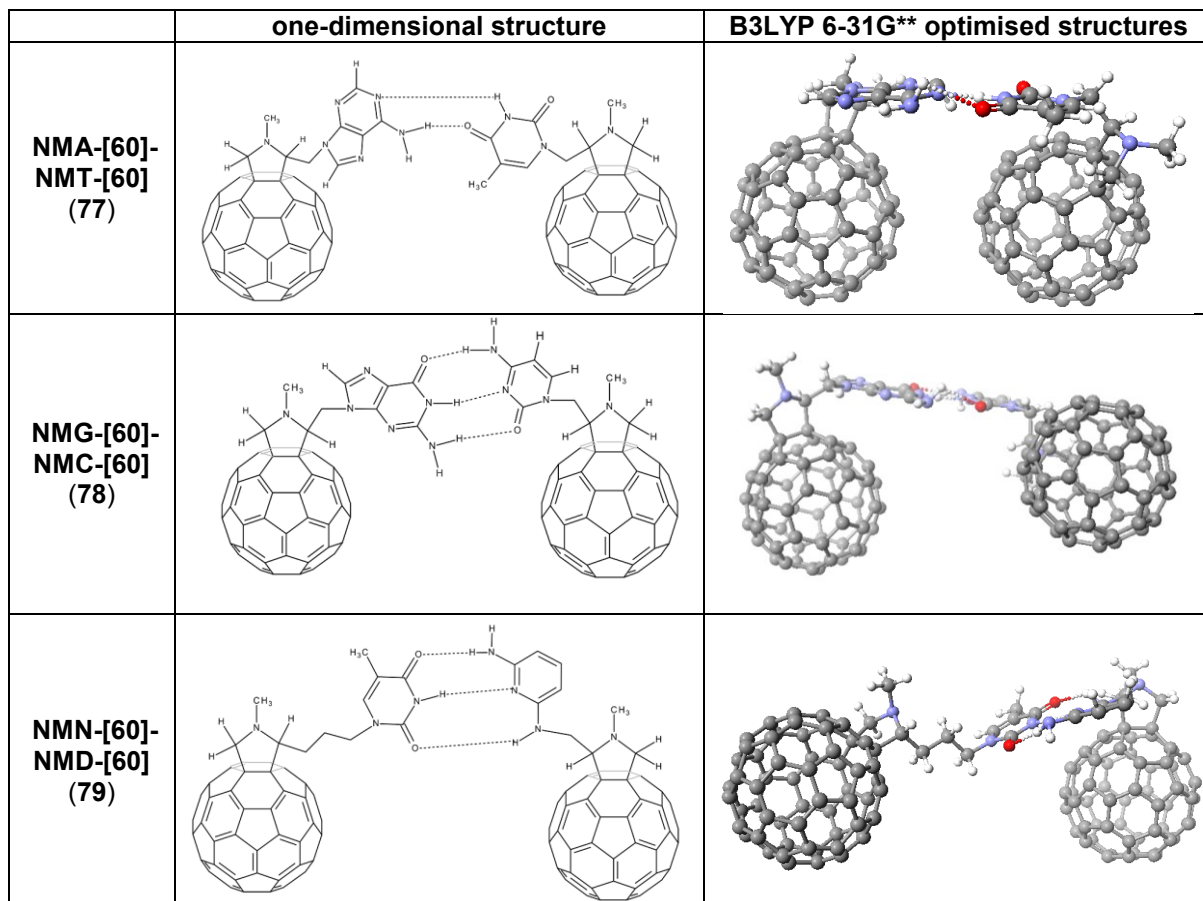
	one-dimensional structure	B3LYP 6-31G** optimised structures
NMA-[60]-NMT-[60] (**77**)		
NMG-[60]-NMC-[60] (**78**)		
NMN-[60]-NMD-[60] (**79**)		

Figure 8.4 The global minimum conformers of the NMA-[60]-NMT-[60] (**77**), NMG-[60]-NMC-[60] (**78**) and NMD-[60]-NMN-[60] (**79**) dimers in the dumbbell conformation after a B3LYP/6-31G** calculation.

In the global minimum conformers for the NMA-[60]-NMT-[60] (**80**), NMG-[60]-NMC-[60] (**81**) and NMN-[60]-NMD-[60] (**82**) dimers in the seesaw formation (**Figures 8.5**) showed the monomers in the NMA-[60]-NMT-[60] and NMG-[60]-NMC-[60] in similar orientations. However, the NMN-[60]-NMD-[60] dimer showed the C_{60} cages on the same plane, as observed in the seesaw orientation (**Figure 8.4**).

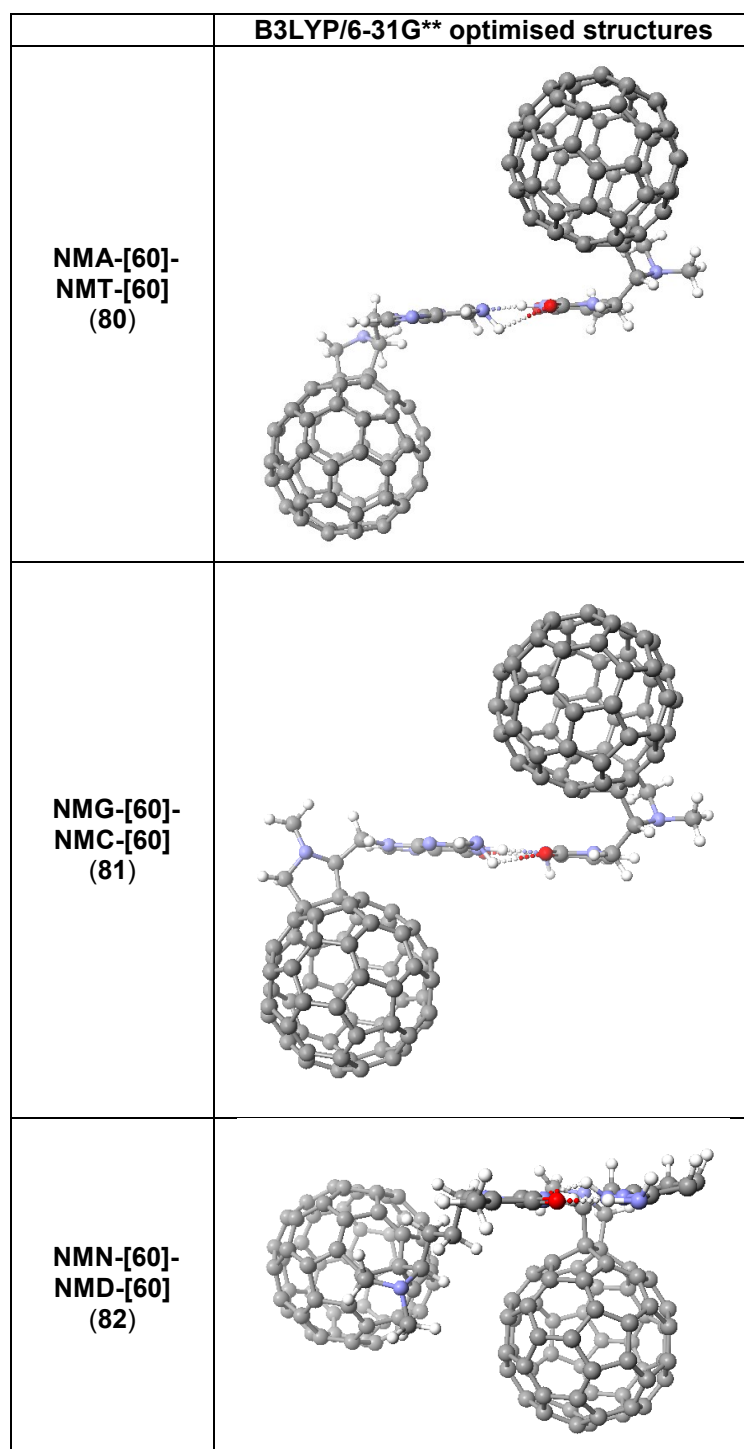|	B3LYP/6-31G** optimised structures
NMA-[60]-NMT-[60] (**80**)	
NMG-[60]-NMC-[60] (**81**)	
NMN-[60]-NMD-[60] (**82**)	

Figure 8.5 The global minimum conformers of the NMA-[60]-NMT-[60] (**80**), NMG-[60]-NMC-[60] (**81**) and NMD-[60]-NMN-[60] (**82**) dimers in the dumbbell conformation after a B3LYP/6-31G** calculation.

8.4 Electronic Properties

8.4.1 Binding Energies

The trend shown in the binding energies of the DNA-based C_{60} derivatives matches the number of H-bonds in each dimer as well as the H-bond lengths (**Table 8.2**). NMG-[60]-NMC-[60] has three H-bonds with an average length of 2.01 Å (dumbbell) and 1.95 Å (seesaw) and the binding energies were -37.6 and -32.2 kcal/mol for the dumbbell and seesaw orientations respectively. The NMA-[60]-NMT-[60] dimer has two H-bonds with an average length of 2.14 Å and 1.95 Å and binding energies of -16.1 and -16.5 kcal/mol for the dumbbell and seesaw orientations respectively (**Table 8.2**). These results show that the shorter the average H-bonds, the more exothermic the binding energy. Furthermore, it is also possible that the shorter H-bonds result in the greater proximity of the C_{60} cages, resulting in more π-π interactions which would lead to higher binding energies. The three H-bonds in NMD-[60]-NMN-[60] had binding energies of -19.2 kcal/mol (1.97 Å) and -19.9 kcal/mol (1.99 Å) for the dumbbell and seesaw orientations. These energies were less exothermic than the NMG-[60]-NMC-[60] dimer but more exothermic than the NMA-[60]-NMT-[60] dimer. These results demonstrate the strength of the three bonds present in NMD-[60]-NMN-[60] and show that these bonds are not as strong as those present in the NMG-[60]-NMC-[60] dimer.

For the NMA-[60]-NMT-[60] and NMD-[60]-NMN-[60] dimers showed very little thermodynamic preference for either orientation. The dumbbell orientation has a slightly lower binding energy (0.4 and 0.7 kcal/mol respectively). A mixture of both orientations is likely to be found when self-assembly occurs. In contrast, the NMG-[60]-NMC-[60] dimer is predicted to prefer the dumbbell orientation (**Table 8.2**).

The binding energy corrected for basis set superposition error (BSSE) showed that the binding energies for the dumbbell orientations were underestimated more than the binding energies for seesaw orientations (**Table 8.3**). The NMA-[60]-NMT-[60] and NMG-[60]-NMC-

[60] dimers were both underestimated by approximately 5 kcal/mol and the NMN-[60]-NMD-[60] dimer by 7 kcal/mol. By contrast, in the seesaw conformations, the NMA-[60]-NMT-[60] dimer was overestimated by 4 kcal/mol while the other two dimers were underestimated by a negligible energy of 0.03 kcal/mol for the NMG-[60]-NMC-[60] dimer and 0.4 kcal/mol for the NMN-[60]-NMD-[60] dimer (**Table 8.3**).

The alternative binding H-binding modes of the bases were modelled for comparison with the standard DNA-base binding modes. Only one orientation, the seesaw orientation, was used in these calculations. Like the NMA-[60]-NMT-[60] dimer, the Hoogsteen, Wobble and self-dimer used two H-bonds for binding. Comparison of the BSSE corrected binding energies of the first three dimers with these structures showed small differences for the Hoogsteen and Wobble binding modes with the self-dimer exhibiting an endothermic binding energy. Since there are small differences between the Watson-Crick binding modes and the Hoogsteen and Wobble binding modes, it is expected that the former binding modes would preferred since they are more dominant in DNA.

	Number of H-bonds	Length of H-bonds (Å)	Binding Energy (kcal/mol)
NMA-[60]-NMT-[60] **(dumbbell)**	**(2)** N6-O4 N1-N3	2.51 1.77	-16.5
NMA-[60]-NMT-[60] **(seesaw)**		3.13 1.77	-16.1
NMG-[60]-NMC-[60] **(dumbbell)**	**(3)** O6-N4 N1-N3 N2-O2	1.76 1.87 2.41	-37.6
NMG-[60]-NMC-[60] **(seesaw)**		1.75 1.87 2.24	-32.2
NMN-[60]-NMD-[60] **(dumbbell)**	**(3)** C2-C2 N3-N1 C4-C6	1.92 1.87 2.12	-19.2
NMN-[60]-NMD-[60] **(seesaw)**		2.15 1.93 1.90	-19.9
NMA-[60]-NMT-[60] **(Hoogsteen)**	**(2)** N3-N7 C4-C6	1.78 1.88	-18.8
NMG-[60]-NMT-[60] **(Wobble)**	**(2)** N1-C2 C6-N3	1.74 1.79	-18.6
NMG-[60]-NMG-[60] **(self-dimer)**	**(2)** N1-C6 C6-N1	1.70 1.71	-29.0

Table 8.2 B3LYP/6-31G** calculated binding energies, H-bond lengths and number of H-bonds for the three dimers in the dumbbell and seesaw conformations and the alternative binding modes.

	Dimer Orientation	BSSE corrected Binding Energy (kcal/mol)	Binding Energy (kcal/mol)	Energy difference (kcal/mol)
NMA-[60]-NMT-[60]	Dumbbell	-21.2	-16.5	-4.7
	Seesaw	-20.0	-16.1	-3.9
NMG-[60]-NMC-[60]	Dumbbell	-42.6	-37.6	-5.0
	Seesaw	-32.2	-32.2	0
NMN-[60]-NMD-[60]	Dumbbell	-26.2	-19.2	-7.0
	Seesaw	-19.4	-19.9	0.5
NMA-[60]-NMT-[60]	Hoogsteen	-19.9	-18.8	-1.1
NMG-[60]-NMT-[60]	Wobble	-24.6	-18.6	-6.0
NMG-[60]-NMG-[60]	Self-Dimer	18.3	-29.0	47.3

Table 8.3 Comparison of the BSSE corrected energies with the uncorrected energies of the three dimers in the dumbbell and seesaw conformations and the alternative binding modes.

	Monomer HOMO LUMO (eV)	Dimer Orientation	HOMO (eV)	LUMO (eV)	HOMO-LUMO gap (eV)	(nm)
C$_{60}$			-5.89	-3.13	2.76	449
NMA-[60]-NMT-[60]	Adenine -5.59 -3.05	Dumb bell	-5.53	-3.01	2.52	492
	Thymine -5.60 -3.06	Seesaw	-5.55	-3.07	2.48	499
NMG-[60]-NMC-[60]	Guanine -5.56 -3.08	Dumb bell	-5.00	-3.04	1.96	632
	Cytosine -5.45 -2.92	Seesaw	-5.01	-3.07	1.94	639
NMD-[60]-NMN-[60]	Diamino pyridine -5.06 -2.90	Dumb bell	-4.84	-3.11	1.73	716
	Npropyl thymine -5.63 -3.08	Seesaw	-4.79	-3.07	1.72	720

Table 8.4 B3LYP/6-31G** calculated HOMO, LUMO and HOMO-LUMO gap energies of the three dimers in the dumbbell and seesaw conformations in the gas phase.

8.4.2 Frontier Orbitals

Table 8.4 shows the electronic properties of the dimers from the B3LYP/6-31G** calculations in the gas phase. The frontier LUMO energies for all three dimers were -3.07 eV for the seesaw conformations. The differences for the dumbbell conformations were from -3.11 eV to -3.04 eV (a range of 0.07 eV), indicating that the LUMOs were almost identical in energy. This can be related to the almost identical electron distribution concentration on the C_{60} cage seen in each LUMO (**Figure 8.6**, **8.7**, **8.8**) and in the LUMO distribution of the monomers. The changes in the LUMO energy from the monomers to the dimers suggest that dimer formation will affect the optical properties of the structures.

In contrast, the HOMO energy levels showed significant differences from dimer to dimer. However, smaller differences were observed between the individual dimers in the dumbbell and seesaw orientations. The NMA-[60]-NMT-[60] dimer had the lowest (most negative) energy in the seesaw conformation (-5.55 eV) (**Figure 8.6**). The similar HOMO energies for the NMD-[60]-NMN-[60] (-4.84 eV and -4.79 eV) and NMG-[60]-NMC-[60] (-5.00 eV and -5.01 eV) dimers can be related to the similar frontier HOMO orbital distribution. In contrast, the HOMO distribution for the NMA-[60]-NMT-[60] in the seesaw configuration which had the highest HOMO energy (-5.55 eV), was localized on the C_{60} cages.

The frontier orbital distributions for the dimers were plotted (**Figure 8.6**, **8.7** and **8.8**). For the dimers that showed degeneracy in their HOMO or LUMO energies and the HOMO -1/LUMO +1 that lay within ± 2 kcal/mol, the HOMO -1 and LUMO +1 orbitals were also plotted.

|	HOMO, HOMO-1	LUMO, LUMO+1
NMA-[60]-NMT-[60] (dumbbell)		
NMA-[60]-NMT-[60] (seesaw)		
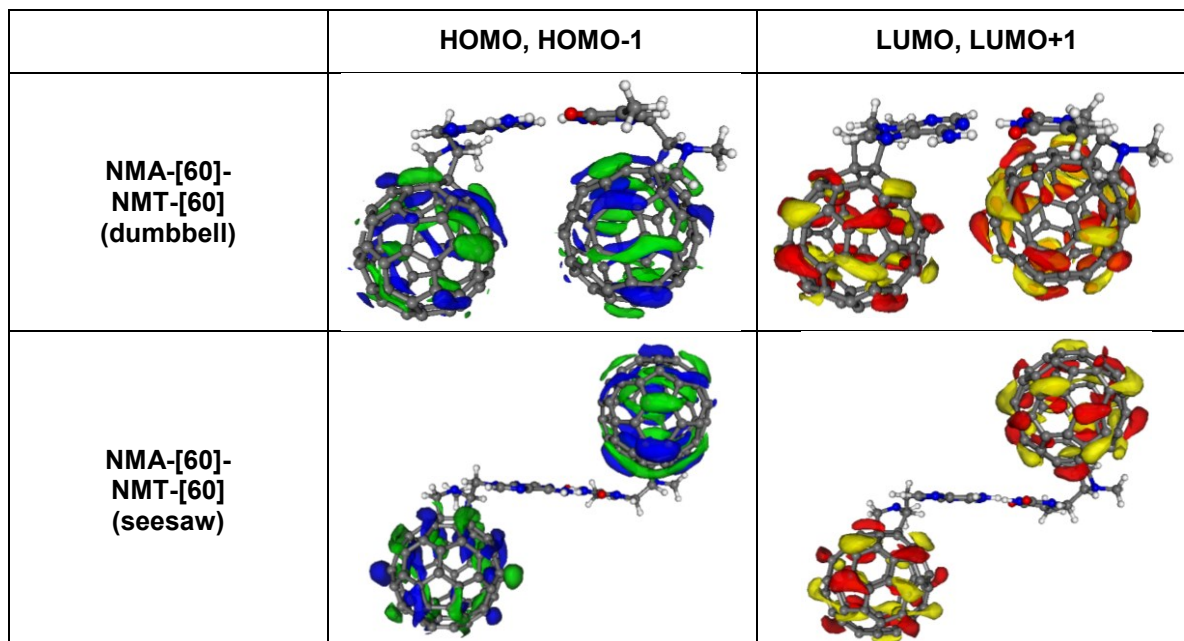
Figure 8.6 B3LYP/6-31G** calculated frontier HOMO, HOMO -1 and LUMO, LUMO +1 orbitals of the NMA-[60]-NMT-[60] dimer in the dumbbell and seesaw conformations in the gas phase.
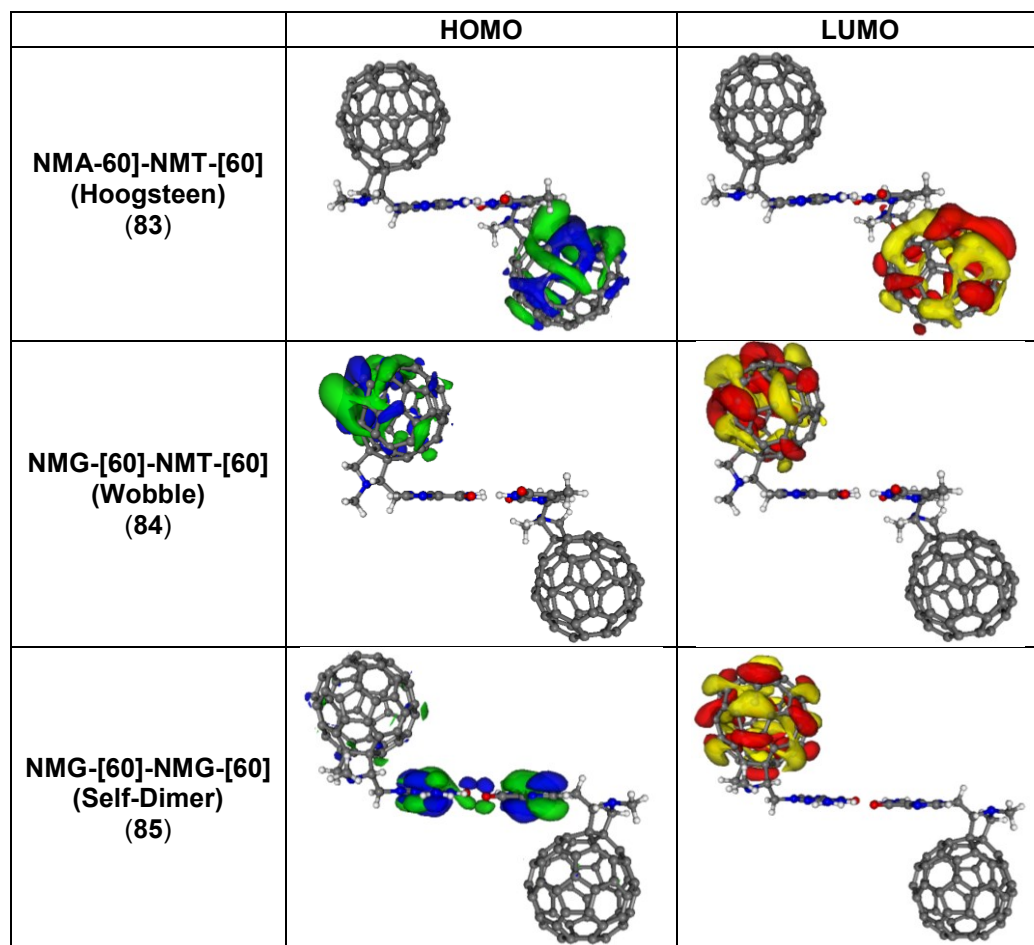
The HOMO and HOMO -1 electron density distributions of the dumbbell and seesaw orientations for the NMA-[60]-NMT-[60] and NMG-[60]-NMC-[60] dimers were localized on NMA-[60] and NMG-[60] (**Figure 8.6** and **8.7** respectively). In the dimers, these monomers where the HOMO -1 electron distribution was concentrated also had the higher HOMO energy. The localization of the distribution also follows the same pattern shown in the individual monomers for NMA-[60] (**Figure 7.2**) and NMG-[60] (**Figure 7.6**). However, for the NMD-[60]-NMN-[60] dimer, the HOMO orbital was localized on the monomer with the lower HOMO energy, NMD-[60] (**Figure 8.8**).

	HOMO	LUMO
NMG-[60]-NMC-[60] **(dumbbell)**		
NMG-[60]-NMC-[60] **(seesaw)**		
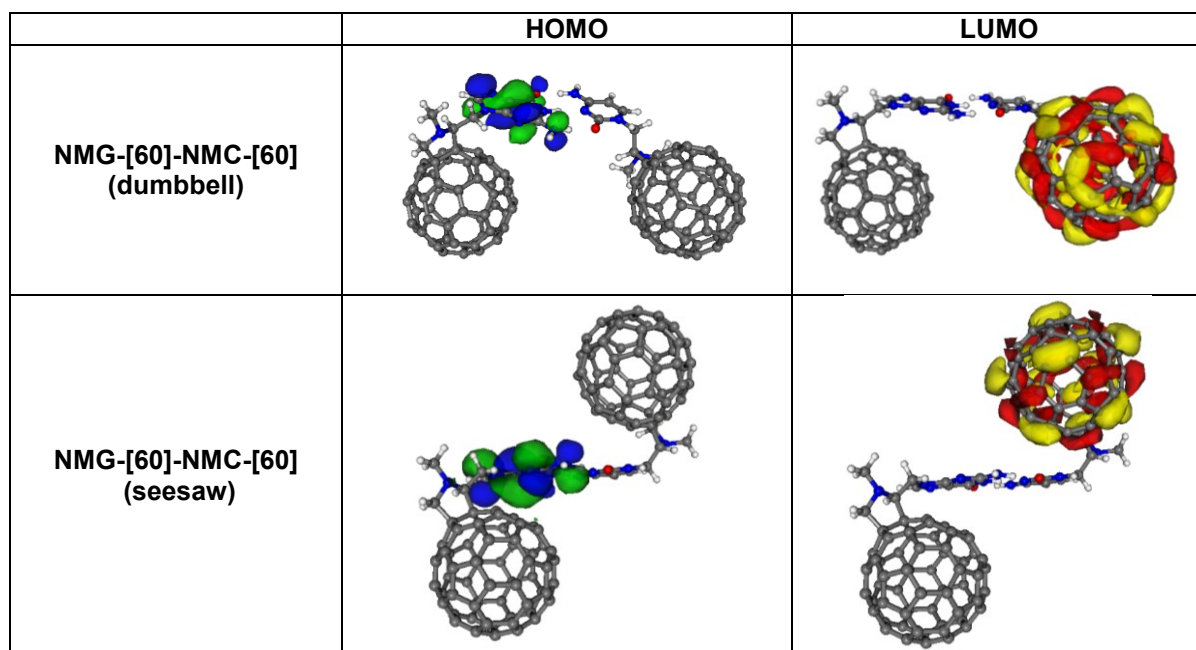
Figure 8.7 B3LYP/6-31G** calculated frontier HOMO and LUMO orbitals of NMG-[60]-NMC-[60] dimer in the dumbbell and seesaw conformations in the gas phase. The LUMO +1 orbitals are also plotted for the dumbbell orientation.

	HOMO	LUMO
NMD-[60]-NMN-[60] **(dumbbell)**		
NMD-[60]-NMN-[60] **(seesaw)**		
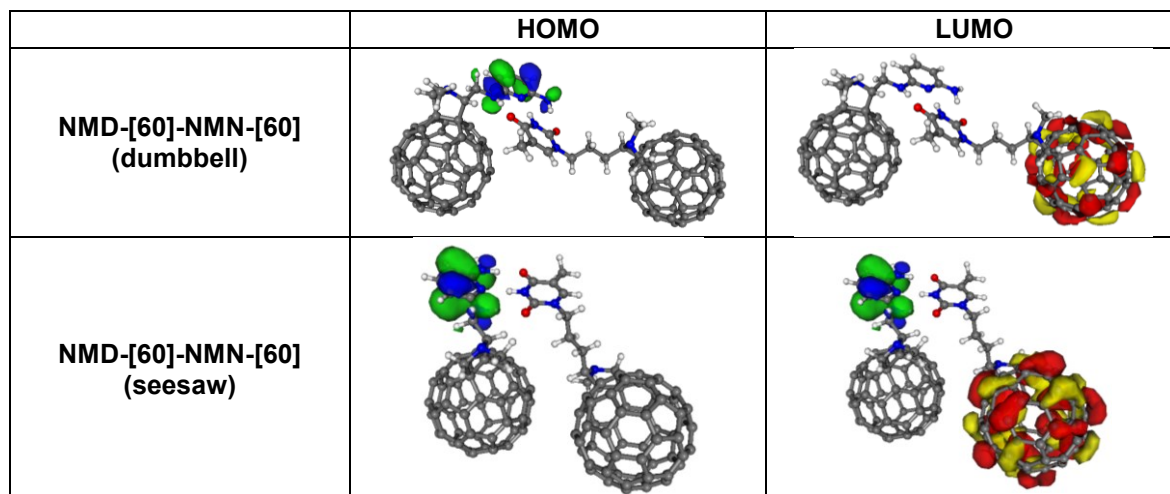
Figure 8.8 B3LYP/6-31G** calculated frontier HOMO and LUMO orbitals of the NMD-[60]-NMN-[60] dimer in the dumbbell and seesaw conformations in the gas phase.

	HOMO	LUMO
NMA-60]-NMT-[60] (Hoogsteen) (83)		
NMG-[60]-NMT-[60] (Wobble) (84)		
NMG-[60]-NMG-[60] (Self-Dimer) (85)		

Figure 8.9 B3LYP/6-31G** calculated frontier HOMO and LUMO orbitals of the Hoogsteen (**83**), Wobble (**84**) and NMG-[60]-NMG-[60] (**85**) self-dimer in the seesaw conformation in the gas phase.

The electronic distribution for the Hoogsteen (**83**), Wobble (**84**) and self-dimer (**85**) binding modes in the HOMO and LUMO frontier orbitals in the seesaw conformation are shown in **Figure 8.9**. With these dimers, only the LUMO orbitals for NMG-[60]-NMG-[60] were degenerate, with 0.35 kcal/mol between the LUMO and LUMO +1 orbitals. The electron distribution of the LUMO +1 orbitals was located on the C_{60} cage of the NMG-[60] derivative, which is not shown in the Table.

In contrast, the HOMO and LUMO orbitals for the Hoogsteen (**83**) and Wobble (**84**) dimers were not degenerate, with energies over 2 kcal/mol between the HOMO, HOMO +1 and

LUMO, LUMO +1 orbitals. In addition, unlike the Watson-Crick NMA-[60]-NMT-[60] dimer which had degenerate HOMO and LUMO orbitals, these orbitals in the Hoogsteen binding mode were not degenerate. The HOMO and LUMO electron distribution were localised on the C_{60} cage of the NMT-[60] derivative (**Figure 8.9**). In the Wobble dimer, the HOMO and LUMO electron distribution was located on the same derivative, the NMG-[60] monomer (**Figure 8.9**). In this dimer, the HOMO distribution on the NMG-[60] monomer was different to the previously observed distribution in the Watson-Crick bound NMG-[60]-NMC-[60] dimer (**Figure 8.7**), and the NMG-[60]-NMG-[60] self-dimer (**Figure 8.9**) where the HOMO electron distribution was located on the guanine entity. For the NMG-[60]-NMG-[60] self-dimer, the HOMO orbitals were mainly concentrated on the guanine entities of both monomers with some orbitals on one of the derivatives. The LUMO electron distribution was located on one of the NMG-[60] derivatives, the same derivative where the HOMO distribution was also located.

8.4.3 HOMO-LUMO Gap

The NMA-[60]-NMT-[60] dimer had a HOMO-LUMO gap of 2.76 eV which is closest to that observed for pristine C_{60}. The HOMO-LUMO gap was lower (2.52 eV) and (2.48 eV) in the dumbbell and seesaw orientations respectively (**Table 8.4**). In contrast, the NMG-[60]-NMC-[60] and NMD-[60]-NMN-[60] dimers had significantly lower band gaps in both orientations. For NMG-[60]-NMC-[60], the band gaps were 29 % and 30 % smaller (1.96 eV and 1.94 eV respectively) than C_{60}. The biggest difference with C_{60} was observed in the NMD-[60]-NMN-[60] dimer, which was 37 % and 38 % smaller for the dumbbell and seesaw orientations respectively (**Table 8.4**).

8.4 Conclusion

The strength of the DNA-base H-bonds was shown by the difference in binding energies between the NMG-[60]-NMC-[60] dimer which had a more exothermic binding energy by an average of 11.93 kcal/mol compared to the NMD-[60]-NMN-[60] dimer even though they

both possess three H-bonds. These differences highlight why Watson-Crick binding modes are an attractive way of constructing stable self-assembling supramolecular structures. In addition, the BSSE corrected binding energies for the Watson-Crick, Hoogsteen and Wobble binding modes and NMN-[60]-NMD-[60] which give a more accurate representation of the binding energies have shown differences within 7 kcal/mol with the uncorrected binding energies.

Chapter 9

Methanofullerenes

9.1 Introduction

Carbenes are neutral, reactive intermediates (**86**) that can form new carbon-carbon bonds.[145] They have six valence electrons, and thus carbenes are electrophilic and can insert themselves into the π bonds of olefins.[145, 146] The ability of C_{60} to react as an olefin means that carbenes are involved in addition reactions at the 6, 6 bonds of C_{60} forming methanofullerene derivatives.[5]

86
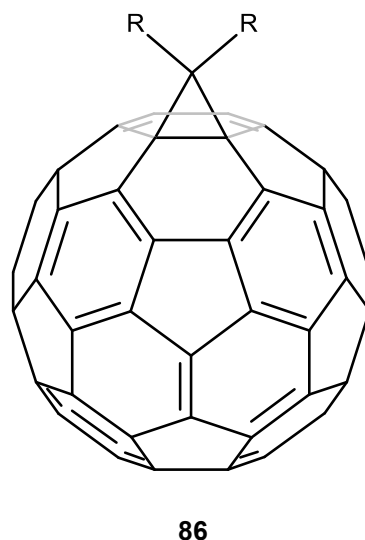
Figure 9.1 Methanofullerenes from [2+1] cycloaddition of a carbene adduct. R = adenine, thymine, guanine, cytosine, diaminopyridine or N-propylthymine.

Methanofullerenes offer the possibility of extending the self-assembly of the modelled derivatives since they have two R groups attached. They are also more symmetrical if the R groups are the same.

9.2 Electronic Properties

9.2.1 HOMO-LUMO Gap

The calculated HOMO-LUMO gap for C_{60} in the gas phase was 2.76 eV and was lowered for all of the methanofullerenes, 1,1-dithymine-1,9-methanofullerene (DTM-[60]), 1,1-diadenine-1,9-methanofullerene (DAM-[60]), 1,1-dicytosine-1,9-methanofullerene (DCM-[60]), 1,1-diN-propylthymine-1,9-methanofullerene (DNM-[60]), 1,1-diguanine-1,9-methanofullerene (DGM-[60]) and 1,1-di-diaminopyridine-1,9-methanofullerene (DDM-[60]) in the gas phase (**Table 9.1**). The smallest decrease was observed in DTM-[60] (2.60 eV) and the biggest decrease was for DDM-[60] (2.11 eV).

	HOMO (eV)	LUMO (eV)	HOMO - LUMO gap (eV)	HOMO - LUMO gap (nm)
C_{60}	-5.89	-3.13	2.76	449
DTM-[60]	-5.89	-3.29	2.60	477
DAM-[60]	-5.74	-3.15	2.59	479
DCM-[60]	-5.59	-3.00	2.59	479
DNM-[60]	-5.81	-3.23	2.58	481
DGM-[60]	-5.76	-3.28	2.48	500
DDM-[60]	-5.11	-3.00	2.11	588

Table 9.1 B3LYP/6-31G** calculated HOMO, LUMO and HOMO-LUMO gap energies of the six monomers in the gas phase. The monomers are arranged according to decreasing HOMO-LUMO gap.

Comparison of the band gaps between the fulleropyrrolidines (in the gas phase) and methanofullerenes showed that the HOMO-LUMO gaps for the fulleropyrrolidines were within 0.06 eV of the methanofullerenes. They are thus considered to be degenerate in energy (**Table 9.2**).

	Thymine	Adenine	Cytosine	Npropyl thymine	Guanine	2,6-Diamino pyridine
Methanofullerenes (eV)	2.60	2.59	2.59	2.58	2.48	2.11
Fulleropyrrolidines (eV)	2.54	2.54	2.53	2.55	2.48	2.16

Table 9.2 Comparison of the HOMO-LUMO gap of the methanofullerenes and fulleropyrrolidines.

The biggest differences between the fulleropyrrolidines and the methanofullerenes were observed in the thymine- and cytosine-functionalized derivatives (0.06 eV) whilst the smallest difference was in the N-propylthymine-functionalized derivatives (0.03 eV) (**Table 9.2**). Therefore, the methanofullerene and fulleropyrrolidine derivatives will have similar photophysical properties.

9.2.2 Frontier Orbital Distribution

The frontier orbital distribution for the methanofullerenes are shown in **Figure 9.2**. The HOMO electron distribution for DTM-[60], DAM-[60], DCM-[60] and DNM-[60] was localized on the fullerene cage whereas on DGM-[60] and DDM-[60], they were localized on the guanine and diaminopyridine entities (**Figure 9.2**). Upon functionalization, the frontier HOMO orbitals lost their degeneracy and the HOMO-1 was more than 2 kcal/mol lower in energy. The same pattern for the frontier HOMO orbital distribution was also observed in the fulleropyrrolidines. This shows that functionalizing the C_{60} cage with these entities, as fulleropyrrolidines or methanofullerenes, produces the same effect in the highest occupied molecular orbitals as might be expected as only one double bond is modified in both derivatives.

	HOMO	LUMO
C$_{60}$		
DTM-[60] (87)		
DAM-[60] (88)		
DCM-[60] (89)		
DNM-[60] (90)		
DGM-[60] (91)		
DDM-[60] (92)		

Figure 9.2 B3LYP/6-31G** calculated HOMO and LUMO orbitals for C$_{60}$ and the methanofullerenes in the gas phase. Isosurface = 1.50 au.

In contrast, the LUMO frontier orbitals were delocalized across the C$_{60}$ cage for all the methanofullerene derivatives. The same frontier LUMO orbital distribution was also observed

for the fulleropyrrolidine derivatives. The similarities show that the nature of the group added to C_{60} determines the electronic distribution and that these properties were maintained when methanofullerenes were modelled.

It is interesting to note that the methanofullerenes with the lowest band gaps, DGM-[60] and DDM-[60], had the HOMO distributions localized on the guanine and diaminopyridine entities.

9.3 Dimer Electronic Properties

9.3.1 Binding Energies

The binding energies and BSSE corrected energies of the three dimers calculated using B3LYP/6-31G** are shown in **Table 9.4**. The BSSE corrected energies show that the adenine- and thymine-functionalized methanofullerene dimer preferred the dumbbell conformation by 4.5 kcal/mol, unlike the fulleropyrrolidine dimer which did not show a preference based on thermodynamic stability (**Table 8.3**). In addition, comparison of the methanofullerene and fulleropyrrolidine dimers showed small differences between the binding energies(dumbbell = 2 kcal/mol and seesaw = 5.5 kcal/mol). In a similar manner, the binding energy for the guanine- and cytosine-functionalized methanofullerene dimer was similar to that of the fulleropyrrolidine dimer with the same moieties in the same conformation, with just 1.2 kcal/mol difference (**Table 8.3**).

Out of the six possible dimers, the results for only three dimers were available. Errors in the SCF mode prevented the calculations for other dimers from converging.

	Binding Energy kcal/mol	HOMO (eV)	LUMO (eV)	HOMO-LUMO Gap (eV)	(nm)
DAM-[60]-DTM-[60] (dumbbell)	-11.5	-5.76	-3.23	2.53	490
DAM-[60]-DTM-[60] (seesaw)	-15.6	-5.75	-3.29	2.46	504
DCM-[60]-DGM-[60] (seesaw)	-29.3	-5.36	-3.22	2.14	579

Table 9.3 B3LYP/6-31G** calculated binding energy, HOMO, LUMO and HOMO-LUMO gap for the two dimers.

	Dimer Orientation	BSSE corrected Binding Energy (kcal/mol)	Binding Energy (kcal/mol)	Energy difference (kcal/mol)
DAM-[60]-DTM-[60]	Dumbbell	-19.2	-11.5	-7.7
	Seesaw	-14.7	-15.6	0.9
DCM-[60]-DGM-[60]	Seesaw	-33.4	-29.3	-4.1

Table 9.4 Comparison of the B3LYP/6-31G** calculated binding energies and the BSSE corrected binding energies.
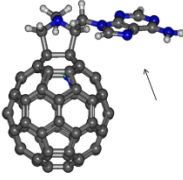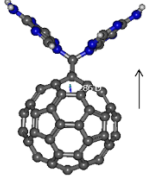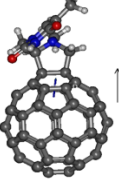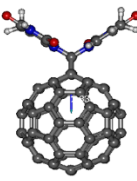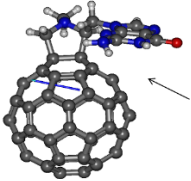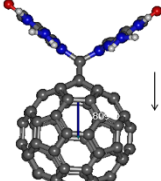
9.3.2 Direction of Dipoles

9.3.2.1 Fulleropyrrolidine and Methanofullerene Monomers

The direction of the dipoles (**Figure 9.3** and **9.4**) give an indication of the orientation the monomers will adopt on the zeolite surface, in particular, the positively charged channels. Comparison of the fulleropyrrolidines and the methanofullerenes shows that the entities affect the direction of the dipoles in a similar manner apart from the thymine and guanine-based derivatives.

The direction of dipoles moves from the most positive to the most negative areas as shown in **Figure 9.3**. For the guanine-cytosine- and N-propylthymine-diaminopyridine-functionalised fullerene dimers, the direction of the dipole was from the monomer where the HOMO electron density was concentrated to the monomer where the LUMO was concentrated, as observed in **Figure 8.7**, **8.8** and **9.5** for the methanofullerene dimers. However, for the adenine-thymine-based dimers, the direction of the dipole was not related to the electron distribution of the HOMO and LUMO orbitals. This may have been because the HOMO and LUMO orbitals for both the fulleropyrrolidine and methanofullerene dimers were degenerate in energy and electron density distribution.

Additionally, apart from the adenine-functionalised monomers and the guanine-cytosine-functionalised dimer (seesaw), there was no pattern observed for the calculated dipole moments.

	Fulleropyrrolidines	Methanofullerenes
Adenine	1.82 D	1.29 D
Thymine	5.21 D	2.84 D
Guanine	6.41 D	10.8 D
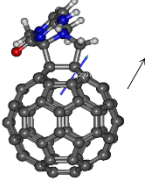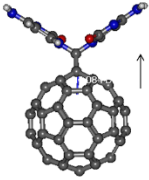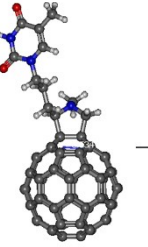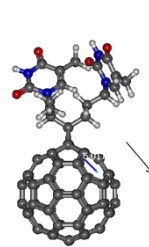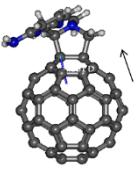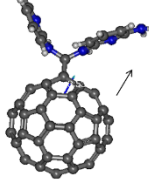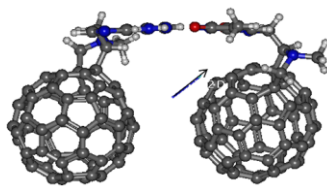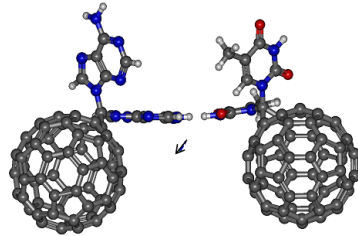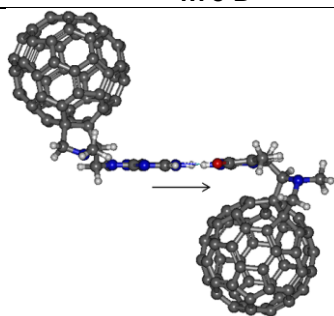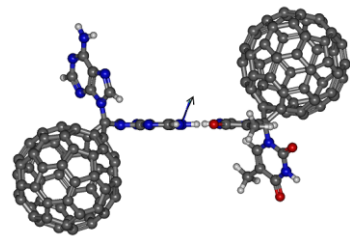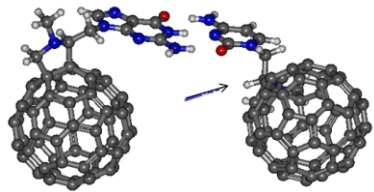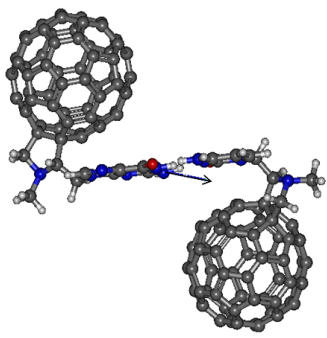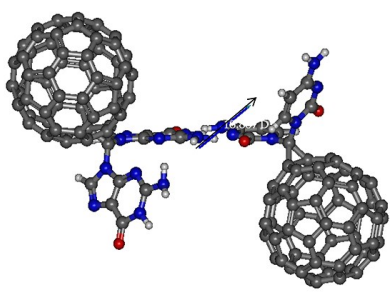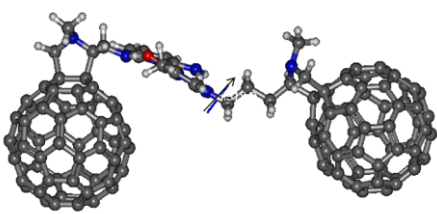
Cytosine	7.24 D	4.08 D
N-propylthymine	3.53 D	6.01 D
Diaminopyridine	4.55 D	3.53 D

Figure 9.3 Comparison of the fulleropyrrolidine and methanofullerene dimer dipole directions.

9.3.2.2 Fulleropyrrolidine and Methanofullerene Dimers

		Fulleropyrrolidines	Methanofullerenes
Adenine-Thymine	**Dumbbell**	 **4.78 D**	 **1.82 D**
	Seesaw	 **3.52 D**	 **3.33 D**
Guanine-Cytosine	**Dumbbell**	 **5.11 D**	
	Seesaw	 **6.21 D**	 **10.9 D**
N-propyl thymine-Diamino pyridine	**Dumbbell**	 **5.31 D**	

	Seesaw	**5.14 D**	
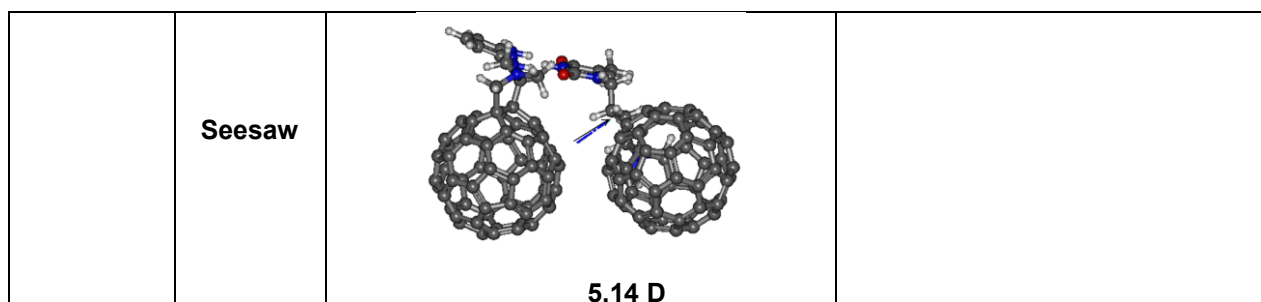
Figure 9.4 Comparison of the fulleropyrrolidine and methanofullerene dimer dipole directions.

9.3.3 Frontier Orbital Distribution

The frontier orbital distribution for the three dimers was similar to the distribution observed for the individual monomers (**Figure 9.5**). The distribution of the HOMO for the DAM-[60]-DTM-[60] dimer was localized on the C_{60} cage of the DTM-[60] monomer whilst for the DGM-[60]-DCM-[60] dimer the orbitals were localized on the guanine entity. As observed in the monomers, the HOMO energy was lower (more negative) when the electron distribution was localized on the C_{60} cage compared to when it was delocalized elsewhere.

In contrast, the frontier LUMO distribution for all three methanofullerene dimers in the different orientations was localized on the C_{60} cage. The similar electronic LUMO distribution on the dimers was reflected in the similar LUMO energies. The difference between the highest and lowest LUMO energy was not significant (0.07 eV) (**Table 9.4**). The similar electronic distribution on the dimer indicates that the C_{60} ring will be the focus of nucleophilic attack on the dimers.

For these methanofullerene dimers, the difference between HOMO and HOMO +1 orbitals was greater than 2 kcal/mol, therefore the orbitals were not degenerate. In contrast, all the LUMO energies were degenerate, and just like the LUMO distribution on the fulleropyrrolidine dimers, the orbitals were also localised on the C_{60} cages (**Figure 9.5**).

	HOMO	LUMO
DAM-[60]-DTM-[60] **(dumbbell)** **(93)**		
DAM-[60]-DTM-[60] **(seesaw)** **(94)**		
DGM-[60]-DCM-[60] **(seesaw)** **(95)**		
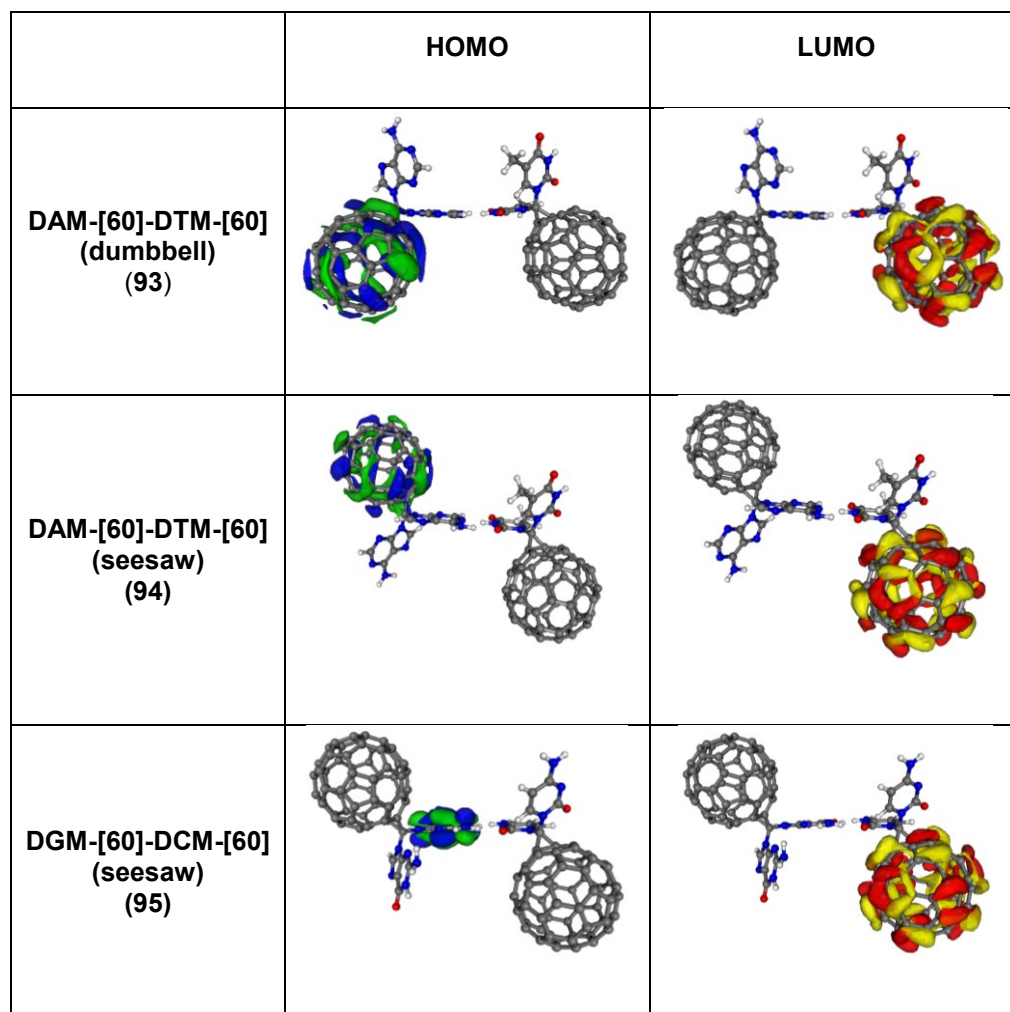
Figure 9.5 B3LYP/6-31G** calculated frontier HOMO and LUMO, LUMO +1 orbitals. Isosurface = 1.50 au.

9.3.4 HOMO-LUMO Gap

The HOMO-LUMO gap for the DAM-[60]-DTM-[60] dimer in the seesaw (2.46 eV) and dumbbell (2.53 eV) orientation was similar to the HOMO-LUMO gap for the fulleropyrrolidine dimer in seesaw (2.48 eV) and dumbbell (2.52 eV) orientation (**Table 8.2**). The band gap for the DGM-[60]-DCM-[60] dimer was lower than that for the DAM-[60]-DTM-[60] dimer (2.14 eV), though higher than the band gap in the NMG-[60]-NMC-[60] dimer in the same orientation (1.94 eV). This pattern was observed in the fulleropyrrolidine dimers as well.

Dimers	Orientation	HOMO-LUMO Band gap	
		Methanofullerenes (eV)	Fulleropyrrolidines (eV)
A-T	Dumbbell	2.53	2.52
	Seesaw	2.46	2.48
G-C	Seesaw	2.14	1.96

Table 9.5 Comparison of the B3LYP/6-31G** calculated HOMO-LUMO gap for the methanofullerene and fulleropyrrolidine dimers.

9.4 Conclusion

Comparison of the HOMO and LUMO energies of the monomers shows that despite the HOMO and LUMO energies of the fulleropyrrolidines being higher (less negative) than those of the methanofullerenes, the band gaps showed very small differences. The biggest difference was for the thymine and cytosine derivatives (0.06 eV), and the smallest difference was for the N-propylthymine derivatives which was 0.03 eV. The electronic distribution of the methanofullerenes matched that of the fulleropyrrolidines, both in the HOMO and LUMO orbitals. As observed in the fulleropyrrolidines, the HOMO distribution was concentrated on the C_{60} cage for four of the monomers, and on the guanine and diaminopyridine entities for their respective derivatives. This distribution pattern indicates that these two monomers would adsorb on the zeolite with the adenine and diaminopyridine entities located inside the zeolite channels, discouraging H-bonded self-assembly. In a similar manner, as found for the fulleropyrrolidines, the LUMO orbitals were localized on the C_{60} cage for all the derivatives.

The similar electron distribution patterns for the methanofullerenes and the fulleropyrrolidines highlight the importance of modelling these derivatives. It also shows that

attaching the same self-assembling entities on the C_{60} cage in two locations does not significantly affect the electronic properties of the derivative.

The methanofullerene dimers did not show a large difference with the electronic properties of the fulleropyrrolidine dimers either. The binding energies were within 5 kcal/mol while the band gaps were within 0.2 eV of the fulleropyrrolidines. These similarities show that the methanofullerenes and the fulleropyrrolidines are likely to adsorb on the zeolite surfaces in a similar manner, as well as undergoing oxidation and reduction reactions in a similar way. The only notable differences were in dipole moments, and this will be because of the two heterocyclic entities contributing towards the polarity of the methanofullerenes, whereas only one is present for the fulleropyrrolidines.

Chapter 10
Biological Applications

10.1 Introduction

The possibility of using fullerenes as HIV-1 protease (PR) inhibitors was first proposed by Friedman *et al*.[28] Steric and chemical complementarity meant that C_{60} was able to dock into the active site of HIV-1 PR.[28, 147, 148] The active site of the HIV-1 PR, has a cylinder shape with a diameter of about 10 Å[148] while the C_{60} cage has one of approximately 7.9 Å.[67-69] These similar radii led to a hypothesis that strong hydrophobic interactions between C_{60} derivatives and the active site were possible, making C_{60} derivatives potential inhibitors of HIV-1 PR.[28] The rigidity of C_{60} cage also meant that hypothetical interactions between the two sites were limited.[28, 148] However, the functional groups attached to C_{60} could form stronger interactions, such as hydrogen bonds, to selected residues.

A computational approach was suggested as an initial step in this field of study, followed by synthesis of the appropriate C_{60} derivatives.[28] With this route, not only would experimentalists save a lot of time, but it would also limit synthesis efforts to the derivatives with the strongest predicted bonding based on docking score.[147, 148] Docking scores are a measure of the interactions such as hydrogen bonding, dispersion or electrostatic forces, between the C_{60} derivative and the active site.[147] Friedman *et al*. suggested that bis-adducts had the potential to have higher binding interactions compared to mono-adduct C_{60} derivatives, because they would be capable of forming more interactions within the active site.[147]

Two water-soluble C_{60} derivatives were modelled and their binding abilities analysed by Marcorin *et al*.[148] The derivatives, which had ammonium groups attached, were then synthesized using the Prato reaction.[17] The difference between the two derivatives was the group attached on the nitrogen atom found on the pyrrolidine ring (**Figure 10.1**, (**96**)).

Computational analysis of the interactions between the C_{60} derivatives and the HIV-1 PR active site showed that a derivative (**Figure 10.1**, (**96b**)) with one group protonated, had a more negative complexation energy (-32.1 kcal/mol) compared to the same derivative with both groups protonated (-25.5 kcal/mol).[148] Using the *cvff* force field, the complexation energy was calculated by subtracting the energies of the optimised (HIV-1 PR)-**96a** or (HIV-1 PR)-**96b** structures from the sum of the isolated enzyme plus the isolated inhibitor.[148] The difference in energies was attributed to the H-bond interactions between one of the amino acids in the active site with the neutral NH_2 rather than the charged NH_3^+ group. The difference in the stability of HIV-1 PR between the C_{60} derivatives was assigned to the interactions of the oligoethylene oxide chain (**96b**) with the Asp25 and Asp125 entities of the active site.[148]

96
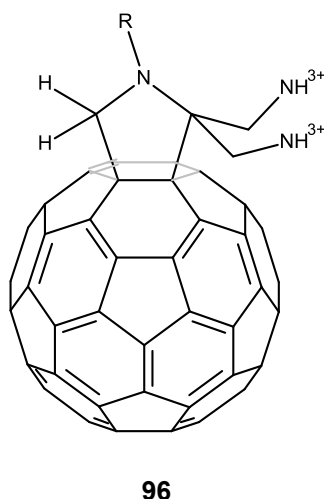
Figure 10.1 Water-soluble ammonium functionalised fullerene derivative (**96**).
(**a**), R = CH_3, (**b**), R = $CH_2CH_2OCH_2CH_2OCH_2CH_2OCH_3$

In a similar study that featured protonated systems, diphenyl C_{60} alcohol (**97**) was docked inside the HIV-1 PR inhibitor in three states, unprotonated, mono-protonated and di-protonated.[149] In contrast to the results obtained by Marcorin *et al.*, their results showed that the di-protonated derivative formed the most stable H-bonds with the active site. In this study, water molecules were left in the active site of the HIV-1 PR protease and were found

to bridge unprotonated and mono-protonated derivatives to the side of the active site. No water molecules were found in the active site of the di-protonated-HIV-1 PR complex.[149] This is consistent with other studies which indicate that the water molecules actually limit the inhibition potential of the ligands and removing water molecules from the active site improves the binding potential of the inhibitors.[28, 147, 150] In fact, the results shown, which state that the water molecules act as a bridge between the C_{60} derivatives and the active site confirm this idea that water molecules limit the inhibition abilities of the derivative since the model without any water molecules formed more stable interactions compared to the models that did.
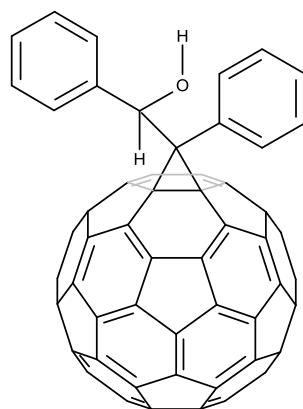
(**97**)

Figure 10.2 Structure of diphenyl C_{60} alcohol (**97**)

Other computational strategies have been developed to screen many potential HIV-1 PR inhibitors and create a database.[151-153] The *in silico* drug screening approach was used to test the inhibitory abilities of C_{60} derivatives which had been synthesized for other biomedical applications computationally. Molecular docking simulations were performed on the C_{60} derivatives and fullerene-based amino-acid derivatives were found to have the best docking scores.[151] These higher docking scores were linked to their ability to form H-bonds with the residues in the active site of HIV-1 PR and the van der Waals interactions between the C_{60} cage and the nonpolar surface of the active site.[151] The amino-acid fullerene derivatives

were then subjected to experimental biological tests and it was found that the calculations predicted the binding affinity of one derivative, whilst over-predicting the binding affinity of another.[151]

The combination of computational and experimental methods in this field of study is clearly very important. Computational studies continue to give insight into the inhibitory abilities of C_{60} derivatives whilst limiting experimental efforts to the compounds with the most potential. However, it is important to model compounds whose synthesis routes are possible and also offer reasonable yields. Thus, it is useful to be able to calculate predicted formation energies as well as other properties.

10.1.1 Computational Method

The DFT B3LYP/6-31G** optimised fullerene derivatives NMA-[60], NMC-[60], NMG-[60], NMT-[60], NMN-[60] and NMD-[60] were docked into HIV-1 PR in order to analyse their binding affinities. The initial HIV-1 PR structure was complexed with a haloperidol molecule (**98**) at 2.2 Å. A maximum of 50 000 generations was used with a mutation rate of 0.3 and the elitism chosen was 7. The procedure chosen docked the fullerene derivative, as a flexible ligand, into the active site of the HIV-1 protein using a genetic docking algorithm that had the PMF scoring function. Fullerene derivatives tested by Durdagi *et al.*[153] were also docked into the HIV protein using the same parameters. In order to compare the conformations of the six docked modelled derivatives in the active site.

(**98**)
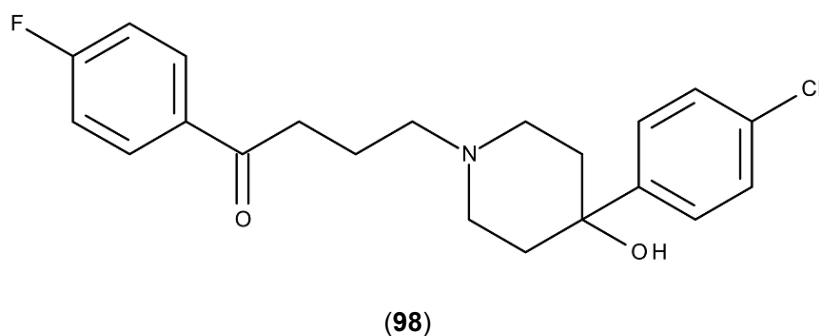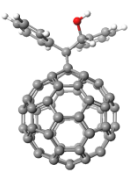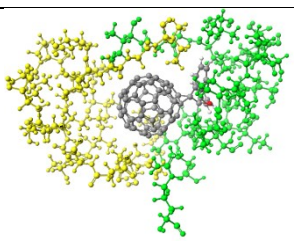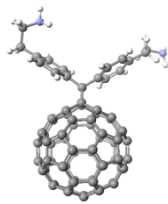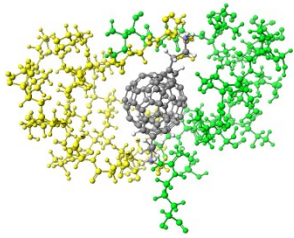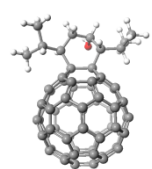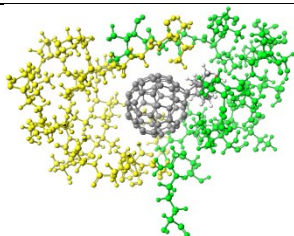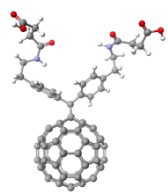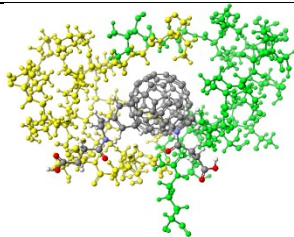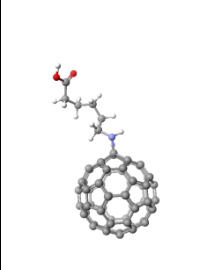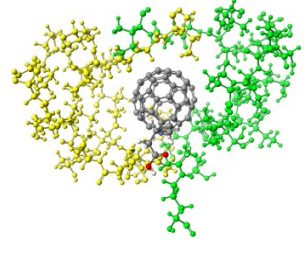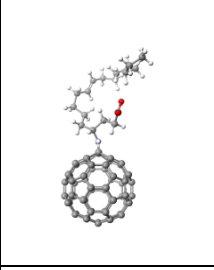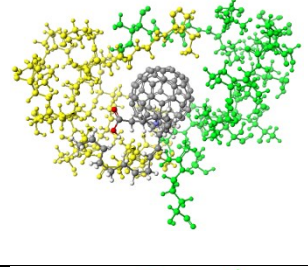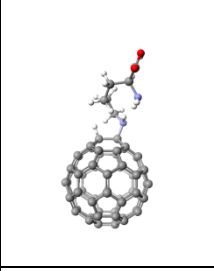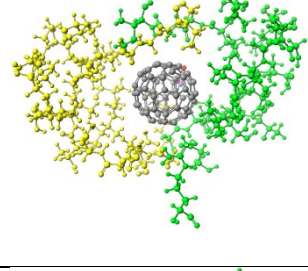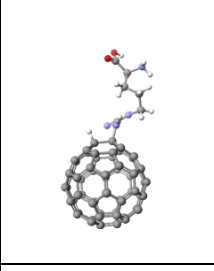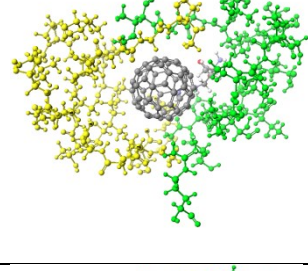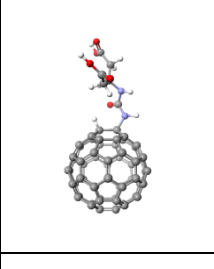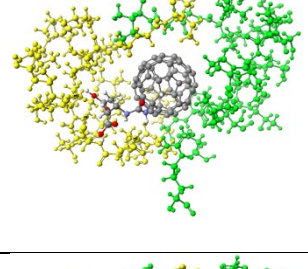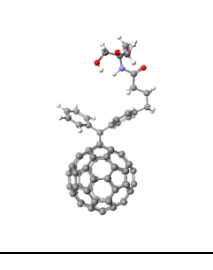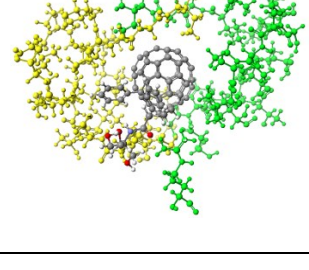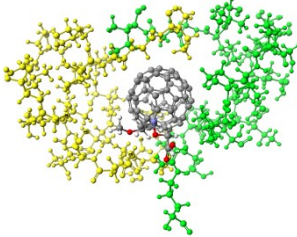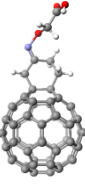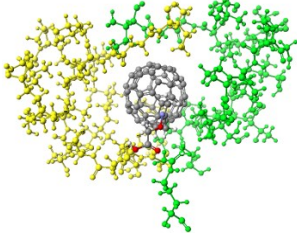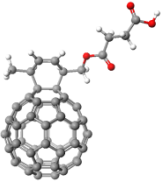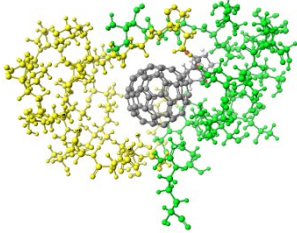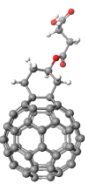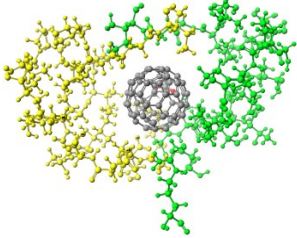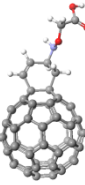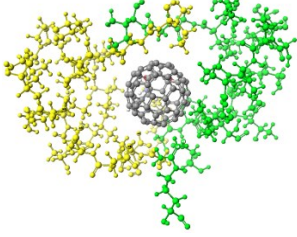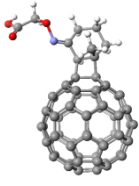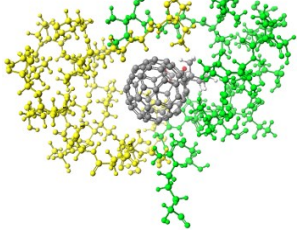
Figure 10.3 Structure of a haloperidol molecule (**98**).
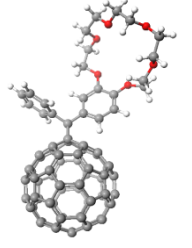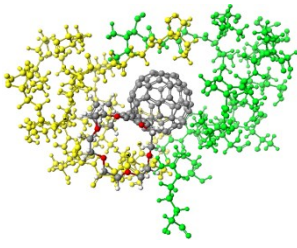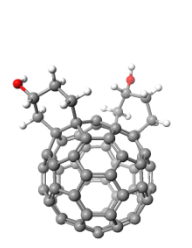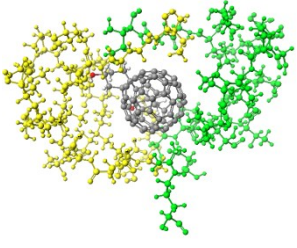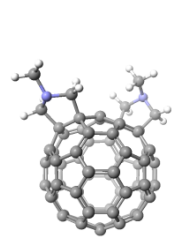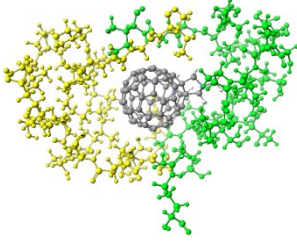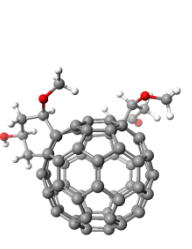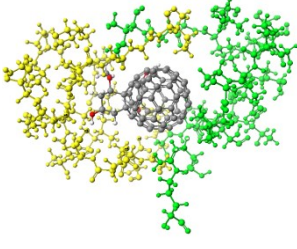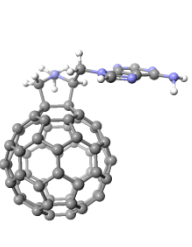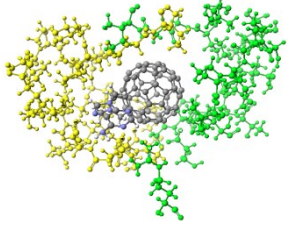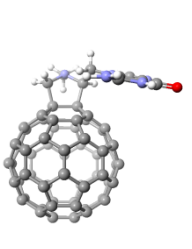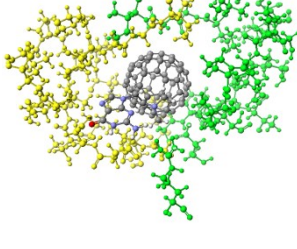
10.1.2 Validation of Algorithm

Results from the docking simulations of the 20 derivatives by Durdagi *et al.* and the modelled fullerene derivatives are shown in **Table 10.1**. The amino acid residues interacting with the ligands were colour-coded according to their structure. Black and red was used to indicate positively and negatively charged amino acids respectively, while blue was for amino acids with hydrophobic side chains and yellow was for special amino acids.

Durdagi *et al.* Model Number /Name	2D Picture	Residue in active site	PMF Docking Score (kcal/mol)	Experimental EC50 (µM)	Interacting Residues in active site
1			-293	0.1	Asp30 Ile47 Pro81
2			-256	5.0	Arg8 Asp29 Ile50
3			-258	0.15	Asp30 Gly48
4			-339	7.3	Lys45 Arg8 Val82 Asp29

#					
5			-274	0.49	Arg8 Asp29
6			-288	1300	Arg8 Asp29 Lys45
7			-276	75	Arg8 Val82 Asp29
8			-291	230	Gly27 Asp29 Ile45
9			-306	-22.1	Asp30 Ile47
10			-314	2.50	Arg8 Asp29 Ile47

11			**-273**	**0.9**	**Arg8** **Ala28** **Asp29**
12			**-252**	**2.9**	**Arg8** **Asp29** **Val82**
13			**-267**	**2.2**	**Asp30** **Lys45**
14			**-272**	**6.30**	**Arg8** **Asp29**
15			**-256**	**7.30**	**Arg8** **Asp29** **Val82**
16			**-261**	**21.7**	**Arg8** **Asp29**

17			-303	137	**Arg8** **Asp29** **Lys45**
18			-258	17.6	**Arg8** **Pro81**
19			-242	72.7	**Arg8** **Ala28** **Ile50**
20			-253	7.70	**Arg8** **Asp30**
NMA-[60]			-308		**Arg8** **Asp29**
NMG-[60]			-296		**Asp30** **Lys45**
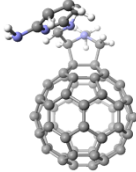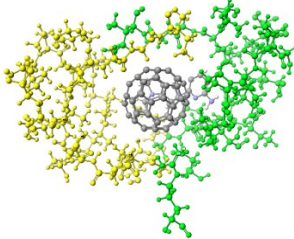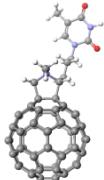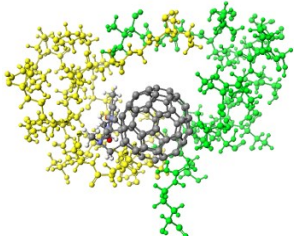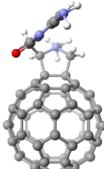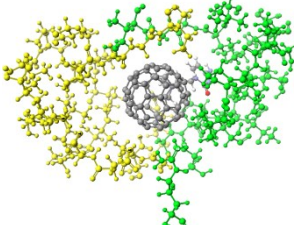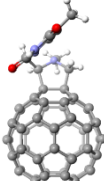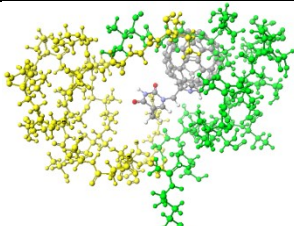
NMD-[60]			**-289**		**Arg8** **Asp30**
NMN-[60]			**-282**		**Asp30** **Val32** **Ile84**
NMC-[60]			**-280**		**Arg8** **Asp30** **Ile47**
NMT-[60]			**-208**		**Arg8** **Asp25** **Val82**

Table 10.1 Figures showing the orientation of the fullerene ligands in the HIV-1 PR binding site with the calculated docking scores for fullerene derivatives calculated by Durdagi *et al*. (**1-20**), and the modelled fullerene derivatives. The interacting amino acid ligands are colour-coded; black - positively charged, red-negatively charged, blue-amino acids with hydrophobic side chains and yellow - special cases. The experimental EC_{50} results were obtained from the paper by Durdagi *et al*.

10.1.3 Discussion

The biological activity of 20 fullerene derivatives described by Durdagi *et al*.[153] are shown in **Table 10.1**. In order to compare the molecular docking scores and orientation of the fullerenes in the active site, the same methods used to obtain the docking scores for the twenty derivatives were used for the six-modelled derivatives also shown in **Table 10.1**. From the twenty derivatives, the structure described as the most potent by Durdagi *et al*., (**1**)

had a docking score of -294 kcal/mol and the calculated EC_{50} of 0.1 µM.[153] The EC_{50} of a compound is the concentration of the fullerene derivative which induces a binding affinity of 50 % of the maximum response of HIV-1 PR.[154] Therefore, the lower the concentration required, the more potent the inhibitor. Although structure (**1**) had the lowest EC_{50} (0.1 µM), it did not have the most negative docking score. PMF docking scores ranged from -242 (19) to -339 kcal/mol (4). From the results in **Table 10.1**, there was no correlation between the Durdagi *et al*. calculated EC_{50} and the PMF docking scores of the Durdagi *et al*. derivatives.[153] These results show the difficulties in calculating computational properties that correlate with experimental data. In particular, even small variations in conformational and flexibility of both protein and ligand can affect the calculated docking scores. The electrostatic, hydrogen bonding and van der Waals interactions between the ligands and the amino acid residues contributed to the docking scores. It is expected that the more negative the docking score observed, the stronger the interaction between the ligand and the active site.

The orientation of structure (**1**) along with the other nineteen derivatives showed the C_{60} cage inside the active site of HIV-1 PR. The attached entities were also inside the cavity or just outside the cavity (**17**) and were still interacting with the residues in the active site. From **Table 10.1**, it is clear that all 19 out of the 20 ligands were close to a negatively charged aspartic acid (Asp 29 or Asp 30). The only structures which were not in close proximity to these amino acid residues were structures (**18**) and (**19**), both of which have relatively high docking scores and EC_{50}.

For the six modelled derivatives, the C_{60} cage was inside the active site for NMA-[60], NMG-[60] NMD-[60], and NMC-[60] (**Table 10.1**). In contrast, for NMN-[60] and NMT-[60], the C_{60} cage was outside the active site with just the attached entities interacting with the active site indicating that the contribution of hydrophobic interactions between C_{60} and amino acids such as alanine (ala), isoleucine (Ile), leucine (leu), phenylalanine (phe), methionine (met), tryptophan (trp), tyrosine (tyr) and valine (val). It is interesting that NMC-[60] (-280 kcal/mol)

had a slightly less negative docking score compared to NMN-[60] (-282 kcal/mol) despite the orientation of the ligands inside the active site. The docking scores indicate that the most potent inhibitor from the six derivatives would be NMA-[60] (-308 kcal/mol). Most of the other derivatives had similar docking scores of -280 to -308 kcal/mol. These docking scores are in comparison to the most potent documented derivative (**1**) with a docking score of -294 kcal/mol. In contrast, NMT-[60] had the lowest docking score of all 26 molecules investigated (-208 kcal/mol) indicating significantly lower binding potential. The most potent derivative, NMA-[60] was 4.8 % higher than (**1**) whilst the least potent, NMT-[60] was 29 % lower.

In a similar manner to the Durdagi *et al*. structures (**Table 10.2**), all six structures interacted with the negatively charged aspartic acid residues (Asp 25, 29 and 30) shown in the red. The first three derivatives, NMA-[60], NMG-[60], NMD-[60] also interacted with positively charged amino acids residues shown in black (Arg 8 and Lys 45). However, NMC-[60], NMN-[60] and NMT-[60] interacted with the charged amino acid residues, and with amino acids with hydrophobic side chains as well (shown in blue). Of the novel derivative studies, NMA-[60], NMG-[60] and NMD-[60] and NMC-[60] appear to have the most potential for further investigation as HIV-1 protease inhibitors.

10.2 Methanofullerenes

DFT functionalized methanofullerenes were also docked into the HIV-1 PR protein. The same DNA base entities used in Section 10.2.2 as well as N-propylthymine and diaminopyridine were used to functionalise the methanofullerene derivatives.

The best docking score for the methanofullerenes (DNM-[60]) (-294 kcal/mol) (**Table 10.2**), was lower than the best docking score for the first six modelled pyrrolidine derivatives NMA-[60] (-308 kcal/mol) (**Table 10.2**). However, both docking scores were similar to the calculated values for Durdagi's best initial structure (**1**). This possibly reflects the high level of interactions between the ligands and the amino acid residues in the active site, Arg8, Asp29 and Ile47, in the active site (representing both charged and hydrophobic residues).
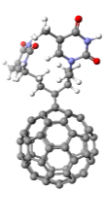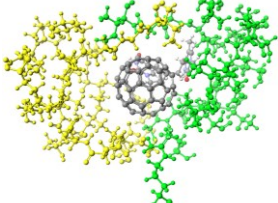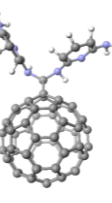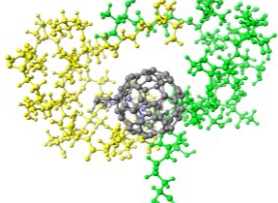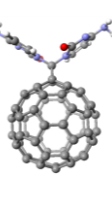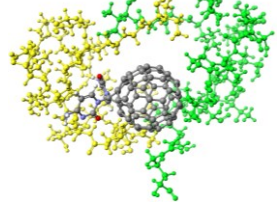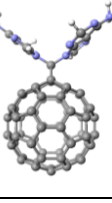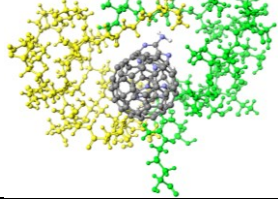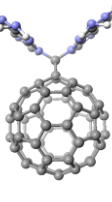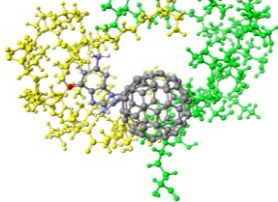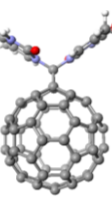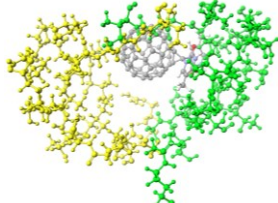
Model Number /Name	2D Picture	Residue in active site	Docking Score (kcal/mol)	Interacting Residues in active site
DNM-[60]			-294	**Arg 8** **Asp29** **Ile47**
DDM-[60]			-223	**Asp30** **Ile84**
DCM-[60]			-214	**Asp30** **Lys45** **Gly48**
DAM-[60]			-201	**Asp25** **Ile50**
DGM-[60]			-200	**Arg8** **Asp29** **Lys45** **Gly48**
DTM-[60]			-186	**Asp29** **Lys45**

Table 10.2 Figures showing the orientation of the methanofullerene ligands in the HIV-1 PR active site with calculated docking scores ranked according to increasing order.

The score is also an indication of the orientation of the ligand, which is fully embedded inside the active site. In comparison, the other five ligands had the C_{60} cage outside the active site (**Table 10.3**). These orientations were possibly the reason for the less negative docking scores observed for the methanofullerenes.

The interactions between all six methanofullerenes and the residues in the active site show that the ligands were all in proximity to negatively charged aspartic acid residues (Asp 25, 29 and 30). Comparison of the results for the fulleropyrrolidines in **Table 10.2** and the methanofullerenes show that the despite them being functionalized by the same entities, they did not interact with the same amino acid residues in the active site. In addition, despite the methanofullerenes having the two entities which would have increased the level of interactions between the ligand and the active site, the docking scores did not reflect this. From these results, the di-substituted methanofullerene derivatives would not be better inhibitors than the mono-substituted pyrrolidine derivatives.

10.3 Conclusion

The experimental biological activity of derivative (**1**), coupled with its docking score, illustrates the importance of comparing computational and experimental data, including the EC_{50} results. From the calculated docking scores alone, it would not be possible to predict the biological activity of the derivatives in the appropriate order. Therefore, not only the structure of the docked protein-ligand complexes but also the experimental biological activity for the modelled fulleropyrrolidines and methanofullerenes as well as their EC_{50} values would be required in order to accurately determine their potency as HIV-1 PR inhibitors. The differences between the docking scores for the methanofullerenes, which would have been expected to be more negative indicating stronger binding than the fulleropyrrolidines are also shown. From both sets of data, the strength of the interactions between the C_{60} cage and the active site are demonstrated by the higher (less negative) docking scores for the derivatives which were not fully embedded inside the active site. Based on both the docking scores and interactions identified, five of the twelve novel derivatives studied could potentially act as

effective inhibitors (NMA-[60], NMG-[60], NMD-[60], NMC-[60] and DNM-[60] should be considered for further investigating both computing and experimentally.

Chapter 11

Future Studies

The work presented in this study presents a wide scope for further studies. Synthesis of the aldehyde intermediates for adenine, cytosine and guanine, which is a key step in the synthesis of the derivatives, has been difficult because of the insolubility of the DNA bases and the possibility of intramolecular interactions within each individual base. Protecting the bases with di-*tert*-butyl dicarbonate has also presented difficulties because of the ease with which the BOC group can be deprotected with a weak acid. For the guanine-based derivative, the key step in the synthesis of the aldehyde intermediate was oxidation of the precursor with osmium tetroxide. The work in Chapter 4 demonstrated the possibility of synthesizing the NMA-[60] derivative. Certainly, further studies carried out to refine this synthesis route could isolate the mono-adduct thus providing the H-bonding partner for NMT-[60]. The key steps identified indicate that successful synthesis of the derivatives is only a step away, and because of the similarities in the synthesis routes for the adenine- and cytosine-based derivatives, if one route is refined and made successful, then the same route can be followed with the other derivative. Further refinement of the synthesis route for the guanine-based derivative is also required. The successful synthesis of all the monomers would enable self-assembly using not only zeolites but also other surfaces as templates.

Computational calculations have been used to predict the electronic properties of the C_{60} derivatives and are subject to the approximations associated with theory, in particular, the underestimation of the HOMO-LUMO gap by DFT calculations. However, the calculations certainly highlight which molecules have properties that are of interest and thus encourage experimentalists to devise synthesis pathways for these molecules. The computational calculations have shown that the fulleropyrrolidine and methanofullerene derivatives have low LUMO energies, small enough to potentially be used as electron acceptors in organic

solar cells. These calculated values could be corroborated by measuring the HOMO and LUMO energies experimentally using cyclic voltammetry.

For experimentalists, it would be particularly useful if it was possible to predict computationally, which synthesis pathways to use. In theory, it is possible to calculate the formation energies of the products of each step in a range of solvents. However, at room temperature, multiple conformations will exist, making it difficult to accurately predict which reactions will be successful. It would be necessary to calculate the Boltzmann weighted formation. The transition states would need to be determined so that activation and reaction energies could be calculated. In addition, it would also be helpful to run calculations analysing the interactions of derivatives with zeolites, but these large-scale systems would require the use of either lower levels of theory or more powerful computational resources.

The Watson-Crick binding mode is likely to be the preferred mode of assembly, with the guanine-cytosine based dimer more likely to occur than the adenine-thymine based dimer, based on the binding energies. The efficacy of the self-assembly of the dimers can be analysed using transmission electron microscopy (TEM) and high resolution SEM.

The BSSE corrected binding energies gave a more accurate representation of the dimer binding energies, therefore, calculating the BSSE corrected binding energies for the methanofullerene dimers as well as the Hoogsteen, Wobble and self-dimer would eliminate the possibility of errors from the basis sets.

The docking calculations showed that all the derivatives, apart from NMT-[60] had high docking scores, possibly a result of high binding affinities with the HIV-1 protease. Five of the fulleropyrrolidine derivatives were also fully docked inside the HIV-1 active site, apart from the NMT-[60] derivative. The interactions identified in the protein-ligand conformations along with the docking scores show that NMA-[60], NMC-[60], NMG-[60], NMD-[60] and NMN-[60] have the potential to be HIV-1 protease inhibitors. The docking interactions could be improved by new algorithms that can dock multiple conformations of a ligand into the

active site followed by quantum scoring of interactions. Biological experiments would be required to measure the EC_{50} and the binding affinities.

Since the N-propylthymine-functionalised methanofullerene had the most exothermic docking score compared to the other methanofullerenes, there is scope for an extension of this work. Linkers to mimic the propyl chain could be inserted into the rest of the methanofullerene monomers as well as the fulleropyrrolidines. This extension might increase the chances for interactions between the attached entities and the active site, thus improving the docking scores. In addition, methanofullerenes with just one functional group and a hydrogen could also be modelled and their docking abilities analysed.

The work on the solvent effects of fullerene derivatives could be extended by calculating the solubility of the derivatives in each solvent. These calculations could be computational and experimental in order to compare the results.

Adsorption of a monomer at the opening of the zeolite channels has been shown in Chapter 3. Further analysis with TEM or SEM can be carried out experimentally to analyse the location of the derivative, inside the channels or at the zeolite opening. If the derivative is located on the zeolite channel opening, the C_{60} cage would act as a stop-cock, blocking the channel. Solvent choice would also play a role in the adsorption since the more soluble the fullerene derivative, the better the adsorption.

Finally, alternative groups could also be used to functionalize C_{60} and the electronic and self-assembling properties of the derivatives studied further. The study of fullerenes, their derivatives and potential applications remains a topic for which further study is not only interesting but also important.

References

1.   H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl and R. E. Smalley, *Nature*, 1985, **318**, 162-163.
2.   W. Kratschmer, L. D. Lamb, K. Fostiropoulos and D. R. Huffman, *Nature*, 1990, **347**, 354-358.
3.   J. M. Hawkings, A. Meyer, T. A. Lewis, S. Loren and F. J. Hollander, *Science*, 1991, **252**, 312-313.
4.   A. Montellano Lopez, A. Mateo-Alonso and M. Prato, *J. Mater. Chem.*, 2011, **21**, 1305-1318.
5.   L. Echegoyen and L. E. Echegoyen, *Acc. Chem. Res.*, 1998, **31**, 593-601.
6.   M. Prato and M. Maggini, *Acc. Chem. Res.*, 1998, **31**, 519-526.
7.   R. C. Haddon, *Science*, 1993, **261**, 1545-1550.
8.   S. H. Hoke, J. Molstad, D. Dilettato, M. J. Jay, D. Carlson, B. Kahr and R. G. Cooks, *J. Org. Chem.*, 1992, **57**, 5069-5071.
9.   M. Ohno, T. Azuma, S. Kojima, Y. Shirakawa and S. Eguchi, *Tetrahedron*, 1996, **52**, 4983-4994.
10.   C. Bingel, *Ber.*, 1993, **126**, 1957-1959.
11.   X. Camps and A. Hirsch, *J. Chem. Soc., Perkin Trans. 1*, 1997, 1595-1596.
12.   F. Diederich, L. Isaacs and D. Philp, *Chem. Soc. Rev.*, 1994, **23**, 243-255.
13.   A. G. Avent, P. R. Birkett, F. Paolucci, S. Roffia, R. Taylor and N. K. Wachter, *J. Chem. Soc., Perkin Trans. 2*, 2000, 1409-1414.
14.   M. Prato, *J. Mater. Chem.*, 1997, **7**, 1097-1109.
15.   M. A. Yurovskaya and I. V. Trushkov, *Russ. Che. Bull, Int. Ed.*, 2002, **51**, 367-443.
16.   M. Prato, M. Maggini, C. Giacometti, G. Scorrano, G. Sandona and G. Farnia, *Tetrahedron*, 1996, **52**, 5221-5234.
17.   M. Maggini, G. Scorrano and M. Prato, *J. Am. Chem. Soc.*, 1993, **115**, 9798-9799.
18.   A. Hirsch, *The Chemistry of Fullerenes*, Thieme, Stuttgart, 1994.
19.   M. Prato, M. Maggini and G. Scorrano, *Synth. Met.*, 1996, **77**, 89-91.
20.   M. Maggini, G. Scorrano, A. Bianco, C. Toniolo, R. P. Sijbesma, F. Wudl and M. Prato, *J. Chem. Soc.*, 1994, 305-306.
21.   A. W. Jensen, S. R. Wilson and D. I. Schuster, *Bioorg. Med. Chem.*, 1996, **4**, 767-779.
22.   S. Bosi, T. Da Ros, G. Spalluto and M. Prato, *Eur. J. Med. Chem.*, 2003, **38**, 913-923.
23.   S. S. Babu, H. Mohwald and T. Nakanishi, *Chem. Soc. Rev.*, 2010, **39**, 4021-4035.
24.   P. Innocenzi and G. Brusatin, *Chem. Mater.*, 2001, **13**, 3126-3139.
25.   Y. He and Y. Li, *PCCP*, 2011, **13**, 1970-1983.
26.   C.-Z. Li, H.-L. Yip and A. K. Y. Jen, *J. Mater. Chem.*, 2012, **22**, 4161-4177.
27.   L. Dou, J. You, Z. Hong, Z. Xu, G. Li, R. A. Street and Y. Yang, *Adv. Mater.*, 2013, **25**, 6642–6671.
28.   S. H. Friedman, D. L. DeCamp, R. P. Sijbesma, G. Srdanov, F. Wudl and G. L. Kenyon, *J. Am. Chem. Soc.*, 1993, **115**, 6506-6509.
29.   P. Morvillo and E. Bobeico, *Sol. Energy Mater. Sol. Cells*, 2008, **92**, 1192-1198.
30.   G. Yu and A. J. Heeger, *J. Appl. Phys.*, 1995, **78**.
31.   J. Y. Gao, J. C. Hummelen, F. Wudl and J. A. Heeger, *Science*, 1995, **270**, 1789-1791.
32.   Y. He, H.-Y. Chen, J. Hou and Y. Li, *J. Am. Chem. Soc.*, 2010, **132**, 1377-1382.
33.   L. Zhang, K. Pei, H. Zhao, S. Wu, Y. Wang and J. Gao, *Chem. Phys. Lett.*, 2012, **543**, 199-204.
34.   B. Qi and J. Wang, *J. Mater. Chem.*, 2012, **22**, 24315-24325.
35.   D. He, X. Du, Z. Xiao and L. Ding, *Org. Lett.*, 2014, **16**, 612-615.
36.   A. S. Shalabi, S. Abdel Aal and M. M. Assem, *Nano Energy*, 2012, **1**, 608-623.
37.   M. Anafcheh, R. Ghafouri and N. L. Hadipour, *Sol. Energy Mater. Sol. Cells*, 2012, **105**, 125-131.
38.   H. Wang, Y. He, Y. Li and H. Su, *J. Phys. Chem. A*, 2011, **116**, 255-262.

39. S. E. Shaheen, C. J. Brabec, S. Sariciftci, F. Padinger, T. Fromherz and J. C. Hummelen, *Appl. Phys. Lett.*, 2001, **78**, 841-843.
40. J. C. Hummelen, B. W. Knight, F. LePeq, F. Wudl, J. Yao and C. L. Wilkins, *J. Org. Chem.*, 1995, **60**, 532-538.
41. M. M. Wienk, J. M. Kroon, W. J. H. Verhees, J. Knol, J. C. Hummelen, P. A. van Hal and R. A. J. Janssen, *Angew. Chem. Int. Ed.*, 2003, **42**, 3371–3375.
42. M. Lenes, G.-J. A. H. Wetzelaer, F. B. Kooistra, S. C. Veenstra, J. C. Hummelen and P. W. M. Blom, *Adv. Mater.*, 2008, **20**, 2116-2119.
43. W. Ma, C. Yang, X. Gong, K. Lee and A. J. Heeger, *Adv. Funct. Mater.*, 2005, **15**, 1617-1622.
44. A. P. Troshin, H. Hoppe, J. Renz, M. Egginger, Y. J. Mayarova, E. A. Goryachev, S. A. Peregudov, N. R. Lyubovskaya, G. Gobsch, N. Sariciftci Sedar and F. V. Razumov, *Adv. Funct. Mater.*, 2009, **19**, 779-788.
45. L. Ye, S. Zhang, W. Ma, B. Fan, X. Guo, Y. Huang, H. Ade and J. Hou, *Adv. Mater.*, 2012, **24**, 6335-6341.
46. M. T. Dang, L. Hirsch and G. Wantz, *Adv. Mater.*, 2011, **23**, 3597-3602.
47. A. Gajewicz, B. Rasulev, D. T. C, P. Urbaszek, T. Puzyn, D. Leszcynska and J. Leszcynksa, *Adv. Drug Deliver. Rev.*, 2012, **15**, 1663-1693.
48. M. Martínez-Herrera and A. n. Rojas, *J. Phys. Chem. C*, 2010, **115**, 1541-1547.
49. W. Zhuang, A. Lundin and M. R. Andersson, *J. Phys. Chem. A*, 2014, **2**, 2202-2212.
50. H. U. Kim, D. Mi, J.-H. Kim, J. B. Park, S. C. Yoon, U. C. Yoon and D.-H. Hwang, *Sol. Energy Mater. Sol. Cells*, 2012, **105**, 6-14.
51. P. Zhang, C. Li, Y. Li, X. Yang, L. Chen, B. Xu, W. Tian and Y. Tu, *Chem. Commun.*, 2013, **49**, 4917-4919.
52. C. J. Brabec, A. Cravino, D. Meissner, N. S. Sariciftci, T. Fromherz, M. T. Rispens, L. Sanchez and J. C. Hummelen, *Adv. Funct. Mater.*, 2001, **11**, 374–380.
53. H. Wang, J. Liu and Y. Han, *Polymer*, 2013, **54**, 948-957.
54. E. A. Khakina, A. A. Yurkova, A. S. Peregudov, S. I. Troyanov, V. V. Trush, A. I. Vovk, A. V. Mumyatov, V. M. Martynenko, J. Balzarini and P. A. Troshin, *Chem. Commun.*, 2012, **48**, 7158-7160.
55. D. Bonifazi, O. Enger and F. Diederich, *Chem. Soc. Rev.*, 2007, **36**.
56. R. J. Baxter, P. Rudolf, G. Teobaldi and F. Zerbetto, *ChemPhysChem*, 2004, **5**.
57. Y. Shirai, L. Cheng, B. Chen and J. M. Tour, *J. Am. Chem. Soc.*, 2006, **128**, 13479-13489.
58. X. Zhang, F. Yin, R. E. Palmer and Q. Guo, *Surf. Sci.*, 2008, **602**, 885-892.
59. L. Tang, Y. Xie and Q. Guo, *J. Chem. Phys.*, 2011, **135**, 114702-114702.
60. M. D. Upward, P. Moriarty, P. H. Beton, P. R. Birkett, H. W. Kroto, D. R. M. Walton and R. Taylor, *Surf. Sci.*, 1998, **405**, L526-L531.
61. A. Gulino, S. Bazzano, G. G. Condorelli, S. Giuffrida, P. Mineo, C. Satriano, E. Scamporrino, G. Ventimiglia, D. Vitalini and I. Fragala, *Chem. Mater.*, 2005, **17**.
62. M. J. Humphry, P. H. Beton, D. L. Keeling, R. H. J. Fawcett, P. Moriaty, M. J. Butcher, P. R. Birkett, D. R. M. Walton, R. Taylor and H. W. Kroto, *J. Phys.: Condens. Matter*, 2006, **18**.
63. G. Sastre, M. L. Cano, A. Corma, H. Garcia, S. Nicolopoulos, J. M. GonzalezCalbet and M. ValletRegi, *J. Phys. Chem. B*, 1997, **101**.
64. G. Gu, W. Ding, G. Cheng, S. Zhang, Y. Du and S. Yang, *Chem. Phys. Lett.*, 1997, **270**, 135-138.
65. A. Lamrabte, J.-M. Janot, A. Elmidaoui, P. Seta, L.-C. de Ménorval, R. Backov, J. Rozière, J.-L. Sauvajol and J. Allègre, *Chem. Phys. Lett.*, 1998, **295**, 257-265.
66. K. Tanaka, C.-K. Choo, S. Sumi, Y. Kamitani, T. Fujii, K. Satoh, K.-i. Fukuda, R. Nakata, M. Yoshimune, Y. Yoshinaga and T. Okuhara, *J. Phys. Chem. B*, 2002, **106**, 4155-4161.
67. J. R. Herance, E. Peris, J. Vidal, J. L. Bourdelande, J. Marquet and H. García, *Chem. Mater.*, 2005, **17**, 4097-4102.
68. E. H. Ellison, *J. Phys. Chem. B*, 2006, **110**, 11406-11414.
69. F. Marquez and A. J. Sabater, *J. Phys. Chem. A*, 2005, **109**, 1559-1563.

70. D. Brühwiler and G. Calzaferri, *Microporous Mesoporous Mater.*, 2004, **72**, 1-23.
71. O.-H. Kwon, H. Yoo, K. Park, B. Tu, R. Ryoo and D.-J. Jang, *J. Phys. Chem. B*, 2001, **105**, 4195-4199.
72. A. Dyer, *An Introduction to Zeolite Molecular Sieves*, Great Britain, 1988.
73. C. K. McLaughlin, G. D. Hamblin and H. F. Sleiman, *Chem. Soc. Rev.*, 2011, **40**, 5647-5656.
74. F. Huang and O. A. Scherman, *Chem. Soc. Rev.*, 2012, **41**, 5879-5880.
75. D. M. Guldi and N. Martin, *J. Mater. Chem.*, 2002, **12**, 1978-1992.
76. L. Sánchez, N. Martín and D. M. Guldi, *Angew. Chem. Int. Ed.*, 2005, **44**, 5374-5382.
77. R. Marczak, V. T. Hoang, K. Noworyta, M. E. Zandler, W. Kutner and F. D'Souza, *J. Mater. Chem.*, 2002, **12**, 2123-2129.
78. S. Q. Xiao, Y. L. Li, H. J. Fang, H. M. Li, H. B. Liu, Z. Q. Shi, L. Jiang and D. B. Zhu, *Org. Lett.*, 2002, **4**, 3063-3066.
79. J. L. Sessler, J. Jayawickramarajah, M. Sathiosatham, C. L. Sherman and J. S. Brodbelt, *Org. Lett.*, 2003, **5**, 2627-2630.
80. J. L. Sessler, J. Jayawickramarajah, A. Gouloumis, T. Torres, D. M. Guldi, S. Maldonado and K. J. Stevenson, *Chem. Commun.*, 2005, 1892-1894.
81. F. D'Souza, S. Gadde, D. M. S. Islam, S.-C. Pang, A. L. Schumacher, M. E. Zandler, R. Horie, Y. Araki and O. Ito, *Chem. Commun.*, 2007, 480-482.
82. V. Nandwana, L. A. Serrano, K. M. Solntsev, B. Ebenhoch, Q. Liu, G. Y. Tonga, I. D. W. Samuel, G. Cooke and V. M. Rotello, *Langmuir*, 2013, **29**, 7534-7537.
83. A. Kumar, J.-H. Hwang, S. Kumar and J.-M. Nam, *Chem. Commun.*, 2013, **49**, 2597-2609.
84. S. Sivakova and S. J. Rowan, *Chem. Soc. Rev.*, 2005, **34**, 9-21.
85. M. Karplus, M. Levitt and A. Warshel, *The Nobel Prize in Chemistry 2013-Advanced Information*, Accessed 20 January, 2014.
86. Accrelys.com, *Microporous and Mesoporous Materials*, Accessed 14 May, 2013.
87. J. C. Meza, R. S. Judson, T. R. Faulkner and A. M. Treasurywala, *J. Comput. Chem.*, 1996, **17**, 1142-1151.
88. P. Atkins and R. Friedman, *Molecular Quantum Mechanics*, 4th edn., Oxford University Press, Oxford, 2005.
89. F. Jensen, *Introduction to Computational Chemistry*, Wiley, New York, 1999.
90. N. L. Allinger, *J. Am. Chem. Soc.*, 1977, **99**, 8127-8134.
91. A. R. Leach, *Molecular Modelling, Principles and Applications*, 2nd edn., Longman Limited, Singapore, 1996.
92. Cosmotherm, COSMOlogic GmbH & Co, KG: Leverkusen, Germany, Editon edn., 2010.
93. A. Alvarez, E. Ochoa, Y. Verdecia, M. Suárez, M. Solá and N. Martín, *J. Org. Chem.*, 2005, **70**, 3256-3262.
94. M. Cases, M. Duran, J. Mestres, N. Martín and M. Solà, *J. Org. Chem.*, 2000, **66**, 433-442.
95. M. Solá, J. Mestres, J. Martí and M. Duran, *Chem. Phys. Lett.*, 1994, **231**, 325-330.
96. S. Guha and K. Nakamoto, *Coord. Chem. Rev.*, 2005, **249**, 1111-1132.
97. P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864-B871.
98. A. Hinchcliffe, *Molecular Modelling for Beginners*, 2nd edn., Wiley, Manchester, United Kingdom, 2008.
99. R. Stowasser and R. Hoffmann, *J. Am. Chem. Soc.*, 1999, **121**, 3414-3420.
100. E. J. Baerends, O. V. Gritsenko and R. van Meer, *PCCP*, 2013, **15**, 16408-16425.
101. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648-5652.
102. J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh and C. Fiolhais, *Phys. Rev. B*, 1992, 6671-6687.
103. A. d. Leon, A. F. Jalbout and V. A. Basiuk, *Chem. Phys. Lett.*, 2008, **452**, 306-314.
104. K. Kavitha and P. Venuvanalingam, *J. Org. Chem.*, 2005, **70**, 5426-5435.
105. S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553-566.
106. A. R. Leach, B. K. Shoichet and C. E. Peishoff, *J. Med. Chem.*, 2006, **49**, 5851-5855.

107. Q. Shen, B. Xiong, M. Zheng, X. Luo, C. Luo, X. Liu, Y. Du, J. Li, W. Zhu, J. Shen and H. Jiang, *J. Chem. Inf. Model.*, 2010, **51**, 386-397.
108. A. Niazi and R. Leardi, *J. Chemom.*, 2012, **26**, 345-351.
109. A. Klamt, *J. Phys. Chem.*, 1995, **99**, 2224-2235.
110. A. Klamt, V. Jonas, T. Bürger and J. C. W. Lohrenz, *J. Phys. Chem. A*, 1998, **102**, 5074-5085.
111. M. K. Shukla and J. Leszczynski, *Chem. Phys. Lett.*, 2006, **428**, 317-320.
112. P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497-1500.
113. A. Zaritsky, C. L. Woldringh, M. Einav and S. Alexeeva, *J. Bacteriol.*, 2006, **188**, 1667-1679.
114. T. Gustavsson, A. Banyasz, E. Lazzarotto, D. Markovitsi, G. Scalmani, M. J. Frisch, V. Barone and R. Improta, *J. Am. Chem. Soc.*, 2006, **128**, 607-619.
115. J. Y. Salpin, S. Guillaumont, J. Tortajada, L. MacAleese, J. Lemaire and P. Maitre, *ChemPhysChem*, 2007, **8**, 2235-2244.
116. Z.-Q. Xu, Y.-L. Qiu, S. Chokekijchai, H. Mitsuya and J. Zemlicka, *J. Med. Chem.*, 1995, **38**, 875-882.
117. A. Pasquarello, M. Schlüter and R. C. Haddon, *Science*, 1992, **257**, 1660-1661.
118. A. M. López, F. Scarel, N. R. Carrero, E. Vázquez, A. Mateo-Alonso, T. D. Ros and M. Prato, *Org. Lett.*, 2012, **14**, 4450-4453.
119. M. Zhang, M. Yudasaka, S. Bandow and S. Iijima, *Chem. Phys. Lett.*, 2003, **369**, 680-683.
120. A. Lamrabte, J. M. Janot, L. C. de Ménorval, R. Backov, J. Rozière, J. L. Sauvajol, J. Allègre and P. Seta, *Synth. Met.*, 1999, **103**, 2426-2427.
121. J. C. Scanlon and L. B. Ebert, *J. Phys. Chem.*, 1993, **97**, 7138-7140.
122. M. Sathish and K. i. Miyazawa, *Molecules*, 2012, **17**, 3858-3865.
123. N. Krainara, F. Illas and J. Limtrakul, *Chem. Phys. Lett.*, 2012, **537**, 88-93.
124. R. N. Goyal, S. Chatterjee and S. Bishnoi, *Electroanalysis*, 2009, **21**, 1369-1378.
125. S. Dey and P. Garner, *J. Org. Chem.*, 2000, **65**, 7697-7699.
126. B. Zinic, I. Krizmanic, D. Vikic-Topic, D. Srzic and M. Zinic, *Croat. Chem. Acta*, 2001, **74**, 399-414.
127. A. Isidro-Llobet, M. Álvarez and F. Albericio, *Chem. Rev.*, 2009, **109**, 2455-2504.
128. F. Wojciechowski and R. H. E. Hudson, *J. Org. Chem.*, 2008, **73**, 3807-3816.
129. A. Porcheddu, G. Giacomelli, I. Piredda, C. Mariolino and N. Giammario, *Eur. J. Org. Chem.*, 2008, 5786-5797.
130. S. A. Sikchi and P. G. Hultin, *J. Org. Chem.*, 2006, **71**, 5888-5891.
131. A. H. St. Amant and R. H. E. Hudson, *Org. Biomol. Chem.*, 2012, **10**, 876-881.
132. J. M. Heemstra and D. R. Liu, *J. Am. Chem. Soc.*, 2009, **131**, 11347-11349.
133. J. A. Joule and K. Mills, *Heterocyclic Chemistry*, 5th edn., Blackwell Publishing Ltd, United Kingdom, 2010.
134. Z.-C. Liu, D.-S. Shin, K.-T. Lee, B.-H. Jun, Y.-K. Kim and Y.-S. Lee, *Tetrahedron*, 2005, **61**, 7967-7973.
135. S. A. Thomson, J. A. Josey, R. Cadilla, M. D. Gaul, C. Fred Hassman, M. J. Luzzio, A. J. Pipe, K. L. Reed, D. J. Ricca, R. W. Wiethe and S. A. Noble, *Tetrahedron*, 1995, **51**, 6179-6194.
136. Glaxo Group Limited. (2009) *New Small-Molecule Synthetic Antimicrobial. European Patent Office. EP 200807610 A1*.
137. N. S. Sariciftci, L. Smilowitz, A. J. Heeger and F. Wudl, *Science*, 1992, **258**, 1474-1476.
138. U. Hahn, F. Cardinali and J.-F. Nierengarten, *New J. Chem.*, 2007, **31**, 1128-1138.
139. A. Sastre-Santos, C. Parejo, L. Martin-Gomis, K. Ohkubo, F. Fernandez-Lazaro and S. Fukuzumi, *J. Mater. Chem.*, 2011, **21**, 1509-1515.
140. Z. Shi, Y. Li, H. Gong, M. Liu, S. Xiao, H. Liu, H. Li, S. Xiao and D. Zhu, *Org. Lett.*, 2002, **4**, 1179-1182.
141. H. Fang, S. Wang, S. Xiao, J. Yang, Y. Li, Z. Shi, H. Li, H. Liu, S. Xiao and D. Zhu, *Chem. Mater.*, 2003, **15**, 1593-1597.

142. M. T. Rispens, L. Sanchez, E. H. A. Beckers, P. A. van Hal, A. Schenning, A. El-ghayoury, E. Peeters, E. W. Meijer, R. A. J. Janssen and J. C. Hummelen, *Synth. Met.*, 2003, **135**, 801-803.
143. M. T. Rispens, L. Sanchez, J. Knol and J. C. Hummelen, *Chem. Commun.*, 2001, 161-162.
144. L. Sánchez, M. T. Rispens and J. C. Hummelen, *Angew. Chem. Int. Ed.*, 2002, **41**, 838-840.
145. M. Yamada, T. Akasaka and S. Nagase, *Chem. Rev.*, 2013, **113**, 7209-7264.
146. J. Clayden, N. Greeves, S. Warren and P. Wothers, *Organic Chemistry*, Oxford University Press, United States, 2001.
147. S. H. Friedman, P. S. Ganapathi, Y. Rubin and G. L. Kenyon, *J. Med. Chem.*, 1998, **41**, 2424-2429.
148. G. L. Marcorin, T. Da Ros, S. Castellano, G. Stefancich, I. Bonin, S. Miertus and M. Prato, *Org. Lett.*, 2000, **2**, 3955-3958.
149. V. S. Lee, P. Nimmanpipug, O. Aruksakunwong, S. Promsri, P. Sompornpisut and S. Hannongbua, *J. Mol. Graphics Modell.*, 2007, **26**, 558-570.
150. Z. Zhu, D. I. Schuster and M. E. Tuckerman, *Biochemistry*, 2003, **42**, 1326-1333.
151. S. Durdagi, C. T. Supuran, T. A. Strom, N. Doostdar, M. K. Kumar, A. R. Barron, T. Mavromoustakos and M. G. Papadopoulos, *J. Chem. Inf. Model.*, 2009, **49**, 1139-1143.
152. H. Tzoupis, G. Leonis, S. Durdagi, V. Mouchlis, T. Mavromoustakos and M. G. Papadopoulos, *J. Comput. Aided Mater. Des.*, 2011, **25**, 959-976.
153. S. Durdagi, T. Mavromoustakos, N. Chronakis and M. G. Papadopoulos, *Bioorg. Med. Chem.*, 2008, **16**, 9957-9974.
154. J. L. Sebaugh, *Pharm. Stat.*, 2011, **10**, 128-134.