QUANTUM CHEMICAL STUDIES ON HYDROGEN-BONDED BASE PAIR AND BACKBONE COMPONENTS OF NOVEL INFORMATION-BEARING MACROMOLECULAR DUPLEXES

Submitted by

SIAMKHANTHANG NEIHSIAL

DEPARTMENT OF CHEMISTRY
SCHOOL OF PHYSICAL SCIENCES

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

CHEMISTRY

TO

NORTH-EASTERN HILL UNIVERSITY
SHILLONG-793 022
INDIA

MAY 2009

NORTH-EASTERN HILL UNIVERSITY

SHILLONG-793 022, INDIA

9th May, 2009

DECLARATION

I, Mr. Siamkhanthang Neihsial, do hereby declare that the subject matter of this Thesis is the record of work done by me, that the contents of this Thesis did not form basis of the award of any previous degree to me or to the best of my knowledge to anybody else, and that the Thesis has not been submitted by me for a research degree to any other University/Institute.

This is being submitted to the North Eastern Hill University for award of the degree of Doctor of Philosophy in Chemistry.

(Siamkhanthang Neihsial)

CANDIDATE

(Prof. B. Myrboh)

HEAD

Head
Department of Chemistry
North-Eastern Hill University
Shillong- 793022.

(Prof. R. H. Duncan Lyngdoh)

SUPERVISOR

Professor
Department of Chemistry
North-Eastern Hill University
Shillong- 793022.

Acknowledgements

First and foremost, it gives me great pleasure to express my sincerest and heartfelt gratitude to **Professor R. H. Duncan Lyngdoh**, my Research Supervisor, who introduced me to the field of macromolecules in particular and Theoretical/ Computational Chemistry in general. I thank him for the valuable guidance, able supervision, cheerful enthusiasm, ever-friendly nature and support that he gave throughout and for making all-out efforts towards the completion of my research work in due time. His great encouragement and untiring support in both academic and personal life, spiritually are of great value to me and cannot be simply forgotten throughout my life. It has been an honor to work with him.

I would like to express my gratitude to Prof. B. Myrboh, Head, Department of Chemistry, NEHU, for providing all the necessary facilities and opportunity to do my research.

I wish to express my warm and sincere thanks to Professor R. K. Poddar, Coordinator, UGC, SAP-DSA-II & III, Department of Chemistry, NEHU, for financial support to enable me to attend some workshops, *viz.*, the "Workshop on Molecular Modeling" (1^{st} - 4^{th} December, 2005) at the Centre for Biotechnology, Anna University, Chennai, and the School on "Numerical Quantum Many-Body Methods in Physics and Chemistry" (29^{th} Oct - 4^{th} November, 2007) at the Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur Campus, Bangalore.

My deep gratitude also goes to Dr. A. K. Chandra for helping me in learning the Linux operating system and for arranging Quantum Chemistry Lectures for the Theoretical/Computational Chemistry Research Scholars of this Department.

I owe my sincere thanks also to Dr. S. Aravamudhan, Senior Lecturer (now retired), who can be approached any time for any problems especially in the theoretical basis of chemistry, programming languages and numerical analysis. His dedication to science heartens me.

I also express my gratitude to Prof. Henry F. Schaefer III, Director, Center for Computational Chemistry, University for Georgia, Athens, Georgia, USA, for his advice in what basis sets should be used for MP2 ab initio calculations. Personally, I would like to thank him for his constant encouragement and motivation in my research career. I thank Professor R. H. Duncan Lyngdoh for introducing me to him.

I express my gratitude to CCL.NET List (Computational Chemistry List), through which I have often cleared up my doubts by putting up questions and receiving the answers in the next 24 hours. Through CCL.NET I learned a lot about Computational Chemistry, as 3 to 4 messages are send to me in my e-mail every day.

I wish to convey my heartfelt thanks and gratitude to my senior research colleagues Dr. Pebam Munindro Singh, Dr. Peter G. S. Dkhar and Dr. Zodinpui Pachuau for the encouragement, inspiration and companionship and help they gave me during my initial research career. Special thanks go to Dr. Zodinpuia who taught me the Gaussian Ssoftware for the first time. I wish them all success in life.

I am much indebted to my research colleagues, Mr. A. Munaf Kharbuli, Mr. Gunajyoti Das, Mr. Kiew Shaprang Kharnaior, Mr. Henry N. Pangamte, Ms. Tejeshwori Salam and Ms. Shougaijam Premila Devi for lending a helping hand, their valuable help, patience and constant encouragement. I wish them success in life too.

I also express my thanks to all the Faculty, Research Scholars and Staff of the Chemistry Department, for their moral support and help. I wish them all the best.

During my PhD, I have collaborated with some colleagues, for whom I have great regard, mention may be made of Mr. Mathew Varghese, Research Scholar, School of Pure and Applied Physics, Mahatma Gandhi University, Kottayam, and Mr. Ranjib Deb, Research Scholar, Department of Biotechnology, Pune University, for their suggestions regarding molecular mechanics and molecular dynamics simulations even though I have not been able to use these yet in my research work.

My thanks also go to the Librarians, NEHU for their help and cooperation.

I also acknowledge my thanks to the University Grants Commission, Government of India, New Delhi, for financial support in the form of **Rajiv Gandhi National Fellowship Schemes for ST Candidates.**

Finally, I wish to place on record my sincerest and heartfelt thanks to ***my beloved mother*** who brought me up in the fear of the Lord since childhood and would never forget me in her daily prayers. She is a model for me as she taught me courage, humility and honesty above other virtues. My thanks also go to my elder brother who stood by me as the source of encouragement and support throughout, and to my elder sister and her husband for their untiring moral support. I dedicated this Thesis to them to honor their love, patience and support.

Above all, I thank God for He is the source of strength and His help is beyond measure.

9th May, 2009

NEHU, Shillong

SIAMKHANTHANG NEIHSIAL

NORTH-EASTERN HILL UNIVERSITY
SHILLONG-793 022, INDIA

DEDICATION

Dedicated to my loving parents

(Late) Mr. SATKHOPAU and Mrs. CHINGKHOCHIIN

"Teach a wise man, and he will be wiser; teach a good man, and he will learn more. For the reverence and fear of God are basic to all wisdom.. Knowing God results in every other kind of understanding."

Proverbs (The Living Bible)
8: 9,10

Contents

Abbreviations

DNA = Deoxynucleic acid

H-bond = Hydrogen bond

AT = adenine-thymine

GC = guanine-cytosine

RNA = Ribonucleic acid

mRNA = messenger Ribonucleic acid

p-RNA = pyranose Ribonucleic acid

TNA =Threose nucleic acid

PNA = Peptide nucleic acid

MP2 = Moller-Plesset second order perturbation theory

CCSD(T) = Coupled-Cluster with Single and Double and Perturbative Triple excitations

MO = Molecular Orbital

MOPAC = Molecular Orbital Package

CHARMM = Chemistry at HARvard Macromolecular Mechanics

AMBER = Assisted Model Building with Energy Refinement

DFT = Density Functional Theory

B3LYP = Becke, three-parameter, Lee-Yang-Parr

BFGS = Broyden–Fletcher–Goldfarb–Shanno

DFP = Davidon-Fletcher-Powell

HF-SCF = Hartree-Fock Self-Consistent Field

Abbreviations

LCAO = Linear Combination of Atomic Orbital

ZDO = zero differential overlap

CNDO = Complete Neglect of Differential Overlap

INDO = Intermediate Neglect of Differential Overlap

MNDO = Modified Neglect of Differential Overlap

AM1 = Austin Model 1

PM3 = Parametrization Method 3

PM6 = Parametrization Method 6

CISD = configuration interaction (CI) calculations at the single and double excitations level

QCISD = quadratic CI (with single and double excitations)

CASSCF = complete active space self-consistent field

MCSCF= Multi-Configurational Self-Consistent Field

MRCI = Multi-Reference CI

STO = Slater Type Orbitals

GTO = Gaussian Type Orbitals

PGTO = primitive Gaussian type orbitals

cc-PVDZ = Correlation-consistent polarized valence double-zeta basis set

cc-pVTZ = Correlation-consistent polarized valence triple-zeta basis set.

cc-pVQZ = Correlation-consistent polarized valence quadruple-zeta basis set.

cc-pV5Z = Correlation-consistent polarized valence quintuple-zeta basis set

Abbreviations

cc-pV6Z=(the correlation-consistent polarized valence Double/Triple/ Quadruple/Quintuple/Sextuple Zeta basis sets).

GGA = generalized gradient approximation

KS = Kohn-Sham

LSDA =Local Spin Density Aprroximation

BSSE = basis set superposition error

ESP = Electrostatic Potential

MEP = Molecular electrostatic potential

NPA= natural population analysis.

CHELP = Charges from Electrostatic Potential

CHELPG = Charges from Electrostatic Potential, Grid method

MK = Merz-Singh-Kollman (MK) scheme

ZPVE = Zero point vibrational energy

CHAPTER ONE

INTRODUCTION

I.1 Discovery of Hydrogen Bonding

The hydrogen bond, weak though it is, plays a key role in chemistry and biology. Its consequences for biology are enormous and we may safely say that life as we know it would not exist without it. Hydrogen bonds play a key role in determining the shapes, properties and functions of biomolecules like DNA and proteins. Structural chemistry is built upon the concept of binding between atoms to form molecules or complexes, and this concept extends also to the hydrogen bond, as will be discussed later. In this Thesis, the abbreviation ***"H-bond"*** will often be used to indicate "hydrogen bond".

It is claimed that the German chemists Werner[1] and Hantzsch[2] along with Pfeiffer discovered the hydrogen bond. Moore and Winmill[3] used the term *weak union* to describe the properties of amines H-bonded in aqueous solution, and these two have been credited with the first cogent discovery of the H-bond. Pauling[4] wrote a general paper on the nature of the chemical bond (a precursor to his now famous book. There he discussed the [H-F...H]- ion (which contains the strongest H-bond known to man) using the term *hydrogen bond* possibly for the first time to describe this phenomenon.

In 1935–36, definitive papers on hydrogen bonding were brought out by Pauling[5] on hydrogen bonds in water and ice, and by Bernal and Megaw[6] on hydroxyl bonds in metallic hydroxides, minerals, and water. It was the chapter on hydrogen bonding in Pauling's book *The Nature of the Chemical Bond*[7] that really introduced the concept of the hydrogen bond to the world of chemistry. The H-bond has since then captured the fascination of chemists down the decades since its discovery and initial study.

I.2 Characteristics of the Hydrogen Bond

Hydrogen bonds are formed in the context of a covalent bond **A-H** behaving as a proton donor to an acceptor system **B** to form a hydrogen bond **A-H...B**. The systems **A** and **B** are most often electronegative atoms, although they may include atoms with weak electronegativity, while **B** may be any proton-accepting system, including even a delocalized pi electron sextet. When the electronegativity of the atom **A** (as defined by Pauling[7]) relative to hydrogen **H** within the **A-H** covalent bond is sufficient to withdraw electrons and leave the proton partially unshielded, the system **A-H** can become a proton donor. To interact with this donor **A-H** bond, the acceptor system **B** must have lone-pairs electrons or polarizable pi electrons. Strong H-bond energies usually range between 15 and 40 kcal/mol; for moderate H-bonds, the energy is 4–15 kcal/mol. For weak H-bonds, the bond energy is only about 1–4 kcal/mol.

The H-bond is thus a non-covalent bond between an electron-deficient hydrogen atom and a region of high electron density. In an H-bond of the **A-H...B** type, **A** is the electronegative element and **B** is the area with an excess of electrons. H-bonds having **A** or **B** as the electronegative first row atoms F, O and N are the most frequent and have been carefully studied, but H-bonds with the second row analogues Cl, S and P are also well-known. Carbon hydrogen bonds (like **C-H...O** H-bonds) have also received attention in recent years due to their abundance and expected role in biomolecules[8-10]. More recently, H-bonds of the type **A-H...Π** (**A** being O or C; and **Π** being a polarizable pi electron system) have been detected, where it has been shown that, contrary to expectation, they significantly contribute to the stability of biomacromolecules and molecular clusters[11-12].

H-bonding attractions can occur between two or more molecules (*inter*-molecular), or within different parts of a single molecule (*intra*-molecular). The H-bond is a strong fixed dipole-dipole van der Waals-Keesom force, but weaker than covalent, ionic and metallic bonds. The H-bond is somewhere between a covalent bond and an electrostatic intermolecular attraction in its broad nature, but not in its strength. Intermolecular hydrogen bonding is responsible for the high boiling point of water (100° C) compared to H_2S and other Group XVI hydrides. This is because of the strong H-bond in water arising out of the high electronegativity of oxygen among the Group XVI elements. Intra-molecular H-bonding is much responsible too for the secondary and tertiary structures of proteins and nucleic acids.

This Thesis occupies itself with the role of hydrogen bonding in the formation of pairs between nitrogenous aromatic heterocyclic bases. The chief cue is taken from the world of biological chemistry, where such base pairs are found in H-bonded form among the nucleic acid macromolecular duplexes typified by the DNA double helix discovered by Watson and Crick[13]. The H-bond is arguably the most thoroughly studied type of weak interaction in the world of chemistry, ranging from simple inorganic complexes like the hydrogen fluoride dimer to the most complex interactions involving the units of proteins and nucleic acids binding together to form these biological macromolecules, for which the pattern of H-bonding is crucial to the biological role of the complex thus generated.

The classical hydrogen bond consists of a weak interaction (of a range appreciably smaller than ionic or covalent bonds) between a proton donor system **A–H** and a proton acceptor **B**, where **A** and **B** are electronegative atoms like nitrogen, oxygen

and fluorine. Atom **B** should have at least one lone pair of electrons. As the two systems approach, the hydrogen atom **H** forms a sort of "bridge" between **A** and **B**. Changes in the covalent **A–H** bond length along with a redistribution of electronic charge takes place, and these serve to result in lowering of energy which goes towards stabilization of the H-bonded complex **A–H...B**. Experimentally measured strengths of hydrogen bonds most commonly reveal a range from about 1 to about 30 kcal/mol. The hydrogen bond itself consists of the interaction between atoms **H** and **B**, written en as **H...B**, with the bond-length R_{hb}. The H-bond is also described by the distance R_{ab} between the two electronegative atoms **A** and **B** which is expected to be smaller than the sum of the van der Waals radii of **A** and **B**. The H-bond angle **A-H...B** is also another determinant of the H-bonded structure, where values near 180° may be deemed as optimal for lowering of energy through H-bonding. Due to the deshielded character of the hydrogen atom (recognizable more or less as a proton), proton NMR spectroscopy can readily characterize an H-bond. The H-bond also has its own characteristics evinced by vibrational spectroscopy corresponding to the vibrational modes associated with the hydrogen bond and its immediate environment.

The potential energy surface associated with the formation of a hydrogen bond is usually very flat and shallow around the minimum. Thus, appreciable changes in the **H...B** distance may take place without correspondingly large changes in the total energy. This calls for a certain degree of rigor during quantum chemical calculations and geometry optimization in order to ensure that a true minimum is arrived at. Rigor or stringency of geometry optimization necessitates very small gradient norms (much smaller than usual), which is quite demanding on computer time..

H-bonds lengths as determined by the **H...B** distance usually range from about 1.7 Å to about 2.2 Å for first row cases. Very strong H-bonds may be shorter in length, as for the hydrogen maleate anion system or for the fluoride-hydrogen fluoride anion system in KHF_2 (the strongest H-bond known). Appreciably longer H-bonds may include those involving the C-H moiety as a proton donor, which are generally weak. Hydrogen bonds of intermediate strength involve N-H, O-H or even S-H as proton donors, where the proton acceptors include nitrogen, oxygen, fluorine, chlorine and sulphur. Finally, it is possible to also have *bifurcated* hydrogen bonds, where two proton acceptors are associated with a single proton donor, or vice versa.

I..2.1 Analysis of Hydrogen Bond Energy

The hydrogen atom with its single *1s* orbital cannot form more than one pure covalent bond. The attraction between two atoms must be due to electrostatic or Coulombic forces. The hydrogen bond is therefore largely electrostatic in nature and is the bonding between an electron-deficient hydrogen and a region of high electron density which includes the electronegative atoms. We deal below with the energy partitioning of the H-bond, of which the electrostatic term usually contributes most to the net lowering of energy. H-bonding interactions are thus thought to be governed mainly by the attractive electrostatic term (see e.g., Kollman[14]; Kollmann & Allen[15]; Umeyama & Morokuma'[16] in addition to other weak interactions like van der Waals and London dispersion forces. In the gas phase, which most quantum calculations mimic, energies of single H-bonds are usually in the range of 2 to 15 kcal/mol. Even though the interaction energy is partitioned among other weak interactions as well, yet the net stability is attributed to the electrostatic force. The central hydrogen figures in some

kind of three-centre chemical bond due to interactions of the **A–H** bond with the **B** atom. The strength of the H-bond depends on the charges located on the three atoms **A, H** and **B**, where **A** and **B** are electronegative atoms. The H-bonds studied here are of the types **N-H...N, N-H...O, N-H...F**, and in some cases **C-H...F**. In most cases, the H-bonds involving ring nitrogens of heterocyles seem among the more stable.

Understanding the nature of hydrogen bonding includes most importantly the *decomposition* of the hydrogen bond energy into component parts. Chemists naturally first think of the electrostatic or Coulombic contributions, but this is by no means the sole feature. At the position of maximum energy lowering, the Morokuma-Kitaura scheme [17] includes the following interactions, so the net H-bond energy consists of

$$E_{hb} = E_{es} + E_{ex} + E_{pl} + E_{ct} + E_{mx}$$

where E_{hb} is the hydrogen-bonding energy, E_{es} is the *electrostatic* interaction (strictly an infinite multipole expansion often truncated to atom-atom, atom-dipole and dipole-dipole terms), E_{pl} the *polarization* term, E_{ct} the *charge transfer* term (in terms of electrons transfer from filled to empty molecular orbitals) and E_{mx} is the *'mixed'* term (a general name for the remaining effects). The role of exchange repulsions, London dispersion forces and van der Waals interactions may all together be incorporated into the 'mixed' term.

Many other decomposition schemes are also known, some being more sophisticated. This Dissertation does not aim at a detailed decomposition or partitioning of the calculated hydrogen bond energies, but rather on the net effect the hydrogen bonds have in stabilizing the complex, with their influence upon the facility and three-dimensional configuration adopted by the H-bonded base pairs studied here.

I.3 Element-Based Classification of H-Bonds

The electronegative elements incorporated as **A** and **B** in the hydrogen bond **A–H...B** may vary, with consequent effects upon the geometry and energetics of the H-bond. The strongest hydrogen bond known in nature is the [F–H...F] anion, where fluorine is known to be the most electronegative element in the periodic table, and the binding effect is enhanced by the negative charge. As we go down the electronegativity scale, the strength of the binding may be expected to decrease, all other things being considered equal. The following types are prominent amongst substituted H-bonded aromatic nitrogen heterocycles, underscored here for the purposes of this Dissertation: the **N–H...O** type between amino and carbonyl groups, the **N-H...N** type between an amino group and an endocyclic imine nitrogen, the **N–H...F** type between an amino group and a fluorine substituent, the **O–H...N** type between a hydroxyl group and an endocyclic imine nitrogen, and lastly, the **C–H...F** type between an aromatic C-H moiety and exocyclic fluorine. These all have rather distinct ranges of H-bond energy, as will be discussed in the course of this Dissertation. The N-H...O and N-H...N types are very commonly found in nature, as in the Watson-Crick DNA base pairs and also in interactions between amino acid units in proteins and polypeptides.

The **C-H...F** type of H-bonding, deserves further comment. Carbon as a proton donor figures especially when the carbon atom is sp^2 or *sp* hybridized, e.g. in alkenes, alkynes and aromatic rings, since this increases the electronegativity of the carbon, enhancing the deshielded character of the H-atom. Discussion on this has proceeded for some time now (Sutor[18]; Green[19]). Crystallographic studies indicate that such C–H protons can definitely participate in H-bonding, and approach O, N and Cl atoms as

proton acceptors (Taylor & Kennard)[20] . Here, the conclusion was based purely upon geometrical considerations (inter-atomic distances). A survey of the crystal structures of some organometallics (Braga *et al*,)[21] has also indicated further evidence that C-H protons could H-bond to oxygen. The order and range of values for such C-H hydrogen bonds cannot be expected to be very large, as will be amply demonstrated by the extensive calculations embodied in the contents of this Dissertation.

I.4 Hydrogen Bonding in Biological Systems

Hydrogen bonding plays a key role in the maintaining structure and specificity of biological systems like DNA and various proteins, and protein-nucleic acid conglomerates. For DNA, the genetic information is stored in the sequence of bases of the nucleic acid and this information is translated during protein synthesis into the amino acid sequence of the protein synthesized. From the point of view of the storage and transfer of genetic information, the most relevant unit of information in the nucleic acids is the *nucleic acid base pair*, of which two distinct and unique types may be characterized, *viz.*, the guanine-cytosine (GC) and the adenine-thymine (AT) pairs. H-bonds are also responsible for nucleobase-pair formation involved in codon-anticodon pairing and in folding of RNA Other biologically relevant H-bonds occur in proteins and determine their secondary and tertiary structure. For instance, in the α-helical secondary structure of coiled proteins like in silk and hair, H-bonding of the N-H...O type occurs at regular intervals between the amino group of one amino acid and the carbonyl group of the fourth amino acid ahead along the primary sequence of the protein single strand. This Dissertation, however, focuses on H-bonds between aromatic nitrogen heterocycles of the broad general type found in DNA and RNA.

I..4.1 Stabilizations of DNA Secondary Structure

The secondary double-helical structure of DNA is the result of a complex set of interactions. One chief problem in molecular biology arises in attempts to provide a reliable description of the interactions which accounts for the stability of DNA structure (Malenkov *et al.*[22]) . There are three main types of interactions responsible for forming the unique DNA structure. These are (a) the *planar* H-bonded interactions

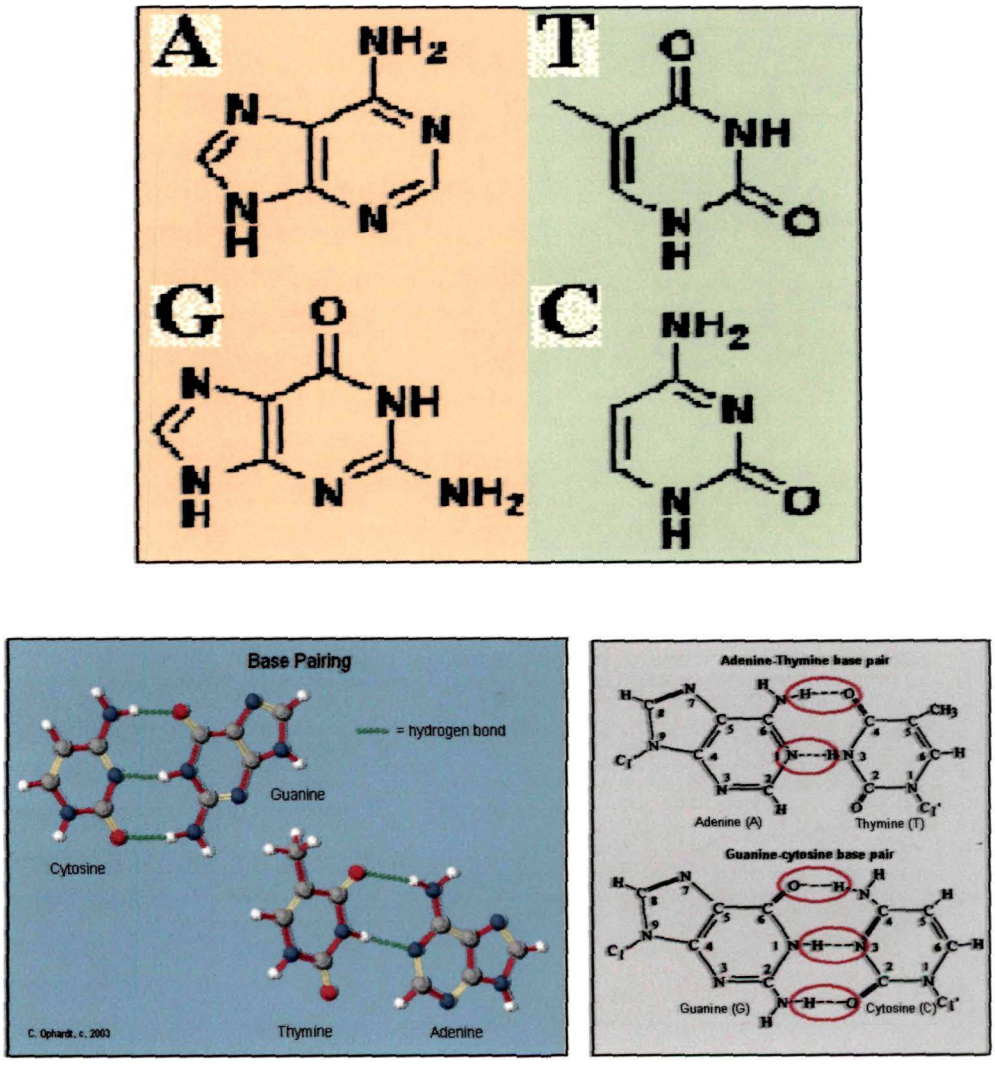


Fig. I.1: The DNA bases and their H-bonded pairs

between bases of the two strands, (b) the *vertical stacking* interactions between the bases within each strand, and (c) the *spatial* interactions amongst the phosphate ions, counter ions and water molecules. The planar H-bonding interactions are believed to play the dominant role, going to form the two unique Watson-Crick base pairs A:T and G:C from the four DNA bases A, T, G and C, as shown in Fig. I.1 (previous page). This coupled with the flexibility of the sugar-phosphate backbone ultimately determines the unique double-helical secondary structure of normal DNA. Portrayed below in Fig. I.2 are two representations of DNA structure, where one shows the incorporated base pairs, and the other shows the double helix.

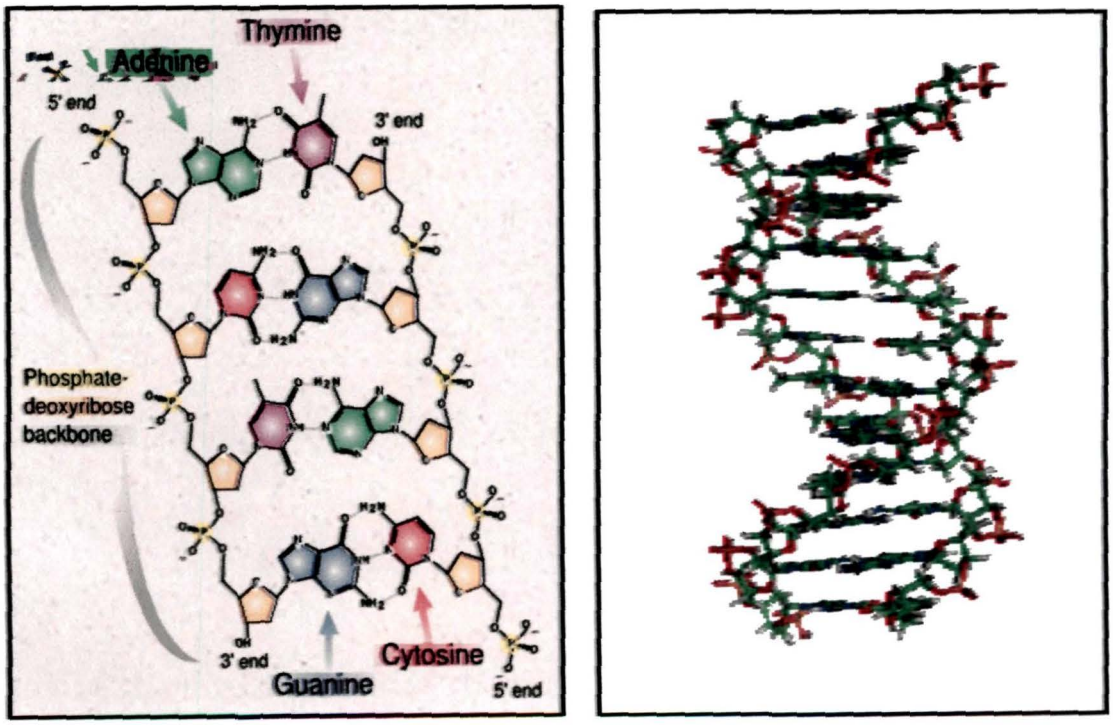


Fig. I.2: The macromolecular duplex of DNA

Apart from the usual Watson-Crick base pairs, there are other constitutional types of base pairs depending on the self-organization of donor and acceptor groups for H-bonds in each base. This understanding of H-bonding as the contributing factor to the stability and structure of DNA (Borer *et al.*,[23]; Kudritskaya & Danilov[24]) is useful to all biochemists and biophysicists who interpret observed facts in terms of the complex biological and biochemical functions (Zhang *et al.*,[25]) , using these concepts for the design of new biochemically functional structures which may be useful as therapeutic agents as well (Uhlmann & Peyman,[26]) . This Dissertation seeks to design *wholly artificial information-bearing duplexes* on the basis of such underlying concepts which govern the formation, geometry and stability of H-bonding between synthetic aromatic nitrogen heterocycles.

I.5 Hydrogen Bonding among Nitrogenous Heterocycles

Nitrogenous heterocycles like pyrroles are known to engage themselves in inter-molecular H-bonding forming long chains of aggregates. Imidazoles and pyrazoles can also form hydrogen-bonded aggregates. Apart from the usual Watson-Crick base-pairs, the four DNA bases may also pair themselves in a variety of other arrangements, of which 29 different structures may be discerned which have two or more H-bonds each (Donohue & Trueblood[27]; Hobza & Sandorfy[28]). Likewise, DNA base analogues like 5-fluorouracil can lend themselves to H-bonded pairing schemes much in the manner of the DNA bases themselves. However, regarding aromatic heterocycles in general, the literature has not yet revealed much extensive study on H-bonded base-pairing between synthetic heteroaromatic systems.

It may be stated that here that one chief aim of this Dissertation is to achieve the design of artificial (and as yet imaginary) base-pairs between aromatic heterocycles which can furnish viable alternatives to the standard DNA base-pairs of nature by way of mimicking their functional (especially the *information-bearing*) aspects. This may be done by reproducing the essential features of the GC and AT pairs, especially with regard to their information-storing capacity and the biological characteristic of semi-conservative replication. It is to be noted that the information-bearing capacity does not arise out of the base pairs considered in isolation, but arises when they are considered in the context of a **restricted** *isomorphic set* of allowed base pairs. Chapters III and IV begin by listing some vital characteristics of such DNA-like base pairs, and then go on to attempt mimicking these features in the base pairs examined.

While the DNA base pairs of nature consist solely of purines and pyrimidines, this Thesis extends the scope of this search to five-membered rings (like pyrroles and imidazoles) and six-membered monocycles (pyridines, pyrimidines and pyrazines). Such artificial base-pairs may be said to successfully mimic the vital characteristics of the natural DNA base-pairs **if** the *essential defining characteristics* are reproduced *qualitatively* (not quantitatively). Such sets of DNA base-pair mimics may be able to lay the foundations for a new order of life itself, if one may be permitted the freedom to engage in some unrealized imagination! For from these solitary pairs can arise the possibility of macromolecular duplexes containing genetic information, and given the appropriate substitutes for the enzymatic machinery required, the possibilities for self-replication of these macromolecules. The next requirement is for a putative equivalent of the translation process and the genetic code, which again requires the equivalent of

the enzymes and proteins involved in protein synthesis. With all his intelligence, man has been so far unable to synthesise life even on the lines that we already know. How much more to think of designing life on the basis of absolutely new and hitherto unexplored systems and processes! This Dissertation is one small and finite initial contribution towards this grand and unfulfilled aim.

I.6 Coding of Genetic Information

Most organisms use the macromolecule DNA to encode genetic information. Some viruses have the DNA molecule replaced by RNA. Both the macromolecules contain three different constituents – (a) the nucleobases, (b) a sugar, either D-(-)-2'-ribose or D (-)-2'-deoxyribose, and (c) a phosphodiester linkage. All three components were and still are prime targets for chemists for manipulation through modification. This Section covers the general principles of nucleic acids with a strong emphasis on the chemical manipulation of the nucleic acid structure and hence of the genetic system.

DNA and RNA are there "just" to encode the genetic information and arrange for its transmission through heredity. For DNA no other function in cells is known. RNA in contrast is known to have, in addition, catalytic functions (splicing, ribosomal) and gene regulatory functions. RNA is also the primary information-bearing macro-molecule in retroviruses. More recently, mRNA molecules have been discovered which encode proteins needed for the biosynthesis of vitamins. The mRNA binds to the produced vitamins, which changes the structure of the mRNA, stopping its translation into protein, and hence facilitating vitamin biosynthesis.

The genetic information is basically contained in the sequence of the four canonical bases adenine, thymine, cytosine and guanine, known as the primary structure of the nucleic acid. They are present, of course, as base residues within the corresponding nucleotides. This base sequence is translated in a complex and highly controlled process into a sequence of amino acids, which folds into a protein. To this end, the information has to be first read from the DNA, and copied onto mRNA, generally one gene at a time, which is called *transcription.* The base (or nucleotide) sequence of the nucleic acid then has to be converted into the amino acid sequence of the synthesized protein, which is termed *translation.* In order to reproduce a cell, the genetic information within the nucleus has itself to be reproduced in the daughter cells. This process is called DNA *replication.* Replication has to be tightly controlled in all unicellular or multicellular organisms in order to maintain genetic integrity. Loss of any integrity here would lead to mutated cells or species, where the general effect of gene mutation is almost always deleterious. Uncontrolled replication of DNA and cell proliferation constitutes the genetic and cytological basis for the cancer groups of diseases.

It is no wonder that chemically modified nucleobases which interfere with the processes of DNA transcription, translation or replication have biologically strong effects. In higher eukaryotic cells, the DNA molecule is present within the cell nucleus. During the process of transcription, the DNA sequence, which needs to be decoded, is "copied" into a mRNA molecule (m = messenger). This mRNA molecule leaves the cell nucleus and travels to the endoplasmatic reticulum where it is bound by ribosomes, which are the actual sites of protein synthesis.

Genetic information is encoded in triplet sequences of DNA and mRNA which are called *codons*. Since there are only 4 major nucleic acid bases to form the nucleic acid script, the number of words possible in the codon dictionary will be a power of 4. A hypothetical single-letter word system will yield only 4^1, *viz.*, four words, while a two-letter system would yield 4^2 or 16 words (codons). Since the number of amino acids in proteins is 20, we would require larger-size codons. This calls for a *triplet* system with three-letter words, so that the total number of codon words becomes 4^3 or 64 codons. Since this is in excess of the minimum of 20 words required, the situation calls for *codon degeneracy*, *viz.*, the possibility that an amino acid may be encoded for by more than one codon, as found for amino acids encoded by 6, 4, 3, 2 or 1 codon, giving a total of 61 codons *actually coding* for amino acids. The 3 *nonsense* codons do not code for any amino acid, but signal the start and the end of a gene for a given protein chain encoded in the DNA. The total number of codons in nature is thus 64, calling for a triplet system to be followed in the genetic code of life.

I.7 Artificial Information-Bearing Macromolecules

The discovery of natural information-bearing macromolecules like DNA and mRNA leads to the possibility of designing and synthesising novel structures that possess the ability to store information even at the molecular level. Storing information at the molecular level means reduction to the smallest possible size scale, much smaller than current devices for storage of information such as magnetic or laser disks. This would allow for greatly reduced space requirements to storing very large amounts of information, something of direct relevance for the modern world of nanotechnology.

Wholly synthetic and artificial systems for storing information at the molecular level may be designed with three principles in mind. The first is to closely follow the pattern already occurring in the naturally occurring DNA and mRNA systems of life itself, where changes may be introduced, chiefly in the *sugar moiety*. The second is to introduce *novel bases* different from the naturally occurring nucleic acid bases Ade, Thy, Gua, Cyt and Ura. The third principle is to design *novel backbones* which are quite different qualitatively from the sugar-phosphate type found in nature.

Several nucleic acid analogs with modified base, sugar, and phosphate backbone have been developed[29]. However, relatively few attempts have been made for the preparation of nucleic acid-like polymers devoid of sugar moiety and phosphodiester linkage[30]. Pitha, Overberger, and other workers have reported the synthesis of several nucleic acid mimics containing nonhydrolyzable polymeric backbones for a variety of applications[31,32] (interference with cell-free protein synthesis[33], anti-viral activity[34], optically active polymers[35], and the base-specific detection of oligo-nucleotides[36]). Nucleobase-functionalized polythiophenes containing adenine and uracil have been synthesized by electrochemical oxidation and their molecular recognition process was studied[37]. without study of their catalytic potential. Komiyama's group demonstrated the ribonuclease activity of a vinyladenine and vinylamine copolymer[38]., noted as considerably higher than poly(vinylamine). Maurel and coworkers conjugated adenine residues to a polyallylamine backbone, followed by the evaluation of its catalytic ability for the hydrolysis of *p*-nitrophenyl acetate[39]. Nucleic acids can coordinate to metal ions through the participation of the base keto oxygen atoms, heterocyclic ring nitrogen atoms, sugar hydroxyl groups, and phosphate oxygen atoms [40].

Eschenmoser and co-workers systematically studied the properties of nucleic acid analogs in which the sugar backbone was replaced by pyranose (p-RNA) or threose sugars (TNA)[41]. The α-hexopyranosyl-(6',4')-linked oligonucleotide analogs synthesized from hexose sugars such as allose, altrose, and glucose displays an inferior base-pairing as compared to natural RNA (attributed to the steric hindrance from the fully hydroxylated hexopyranosyl sugar units[42]). Eschenmoser's group turned their focus toward sterically less hindered pentopyranose sugars. Interestingly, the diastereomeric pentopyranosyl-(4',2')-linked oligonucleotide systems formed Watson–Crick paired double helices more stable than RNA[43]. The threose analogues, *viz.*, threofuranosyl oligonucleotides (TNAs), contain a vicinally connected 3',2'-phosphodiester bond and display efficient base-pairing similar to p-RNA [44]. Strikingly, TNA oligonucleotides were found to cross-pair with RNA and DNA and were also hydrolytically more stable than RNA. Thus, TNA could potentially serve as a template in nonenzymatic template-directed oligomerization of RNA, although this is yet to be tested.

This means that a transition from a TNA world to an "RNA world" is possible. Though prebiotic synthesis of nucleic acid analogs is not likely facile, TNA may be considered a potential primitive replicating system with a relatively simple chemistry and equipped with a stable phosphodiester backbone. Peptide nucleic acid (PNA) is another potential RNA predecessor[45], being an uncharged, achiral analog of RNA or DNA with the ribose-phosphate backbone replaced by aminoethyl glycine units. PNA forms stable double helices with RNA or DNA[46,47], and information is transferred from PNA to RNA in a template-directed fashion [48]. However, PNA monomers are susceptible to an intramolecular *N*-acyl transfer reaction, which would make oligomer

formation difficult under prebiotic conditions[49]. Considering the chemical nature of PNA monomers, it is unlikely that PNA would have been very important in the early existence of life. It is also important to mention that the catalytic potential of PNA, TNA, and other possible RNA analogs has not yet been explored in detail.

We have drawn inspiration from the above work on nucleic acid analogues. Firstly, we have conducted a systematic search for suitable base pairs built up from wholly synthetic heteroaromatic systems, as described in Chapters Three and Four. Then, this Thesis uses a *six-membered ring sugar* (a hexose in pyranose ring form) as the sugar moiety (instead of D-ribose), which idea is advanced in Chapter Five to design a sugar-phosphate backbone for the information-bearing macroduplex. We have also gone far beyond the sugar-phosphate idea for a backbone by designing a completely novel *polyamide backbone* for the macroduplex , also described in Chapter Five.

References

1. Werner, A. *Liebigs Ann*, 1902, **322**, 261.

2. Antzsch, A. *Berichte*, 1910, **43**, 3049.

3. Moore, T. S; Winmill, T. F. *J Comp Soc*, 1912, **101**, 1635.

4. Pauling, L. *J Am Chem Soc*, 1929, **51**, 1010.

5. Pauling, L. *J Am Chem Soc*, 1935, **57**, 2680.

6. Bernal, J. D; Megaw. H. D. *Proc Roy Soc (London)*, 1935, **151A**, 384.

7. Pauling, L. *The Nature of the Chemical Bond. Ithaca*, Cornell University Press, New York, 1939.

8. Jeffrey, G. A. *An Introduction to Hydrogen Bonding*, Oxford University Press, New York, 1997.

9. Desiraju, G. R; Steiner, T. *The Weak Hydrogen Bond*, Oxford University Press, Oxford, 1999.

10. Scheiner, S. *Hydrogen Bonding*, University Press, New York, 1997.

11. Klusak, V; Havlas, Z; Rilisek, L; Vondrasek, J; Svatos. *A Chem Biol*, 2003, **10**, 331.

12. Braun, J; Neusser, H. J; Hobza, P. *J Phys Chem A*, 2003, **107**, 3918.

13. Watson, J. D; Crick, F. H. C. Nature, 1953, **171**, 737.

14. Kollmann, P. A. In *Modern Theoretical Chemistry: Applications of Electronic Structure Theory*, Vol. 4, Part B, 1977.

15. Kollmann, P. A; Allen, L. C. *Chem Rev*, 1972, **72**, 283.

16. Umeyama, H; Morokuma, K. *J Am Chem Soc*, 1977, **99**, 1316.

17. Morokuma, K. and Kitaura, K. In *Chemical Applications of Atomic and Molecular Electrostatic Potentials* (Politzer, P; Truhlar, D. G, eds.), Plenum, New York, 1981

18. Sutor, D. J. *Nature*, 1962, **195**, 68.

19. Green, R. D. *Hydrogen Bonding by C-H Groups*, Wiley Interscience, New York, 1974.

20. Taylor, R; Kennard, O. *J Am Chem Soc*, 1982, **104**, 5063.

21. Braga, D; Grepioni, F; Biradha, K; Pedireddi, V. R; Desiraju, G. R. *J Am Chem Soc*, 1995, **117**, 3156.

22. Malenkov G, G; Gagua, A. V; Timofeev, V. P. *Int J Quant Chem*, 1979, **16**, 655.

23. Borer, P. N; Denyler, B; Tinoco, I; Uhlenbeck, O. C. *J Mol Biol*, 1974, **86**, 843.

24. Kudritskaya, Z. G; Danilov, V. *J Theor Biol*, 1976, **59**, 303.

25. Zhang, Ya Jun, Merz, Kenneth, *J Comput Chem*, 1992, **13**, 1151.

26. Uhlmann, E, Peyman, A. *Chem Rev*, 1990, **90**, 543.

27. Donohue, J; Trueblood, K, N. *J Mol Biol*, 1960, **2**, 363.

28. Hobza, P; Sandorfy, C. *J Am Chem Soc*, 1987, **109**, 1302.

29. Verma, S; Eckstein, F. *Ann Rev Biochem*, 1998, **67**, 99.

30. Nielsen, P. E. *Acc Chem Res*, 1999, **32**, 624.

31. Pitha, P. M; Pitha, J. *Biopolymers, 1970,* **9**, 965.

32. Lan, M. J, Overberger, C. G. *J Polym Sci Polym Chem Ed*, 1987, **25**, 1909.

33. Pitha, J; Pitha, P. M. *Science, 1971,* **172**, 1146.

34. Reynolds, F; Grunberger, D; Pitha, J; Pitha, P. M. *Biochemistry, 1972,* **11**, 3261.

35. Overberger, C. G; Chang, J. Y. *J Polym Sci Polym Chem Ed*, 1989, **27**, 3589.

36.Yashima, E; Suehiro, N; Miyauchi, N; Akashi. M. *J Chromatogr A*,1993, **654**, 151.

37. Baeuerle, P; Emge, A. *Adv Mater*,1998, **10**, 324.

38. Shiiba, T; Komiyama, M; Yashima, E; Akashi. M. *Nucleic Acids Symp Ser*, 1991, **25**, 71.

39. Ricard, J; Vergne, J; Decout, J. L; Maurel. M. C. *J Mol Evol*,1996, **43**, 315.

40. Saenger, W. In *Principles of Nucleic Acid Structure*, pp. 201–219, Springer-Verlag, New York, 1984.

41. Eschenmoser, A. *Science*,1999, **284**, 2118.

42. Groebke, K; Hunziker, J; Fraser, W; Peng, L; Diederichsen, U; Zimmermann, K; Holzner, A; Leumann, C; Eschenmoser, A. *Helv Chim Acta*,*1998*, **81**, 375.

43. Beier, M; Reck, F; Wagner, T; Krishnamurthy, R; Eschenmoser, A. *Science*, 1999, **283**, 699.

44. Schöning, K. U; Scholz, P; Guntha, S; Wu, X; Krishnamurthy, R; Eschenmoser, A. *Science*,2000, **290**, 1347.

45. Nielsen, P. E; Egholm, M; Berg, R. H; Buchardt, O. *Science*,1991, **254**, 1497.

46. Egholm, M; Buchardt, O; Nielsen, P. E; Berg, R. H. *J Am Chem Soc*,1992, **114**, 1895.

47. Egholm, M; Buchardt, O; Christensen, L; Behrens, C; Freier, S.M; Driver, R. H. *Tetrahedron Lett*,1999, **37**, 1413

48. Schmidt, J. G, Nielsen, P. E. Orgel, L. E. *Nucleic Acid Res*,1997, **25**, 4797.

49. Eriksson, M; Christensen, L; Schmidt, J. G; Haaima, G; Orgel, L. E; Nielsen, P. E. *New J Chem*,1998, **22**, 1055.

A chemist walks into a pharmacy and asks the pharmacist for some acetylsalicylic acid.
The pharmacist replies, "You mean asprin?"
The chemist answers, "That's it... I can never remember that word."

CHAPTER TWO

METHODOLOGY AND APPROACH

II.1 Theoretical Treatment of Hydrogen Bonding

Numerous texts and monographs have dealt with the subject of hydrogen bonding from the experimental side. Authors of some well-known titles may be quoted as follows: Pimental and McClellan[1], Vinogradov and Linnell[2], Schuster[3] and Smith[4]. However, few comprehensive treatises have been written which devote themselves sufficiently to the theoretical aspects of the matter. One notable contribution here is *Hydrogen Bonding: A Theoretical Perspective* by Scheiner[5]. This monograph deals exclusively with *ab initio* studies only, still falling short of providing a good overall view, as evinced by the large amount of theoretical work done on hydrogen bonding using semi-empirical MO regimes (even though these are largely outdated by now).

Adequate theoretical treatment of the phenomenon of hydrogen bonding requires a full knowledge of the nature of the *interactions* involved. Various energy partitioning schemes have been devised, as mentioned in the previous Chapter. The accurate modeling of all these interactions within one effective and unified theoretical frame-work is yet the goal of the theoretical chemist. To date, the best efforts are represented by the sophisticated *ab initio* methods, with basis sets ranging up to the triple zeta polarization basis sets and correlation-consistent extended basis sets, along with the incorporation of electron correlation by many-body perturbation theory, Moller-Plesset theory at higher levels (MP2, MP3 and MP4), configuration interaction (CI), coupled cluster methods like CCSD(T), etc. While numerous accounts of such post-Hartree-Fock work have been incorporated into reviews and monographs like that of Scheiner[6],

most high-level work deals mainly with rather small inorganic and simple organic systems. The *ab initio* treatment of the complex interactions involved in biological macromolecules has been limited chiefly to solitary base-pairs and amino acids or small peptides. Extensive use has been made of accurate quantum chemical theory to treat nucleic acid base pairs of various kinds by the groups of Hobza, Sponer and others.[7-17] Larger systems than these would require the lower and less accurate level of semi-empirical MO theory, and these often fall short in most cases to furnish a reliable picture of hydrogen-bonding in biomolecules. The recently developed semi-empirical PM6 formalism of the MOPAC 2009 package shows especial promise here. At the macromolecular level, molecular mechanics and molecular dynamics simulation methods based on classical potentials and parametrized force fields (as found in software programmes like CHARm and AMBER) have been applied and prove to be of lasting usefulness and applicability.

Given the importance of H-bonding in biological systems, considerable theoretical attention has been focused on exploring the nature and strength of these interactions. Experimentally, solitary nucleic acid base pairs have been difficult to study *in vitro*, and there is little data with which theoretical results may be compared. Extensive theoretical calculations have been carried out on H-bonded nucleic acid base pairs using various semi-empirical methods, e.g. by Kudritskaya and Danilov[18], Poltev and Shulyupina[19], Stamatiadou *et al.*[20] and many others. Such approaches have also been applied to chemically modified DNA base pairs in the context of mutagenesis and cancer[21,22]. Use of the *ab initio* regimes for such systems has been performed starting from the pioneer work of Clementi *et al.*[23] and other workers. Several research groups

are now devoted to accurate *ab initio* and post-Hartree-Fock calculations on nucleic acid systems, notably those of Hobza, Sponer and co-worlers. The research group of Schaefer[24,25] has used DFT methods to study H-bonding in the radical anion forms of nucleic acid base pairs. RNA base pairs have been studied here[26] in the context of codon-anticodon pairing and the genetic code using the PM3 method. Other notable studies on DNA and RNA base pairs includes work by Del Bene[27], Hobza and Sandorfy[28], Hrouda et al.[29] and many others. The work of Kollmann's group[30] represents efforts to study DNA and RNA using classical potential methods.

II.2 Approach to Study of Base-Pairing

This Dissertation comprises a set of molecular orbital calculations on various hydrogen-bonded hetero-aromatic systems variously using *(a)* the semi-empirical PM3 SCF-MO method, as well as *(b)* Density Functional Theory (DFT) and *(c)* Moller-Plesset second order perturbation theory (MP2). The first step in this approach consists of *designing* the **solitary** base monomers. For instance, in the case of the *self-associative* base pairs of Chapter III, it is imperative that each monomer base possesses within itself a proton donor system **X-H** as well as a proton acceptor **Y** to allow for formation of a hydrogen bond **X-H...Y** between the two identical monomers. Likewise, for the *hetero*-associative base-pair systems of Chapter IV, the proton donor **X-H** group(s) in one base of the pair has to be complemented by the counterpart proton acceptor **Y** group(s) in the other base, and vice versa. The first step is to design well-defined sets of bases which can pair through H-bonding with each other in the manner just described, notably, within a well-specified configuration. The aim is to find a set of

bases which furnish a number of unique pairs which possess the *same* H-bonded pairing configuration, so that only a limited number of base pairs may be allowed, and all others disallowed due to the configurational constraints. It is this *limitation* in the number and identity of allowed base pairs which is the *real source* of the **information-bearing capacity** of the H-bonded macromolecular duplex envisaged here.

The solitary bases are subjected to *full geometry optimization* by various algorithms like the Davidon-Fletcher-Powell algorithm[31,32,33], the BFGS algorithm[34], as well as the Berny optimization algorithm[35,36] at higher levels of theory. These optimized bases are then arranged in various H-bonded base pairing motifs. Each such pairing motif is also subjected to fully unconstrained geometry optimization. Finally, a particular **unique** given intermolecular pairing configuration is adopted as the standard one for each set of bases, providing the *configurational constraint* to base-pairing. All allowed base pairs arising from a given set of monomer bases must exist within this configuration.

The procedure is as follows. Firstly, all possible base-pairs resulting from the given set are aligned together and a preliminary *screening test* is performed to rule out any pairs characterized by strong repulsive interactions, like close contact within the van der Waals distance between electronegative atoms, or between hydrogen atoms. The only close contacts allowed are those which would be expected to favoring hydrogen-bonding (as gauged from a semi-quantitative viewpoint). This is done by assigning permissible limits for atom-atom contact, as assessed from the van der Waals radii of the concerned atoms covering various cases. By this screening, many of the putative base-pairs are eliminated as being not feasible within the configurational constraint imposed, the remaining few pairs being screened in and then subjected to full geometry

optimization by the chosen MO method without any symmetry restraints. Data noted down for the energy-minimized base-pair concerns the geometry around the hydrogen-bond, the total electronic energy or heat of formation and the intermolecular pairing configuration. It should be noted that the standard pairing configuration as sketched schematically by the preliminary screening procedure assumes coplanarity of the bases in each pair, and may be subject to considerable deviation when actually calculated by theoretical methods with full optimization of geometry.

II.3 Optimisation of Molecular Geometry

Optimisation of molecular geometry means that the determinants of geometry (based on internal, cartesian or other coordinate systems) all attain values that correspond to an energy minimum for the molecule on the potential energy surface generated by the method chosen. The energy-minimized configuration in space is called the *equilibrium geometry*, occurring at the *bottom* of a potential well. The multi-dimensional potential energy surface for a molecular system may have a several points on it in which the forces (or energy first derivatives) vanish; these are called the *stationary points*. An energy minimum (equilibrium geometry) is one such stationary point, as also is the transition state. Vanishing of all the forces leads to a zero gradient norm G_n, (the sum of the squares of all the forces). In practice, the optimization procedure is stopped when the value of the gradient norm falls below a prescribed acceptable threshold value.

The most thorough and rigorous way to find the equilibrium geometry would be to laboriously construct the entire potential energy surface as a function of all coordinates, searching the entire surface for the presence of an energy minimum. The most straight-forward way for optimization is to vary all the variables one at a time, until the function

reaches a minimum, and then extend the same procedure to the other variables. This is a very time-consuming process. For e.g, if *"f"* is a function of x_1, x_2, ... x_i, the task is to find values of all the x_i for which *"f"* is at a minimum. In other words, the condition is that $\delta f/\delta x_i = 0$ for all the x_i. In addition, all second derivatives should be positive for an energy minimum on the potential energy surface. This additional condition imposes even more computational labour. Normally this rigorous procedure is not practical, as the number of variables (internal coordinates) involved in an *N* atom system is *3N - 6* (or *3N–5* in the case of linear molecules). Since the coordinates are coupled, many iterations over all the variables would be necessary and this becomes highly cumbersome as a greater number of variables are involved.

Happily, there are more viable alternatives. By using *energy gradients* (energy first derivatives with respect to molecular coordinates), the presence of an energy minimum may be located more efficiently. Gradient evaluation may be numerical or analytical, the former being more time-consuming. Semi-empirical MO methods allow for a more rapid search for energy minima using gradients which can be evaluated *analytically* without much expense of computer time.

Optimisation algorithms using analytical gradients include the Newton-Raphson[37] method, the Davidon-Fletcher-Powell (DFP)[31,32,33] method, the Broyden-Fletcher-Goldfarb-Shanno (BFGS)[34] algorithm and the Berny optimization algorithm[35,36]. In the DFP, BFGS and other methods, a more facile approach to the energy minimum is possible through use of the energy *second* derivatives contained in the Hessian (or force constant) matrix. These give a more sensitive estimation of the size of the step required to be taken by atoms (or internal coordinates) as they descend towards the minimum.

Several optimization procedures are available, classified into *three* main types, *viz.*, *(a)* non-derivative methods, *(b)* first order derivative methods and *(c)* second order derivative methods,[38,39] as described briefly below:

(a) Non-Derivative Methods: The compute-intensive *grid search* method is an example of a non-derivative minimization algorithm. In this method a cubic grid is placed upon the surface and the value of the function at each node is calculated. The grid point with the lowest energy is chosen as minimum. Clearly, the quality of a grid search depends on the density of the grid mesh; the higher the density of the grid, the higher is the computational expense and accuracy. Initially, a grid with less density can be used and once a minimum energy area is located, the minimization can be refined by increasing the density of the grid. Since we are already in the vicinity of the best point, the convergence of the method further improves.

(b) First-order derivative methods: These procedures use the *first derivatives* of the multi-dimensional potential energy surface to direct the search towards the nearest local minimum. In other words, the information about the slope (but not about the curvature which is given by the second derivative) is used during the optimization procedure. The steepest descent and the conjugated gradient methods are examples of these methods. In general, these methods iterate over the following equation in order to perform the minimization:

$$R_k = R_{k-1} + l_k S_k \qquad \text{..... (2.1)}$$

where R_k is the new position at step k, R_{k-1} in the position at the previous step k-1, while l_k is the size of the step to be taken at step k and S_k is the direction of that step

Steepest-descents method. In each step of the steepest-descents method, the energy gradient g_k is calculated and a displacement S_i added to all the coordinates in a direction opposite to the gradient (i.e., in the direction of the force). In terms of the scheme outlined above this means:

$$S_k = -g_k L \qquad \ldots\ldots\ldots\ldots (2.2)$$

The step size L may be increased if the new geometry has a lower energy and decreased otherwise. This is continued till the gradients become acceptably small.

Conjugated gradients method. This is another first-order derivative method for geometry optimization. Unlike the steepest descent algorithm, which uses only the current gradient, this procedure uses also information of the gradients from the previous steps. The first step is similar to the steepest descent algorithm in the sense that a certain displacement is added to all the coordinates from the information derived from the first derivative.

i.e., $S_k = -g_k$; only for $k = 1$

For all steps $k > 1$, the direction of the step is a weighted average of the current gradient and the previous step direction, i.e.,

$$S_k = -g_k + b_k S_{k-1} \qquad \ldots\ldots\ldots\ldots (2.3)$$

where b_k is the ratio of the magnitudes of the current and the previous gradients. This information compensates for the lack of information about the curvature of the surface, i.e., the second derivatives. Many algorithms based on conjugate gradient techniques, like the Fletcher-Reeves, Polak-Ribiere and Hestenes-Stiefel methods are available. This method is generally an improved one compared to the steepest descent method.

(c) Second-order derivative methods: These use both the first and the second derivatives of the energy with respect to nuclear coordinates during the minimization process. This means that, for a molecule of N atoms, it requires not only the vector of the $3N$ first derivatives to be calculated, but also the Hessian matrix of $(3N)^2$ second derivatives. The Newton-Raphson method is a second-order derivative method, as is also the Davidon-Fletcher-Powell method.

Newton-Raphson method. The basic idea in the Newton-Raphson minimization for a one-dimensional case can be represented as follows:

$$X_{k+1} - X_k = -\frac{F'(Xk)}{F''(Xk)} \qquad \text{..... (2.4)}$$

where X_{k+1} is the next position, X_k is the current position, while $F'(X_k)$ and $F''(X_k)$ are the first and second derivatives at X_k. Near a minimum, all Hessian eigenvalues are positive and the step direction is taken as opposite to the gradient direction to reach the minima. Although this optimization procedure is very accurate and converges very well, it is computationally expensive to apply to large systems. The need to calculate the Hessian matrix as well as its inverse at every iteration makes this algorithm computationally more expensive. The use of analytical derivatives speeds up the process, though.

In principle, geometry optimization should yield stationary points, where all the gradients (first derivatives of energy with respect to the nuclear coordinates) should vanish. This applies to minima, saddle points and hilltops as well. A threshold value for the gradient norm (or the root mean square force) is fixed so that the optimization process can be terminated upon crossing the threshold value to yield an acceptable stationary point.

All these methods to locate energy minima would, in general, lead to that minimum which is *closest* to the starting input geometry. Care should be taken that the input geometry is not too far from the expected equilibrium geometry. A multi-dimensional potential energy surface may have several *local* minima on it, and which of these is the one actually sought is to be determined beforehand by an educated guess. The local minimum which is the lowest of all in energy is termed as the *global* minimum.

A true minimum (equilibrium geometry), whether local or global, on the potential energy surface is confirmed by setting up the $M X M$ energy second derivative matrix (the Hessian matrix, where M is the number of degrees of freedom for the molecule), and diagonalizing it. A true minimum will yield M positive Hessian eigenvalues. Thus, a minimum point on an energy surface is characterised by two criteria - a zero gradient form and the emergence of M positive eigenvalues. If Cartesian coordinates are used for a system with N atoms, the Hessian matrix becomes $3N X 3N$. A minimum will have 6 zero (or near zero) eigenvalues and $3N-6$ positive Hessian eigenvalues in the Cartesian framework. These 6 eigenvalues correspond to translational and rotational degrees of freedom, not being concerned with the internal geometry.

II.4 Hartree-Fock Self-Consistent Field Theory

The Hartree-Fock theory is fundamental to electronic structure theory, applicable to atoms as well as molecules. In the context of molecular orbital theory, the Hartree-Fock Self-Consistent Field (HF-SCF) theory is one way to treat interactions between electrons in a multi-electron system. Both coulombic and spin interactions are treated. Here, it is assumed that each electron feels the other electrons only as an average

charge cloud, not as individual electrons. According to this theory, the total energy of the system is given as a sum of five components:

$$E_{HF} = E_{NN} + E_T + E_V + E_{coul} + E_{exch} \qquad \text{.....(2.5)}$$

where,

E_{NN} = the nuclear-nuclear repulsion,

E_T = the kinetic energy of the electrons,

E_v = the nuclear-electron attraction energy

E_{coul} = the classical electron-electron coulomb repulsion energy,

E_{exch} = the non-classical electron-electron exchange energy.

Calculation of the E_{coul} and E_{exch} terms constitutes the main effort of Hartree-Fock calculations. The molecular electronic wave function in Hartree-Fock theory may be based on the LCAO scheme describing each MO as a linear combination of basis functions, as proposed by Roothaan.

$$\Phi_i = \sum C_{\mu i}\, \chi_\mu \qquad \text{.....(2.6)}$$

where,

Φ_i = the ith molecular orbital

$C_{\mu i}$ = coefficient of the μth basis function in the ith molecular orbital

χ_μ =the μth basis function (summed over the full set of basis functions)

The wave function is given by a single Slater determinant of N spin orbitals as written below:

$$\Psi = \begin{vmatrix} \phi_1(\mathbf{x}_1) & \phi_1(\mathbf{x}_2) & \cdots & \phi_1(\mathbf{x}_N) \\ \phi_2(\mathbf{x}_1) & \phi_2(\mathbf{x}_2) & \cdots & \phi_2(\mathbf{x}_N) \\ \vdots & \vdots & & \\ \phi_N(\mathbf{x}_1) & \phi_N(\mathbf{x}_2) & \cdots & \phi_N(\mathbf{x}_N) \end{vmatrix} \qquad \text{.....(2.7)}$$

where χ covers the variables, which include the coordinates of both space and spin. The overall electronic wavefunction Ψ of the system is constructed as an antisymmetrized product of the individual molecular orbitals (a determinant) in order to fulfill the Pauli exclusion principle since electrons are of half-integral spin (fermions).

The key elements of the Fock operator are the *coulomb* operator and the *exchange* operator. Now it so happens that these operators require a form of the wavefunction itself to be present within the operator. In other words, the process of solving the HF equations requires a good initial guess for the starting wavefunction. Thus the Hartree-Fock equations require an iterative solution, where a new wavefunction is obtained from the old by diagonalizing the energy matrix, the eigenvectors of which are the molecular orbitals, and the eigenvalues are the MO energies. The iterative procedure is halted when successive wave-functions differ from each other within a prescribed limit. The wavefunction is then said to be self-consistent with the field it generates.

In practice, it is the coefficients of the basis functions within each MO which change from iteration to iteration as per the LCAO formalism. As SCF convergence is approached and the solution is near at hand, the difference between successive values of these coefficients decreases steadily until the acceptable threshold is crossed.

Now the Hartree-Fock solution falls short of the exact solution of the Schrodinger equation for the multi-electronic system. The Hartree-Fock limit is said to be reached when the basis set becomes infinite. Yet there remains one factor further to be taken into consideration – the *instantaneous* correlation between electrons as they move around. The varied approaches which treat this aspect are referred to as "post-Hartree-Fock", treating in often ingenious ways the phenomenon of "electron correlation".

II.5 Semi-Empirical SCF-MO Theory

Semi-empirical SCF-MO methods are simplified versions of Hartree-Fock SCF theory using drastic reductions in integral evaluation along with various empirical corrections (derived from experimental data) in order to improve performance. Apart from these, the basic HF formalism holds. One of the major approximations of semi-empirical methods is the "neglect of differential overlap" when evaluating electron-repulsion integrals. By eliminating or reducing these two electron integrals, the time required for the calculation decreases by a substantial amount. To do this, semi-empirical MO methods use an approximation known as the *zero differential overlap* or ZDO approximation[40]. This means that all integrals involving the overlap between two basis functions are neglected if the two functions are not identical. The ZDO approximation may be applied at varying levels of rigor. These methods also *parametrize* some of the calculated terms in the energy matrix. Specifically, semi-empirical methods replace calculation of the two-electron integrals with data from spectroscopic experimental data or other sources, or else from standard values obtained from rigorous theory.

Parametrization means that this empirical data is used to help speedily generate the elements of the energy matrix as datasets that are stored in the computer code, and accessed at the appropriate point of the semi-empirical calculation.

1. **Older methods**: These were developed by John Pople, who also developed the Gaussian software package. In these methods, data generated by ab initio calculations are analyzed using various data fitting algorithms. The results of these data fittings are stored in the software for use during the calculation.

(a) Complete Neglect of Differential Overlap (CNDO): Here all two-centered electron repulsion integrals are assigned zero values, the most rigorous application of the ZDO approximation. This method does not consider that there are bonds of varying degrees between atoms, and calculates a topological wavefunction based on the type of atom and its location. The CNDO/2 method is a popular variant of CNDO, but now obsolete.

(b) Intermediate Neglect of Differential Overlap (INDO): In this method, some electron-electron repulsions are ignored, but not those that are centered over the same atom. The original INDO method does not have any data (parameters) for atoms with atomic number greater than 9, so it cannot be used for molecules containing those atoms, although updates have been evolved for second and third row elements as well.

2. **Newer Methods**: All these methods use a variant of ZDO known as "Neglect of Differential Diatomic Overlap" or NDDO, along with various parametrization schemes designed to reproduce experimental heats of formation or other reference quantities. These methods are attributed primarily to M. J. S. Dewar's group, who also introduced the well-used MINDO/3 method. Other methods are described briefly below:

(a) Modified Neglect of Differential Overlap (MNDO): Parameters for this methods come from a statistical analysis (a linear least squares regression fit) of enthalpies of formation and well-known molecular geometries. The MNDO method tends to overestimate repulsion between atoms.

(b) Austin Model 1 (AM1)[37]: The AM1 method addresses the overestimation of repulsive forces by recalculation of the atom-to-atom forces. It does so by scaling these forces by a factor obtained from Gaussian STO calculations. In the AM1 method, there are between 10 to 19 parameters for an individual atom.

(c) Parametrization Method 3 (PM3)[41]: This method was developed by J. P. P. Stewart in the late 1980s. The three '3' comes from the fact that this is the third NDDO method, following MNDO and AM1. The PM3 method contains approximately 18 different parameters for each atom type.

(d) Parametrization Method 6 (PM6)[42,43]: This method is the latest parametrization method available in the recently developed MOPAC 2009. It has the following features: *(i)* More accurate heat of formation and geometries; *(ii)* All main group elements and transitions metals are parametrized; *(iii)* Serious errors from PM3 and AM1 corrected; *(iv)* Crystals, surfaces and polymers with periodic boundaries can be treated; and *(v)* Very large systems may be calculated.

II.6 Higher Level Quantum Chemical Theories

The Hartree-Fock (HF) limit with an infinite (or complete) basis set represents the furthest extent of accuracy achievable without taking electron correlation into explicit consideration. This still does not meet the demand of modeling nature as it really is for a chemical system and can be lead to various errors when compared with experimental results. The following methodologies represent some efforts to treat electron correlation by the use of various models:

II.6.1 Configuration Interaction

The configuration interaction (CI) method expands the ground state wavefunction by including *excited state* Slater determinants.[44,45] This is achieved by expressing the wave-function as a linear combination Ψ_{CI} of the Slater determinants corresponding to ground and all possible excited states. The molecular orbitals derived from the HF SCF

method are used to generate the excited state determinants. The coefficients of each Slater determinant are optimized with respect to energy, the MO coefficients being kept fixed. The determinants are classified as including the HF ground state wave-function along with the singly, doubly, triply,.... *N*-tuply excited state determinants (numbered with reference to the ground state determinant). Given a trial function for Ψ_{CI}, the exact wave-function can be computed using the variational method, and expressed as

$$\Psi_{CI} = C_0\Psi_{HF} + \sum_{a}\sum_{r} C_a^r \Psi_a^r + \sum_{a>b}\sum_{r>s} C_{ab}^{rs}\Psi_{ab}^{rs} + \sum_{a>b>c}\sum_{r>s>t} C_{abc}^{rst} \Psi_{abc}^{rst} - \quad \text{.....(2.8)}$$

Truncated CI: In the full CI, the number of determinants in the wave-function exponentially increases with the size of the system and hence is computationally demanding even for small molecules. To overcome this problem, one has to *truncate* the CI expansion for the wavefunction, assuming that the determinants corresponding to the higher excited states do not contribute to the wavefunction significantly; this procedure is referred to as *truncated CI*. The procedure in which only single and double excitations along with the ground state determinant are included in the total wave-function is referred to as CI with singles and doubles substitutions (CISD). The contribution from single excitation to the correlation energy is, however, very small compared to that from double excitation for closed shell ground state systems. This is due to the fact that the ground state determinant does not mix with the singly excited Slater determinants (Brillouin's theorem) due to symmetry constraints. CID is another truncated CI procedure, where only the doubly excited state determinants are included in the CI expansion.

Size-consistency and truncated CI. A theoretical procedure is *size-consistent* if the energy of a many-particle system is proportional to the number of particles (P) in the limit P→ ∞. In other words, the energy of a dimer made of non-interacting monomers should be the sum of the individual energies of the monomers. A full CI calculation is size-consistent, while a truncated CI is not[46]. Many schemes for correcting energies obtained by CID and CISD are reported to solve the size-consistency problem of which Davidson's correction has been widely used.[47,48]

II.6.2 Quadratic Configuration Interaction

Quadratic CI is a procedure formulated by Pople's group to make the CISD method size consistent.[49] In the CISD method, the linear equations consisting of the configuration expansion coefficients are solved iteratively. In QCISD, these equations are modified by including some quadratic terms, which is equivalent to including Slater determinants corresponding to higher excited states, which makes QCISD size-extensive. It was shown that it is equivalent to the CCSD procedure, where not-so-important integrals are neglected. Similar to the CCSD(T) approach, *triples* contributions to QCISD may be included using a non-iterative scheme, the QCISD(T) method.

II.6.3 Gaussian-N Approaches

The Gaussian-*N* approaches refer to a set of methods that have been tested to reproduce 125 experimentally determined simple reaction energies (atomization and ionization energies, proton and electron affinities) to an accuracy of ±1 kcal/mol. In these procedures, a series of quantum mechanical calculations is done to obtain the final energy. The energy using the G1 procedure of Eqn. (2.9) below is the additive of the

MP4/6-311G(d,p) energy and the correction factors for addition of diffuse and higher polarization functions and correlation effects. The GI energy E^{GI} is obtained at the QCISD(T)/6-311+ G(2df,p) level. The steps involved are given below:

$$E^{GI} = E_{MP4} + \Delta E^{+} + \Delta E^{2df} + \Delta E^{QCI} + \Delta E^{HLC} + (0.8929 * ZPE) \quad \ldots\ldots\ldots (2.9)$$

where $\Delta E^{+} = E_{MP4/6\text{-}311+G(d,p)} - E_{MP4/6\text{-}311G(d,p)}$ ..... (2.10)

$\Delta E^{2df} = E_{MP4/6\text{-}311+G(2df,p)} - E_{MP4/6\text{-}311G(d,p)}$ ..... (2.11)

$\Delta E^{QCI} = E_{QCISD(T)/6\text{-}311G(d,p)} - E_{MP4/6\text{-}311G(d,p)}$ ..... (2.12)

$\Delta E^{HLC} = -0.19_{n\alpha} - 5.95_{n\beta}$ ..... (2.13)

MP4 calculations with large basis sets and QCISD(T) are run as single-point calculations on MP2/6-31G(d) optimized geometries. The zero point energy is obtained at the HF/6-31G(d) level. For the ΔE^{HLC} term (Eqn. 2.13), $n\alpha$ and $n\beta$ are the number of α and β valence electrons respectively. The G1 theory fails to quantitatively predict dissociation energies of ionic molecules, the singlet-triplet splittings of carbenes, SiH_2 and NH_3, and the atomization energies of hypervalent molecules. To rectify these, the Gaussian-2 (G2) theory was developed. It involves the same steps as the G1 procedure and includes more correction terms to the computed energy than the G1 theory, as shown below:

$$E_{G2} = E_{G1} + \Delta_1 + \Delta_2 + 1.14_{npair} \quad \ldots\ldots\ldots (2.14)$$

where $\Delta_1 = \{E_{(MP2/6\text{-}311+G(2df,p)} - E_{(MP2/6\text{-}311G(d,p)}\} - \{E_{(MP2/6\text{-}311+G(d,p)} - E_{(MP2/6\text{-}311G(d,p)}\} - \{E_{(MP2/6\text{-}311G(2df,p)} - E_{(MP2/6\text{-}311G(d,p)}\}$ ..... (2.15)

and $\Delta_2 = \{E_{(MP2/6\text{-}311G(3df,p)} - E_{(MP2/6\text{-}311G(2df,p)}\}$ ..... (2.16)

where n_{pair} is the number of paired electrons. Extensions of the Gaussian-N approaches (such as the G3 stratagem) are designed for highly sophisticated, accurate calculations, although geometry optimization at these levels is usually prohibitively expensive.

II.6.4 Other Approaches

In the *Moller-Plesset perturbation theory* (MPn, where n = 2, 3, 4 ...), the HF wave-function is taken as the zeroth-order solution and electron correlation treated as a perturbation (second, third, fourth order etc.)[50-53]. In contrast to truncated CI methods, MPn procedures are size-extensive, forming a well-defined hierarchal method providing accuracy with increasing order of n. The MPn approaches are dealt with further in Section II.9. They are compute-intensive, as indeed all post-HF methods are.

In the *coupled cluster* method (CC), developed for prediction of non-multideterminantal wavefunctions,[54-56] exponential functions generate excited states (unlike the CI where the excited states are unequivocally considered). The main differences between the CI and the CC approaches are the manner in which the excitations are treated.

In the *Multi-Configurational Self-Consistent Field* (MCSCF) method, the orbital coefficients in the molecular orbitals are varied to minimize the energy as per the variational principle. Here, the MO's are fixed and only the configurations are varied. One widely used MCSCF method is the *complete active space self-consistent field* (CASSCF) methodology.[57] The active space includes a selected window comprising a few high-lying filled molecular orbitals and a few low-lying virtual molecular orbitals; the remaining MO's are in the inactive space. The crucial task in using the CASSCF method is to select the appropriate orbitals to be included in the active space.

The MCSCF method is widely used for excited state calculations; however, it often suffers from convergence problems. CASSCF procedures overestimate diradicaloid character which can be minimized by including all valence electrons in the active space. When many configurations are chosen as a reference in which other excited state determinants are produced, it is called the Multi-Reference CI (MRCI) method.[58]

II.7 Atomic Orbital Basis Sets

Molecular orbitals may be expressed as a linear combination of the atomic orbitals (LCAO) which constitute the basis functions. The basis set is a finite collection of the basis functions. Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO) are the two types of basis functions commonly used in electronic structure calculations. The functional forms of Slater Type Orbitals (STO)[59] and Gaussian Type Orbitlas (GTO)[60] are given in eqns. (2.17) and (2.18) respectively.

$$\chi_{\zeta,n,l,}(r, \theta,\varphi) = NY_{l,m,}(\theta,\varphi)\, r^{n-l}e^{-\zeta r2} \qquad \text{..... (2.17)}$$

where N is the normalization constant and $Y_{l,m}$ are the spherical harmonic functions

$$\chi_{\zeta,lx,ly,lz\,,}(x,y,z) = N_x^{lx}y^{ly}z^{lz}e^{-\zeta r} \qquad \text{..... (2.18)}$$

where the sum of l_x,l_y,l_z determines the type of orbital (e. g. $l_x+l_y+l_z= 1$ is a p-orbital).

STO's have no radial nodes, which may be introduced by a linear combination of STO's. However, they are not amenable to implementation in MO calculations since the STO integrals are difficult to evaluate particularly when the atomic orbitals are centered on different nuclei. GTO's are however, facile to use since the products of two GTO's is itself a GTO. Both STOs and GTOs can be chosen to form a good basis set, but GTOs are preferred due to the efficiency of integral evaluation when they are used.

The important factor in classifying basis sets is the number of functions used. The smallest number of functions possible is a minimal basis set. In the minimal basis sets of STO-nG, each STO consists of '*n*' primitive Gaussian type orbitals (PGTO)[61,62] and the exponents of the GTOs are determined by fitting to the STO. Although basis sets with $n = 2$ - 6 have been known, the STO-3G is the most general minimal basis set.

A basis set which doubles the number of functions in the minimal basis set is termed a double zeta basis set. When doubling or tripling is restricted only to the valence shell, we have the *split valence* basis sets. Pople's group has designed split valence basis sets of the type *k-nlm*G, where *k* indicates how many PGTOs are used to represent the core orbitals, while *n*, *l* and *m* indicate both how many functions the valence orbitals are split into and also how many PGTOs are used for their representation. The 3-21G and 6-31G basis sets are examples of split valence basis sets.[63,64] The 6-311g set is a split valence triple-ζ basis set where core orbitals are represented by 6 PGTOs and valence orbitals are split into 3 functions (represented by 3, 1 and 1 PGTOs, respectively.)[65]

Orbital functions of higher angular momentum than the occupied atomic orbitals (known as *polarization* functions) may be added to the basis sets. The polarization functions are indicated in brackets after G in the notation for the basis sets, with a separate designation for heavy atoms and hydrogens, e.g. the 6-31G(d) basis set has *d*-functions on the heavy atoms, while the 6-31G(d,p) set has *d*-functions on the heavy atoms and *p*-functions on hydrogen. Some cases, particularly excited states and anions where the electron density is spread over the molecule, may be modeled more correctly by adding *diffuse* functions. The diffuse functions are denoted by + or ++, the first indicates one set of diffuse s- and p-functions on all heavy atoms, and the second + indicates that diffuse s-functions are also added to hydrogen. Diffuse functions can also be added to polarization functions, e.g., the 6-311++G(3df,3pd) set represents a triple split valence basis with added *s*- and *p*-functions, three *d*- and one *f*-type polarization functions on the heavy atoms, and diffuse *s*- and three *p*- and d-type polarization functions on the hydrogens.

The basis sets designed by Dunning specifically for calculations using correlation methods are referred to as *correlation-consistent* basis sets.[66] Several different sizes of correlation-consistent basis sets are available, *viz.*, the cc-PVDZ, cc-pVTZ, cc-pVQZ, cc-pV5Z and cc-pV6Z (the correlation-consistent polarized valence Double/Triple/ Quadruple/Quintuple/Sextuple Zeta basis sets). An appropriate number of polarization and diffuse functions are added to produce a balanced basis set.

In general it may be assumed that larger basis sets give rise to better results when used within any particular type of method. Some basis sets are, moreover, particularly suited to a particular method, or to a particular kind of property required to be studied.

II.8 Density Functional Theory

While Schrödinger had postulated the wave-function Ψ as the state function from which one may derive all observable properties, Hohenberg and Kohn[67] had shown that the *electron density* can also determine the energies and all other properties of the ground state by proposing the novel *density functional theory* (DFT) approaches[68-71]. The electron density is a function of three variables not depending on the system electrons, unlike the unconventional SCF procedure which uses a *3N* variable wave-function. Accurate functionals which connect the electron density to all the molecular properties are an important theoretical milestone required to be formulated clearly, as in the DFT method. The revolutionary Kohn-Sham DFT procedure[72] provided the breakthrough here. The state function here is analogous to the one-electron wave function of the HF procedure and requires an iterative method to achieve self-consistency. The coulomb energy and the exchange-correlational functional depend on the total electron density,

and the minimization of electron density with respect to energy generates a set of orthonormal Kohn-Sham (KS) molecular orbitals. Among the various proposals for the exchange and correlation functional, the introduction of non-local methods or the generalized gradient approximation (GGA) paved the way for practical application of the DFT formalism to molecular systems. Here, the independency of the exchange and the correlation functional on electron densities and gradients explains the interacting nature of the electrons.

Lee, Yang and Parr proposed a functional for the correlation part called the LYP correlation functional[73], which is a pure (non-hybrid) density functional approach. The exchange contributions are comparatively larger than the electron correlation effects implying the need of a precise definition for the exchange functional. Thus, the hybrid methods acquire an important position in the DFT method as the exact part can be derived from the HF formalism whereas the correlation part may be expressed in terms of appropriate functional(s). Becke gave a popular three-parameter functional (Becke3 or B3)[74] where some of the functionals are in accordance with hybrid methods:

$$E_{xc}^{H+H} = \frac{1}{2}E_x^{exact} + \frac{1}{2}\left(E_x^{LSDA} + E_C^{LSDA}\right) \quad \ldots\ldots\ldots (2.19)$$

The above incorporates the exchange functional derived from HF methodology. Eqn. (2.19) is the H+H functional[75], where the B3 functional contains both the exchange and correlation functional terms. The exchange functional consists of a combination of exchange terms from HF theory, the Local Spin Density Approximation (LSDA) and also a gradient correction term, whereas the correlation functional contains a LSDA form and a gradient correction term, as given below:

$$E_{XC}^{B3} = (1-a)\,E_X^{LSDA} + aE_X^{exact} + b\Delta E_X^{B88} + E_c^{LSDA} + cE_c^{GGA} \quad \text{.....(2.20)}$$

This is called the Becke3 parameter functional, since the three empirical parameters *a*, *b* and c are computed by best fit to the experimental data (indicating the semi-empirical character of DFT).

The above constitute just a few of the many models within the framework of DFT. We may note that the empirical aspect is never missing from DFT in general, where best fits are made to either experimental data or the results of highly accurate ab initio calculations. The treatment of excited states has presented a challenge to the DFT approach. So also has the treatment of ionic systems where the negative or positive charge is greater than unity. Hypervalent molecules also require special treatment. Unrestricted DFT formalisms have been devised to treat multiple spin states.

Although the success of DFT in solving a wide-range of chemical problems is truly spectacular, when a particular exchange or correlation functional fails to reproduce established experimental results, there is no systematic way to improve it, which also occurs in the semi-empirical methods. Traditional *ab initio* procedures, however, do have a mechanism to systematically improve the reliability of the computed numbers, although they are not computationally economic. The CI and MP methodologies are two main routes to include electron correlation taking the HF wave-function as a reference. However, when a single determinantal scheme is inadequate, e.g. for a biradicaloid, one has to resort to the CASSCF methology.

II.9 Perturbation Theory

Møller-Plesset (MP) perturbation theory is one of several quantum chemistry post-Hartree-Fock *ab initio* methods in the field of computational chemistry. It improves on the Hartree-Fock method by adding electron correlation effects by means of Rayleigh-Schrödinger perturbation theory (RS-PT), usually to second (MP2), third (MP3) or fourth (MP4) order. In RS-PT, we consider an unperturbed Hamiltonian operator $\hat{H}_0$, to which is added a small (often external) perturbation $\hat{V}$:

$$\hat{H} = \hat{H}_0 + \lambda\hat{V} \qquad \text{.....(2.21)}$$

where λ is an arbitrary real parameter. In MP theory, the zeroth-order wave function is an exact eigenfunction of the Fock operator, which thus serves as the unperturbed operator. The perturbation is the correlation potential involving the instantaneous effect of the electrons on each other.

In RS-PT the perturbed wave function and perturbed energy are expressed as a power series in λ:

$$\Psi = \lim_{n\to\infty} \sum_{i=0}^{n} \lambda^i \Psi^{(i)}$$

,

$$E = \lim_{n\to\infty} \sum_{i=0}^{n} \lambda^i E^{(i)}$$

Substitution of these series into the time-independent Schrödinger equation gives a new equation

$$\left(\hat{H}_0 + \lambda V\right)\left(\sum_{i=0}^{n} \lambda^i \Psi^{(i)}\right) = \left(\sum_{i=0}^{n} \lambda^i E^{(i)}\right)\left(\sum_{i=0}^{n} \lambda^i \Psi^{(i)}\right) \qquad \text{.....(2.22)}$$

Equating the factors of λ^k in this equation gives an kth-order perturbation equation (k = 0,1,2, ..., n). The MP-energy corrections are obtained from RSPT with the perturbation (*correlation potential*) as:

$$\hat{V} \equiv H - F - \langle \Phi_0 | H - F | \Phi_0 \rangle, \qquad \text{.....(2.23)}$$

where the normalized Slater determinant Φ_0 is the lowest eigenfunction of the Fock operator, so that

$$F\Phi_0 \equiv \left(\sum_{k-1}^{N} f(k) \right) \Phi_0 = 2 \left(\sum_{i-1}^{N/2} \varepsilon_i \right) \Phi_0. \qquad \text{.....(2.24)}$$

Here N is the number of electrons of the molecule under consideration, H is the usual electronic Hamiltonian, $f(1)$ is the one-electron Fock operator, and ε_i is the orbital energy belonging to the doubly-occupied spatial orbital φ_i. The shifted Fock operator

$$\hat{H}_0 \equiv F + \langle \Phi_0 | H - F | \Phi_0 \rangle$$

serves as the unperturbed (zeroth-order) operator.

The Slater determinant Φ_0 being an eigenfunction of F, it follows readily that

$$F\Phi_0 - \langle \Phi_0 | F | \Phi_0 \rangle \Phi_0 = 0 \Longrightarrow H_0 \Phi_0 = \langle \Phi_0 | H | \Phi_0 \rangle \Phi_0, \qquad \text{.....(2.25)}$$

so that the zeroth-order energy is the expectation value of H with respect to Φ_0, *i.e.*, the Hartree-Fock energy:

$$E_{\text{MP0}} \equiv E_{\text{HF}} = \langle \Phi_0 | H | \Phi_0 \rangle. \qquad \text{.....(2.26)}$$

Since the first-order MP energy correction for electron correlation

$$E_{\text{MP1}} \equiv \langle \Phi_0 | V | \Phi_0 \rangle = 0 \qquad \text{.....(2.27)}$$

is obviously zero, the lowest-order MP correlation energy appears in second order. This result is the Møller-Plesset theorem[76] which states that *the correlation potential does not contribute in first-order to the exact electronic energy.*

In order to obtain the MP2 formula for a closed-shell molecule, the second order RS-PT formula is written on the basis of doubly-excited Slater determinants. (Singly-excited Slater determinants do not contribute because of the Brillouin theorem). After application of the Slater-Condon rules for the simplification of *N*-electron matrix elements with Slater determinants in *bra* and *ket* and integrating out spin, the MP2 energy correction becomes

$$E_{\mathrm{MP2}} = \sum_{i,j,a,b} \langle \varphi_i(1)\varphi_j(2)|r_{12}^{-1}|\varphi_a(1)\varphi_b(2)\rangle \times \frac{2\langle \varphi_a(1)\varphi_b(2)|r_{12}^{-1}|\varphi_i(1)\varphi_j(2)\rangle - \langle \varphi_a(1)\varphi_b(2)|r_{12}^{-1}|\varphi_j(1)\varphi_i(2)\rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b},$$

.....(2.28)

where φ_i and φ_j are canonical occupied orbitals and φ_a and φ_b are canonical virtual orbitals. The quantities ε_i, ε_j, ε_a, and ε_b are the corresponding orbital energies. Clearly, through second-order in the correlation potential, the total electronic energy is given by the Hartree-Fock energy plus the second-order MP correction: $E \approx E_{\mathrm{HF}} + E_{\mathrm{MP2}}$. The solution of the zeroth-order MP equation (which by definition is the Hartree-Fock equation) gives the Hartree-Fock energy. The first non-vanishing perturbation correction beyond the Hartree-Fock treatment is the second-order energy. In analogous manner, MP correlation energy corrections to the third, fourth etc. order may be derived and implemented, depending on the degree of rigor called for.

II.10 Basis Set Superposition Error

One decisive point for the theoretical treatment of H-bonded systems is the choice of basis set. The strength of H-bonds computed by means of traditional *ab initio* theory requires highly flexible basis sets including diffuse functions and an explicit recovery of electron correlation effects. The fact that the basis sets employed are practically always incomplete is troublesome with respect to a balanced description of the molecular complex as well as its constituting fragments. All this arises from the usually small magnitude of the hydrogen bond energy as well as the rather flat nature of the potential well within which the H-bonded complex lies.

There are two aspects of this problem from the perspective of post-HF theory. Quite a vexing one is the *basis set superposition error* (BSSE), arising from use of finite-sized basis sets in the super-molecular approach, which is usually adopted to compute the interaction energy between **A-H** and **B** as the difference between the total energy of the **A-H...B** complex and the sum of the total energies of the fragments **A-H** and **B**. While the isolated fragments are described in their own basis sets, in the interacting complex each of them will expand its respective wavefunction, using virtual orbitals of the other. This will lead to a spurious lowering of the total energy of the complex with respect to its fragments and thus to an artificial overestimation of the complexation energy.

An approximate and popular way to estimate the BSSE *a posteriori* is the *counter-poise* correction of Boys and Bernardi[77] in which the fragment energies are computed using the basis functions of the *entire* complex but considering the atoms of each respective fragment only. The fragment energies are lowered since their basis sets after expansion become more complete. As an example, let us consider the calculation of the

dimerization energy of H_2O using the 3-21G basis set. The natural procedure would be to calculate the energy of the dimer $(H_2O)_2$ from its 3-21G equilibrium geometry using a 3-21G basis set on each of the six atoms of $(H_2O)_2$ and to calculate the energy of each H_2O monomer at its 3-21G equilibrium geometry using a 3-21G basis set on each of the three atoms of the monomer, and to take the dimerization energy as

$$\Delta\varepsilon = \varepsilon_{AB}(\{\chi_A + \chi_B\}) - \varepsilon_A(\{\chi_A\}) - \varepsilon_B(\{\chi_B\}) \quad \ldots\ldots\ldots\ldots (2.29)$$

Here **A** and **B** stand for the monomer molecules and **AB** for the dimer in general. In this case, **A** = **B** = H_2O, and **AB** = $(H_2O)_2$. Here, $\{\chi_A\}$ symbolizes the 3-21G basis set centered on the atoms of **A** alone, and similarly for $\{\chi_B\}$. Thus, $\varepsilon_A(\{\chi_A\}$ is the equilibrium-geometry energy of monomer **A** calculated with the $\{\chi_A\}$ basis set, with a similar meaning for $\varepsilon_B(\{\chi_B\})$. For $(H_2O)_2$, $\varepsilon_A(\{\chi_A\} = \varepsilon_B(\{\chi_B\})$, but this is not true for a mixed dimer such as HF-H_2O. The quantity $\varepsilon_{AB}(\{\chi_A + \chi_B\}$ is the energy of **AB** calculated with 3-21G basis orbitals on *all* atoms of **AB**.

However, this procedure involves an inconsistency. When the monomer energy $\varepsilon A(\{\chi_A\}$ is calculated, the electrons of **A** have available to themselves only the 3-21G orbitals on the three atoms of **A**. However, when $\varepsilon_{AB}(\{\chi_A + \chi_B\}$ is calculated, the electrons of each H_2O molecule within the dimer have available not only the orbitals on their own nuclei, but also the orbitals on the nuclei of the other H_2O molecule. In effect, the dimer basis set is larger than that of each monomer, and this produces an artificial lowering of the dimer energy relative to that of the separated monomers. This artificial lowering is called the basis-set superposition error (BSSE). The BSSE would vanish inthe limit of using a complete set for each monomer. The most often used procedure to correct for the BSSE energy is to calculate the dimerization energy as

$$\Delta\varepsilon = \varepsilon_{AB}(\chi_A + \chi_B\} - [\varepsilon_A(\chi_A + \chi_B\}) - \varepsilon_B(\{\chi_A + \chi_B\}] \quad \text{.....(2.30)}$$

where $\varepsilon_A(\chi_A + \chi_B)$ is calculated with a basis set that contains of 3-21G orbitals on each nucleus of the monomer A along with the appropriate 3-21G orbitals centered at the equilibrium positions of the other three nuclei of **B** in the dimer. Likewise, for the calculation of the energy of **B**. This procedure, called the counterpoise correction, has gone through a period of denunciation by many researchers, but is now recognized as probably the best way to reduce the BSSE error[78,79]

As an illustration, Hartree-Fock STO-3G, 3-21G, 6-31G*, and 6-31G** calculations give H_2O dimerization electronic energies of –5.9, -11.0, -5.6, and -5.5 kcal/mol respectively. Counterpoise-corrected dimerization energies were evaluated as -0.2, -6.2, -4.6, and –4.5 kcal/mol respectively[80]. These values may be compared with non-counterpoised SCF calculations with huge basis sets giving the dimerization energy at this Hartree-Fock limit as –3.73 ± 0.05 kcal/mol. Clearly, the 6-31G** method with counterpoise correction approaches nearest to the Hartree-Fock limit used. In this way, BSSE corrections can help arrive at improved values for dimerization energy.

II.11 Vibrational Frequency Analysis[81,82]

Frequency calculation is done by constructing the Hessian matrix and subsequently diagonalising it. This procedure is done to ascertain what kind of a stationary point the given structure is on the potential energy surface. For a molecular system with *M* degrees of freedom, a true minimum will yield *M* positive Hessian eigenvalues, while a saddle point or transition state will yield *M-1* positive eigenvalues and one single negative eigenvalue (corresponding to the reaction coordinate itself).

The vibrational frequencies of a molecular system correspond to the *3N - 6* internal (or normal) coordinates for the general case. The *Cartesian* Hessian matrix (a *3N X 3N* two-dimensional matrix for an *N*-atom system) contains the force constants as its elements. The first step after obtaining the Hessian matrix is to convert the matrix elements to mass-weighted coordinates. Diagonalizing the mass-weighted coordinates gives *3N* eigenvalues and *3N* eigenvectors. These *3N* eigenvectors correspond to the *3N-6* modes of vibration, 3 modes of rotation and 3 modes of translation. These *3N* eigenvalues correspond to the fundamental frequencies of the given modes for the translational, rotational and vibrational motions.

Then the next step involves identifying the modes corresponding to the rotational and translational motions and separating them from the vibrational modes. Only the *3N-6* vibrational modes have anything to do as such with the actual shape or structure of the molecule. For a stationary point, the frequencies corresponding to the three rotational and three translational modes should be *zero*, i.e., six of the Cartesian Hessian eigen-values should vanish for a non-linear molecule. These six modes are identified by calculating the moments of inertia and diagonalizing the resulting tensor, from which the vectors corresponding to the rotational and translational; motions can be identified. After these are separated out, the Hessian matrix in the mass-weighted coordinates is transformed to the internal coordinates. Then, the vibrational part corresponding to the internal coordinates is diagonalized, and the eigenvalues (as $4\pi^2\nu^2$) and eigenvectors (the normal modes of vibration) are obtained.

Once obtained, the Hessian eigenvalues indicate a true minimum if all eigenvalues are positive, and a transition state (saddle point) if only one eigenvalue is negative.

II.12 Molecular Electrostatic Potential

The electric potential Φ at a point P in space is defined as the reversible work per unit charge needed to move an infinitesimal test charge Q_t from infinity to P, which we write as $\Phi_P = w_{\infty \rightarrow P} / Q_t$. The SI unit of Φ is the volt (V), where 1V = 1J/C. When we do reversible work w on the test charge, we change its potential energy by w (just as reversibly raising or lowering a mass in the earth's gravitational field changes its potential energy). If we take the potential energy of Q_t as zero at infinity, we therefore have $V_p = w_{\infty \rightarrow P} = \Phi_P Q_t$. The electrical potential energy V of a charge at point P (where the electric potential is Φ_P) is $\Phi_P Q_t$. From the definition of Φ_P, it readily follows that in the space around a point charge Q, the electric potential (in SI units) is given as $\Phi_P = Q/4\pi\varepsilon_0 rd$, where d is the distance between point P and the charge. The electrical potential is a function of the location (x, y, z) of point P in space: $\Phi = \Phi(x, y, z)$.

If our system consists of a single point charge Q_A located at (x_A, y_A, z_A), then $\Phi 1 = Q_A/4\pi\varepsilon_0 r_{1A}$, where r_{1A} is the distance between point A and point 1 with coordinates (x_1, y_1, z_1). If our system consists of several point charges, then each contributes to Φ and

$$\Phi_1 = \sum_i \frac{Q_i}{4\pi\varepsilon_0 r_{1i}} \qquad \text{..... (2.31)}$$

If our system is a molecule, it is viewed as a collection of point-charge nuclei and electronic charge smeared out into a continuous distribution. Now, electrons are point charges and are not actually smeared out into a continuous distribution; however, the electronic charge-cloud picture is a reasonable approximation when considering inter-actions between two molecules that are not too close to each other. The probability of finding an electron within a tiny volume $dV = dxdydz$ is ρdV, where ρ is the electron probability density. Therefore, the amount of electronic charge in dV is $-e\rho\, dV$.

Addition of the contributions of the molecular electronic charge and of the nuclei α then gives the molecular electric potential φ at each point (x, y, z) as

$$\phi(x_1,y_1,z_1) = \sum_\alpha \frac{Z_\alpha e}{4\pi\varepsilon_0 r_{1\alpha}} - e\iiint \frac{\rho(x_2,y_2,z_2)}{4\pi\varepsilon_0 r_{12}} dx_2 dy_2 dz_2) \quad \text{.....(2.32)}$$

The MEP is a useful property to determine the electrostatic component of chemical reactivity as two molecules interact with each other. It can also be exploited to derive a more reliable *charge distribution* for the molecule of interest (see next Section).

II.13 Electrostatic Potential-Derived Charges

The charge distribution around a molecule is actually an electron cloud, but it is often useful to conceive of it as located or concentrated at various points within the molecule, notably around the atomic nuclei (atom charges) or on the bonds between atoms (bond densities). Net atomic charges are calculated from the electron density around the atom minus the positive charge of the nuclei (or sub-valence shell core as the case may be). Since it has often been difficult to estimate the charge distribution directly by the use of experimental means, theoretical analysis is of special utility here.

The MEP is a well-defined, physically significant quantity if it is calculated from an accurate wave function. The question of how to estimate *atomic charge* is, however, not easy to answer. The terms atomic charge, net atomic charge, and partial atomic charge are variously used synonymously to denote the electronic charge on a particular atom within a molecule. The *Mulliken population analysis* consists of summing the density matrix elements over an atom or between two atoms. It gives formal atomic charges and bond orders that vary erratically as the basis set is changed or improved. A better estimate of atomic point charges is obtained from *natural population analysis*.

A popular alternative way to get a reasonable estimate of the atomic charges Q_t is by fitting them to the molecular electrostatic potential (MEP) Φ. One first uses a molecular wave function to calculate values of Φ by filling a grid of many points in the region outside the molecule's van der Waals surface. One then places a charge Q_a at each nucleus α, and calculates the quantity $\Phi_i^{approx} = \Sigma_\alpha Q_\alpha / 4\pi\varepsilon_0 r_{i\alpha}$ at each grid point. One then varies the Q_a so as to minimize the sum of the squares of the deviations $[\Phi_i^{approx} - \Phi_i]$ for the grid points. There are various ways to define the grid and different schemes for finding the atomic charges Q_a, but all may be called as *MEP-derived atom charges.* Methods for deriving partial atomic charges from the quantum chemical electrostatic potential Φ are indicated here as the CHELP[83], CHELPG[84], and Merz-Kollman[85,86] approaches. These three charge-fitting routines are implementations of the algorithms derived respectively by Chirlian and Francl (CHELP), by Brenneman and Wiberg (CHELPG) and by Merz and Kollman (MK), briefly enumerated as follows:

1) CHELP uses concentric spheres starting at the van der Waals radii with interior points all excluded.

2) CHELPG uses a regular rectangular grid with points within the van der Waals radii excluded, along with criteria for the maximum distance from the atomic centers.

3) MK uses nested solvent-excluded surfaces, again excluding the volume within the innermost surface which coincides in large part with the van der Waals radii.

The suggested van der Waals radii for each element vary as per the authors' choice, and so the exclusion zone varies slightly between the methods. These three types of MEP-derived charges have their individual characteristics, each with their strengths and their weaknesses, and so have different areas of optimal applicability.

II.14 Design of Novel H-Bonded Macromolecular Duplexes

This Dissertation takes a break away from the study of the traditional nucleic acid base-pairs found in nature, and seeks to delve into *hitherto unstudied* systems. One grand aim and objective of this search is too arrive at *suitable sets of nitrogenous bases* which can pair among themselves in a way parallel to the fashion adopted by the normal DNA pairs GC and AT, such that the basic functional aspects of the DNA duplex may be mimicked by these artificially designed alternatives.

First of all, a study of *hetero-associative* (complementary) base-pairing is taken up with this same objective in mind. Chapter III focuses on hetero-associative base-paring between various substituted nitrogenous base systems (pyridines, pyrimidines and pyrazines) to look out for a well-defined set of uniquely designed bases which satisfy the prescribed criteria for constituting DNA base-pair mimics.

Following this, the search then takes us away from the world of complementary base-pairing and takes up the study of the possibilities for *self-associative* base-pairing to eventually create a dimeric duplex structure whose second strand is *identical* with the first one. Substituted five-membered rings (pyrrole and imidazole systems) along with some six-membered monocycles (substituted pyridines, pyrazines and pyrimidines) are adopted for this purpose in Chapter IV, and the best set is chosen as representative.

Careful note is to be taken to recognize the difference between DNA base analogues and the DNA base-pair mimics studied here. DNA base *analogues* are molecules which by similarity of topology and structure are able to pair in a manner close to any one of the four nucleic acid bases. By this property of imitation, they can serve as artificial substitutes for any one of the four nucleic acid bases, and thereby fit into a position in

the double-helix of DNA. DNA *base-pair mimics*, however, do not exactly imitate the base-pairing properties of any of the natural DNA bases, but go to form sets of their own which mimic the general features as a whole of the set of four DNA bases forming two exclusive pairs. The aim here is to design alternatives schemes which eventually will lay the foundation for new macromolecules which mimic the natural DNA in the salient structural and functional aspects like semi-conservative replication and bearing of genetic information.

Specificity and non-ambiguity of molecular recognition through hydrogen-bonding lies at the crux of the whole matter. Just as the biomolecules of nature suffer from no error or ambiguity while recognizing one another (at least to a large extent), so is it also the aim of this Dissertation to reproduce such features in the systems designed here.

Ultimately, the goal here is to achieve the complete design of artificial information-bearing macromolecules which are as specific and non-ambiguous as can be. This grand aim is still not yet attained in the context of this Dissertation, and awaits more time, work and facilities to make it a reality. This work here takes the first step towards the goal, *viz.*, the design of the component base-pairs of such an artificial information-bearing macromolecule, bearing in mind the assumption that the configuration and specificity of the designed base-pair will continue preserved relatively intact when incorporated into the macromolecular backbone itself. Furthermore, the work here takes a second step forward by designing viable backbones for the chosen pair systems, where a novel polynucleotide backbone and a novel polyamide backbone are taken for consideration. This work leads to design of the complete monomeric repeat units for the putative double-stranded and H-bonded macromolecular duplex itself.

II.15 Philosophy of Approach Utilized Here

This Dissertation makes an attempt to design novel hydrogen-bonded base pairs built up from nitrogenous bases other than the bases of natural DNA by using *theoretical* models. The novel base pairs should prove themselves to be *isomorphic* on the basis of their molecular descriptors, *viz.*, pairing configuration and hydrogen bond geometries. The isomorphic set of base pairs is then subjected to further study, where a suitable backbone is designed so as to construct coherent monomer repeat units for the macromolecular H-bonded duplex (the information-bearing macromolecule).

Drawing from quantum mechanics the concept of the wave-function to describe all properties of a system, the application of this axiom requires a successive series of *approximations*, since the Schrödinger equation for a molecular system other than a one-electron one cannot be exactly solved. The molecular orbital method itself is an approximation, where the concept of two electron molecular orbitals spanning the molecular framework is an approximation of the true wave-function. The Pauli exclusion principle leads to the approximate wave-function taking the form of a Slater determinantal combination of two-electron molecular orbitals, which is itself approximate, since only one electronic configuration is envisioned. The Roothan equations are also a further approximation, since the two electron MO's are expressed as linear combinations of atomic orbitals, themselves also expressed in an approximated form. Finally, the very concept of the potential energy surface itself of a molecular system (having equilibrium geometries and transition states) is tenable only within the framework of the Born-Oppenheimer approximation separating nuclear and electronic motions, without which analysis of any stationary point becomes meaningless.

This Dissertation consistently employs only the *gas phase* for studying the structure and geometry of the model systems, employing standard DFT and ab initio methods as available in the Gaussian 2003 suite of programs. Gas phase calculation is sufficient (at the moment!) for our chosen systems, for the simple reason that this Dissertation deals with the initial work on novel H-bonded base pairs which may be used as repeat units for novel information-bearing macromolecules. And no experimental datas are available for the moment. In the DFT model, the B3LYP level of calculation with a 6-31G** basis set has been utilized, which consisits of polarisation functions for the heavy elements and for the hydrogen atom respectively. A density functional theory at B3LYP (Becke 3 LeeYang Parr) level is a good method , for the following reasons.

The advantage) is that DFT calculations contain almost all of the correlation portion missing in HF and semi-empirical method; therefore, it requires of polarization functions for its correct calculation. Calculations without polarization functions yield qualitatively wrong results . Basically DFT scales as N**3 as opposed to HF who scales as N**4. For large systems both exponents decrease. For small systems, it is commom that the N**3 is not that advantage because the numerical integrations involved in DFT codes. When you have ten or more heavy atoms DFT is the only choice for good energetics.

At the MP2 level, the 6-31G* and 6-311++G** basis sets have been employed, which include the diffuse function as well as the polarization basis function for the heavy elements and for the hydrogen atom. The diffuse functions '++' take care of the correlation between the noncovalent bonded atoms of the H-bonded systems, which makes the calculation more efficient. This is done in order to treat H-bonding in a sounder fashion.

References

1. Pimental, G. C; McClellan, A. L. *The Hydrogen Bond*, Freeman, San Francisco, 1960

2. Vinogradov, S. N; Linell, R. H. In *Hydrogen Bonding*, Van Nostrand Reinhold, 1971

3. Schuster, P; Zundel, G; Sandorfy, C. (eds). *The Hydrogen Bond*, Vol 120, Springer-Verlag, Berlin, 1984.

4. Smith, D. A. (ed.). *Modelling the Hydrogen Bond*, Vol. 569, American Chemical Society, Washington DC, 1984

5. Scheiner, S. *Hydrogen Bonding: A Theoretical Perspective*, New York, 1997.

6. Scheiner, S. *Hydrogen Bonding: A Theoretical Perspective*, New York, Oxford University Press, 1997.

7. Sponer, J; Lankas, F. (eds.) *Computational Studies of DNA and RNA*, Springer, Dordrecht, 2006.

8. Leszczynski, J. (ed.) *Computational Molecular Biology*, Elsevier Science, 1999.

9. Sponer, J; Florian, J; Hobza, P; Leszezynski, J. *J Biomol Struct Dyn*, 1996, **13**, 827.

10. Sponer, J; Leszczynski, J; Hobza, P. *J Biomol Struct Dyn*, 1996, **14**, 117.

11. Giese, T. J; Shere, E. C; Cramer, C. J; York, D. M. *J Chem Theory Comput*, 2005, **1**, 1275.

12. Sponer, J; Jurecka, P; Hobza, P. *J Am Chem Soc*, 2004, **126**, 10142.

13. Danilov, V. I; Anisimov, V. M. *J Biomol Struct Dyn*, 2005, **22**, 471.

14. Machado, M; Ordejon, P; Artacho, E; Sanchez-Portal, D. J. M. Soler, Online at

oai: arXiv. Org: physics/9908022, 1999

15. Elstner, M; Hobza, P; Fravenheim, T; Suhai, S; Kaxiras, E. *J Chem Phys*, 2001, **114**, 5149.
16. Monajjemi, M; Chahkandi, B; Zare K; Amiri, A. *Biochem* (Moscow), 2005, **70**, 447.
17. Gould, I. R; Kollman, P. A. *J Am Chem Soc*, 1994, **116**, 2493.
18. Kudritskaya, Z. G; Danilov, V. J. *Theor Biol*, 1976, **59**, 199.
19. Poltev, V. I; Shulyupina, N. V. *J Biol Struc Dyn*, 1986, **3**, 739.
20. Stamatiadou, M. N; Swissler, T. J; Rabinowitz, J. R; Rein, R. *Biopolymers*, 1972, **11**, 1217.
21. Venkateswarlu, D; Lyngdoh, R. H. D. *J Chem Soc Perkin Trans 2*, 1995, 839
22. Venkateswarlu, D; Lyngdoh, R. H. D. Bansal, M. *J Chem Soc Perkin Trans 2*, 1997, 621
23. Clementi, E, Mehl. J, von Niessen, W. *J Chem Phys*, 1971, **54**, 508.
24. Gu, j; Xie, Y, Schaefer, H. F. *J Phys Chem B*, 2005, **109**, 13067.
25. Richardson, N. A; Wang, S; Schaefer, H. F. *J Am Chem*, 2004, **126**, 4404
26. Das, G; Lyngdoh, R. H. D. *J Mol Struct (THEOCHEM)*, 2008, **851**, 319
27. Del Bene, J. *J Mol Struct (THEOCHEM)*, 1985, **124**, 201.
28. Hobza, P; Sandorfy, C. *J Am Chem Soc*, 1987, **109**, 1302.
29. Hrouda, S; Florian, J; Hobza, P. *J Phys Chem*, 1993, **97**, 1542.
30. Gould, I. R; Kollmann, P. A. *J Am Chem Soc*, 1994, **116**, 2493.
31. Fletcher, R; Powell, M. J. D. *Computer J*, 1963, **6**, 163.
32. Davidon, W. C. *Computer J*, 1968, **10**, 406.

33. Poppinger, D. *Chem Phys Lett*, 1975, **34**, 332.

34. Fletcher, R, *Computer J*, 1970, **13**, 317.

35. Peng, C; Ayala, P, Y; Schlegel, H, B; Frisch, M, J. *J Comp Chem*, 1996. **17**, 4935.

36. Reed, A. E; Weinhold, F. *J Chem Phys*, 1983, **78**, 4066.

37. Clark, T. *A Handbook of Computational Chemistry*, Erlangen, Germany, 1985, pp. 140-151.

38. Nocedal, J; Wright, S. J. *Numerical Optimization*, Springer, 1999.

39. Walsh, G. N. *Methods of Optimisation*, Wiley, 1975.

40. Levine, I. N. *Quantum Chemistry*, 4th Edition, Prentice Hall of India, New Delhi,-2002, pp. 572-

41. Stewart, J. J. P. *J Comput Chem*, 1989, **10**, 209.

42. Stewart, J. J. P. *J Mol Modelling*, 2004, **10**, 6.

43. Stewart, J. J. P. *J Phys Ref Data*, 2004, **33**, 713.

44. Shavitt, I. *J Chem Phys*, 1986, **49**, 4048.

45. Handy, N. C; Pople, J. A; Shavitt, I. *J Phys Chem* 1996,100, 6007.

46. Ahlrichs, R; Lischka, H; Staemmler, V; Kutzelnigg, W. *J Chem Phys*, 1975, **62**, 1225.

47. Langhoff, S. R; Davidson, E. R. *Int J Quantum Chem*, 1974, **8**, 61.

48. Ahlrichs, R; Scharf, P; Ehrhardt, C. *J Chem Phys*, 1985, **82**, 890.

49. Scuseria, G. E; Schaefer, H. F. *J Phys Chem*, 1989, **90**, 3700.

50. Raghavachari, K; Pople, J. A; Replogfle, E. S; Head-Gordon, M. *J Phys Chem*, 1990, **94**, 5579.

51. Raghavachari, K; Pople, J. A. *Int Quantum Chem*, 1978, **14**, 91.

52. Kucharski, S; Bartlett, R. J. *Adv Quantum Chem*, 1986, **18**, 281.

53. Bartlett, R. J; Sekino, H; Purvis, G. D. *Chem Phys Lett*, 1983, **98**, 66.

54. Taylor, P. R; Bacskay, G. B; Hush, N. S; Hurley, A. C. *Chem Phys Lett*, 1976, **41**, 444.

55. Scuseria, G. E ; Janssen, C. L; Schaefer, H. F. *J Chem Phys*, 1988, **89**, 7382.

56. Bartlett, R. J. *J Chem Phys*, 1989, **93**, 1697.

57. Roos, B. *Adv Chem Phys*, 1987, **69**, 399.

58. Lawley, K. P. *Adv Chem Phys*, 1987, **67**, 249.

59. Cizek, J. *Adv Chem Phys*, 1969, **14**, 35.

60. Bartlett, R. J. *J Chem Phys*, 1989, **93**, 1697.

61. Hurley, A. C. *Chem Phys Lett*, **1976**, 41, 444.

62. Chiles, R. A; Dykstra, C. E. *J Chem Phys*, **1981**, 74, 4544.

63. Scuseria, G. E; Janssen, C. L; Schaefer, H. F. *J Chem Phys*, 1988, **89**, 7382.

64. Lee, Y. S; Kucharski, S. A; Bartlett, R. J. *J Chem Phys*, 1984, **81**, 5906.

65. Raghavachari, K. *J Chem Phys*, 1985, **82**, 4607.

66. Stanlon, J. F. *Chem Phys Lett*, 1997, **281**, 131.

67. Hohenberg, P; Kohn, W. *Phys Rev*, 1964, **136**, B864.

68. Dreizler, R. M; Gross, E. K. V. *Density Functional Theory*, Springer, Berlin

69. Koch, W; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, Wiley and Weinheim University Press, New York, 1989.

70. Johnson, B. G; Gill, P. M. W; Pople, J. A. *J Chem Phys*, 1993, **98**, 5612.

71. Kohn, W; Becke, A. D; Parr, R. G. *J Phys Chem*, 1996, **100**, 12974.

72. Kohn, W; Sham, L. *J Phys Rev A*, 1965, **140**, 1133.

73. Lee, C .Yang, W; Parr, R. G. *Phys Rev B*, 1988, **37**, 785.

74. Becke, A. D. *J Chem Phys*, **1997, 107**, 8554.

75 . Becke, A. D. *J Chem Phys*, 1993, **98**, 1372.

76. Møller C, Plesset M. S. *Phys Rev*, 1934, **46**, 618.

77. Boys, S. F; Bernadi, F. *Mol Phys*, 1985, **19**, 553.

78. Barlett, R. J. *Ann Rev Phys Chem*, 1981, **32**, 359.

79. Chalasinski, G; Szczesniak, M. M. *Chem Rev*, 1994, **94**, 1723.

80. Van Duijneveldt, F. B. et al, *Chem Rev*, 1994, **94**, 1873.

81.Hehre, W. J; Radom, R; Schleyer, P. v. R, Pople, J. A. *Ab Initio Molecular Orbital Theory*, Wiley, New York, 1986, 253-255.

82. Curtiss, L. A; Raghavachari, K; Redfern, P. C; Rassolov, V, Pople, L. A. 1998, *J. Chem Phys*, 1998, **109**, 7764

83. Chirlian, L. E; Francl, M. M. *J Comp Chem*, 1987, **8**, 894

84. Breneman, C. M; Wiberg, K. B. *J Comp Chem*, 1990, **11**, 361.

85. Besler, B. H; Merz, K. M; Kollman, P. A. *J Comp Chem*, 1990, 11, 431.

86. Singh, U. C; Kollman, P. A. *J Comp Chem*, 1984, **5**, 129

Q: What's the most important lesson in chemistry?

A: Never lick the spoon.

CHAPTER THREE

HETERO-ASSOCIATIVE BASE PAIRS AS REPEAT UNITS FOR NOVEL INFORMATION-BEARING MACROMOLECULAR DUPLEXES

III.1 Introduction

The discovery of DNA gives rise to the possibility of designing synthetic alternatives with the potential to bear information at the molecular level. Such systems may one day be of relevance for the developing world of nanotechnology with regard to information technology at the molecular level, where the information-bearing capacity arises out of how specific units are arranged sequentially.

DNA has two such units (the Watson-Crick base pairs) which serve as a model for creating other alternatives. DNA base pairs have received much attention from quantum chemists[1,2]. Some pertinent work includes studies on non-planar pairs[3], stacking interactions[4], use of semi-empirical models[5], accurate pairing energies[6], use of post Hartree-Fock methods[7], use of density functional methods[8,9], focus on H-bond strengths[10,11], and modified DNA mispairs as a source of mutations[12,13].

The theoretical study of novel base pairs as potential repeat units for synthetic information-bearing macromolecules is, however, still in need of attention. In this study, we use *ab initio* density functional theory (DFT) to design wholly synthetic hydrogen-bonded nitrogenous base-pair systems which serve as the repeat units for an artificial macromolecular H-bonded duplex which qualitatively mimics the information-bearing characteristics of natural DNA, given as follows:

1. The duplex is a polymer incorporating repetitive sequences comprised only of *M* chosen base pairs (*M* equals two for DNA)
2. These *M* base pairs arise out of hydrogen-bonded association among a limited set of *N* bases (*N* equals 4 in DNA) which may pair in various ways, of which only *M* pairs are chosen for the duplex.
3. Choice of the *M* select base pairs rests upon the *pairing configuration*, which should be basically the *same* for all *M* chosen pairs. Within this configurational constraint, each base may pair only with one other base and no other.
4. To ensure configurational and conformational stability, at least two hydrogen bonds between the component bases are required.

The *N* bases involved in the *M* chosen base pairs are defined here as a *DNA base mimic set*, defined as a well-defined set of bases which, although differing much in actual structure from the DNA base set, bear H-bonding properties and information-bearing potential in a manner similar to the DNA base set. These are *not* the same as DNA base analogues, which are non-DNA bases that can fit into the DNA duplex because of their topological and base pairing properties.

Buam and Lyngdoh[14,15] used the PM3 SCF-MO method to study the H-bonded pairing properties of numerous sets of nitrogenous bases with a view to determining their suitability as repeat units for information-bearing

macromolecules. The first study[14] dealt with self-associative pairs and the other[15] dealt with hetero-associative pairs.

III.1.1 Criteria for suitable DNA base mimic sets

Features for selection of any set of bases as suitable for incorporation into the macromolecular duplex include *(a)* facility of H-bonded base pairing leading to *stable* structures, and *(b)* similarity or *isomorphism* of pairing configuration.

Buam and Lyngdoh[14,15] delineated the following features concerning facility of hydrogen-bonded pairing between nitrogenous bases of the kind studied here :

1. The *number* of hydrogen bonds present in the base pair.
2. The *type* of H-bond occurring, where, for instance, H-bonds of the C=O----H–N type are stronger than those of the F----H–N type[10,11]. High electronegativity of the atoms **X** and **Y** involved in the H-bond promotes strength of the H-bond.
3. The *linearity* of the H-bonds present, where non-linearity detracts from base-pairing facility.
4. The *co-planarity* of the rings of the two component bases, since this augments stacking interactions in the duplex.

The pairing configuration of the base pairs is also important, and has been monitored through the use of appropriate interatomic distances, angles and dihedral angles[12-15]. For any select base pairs to fulfill the suitability criteria for

being chosen as H-bonded repeat units of the information-bearing macromolecule, the pairing configuration should be closely similar for all pairs. It is assumed here that the pairing configuration of the solitary base pairs in gas phase would also be largely similar to that adopted when present within the macromolecular duplex itself. This assumption appears to be true for the DNA base pairs, whose gas phase pairing configuration as studied theoretically is quite close to that obtained from the crystal structure of the DNA double helix itself[8,16-21].

III.1.2 Candidate DNA base mimic sets

We choose certain sets of nitrogenous bases (substituted pyrimidines, pyridines and pyrazines) as candidate units for the formation of novel, synthetic DNA-type base pairs which may be incorporated into a macromolecular duplex by covalent attachment to a polymeric backbone. The pairs are studied in gas phase without incorporation into the duplex since it is known that the DNA base pairs have almost the same configuration in gas phase and within the actual DNA double-helix[8,16-21]. The structure of the backbone for the DNA mimic sets studied here is discussed in Chapter V, where sugar phosphate and polyamide backbones are taken for consideration. The bases would then be present as their nucleosides or as amide derivatives. Here, the point of attachment between the base moiety and the backbone is modeled by attachment of a methyl group to one of the ring nitrogens of the base (numbered as N1). Five candidates for such DNA base mimic sets are chosen for study here, each expressed as the 1-methylbase, and given as follows:

1. Four substituted pyrimidines: **A1** (1-methylpyrimid-2,4-dione), **A2** (1-methyl-4,6-diaminopyrimid-2-one), **A3** (1-methyl-4-aminopyrimid-2,6-dione) and **A4** (1-methyl-4-aminopyrimid-2-one).

2. Three substituted pyrazines: **B1** (1-methyl-5-fluoropyraz-2-one), **B2** (1-methyl-3-aminopyrazine, and **B3** (1-methyl-5-aminopyraz-2-one).

3. Two substituted pyridines and two substituted pyrimidines: **C1** (1-methyl-4-amino-5-fluoropyrid-2-one), **C2** (1-methyl-4-aldehydo-5-fluoropyrid-2-one, **A5** (1-methyl-4-aldehydo-pyrimid-2,6-dione) and **A6** (1-methyl-4-aminopyrimid-2,6-dione).

4. Two substituted pyridines and two substituted pyrazines: **C2** (1-methyl-4-aldehydo-5-fluoropyrid-2-one), **C1** (1-methyl-4-amino-5-fluoropyrid-2-one), **B2** (1-methyl-3-aminopyrazine) and **B4** (1-methyl-2-oxo-5-amino-pyrazine-4-oxide).

5. Two substituted pyrazines and two substituted pyrimidines: **B2** (1-methyl-3-amino-pyrazine), **B1** (1-methyl-5-fluoropyraz-2-one), **A6** (1-methyl-4-fluoro-pyrimid-2-one) and **A3** (1-methyl-4-aminopyrimid-2,6-dione).

From each of the above sets, a large number of pairs in principle may arise due to the different pairing combinations possible. The only ones selected here for study are those which fall into a well-defined pairing configuration which should be approximately *identical* for two different base pairs arising from the set.

Fig. III.1(a) portrays the four pyrimidines **A1**, **A2**, **A3** and **A4** comprising the **Set I** along with the two pyrimidine-pyrimidine pairs **A1:A2** and **A3:A4** formed

through H-bonding. Fig. III.1 (b) depicts the three pyrazines **B1**, **B2**, and **B3** comprising the **Set II** along with the two pyrazine-pyrazine pairs **B1:B2** and **B3:B3** they form through H-bonding. **Set II** is exceptional in that it includes the scope for self-association in the **B3:B3** pair. Fig. III.2(a) gives the two pyridines **C1** and **C2** and the two pyrimidines **A5** and **A3** comprising the **Set III** along with the two pyridine-pyrimidine pairs **C1:A5** and **C2:A3** formed through H-bonding. Fig. III.2(b) shows the two pyridines **C2** and **C1** and the two pyrazines **B2** and **B4** comprising the **Set IV** along with the two pyridine-pyrimidine pairs **C2:B2** and **C1:B4** they form through H-bonding. Finally, Fig. III.3 portrays the two pyrazines **B2** and **B1** and the two pyrimidines **A6** and **A3** comprising the **Set V** along with the two pyrazine-pyrimidine pairs they form through H-bonding, *viz.*, **B2:A6** and **B1:A3**. These diagrams are only schematic and do not really represent the optimized three-dimensional structures obtained through quantum chemical calculations.

III.2 Methodology

The B3LYP density functional theory (DFT) model[22-24] was used with the 6-31G* basis set for all molecular species (solitary bases and base pairs) with full optimization of geometry using initial structures from PM3 SCF-MO calculations as input. The pairing energy E_p was calculated from the difference in total energies between the pair and the component bases. The basis set

superimposition error (BSSE) energy was estimated by using the Boys-Bernardi counterpoise method[25] involving re-calculation of the two monomers within the dimer using the basis set for the dimer. This method works well for hydrogen-bonded base pairs since the monomers do not change very much in geometry when incorporated into the dimer. The BSSE corrected pairing energy is denoted here as E_p(cr). All calculations were carried out using the GAUSSIAN 2003 program[26].

This study adopts the following descriptors of pairing configuration, which correspond to those commonly used for DNA base pairs :

1. The distance R_{cc} between the carbons of the two methyl groups attached to the two component bases of the pair.
2. The angles θ_1 and θ_2 spanning the C1-N1-N2 and C2-N2-N1 atoms in the two bases of the pair, where C1 and C2 are the two methyl carbons, while N1 and N2 are the nitrogen termini of the would-be glycoside C-N bonds.
3. The dihedral φ which spans the C1-N1-N2-C2 atoms, serving as an indicator of co-planarity (or the lack of it) between the two bases within the pair.

A hydrogen bond within a base pair is represented as **X....H-Y** or **X-H....Y**, where **X** and **Y** are electronegative atoms belonging to two different component bases. Geometry around the hydrogen bonds was studied using the following determinants :

1. The length R_{hb} of the actual hydrogen bond **X....H** or **H....Y** for the two cases.
2. The length R_{xy} between the two electronegative atoms **X** and **Y**.
3. The hydrogen bond angle θ_{hb} of the moiety **X....H-Y** or **X-H...Y**.

The convention here is that the atom **X** belongs to the base on the left, while atom **Y** belongs to the base on the right, as appearing in Figs. III.1, 2 and 3, each atom being numbered as per the conventions of heterocyclic chemistry. The hydrogen atom in the middle is numbered as per the atom **X** or **Y** to which it is bonded covalently.

Note is also taken of the change in charge distribution occurring upon formation of the base pair in each case. This is done by correlating the direction of net charge transfer with the number, type and directionality of the H-bonds involved. The Mulliken monopole (point charge) model is used here to calculate the charge distribution. The dipole moment of the base pair is also compared with those of the individual bases.

For a given H-bond **X....H-Y**, it is expected that charge transfer occurs from the base with the atom **X** to the base with the atoms **H** and **Y**. Likewise, for an H-bond **X-H....Y** in the reverse direction, the direction of charge transfer would be reversed. The net direction of charge transfer from one base to the other is expected to depend on the *number* of H-bonds pointing one way compared with the number pointing the other way. The *type* of H-bond involved is also a factor, since H-bonds involving different elements as **X** and **Y** would exhibit differing

degrees of charge transfer associated with each. The *geometry* of the H-bonds and of the pair as a whole would also affect the charge transfer.

III.3 Results and Discussion

Table III.1 presents the B3LYP/6-31G* optimized results for the pairing energy E_p, and the BSSE-compensated pairing energy E_p(cr) for each pair of each of the Sets **I** to **V**. Table III.1 also presents the configurational data for the base pairs of the five Sets, using the configuration markers R_{cc}, θ_1, θ_2 and φ described earlier. Table III.2 gives data about the geometry of the hydrogen bonds present in each base pair, while Table III.3 presents data concerning charge transfer in each base pair.

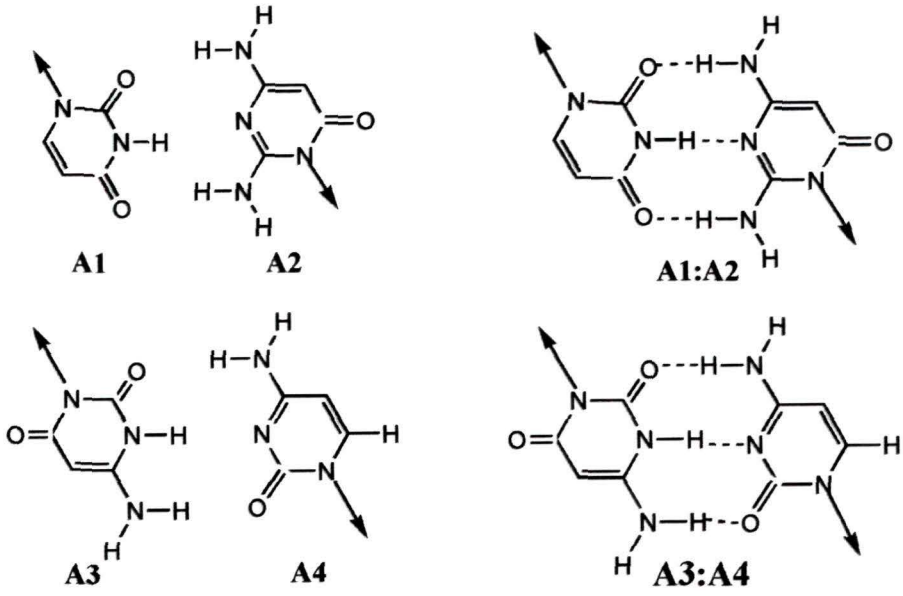
The BSSE energy is only a small fraction of the total energies of the bases and base pairs, ranging from 0.0025951 to 0.0074581 hartree (0.88 to 4.40 kcal/mol). However, it does appreciably alter the values of the pairing energies E_p.

III.3.1 Pyrimidine-pyrimidine pairs (Set I)

The **Set I** candidate mimic set comprises the four pyrimidines **A1**, **A2**, **A3**, and **A4**, which can pair to form the two H-bonded pyrimidine-pyrimidine pairs **A1:A2** and **A3:A4** as shown in Fig. III.1(a). The pairing energies and configurational data are given in Table III.1, while data concerning the H-bond geometries are given in Table III.2.

Schematic Representation

(a) **Set I** (substituted pyrimidine-pyrimidine pairs)



(b) **Set II** (substituted pyrazine-pyrazine pairs)

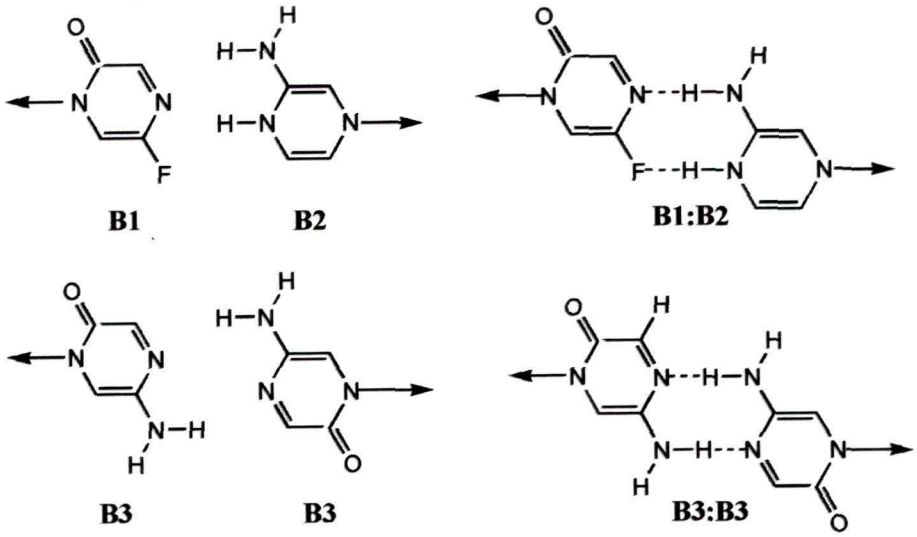


Fig. III.1 : Components bases and resulting base pairs built out of
(a) the pyrimidine-pyrimidine base-pairing **Set I**, and
(b) the pyrazine-pyrazine base-pairing **Set II**

Table III.1 Uncorrected pairing energy E_p and BSSE corrected pairing energy E_p(cr) along with the configurational data for H-bonded base pairs of **Sets I** and **II** as obtained from B3LYP/6-31G* optimized geometries and energies*

Base Pair	E_p	E_p(cr)	R_{cc}	θ_1	θ_2	ϕ
Set I *Pyrimidine-pyrimidine pairs*						
A1:A2	-20.08	-15.68	9.736	136.8	137.2	173.5
A3:A4	-29.49	-25.10	9.697	137.8	135.7	-171.3
Set II *Pyrazine-pyrazine pairs*						
B1:B2	-3.76	-1.88	11.575	156.7	142.8	-137.4
B3:B3	-10.66	-8.15	11.776	166.3	164.0	-179.6

*Pairing energies in kcal/mol; bond lengths in angstrom; bond angles and dihedrals in degrees

Table III.2 H-bond geometry data* for the base pairs of **Sets I** and **II** as obtained from B3LYP/6-31G* optimized geometries

Base Pair	H-bond	R_{hb}	R_{ab}	θ_{hb}	E_p(cr)
Set I *Pyrimidine-pyrimidine pairs*					
A1:A2	O2-H4	1.878	2.899	176.3	-15.68
	H3-N3	1.910	2.955	178.1	
	O4-H2	1.967	2.984	178.8	
A3:A4	O2-H4	1.814	2.845	177.9	-25.10
	H3-N3	1.914	2.945	176.5	
	H4-O2	1.909	2.933	177.5	
Set II *Pyrazine-pyrazine pairs*					
B1:B2	N4-H5	2.720	3.496	134.8	-1.88
	F5-H4	1.959	2.929	160.3	
B3:B3	N4-H5	2.120	3.138	173.8	-8.15
	H5-N4	2.127	3.145	173.5	

*Bond lengths in angstrom; bond angles in degrees

The two pyrimidine-pyrimidine pairs **A1:A2** and **A3:A4** are marked by the largest pairing energies among all the systems studied here. The BSSE-corrected pairing energy E_p(cr) values are respectively -15.68 and -25.10 kcal/mol (Table III.1), attributed to *(a)* the three H-bonds present in each pair (the others having only two), and *(b)* the linearity of the H-bonds (Table III.2), where the various H-bond angles θ_{hb} closely approach 180°. The relatively short lengths R_{hb} (1.878 to 1.967 Å) and R_{xy} (2.845 to 2.984 Å) for the H-bonds also testify further to their stability. It may be also noted that H-bonds containing oxygen (of the C=O...H-N type) are two in number (out of the three present) in each of the pairs **A1:A2** and **A2:A3**. Such bonds had been deemed as relatively stronger than the others occurring in this context by earlier PM3 studies [14, 15], and contribute to the overall stability of H-bonding in these two pairs.

The data of Table III.1 predicts that the two pairs **A1:A2** and **A2:A3** are remarkably close to each other in pairing configuration. The interatomic distances R_{cc} (9.736 and 9.697 Å for **A1:A2** and **A3:A4** respectively) differ by only 0.089 Å. The θ_1 and θ_2 values range from 135.7 to 137.8° (a fairly narrow range) and show that these pairs retain their basic pairing configuration even upon reversal. Although this feature of reversibility may not be an absolute criterion for suitability of a DNA mimic set here, it is certainly true of the two DNA base pairs A:T and G:C. The dihedral φ has values of 173.5 and -171.3°, both close to the ideal value of 180° pointing to base co-planarity. This co-planarity of the

bases within these two pairs predicts that pi-pi stacking interactions may be expected to be on the high side, further contributing to stability of the macromolecular duplex.

The planarity of the **A1:A2** and **A3:A4** pairs concerns primarily the heavy C, N and O atoms only. Apart from the methyl hydrogens, the amino hydrogens in the solitary bases **A2**, **A3** and **A4** are not coplanar with the rings. This pyramidalisation of the amino hydrogens is not well-attested to in the earlier PM3 calculations[14,15] on these systems. Amino group pyramidalisation is reduced, though, when the bases **A2**, **A3** and **A4** are present in the pairs **A1:A2** and **A3:A4**, being directly involved in the H-bonding.

III.3.2 Pyrazine-pyrazine pairs (Set II)

The DNA mimic **Set II** comprises the three pyrazines **B1**, **B2** and **B3**, shown in Fig. III.1(b) along with the two pairs pyrazine-pyrazine arising out of these, *viz.*, **B1:B2** and **B3:B3**. The pairing energies E_p (Table III.1) are respectively -3.76 and -10.66 kcal/mol, the corresponding E_p(cr) values being -1.88 and -8.15 kcal/mol. The small pairing energy for **B1:B2** may be attributed to *(a)* absence of the C=O...H-N type of H-bond, where there are instead the N...H-N and F...H-N types of bonds, both (especially the latter) being deemed as of the weaker H-bond category by earlier PM3 studies[14,15] *(b)* the non-linearity of the N4-H5 H-bond (θ_{hb} being 134.8°), and *(c)* the correspondingly long H-bond lengths R_{hb} and R_{xy} of 2.720 and 3.496 Å respectively for this H-bond (Table III.2). The pair **B3:B3**, in

contrast, has a larger pairing energy (-8.15 kcal/mol) because it does not contain the weak F...H-N type of H-bond, and no markedly long or non-linear H-bonds. However, its pairing energy still does not match that of the pyridimine-pyrimidine pairs **A1:A2** and **A3:A4** since it contains no stronger H-bonds of the C=O...H-N type. Furthermore, these pairs are both linked by only two H-bonds each.

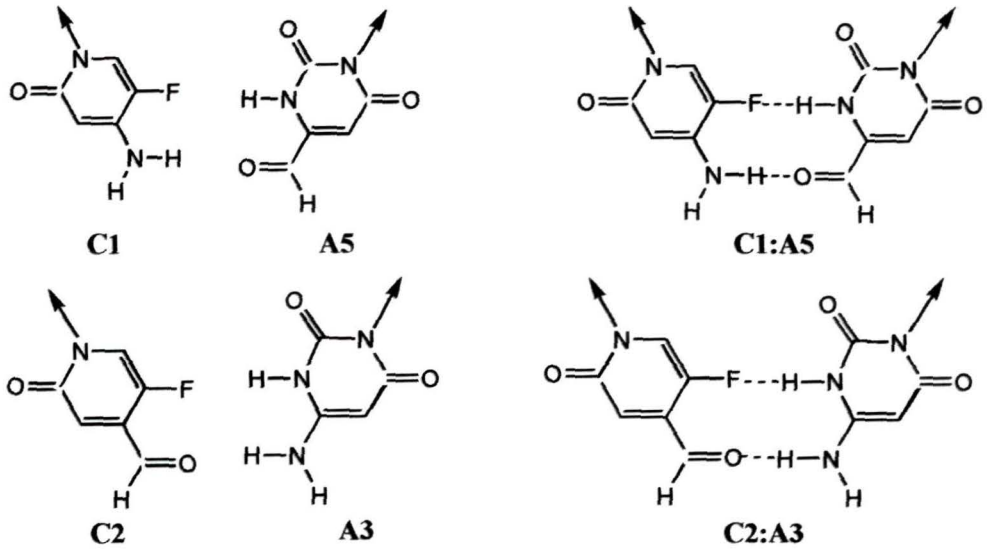
Of these two pairs, the data of Table III.1 predicts a non-planar configuration for **B1:B2** (φ = -137.4°), while **B3:B3** is still fairly planar (φ = -179.6°). Non-planarity of **B1:B2** may be linked to the non-planarity of the component pyrazine base **B2**, and may be expected to reduce pi-pi stacking contributions when present in the duplex. The dihedral φ also predicts that the DNA mimic **Set II** would yield pairs whose pairing configurations are not very similar. Although the R_{cc} values (11.575 and 11.776 Å) for **B1:B2** and **B3:B3** correspond fairly closely, the φ values are dissimilar (differing by 42.2°). Furthermore, the θ_1 and θ_2 values (ranging from 142.8 to 166.3°) show the base pairs are not reversible. Since the two pairs do not correspond closely in configuration, the mimic **Set II** does not qualify as a satisfactory candidate for a DNA base mimic set.

III.3.3 Pyridine-pyrimidine pairing (Set III)

The two substituted pyridines **C1** and **C2** and the two substituted pyrimidines **A5** and **A3** constitute the candidate mimic **Set III**, yielding the two pyridine-pyrimidine pairs **C1:A5** and **C2:A3** as shown in Fig. 2(a). Table III.1 predicts the values of their BSSE-corrected pairing energies E_p(cr) as -3.13 and -7.52

Schematic Representation

(c) **Set III** (substituted pyridine-pyrimidine pairs)



(d) **Set IV** (substituted pyridine-pyrazine pairs)

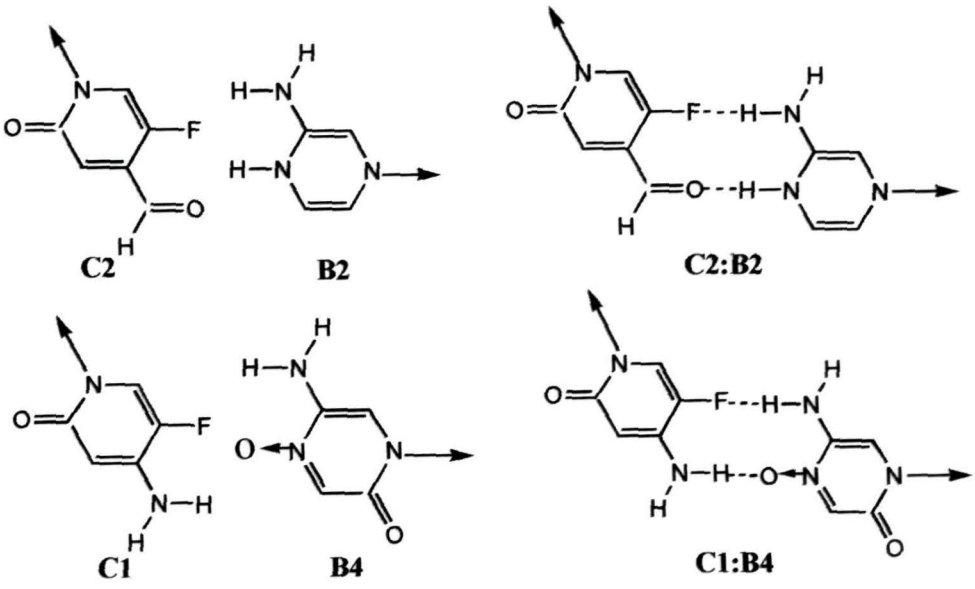


Fig. III. 2 : Component bases and resulting base pairs built out of
(c) the pyridine-pyrimidine base-pairing **Set III,** and
(d) the pyridine-pyrazine base-pairing **Set IV**

Table III.3 Uncorrected pairing energy E_p and BSSE corrected pairing energy E_p(cr) along with the configurational data for H-bonded base pairs of **Sets III** and **IV** obtained from B3LYP/6-31G* optimized geometries and energies a

Base Pair	E	E_p(cr)	R_{cc}	θ_1	θ_2	ϕ
Set III *Pyridine-pyrimidine pairs*						
C1:A5	-6.27	-3.13	7.081	99.4	107.2	-11.9
C2:A3	-11.29	-7.52	7.666	117.6	97.8	12.8
Set IV *Pyridine-pyrazine pairs*						
C2:B2	-6.27	-4.39	11.164	142.4	123.8	-96.5
C1:B4	-8.78	-6.27	11.758	153.5	142.4	51.3

*Pairing energies in kcal/mol; bond lengths in angstrom; bond angles and dihedrals in degrees

Table III.4 H-bond geometry data* for the base pairs of Sets III and IV as obtained from B3LYP/6-31G* optimized geometries

Base Pair	H-bond	R_{hb}	R_{ab}	θ_{hb}	E_p(cr)
Set III *Pyridine-pyrimidine pairs*					
C1:A5	F5-H3	1.888	2.894	168.7	-3.13
	H4-O4	2.448	3.344	147.1	
C2:A3	F5-H3	2.073	2.994	150.1	-7.52
	O4-H4	2.168	3.163	151.3	
Set IV *Pyridine-pyrazine pairs*					
C2:B2	F5-H3	2.292	3.119	137.5	-4.39
	O4-H4	2.037	3.047	169.4	
C1:B4	F5-H5	2.410	2.929	110.8	-6.27
	H4-O4	2.001	2.752	111.9	

*Bond lengths in angstrom; bond angles in degrees

kcal/mol respectively. Although both pairs contain an H-bond of the F...H-N type, they display E_p and E_p(cr) values larger than the previously treated pair **B1:B2** because of the relatively strong C=O...H-N type of H-bond present in both **C1:A5** and **C2:A3**. From Table III.2 we see that, in **C1:A5**, the H4-O4 bond (of the C=O...H-N type) is weakened due to its length (R_{hb} and R_{xy} equal 2.448 and 3.344 Å) and non-linearity (θ_{hb} equals 147.1°). In the **C2:A3** pair, the pairing energy is somewhat larger than in **C1:A5** because the two H-bonds are not so long, although still fairly non-linear.

The two pairs **C1:A5** and **C2:A3** differ from each other in their pairing configurations (Table III.1) mainly with respect to the inter-nuclear distance R_{cc}, which is 7.081 Å for **C1:A5** and 7.666 Å for **C2:A3**. This difference of 0.585 Å may be deemed as quite significant in this context. The θ_1 and θ_2 angles vary from 97.8 to 117.6°, which indicates the pairs as being not of the reversible type. Values of the dihedral φ (-11.9 and 12.8°) indicate that the pairs are more or less of the planar type. Although the pairing energies E_p and E_p(cr) for these pairs are quite substantial, the divergence in their R_{cc} values would seem to indicate that the pyridine-pyrimidine pairs **C1:A5** and **C2:A3** arising from the candidate mimic **Set III** are not isomorphic and hence not suitable for incorporation as the repeat units for an information-bearing macromolecule of the type envisaged here.

III.3.4 Pyridine-pyrazine pairing (Set IV)

Fig. III.2(b) portrays the pyridine bases **C2** and **B2** and the pyrazine bases **C1** and **B4** which belong to the DNA base mimic **Set IV**, yielding the two pyridine-pyrazine pairs **C2:B2** and **C1:B4**. Values of the pairing energy E_p (Table III.1) for **C2:B2** and **C1:B4** are -6.27 and -8.78 kcal/mol respectively. The BSSE-corrected E_p(cr) values are -4.39 and -6.27 kcal/mol. These values are appreciable and point to substantial strength for the H-bonds present in each pair. The H-bonds of the F...H-N type are not expected to be strong, and the F5-H3 H-bond of **C2:B2** together with the F5-H5 H-bond of **C1:B4** (Table III.2) are both markedly non-linear (θ_{hb} = 137.5 and 110.8° respectively) and long as well (R_{hb} = 2.292 and 2.410 Å respectively). However, the C=O...H-N hydrogen bond in **C2:B2** and the N-H....O←N H-bond in **C1:B4** are expected to be strong, and contribute much to the pairing energies. The base **B4** is noteworthy in having an N-oxide oxygen participating in H-bonding in the pair **C1:B4**. The formal unit negative charge on this oxygen makes it even more electronegative than the oxygen of the carbonyl group in **C2:B2**, and explains the larger pairing energy of **C1:B4** compared to **C2:B2**, despite the marked non-linearity of both H-bonds in **C1:B4** (θ_{hb} = 110.8 and 111.9° respectively).

The configuration data of Table III.1 for pairs **C2:B2** and **C1:B4** points to their obvious unsuitability for serving as repeat units in the proposed information-bearing duplexes. The R_{cc} values (11.164 and 11.758 Å respectively) differ by as

much as 0.494 Å. The θ_1 and θ_2 angles (varying from 123.8 to 153.5°) indicate the pairs as not being of the DNA reversible type. More importantly, the dihedral φ values (-96.5 and 51.3° respectively) are widely divergent and point to extensive departure from planarity in both the pairs. This describes the pairs **C2:B2** and **C1:B4** arising from the pyridine-pyrazine mimic **Set IV** as being quite unsuitable for inclusion in the putative information-bearing duplex.

III.3.5 Pyrimidine-pyrazine pairs (Set V)

The DNA base mimic **Set V** consists of the pyrimidine bases **B2** and **B1** along with the pyrazine bases **A6** and **A3** which form the two pyrimidine-pyrazine pairs **B2:A6** and **B1:A3** shown in Fig. III.3. The BSSE-corrected pairing energy E_p(cr) for **B2:A6** (-0.62 kcal/mol) is markedly smaller than for **B1:A3** (-6.90 kcal/mol), even though both pairs are each linked by two H-bonds of the same type, *viz.*, the N-H...F and N-H...N types. However, the H4-F4 H-bond in **B2:A6** (of the N-H...F type) is markedly long (R_{hb} = 2.415 Å; R_{xy} = 3.432 Å) and thereby predicted to be weak, despite being fairly linear.

The configurational data of Table III.1 for **B2:A6** and **B1:A3** points to their unsuit-ability for incorporation into an information-bearing duplex. Not only do their inter-nuclear R_{cc} distances differ appreciably (by 0.712 Å), but **B2:A6** is relatively non-planar (φ = -31.7°) compared to **B1:A3** (φ = 0.2°). The small pairing energy for **B2:A6** is also another factor which disqualifies **Set V** from serving as a good DNA base mimic set.

Schematic Representation

(e) **Set V** (substituted pyrazine-pyrimidine pairs)

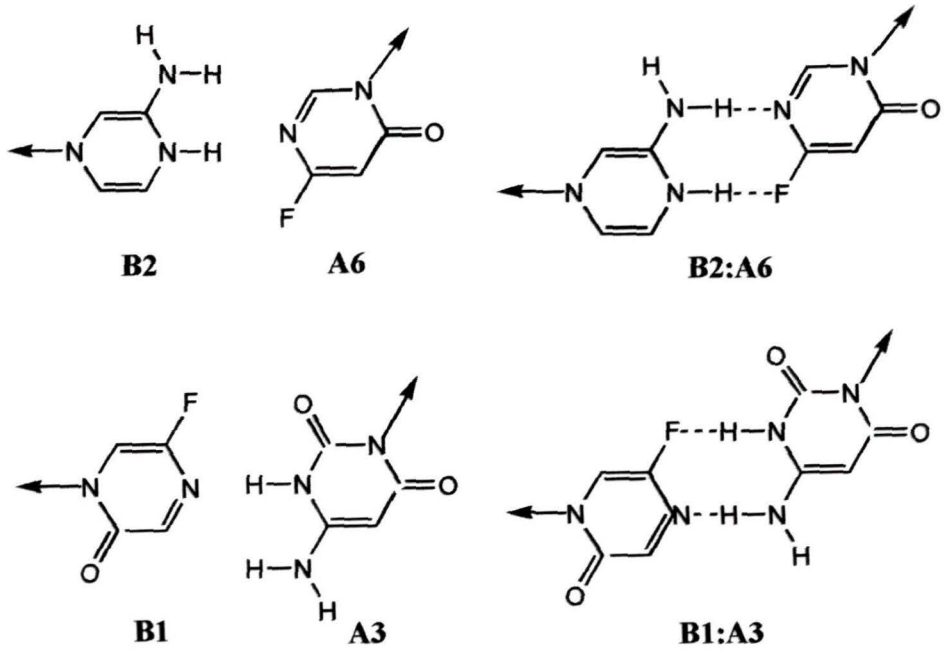


Fig. III. 3 : Component bases and resulting base pairs built out of (e) the pyrazine-pyrimidine base-pairing **Set V**

Table III.5 Uncorrected pairing energy E_p and BSSE corrected pairing energy E_p(cr) along with the configurational data for H-bonded base pairs of **Set V** as obtained from B3LYP/6-31G* optimized geometries and energies [a]

Base Pair	E_p	E_p(cr)	R_{cc}	θ_1	θ_2	ϕ
Set V *Pyrimidine-pyrazine pairs*						
B2:A6	-2.51	-0.62	10.828	134.0	141.0	-31.7
B1:A3	-9.41	-6.90	10.116	144.1	132.4	0.2

*Pairing energies in kcal/mol; bond lengths in angstrom; bond angles and dihedrals in degrees

Table III.6 H-bond geometry data* for the base pairs of **Set V** as obtained from B3LYP/6-31G* optimized geometries)

Base Pair	H-bond	R_{hb}	R_{ab}	θ_{hb}	E_p(cr)
Set V *Pyrimidine-pyrazine pairs*					
B2:A6	H3-N3	2.255	3.245	163.3	-0.62
	H4-F4	2.415	3.432	177.1	
B1:A3	F5-H3	2.101	3.103	170.2	-6.90
	N4-H4	2.205	3.215	170.8	

*Bond lengths in angstrom; bond angles in degrees

III.3.6 Comparison with PM3 SCF-MO results

These B3LYP/6-31G* results may be compared with the earlier PM3 SCF-MO work done on analogous systems[14,15]. The earlier studies did not incorporate an N-methyl group to model the glycoside, using only the free bases. The first observation that comes to mind is that the pairing energies E_p are much larger in the B3LYP/6-31G* method than in the PM3 method. The B3LYP values for E_p range from -2.51 to -29.49 kcal/mol, while the PM3 values range from only -0.68 to -5.11 kcal/mol.

The pairing configurations also differ appreciably in many cases. The co-planarity and isomorphism of the pairs **A1:A2** and **A3:A4** and of the pairs **C1:A5** and **C2:A3** are predicted by both methods. However, the pairs **B1:B2** and **B3:B3**, the pairs **C2:B2** and **C1:B4** and the pairs **B2:A6** and **B1:A3** which are predicted as quite co-planar and isomorphic by the PM3 method turn out to have noticeable disparities in configuration when calculated by the B3LYP strategy used here, so that co-planarity and isomorphism are lost. This departure from isomorphism predicted by the B3LYP method also results in H-bonds which are long and non-linear for pairs **B1:B2**, **C1:A5**, **C2:B2**, **C1:B4** and **B2:A6**, all of which are predicted to have normal, linear H-bonds by the PM3 method.

The key reason here is that the PM3 method predicts largely planar geometries for all the solitary bases, while the B3LYP method predicts non-planar geometries for some of them. In view of the greater sophistication and accuracy of the *ab*

initio B3LYP/6-31G* strategy over the semi-empirical PM3 SCF-MO method, the results arising from the former are perhaps more to be trusted in this context.

III.3.7 Charge transfer during base pairing

Table III.3 lists for each of the 10 base pairs the amount of charge transferred from one base to the other and the direction of charge transfer when H-bonded pairing occurs, with charges given in atomic units. For each base pair, the dipole moments of the two individual bases and of the pair are also given in Table III.3.

The amounts of charge transferred are small, ranging from -0.0006 to -0.0710 a.u. It is possible from the data to correlate the direction of charge transfer with the number of the H-bonds pointing in a given direction. Increase or decrease in dipole moment upon pairing may be correlated with the scope for charge polarisation occurring in each pair.

Set I: The pair **A1:A2** has net charge transferred in the direction **A1→A2** since it has two H-bonds (of the O...H-N type) expected to donate charge in this direction and only one H-bond (of the type N-H...N) donating charge in the opposite direction. The pair **A3:A4**, however, has two H-bonds (of the N-H...O type) donating charge in the **A4→A3** direction, and only one (of the O---H-N type) donating charge in the opposite direction, hence the direction **A4→A3** calculated for the net charge transfer. For both these pairs, the appreciable net charge transfer results in a larger dipole moment for the pair as compared to either of the individual bases.

Table III.7 Net charge transfer occurring during base pairing and dipole moments for each individual base and for the resultant base pair (B3LYP/6-31G* values)

No	Base pair	Charge transferred	Direction of transfer	Dipole Moments		
				1st Base	*2nd base*	*Base pair*
1	**A1:A2**	-0.0242	**A1→A2**	+4.62	+4.51	+9.22
2	**A3:A4**	-0.0693	**A4→A3**	+5.43	+5.88	+10.86
3	**B1:B2**	-0.0326	**B1→B2**	+3.63	+0.83	+4.41
4	**B3:B3**	-0.0006	**B3→B3**	+4.67	+4.67	+0.13
5	**C1:A5**	-0.0248	**C1→A5**	+3.63	+1.85	+4.44
6	**C2:A3**	-0.0619	**C2→A3**	+4.91	+5.38	+4.84
7	**C2:B2**	-0.0010	**C2→B2**	+4.91	+2.06	+3.39
8	**C1:B4**	-0.0344	**B4→C1**	+3.67	+5.06	+6.62
9	**B2:A6**	-0.0571	**A6→B2**	+1.12	+3.09	+4.57
10	**B1:A3**	-0.0710	**B1→A3**	+3.63	+5.39	+8.69

* Charges in atomic units; dipole moments in Debye

Set II: The pair **B1:B2** has two H-bonds (of the N---H-N and F---H-N types) both expected to donate charge in the direction **B1→B2**, which is the direction for net charge transfer noted. The charge transfer in **B1:B2** results in a dipole moment for the pair larger than either of the individual bases. The pair **B3:B3** has virtually no net transfer (only -0.0006 a.u.) since it has two H-bonds (of the N---H-N and N-H---N types) donating charge in opposite directions to each other. In principle, the **B3:B3** pair should have perfect C_{2v} symmetry (D_{2h} if planar as well), but it was optimised without any symmetry constraints. As such, it would not be expected to undergo any net charge transfer. The small dipole moment of only +0.13 D calculated for **B3:B3** (which in principle should be zero) also testifies to the symmetry and lack of charge transfer here.

Set III: The pair **C1:A5** has an H-bond of the F---H-N type expected to donate charge in the **C1→A5** direction, but another stronger N-H---O bond donating charge in the reverse direction, leading to net charge transfer in the **A5→C1** direction. The result is a larger dipole moment for the **C1:A5** pair as compared to either **C1** or **A5**. The **C2:A3** pair has two H-bonds (F---H-N and O---H-N types), both expected to donate charge in the direction **C2→A3** which is as calculated for the pair as a whole. The non-planarity of this pair, coupled with the non-linearity of its H-bonds, results in a dipole moment for **C2:A3** which is somewhat smaller than that for either **C2** or **A3**.

Set IV: The **C2:B2** pair has two H-bonds (F---H-N and O---H-N types) both expected to donate charge in the same direction (**C2→B2**), which is indeed the direction predicted by these calculations. As for **C2:A3** above, the marked non-planarity of the pair and the non-linearity of its H-bonds results in a dipole moment for **C2:B2** which is smaller than that for either **C2** or **B2**. The **C1:B4** pair has a strong H-bond of the N-H---O type donating charge in the **B4→C1** direction, and another weaker F---H-N bond donating charge in the opposite direction. Here, the effect of the stronger H-bond prevails to result in net charge transfer in the **B4→C1** direction as calculated, and also in an appreciably increased dipole moment for **C1:B4**.

Set V: The **B2:A6** pair has two H-bonds (of the N-H---N and N-H---F types) which are both expected to donate charge in the direction **A6→B2**, which is the direction predicted by the calculations. The **B1:A3** pair has two H-bonds (F---H-N and N---H-N types) both expected to donate charge in the **B1→A3** direction, which is as predicted by the calculations. The appreciable charge transfer for both these pairs results in a dipole moment which is noticeably larger for the pair than for the individual bases in each case.

III.4 Conclusions

1. The wide range in facility of H-bonded pairing among the bases of the five sets studied may be explained by referring to the number and type of H-bonds involved, as well as the geometry of the H-bonds and of the pair as a whole.

2. The direction and extent of charge transfer occurring upon formation of the various base pairs may be correlated with the number and type of H-bonds involved, along with their directionality and geometry.

3. Out of the five candidate DNA base mimic sets chosen here for study, only *one* set – the **Set I** comprised of the pyrimidine bases **A1**, **A2**, **A3** and **A4** – may be deemed as truly suitable for furnishing H-bonded base pairs that can serve as repeat units for an information-bearing macromolecular duplex.

References

1. Sponer, J; Lankas, F (eds.), *Computational Studies of DNA and RNA*, Springer, Dordrecht, 2006.
2. Leszczynski, J (ed.), *Computational Molecular Biology*, Elsevier Science, 1999.
3. Sponer, J; Florian, J; Hobza, P; Leszezynski, J. *J Biomol Struct Dyn*,1996, **13**, 827
4 .Sponer, J; Leszczynski, J; Hobza, P. *J Biomol Struct Dyn*,1996, **14**, 117
5. Giese, T. J; Shere, E. C; Cramer, C, J; York, D. M. *J Chem Theory Comput*, 2005, **1**, 1275.
6. Sponer, J; Jurecka, P; Hobza, P. *J Am Chem Soc*,2004, **126**, 10142
7. Danilov, V.I; Anisimov, V, M. *J Biomol Struct Dyn*,2005, **22**, 471
8. Machado, M; Ordejon, P; Artacho, E; Sanchez-Portal, D; Soler, J, M. Online at oai : arXiv. Org: physics/9908022 (1999)
9. Elstner, M; Hobza, P; Fravenheim, T; Suhai, S; Kaxiras, E; *J Chem Phys*,2001, **114**, 5149
10. Monajjemi, M; Chahkandi, B; Zare, K; Amiri, A. *Biochem* (Moscow), 2005, **70**, 447
11. Gould, I. R; Kollman, P. A. *J Am Chem Soc*,1994, **116**, 2493
12. Venkateswarlu, D; Lyngdoh, R. H. D. *J Chem Soc Perkin Trans 2*,1995, 839
13. Venkateswarlu, D; Lyngdoh, R. H. D; Bansal, M. *J Chem Soc Perkin Trans 2*, 1997, 621
14. Buam, D, M. L; Lyngdoh, R. H. D. *J Mol Struct (THEOCHEM)*,2000, **505**, 149
15. Buam, D. M. L; Lyngdoh, R. H. D. *Indian J Chem*,2002, **41B**, 2346

16. Guerra, C. F; Bickelhaupt, F. M; Snijders, J. G; Baerends, E. J. *J Am Chem Soc*, 2000, **122**, 4117

17. Watson, J. D; Crick, F. H. C. *Nature*, 1953, **171**, 737

18. Crick, F. H. C; Watson, J. D. *Proc Roy Soc (London) Ser A*, 1954, **223**, 80

19. Seeman, N. C; Rosenberg, J. M; Suddath, F. L; Kim, J. J. P; Rich, A. *J Mol Biol*, 1976, **104**, 109

20. Rosenberg, J. M; Seeman, N. C; Day, R. O. Rich, R .A. *J Mol Biol*, 1976,**104**, 145

21. Pauling, L; Corey, R. B. *Archiv Biochem Biophys*, 1956, **65**, 164

22. Becke, A. D. *Phys Rev B*, 1998, **38**, 3093

23. Becke, A. D. *J Chem Phys*, 1993, **98**, 5648

24. Lee, C; Yang, W; Parr, R. G. *Phys Rev B*, 1998, **37**, 785

25. Boys, S. F; Bernardi, F. *Mol Phys*, 1985, **19**, 553

26. Frisch, M. J; Trucks, G W; Schlegel, H.B; Scuseria, G.E; Robb, M.A; Cheeseman, J. R; Montgomery Jr, J. A; Vreven, T; Kudin, K. N; Burant, J. C;Millam, J. M; Iyengar, S. S; Tomasi, J; Barone, V; Mennucci, B; Cossi, M; Scalmani, G; Rega, N; Petersson, G. A; Nakatsuji, H; Hada, M; Ehara, M; Toyota, K; Fukuda, R; Hasega-wa, J; Ishida, M; Nakajima, T; Honda, Y; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V; Adamo, C; Jaramillo, J; Gomperts, R; Stratmann, R. E; Yazyev, O; Austin, A. J; Cammi, R; Pomelli, C; Ochterksi, J. W; Ayala, P. Y; Morokuma, K; Voth, G. A; Salvador, P; Dannenberg, J. J; Zakrzewski, V. G; Dapprich, S; Daniels, A. D; Strain, M. C; Farkas, O; Malick, D. K; Rabuck, A. D; Raghavachari, K; Foresman, J. B; Ortiz, J. V; Cui, Q; Baboul, A. G; Clifford, S; Cioslowski, J; Stefanov, B. B; Liu, G; Liashenko, A;

Piskorz, P; Komaromi, I; Martin, R. L; Fox, D. J; Keith, T; Al-Laham, M. A; Peng, C. Y; Nanayakkara, A; Challacombe, M; Gill, P. M. W; Johnson, B; Chen, W; Wong, M. W; Gonzalez, C; Pople, J. A. Gaussian 03 Revision A.1, Gaussian, Inc., Pittsburgh PA, 2003.

A chemistry professor couldn't resist interjecting a little philosophy into a class lecture. He interrupted his discussion on balancing chemical equations, saying, "Remember, if you're not part of the solution, you're part of the precipitate!"

CHAPTER FOUR

SELF-ASSOCIATIVE BASE PAIRS AS REPEAT UNITS FOR NOVEL H-BONDED INFORMATION-BEARING MACROMOLECULAR DUPLEXES

IV.1 Introduction

The concept of H-bonded macromolecular duplexes with the capacity to store and replicate information at the molecular level began with the discovery of the structure and function of DNA.[1] Such H-bonded double-stranded macromolecules may be also designed artificially, and may one day be of relevance to information technology at the molecular level. The information-bearing capacity arises out of the manner in which constraints are imposed upon the allowed H-bonded pairing units. Only certain pairing motifs are possible due to the configurational constraints imposed. These macromolecules should also possess, in principle, the potential to replicate or be copied, which arises out of their double-stranded structure.

In DNA, the two strands of the duplex are not identical but *complementary*, and arise from the two *hetero-associative* base pairs AT and GC. DNA base pairs have received much attention from quantum chemists which has been amply reviewed.[2-5] Such work is not directly related to the subject of this Thesis (which deals with novel H-bonded pairing systems not found in DNA or in nature). However, the general approaches and methodologies used in these studies are relevant to this work. Some of this work includes the study of stacking interactions,[6] highly accurate pairing energies,[7] cation binding to base pairs,[8] radical ion forms of DNA base pairs,[9-11] and modified DNA mispairs as a source of mutations.[12, 13]

IV.1.1 Characteristics of DNA base mimic set

The studies of Chapter Three were conducted in the context of hetero-associative base pair systems which could go to form a macromolecular duplex where the two strands are not identical but complementary. It is also possible to conceive of macromolecular duplexes where the two strands are *identical.* The repeat units would then have to consist of *self-associative* base pairs (or dimers). In order for such H-bonded dimers to serve as viable repeat units for a self-associative macromolecular duplex, they would have to satisfy the following criteria:

1. The duplex is a polymer incorporating repetitive sequences comprised of *M* chosen base pairs, where each base pair is an H-bonded dimer of a particular member of the set of *M* chosen solitary bases.

2. Choice of the *M* select base pairs rests primarily upon the *pairing configuration,* which should be basically the *same* for all the *M* pairs. Within this configurational constraint, each base may pair *only with itself* and no other base in the set.

4. To ensure configurational and conformational stability, at least two hydrogen bonds between the component bases are required.

5. There should be one atom in each base which serves as a point of attachment to the backbone of the macromolecule, and which should not be involved in the H-bonding.

The pioneering work of Eschenmoser and his group[14-16] has dealt with alternative schemes for constructing H-bonded duplexes based on the natural DNA and RNA base pairs, but incorporating other sugars besides D-ribose (including various tetroses, hexoses etc.) so as to build up a variety of viable duplex structures. This study aims to

explore the potential of *wholly synthetic* bases and base pairs to serve as repeat units for information-bearing H-bonded macromolecular duplexes, as was the focus of the earlier theoretical studies.[17-19] The work of Eschenmnoser's group utilized only the normally occurring nucleic acid bases of nature as the fundamental repeat units.

Buam and Lyngdoh[17, 18] used the PM3 SCF-MO method to study H-bonded pairing among various nitrogenous bases to determine the suitability of such pairs to serve as repeat units for information-bearing macromolecules, dealing with self-associative dimers of nitrogenous heterocycles as well as hetero-associative pairs. The use of the B3LYP/6-31G* approach to design hetero-associative base pairs along similar lines has formed the content of Chapter Three .[19]

The features required to select a group of base pairs as suitable for incorporation into a self-associative information-bearing macromolecular duplex include *(a)* facility of H-bonded pairing giving stable pairs, *(b)* similarity (isomorphism) of pairing configuration, and *(c)* specificity of base-pairing so that each base in the set pairs only with itself and with no other member of the set within the prescribed configuration.

Earlier computational studies[17-18] had delineated the following features influencing the facility of H-bonded pairing between nitrogenous bases of the kind studied here:

1. The *number* of hydrogen bonds present in the base pair.

2. The *type* of H-bond occurring, where, for instance, H-bonds of the C=O----H–N type are stronger than those of the F----H–N type.

3. The *linearity* of the H-bonds, where non-linearity diminishes pairing facility.

4. The *co-planarity* of the rings of the two component bases within the pair, since this can augment stacking interactions in the duplex.

The pairing configuration of the base pairs is also important. For any set of base pairs to be chosen as repeat units of the information-bearing macromolecule, the pairing configuration should be closely similar for all the constituent pairs. We assume that the pairing configuration of a solitary base pair in gas phase would be largely similar to that when present within the macromolecular duplex itself. This assumption appears to be true for the DNA base pairs, whose gas phase pairing configuration as studied theoretically is close to that obtained from the crystal structure of the DNA double helix itself.[20-24]

IV.1.2 Self-associative base pairs

Unlike in Chapter Three,[19] we now focus on *self-association* as a basis for designing the repeat units of the macromolecular duplex. For a base monomer to be able to H-bond with itself, the basic requirement is that the base should contain both a proton donor and a proton acceptor within itself. Proton donors for H-bonding include the exocyclic $-NH_2$ and -OH functionalities and the endocyclic –NH- functionality. Proton acceptors include endocylic nitrogen atoms (-N=) and exocyclic carbonyl oxygen atoms (-C=O). In order for H-bonding to occur successfully, the proton donor and proton acceptor functionalities should be adjacent to each other and face the same direction. This then calls for difunctional moieties like the O=C-NH-, $-N=C-NH_2$ and the -N=C-OH groups within each base. We do not consider the OH group as a proton donor here since it often undergoes tautomerism to a C=O group when present in an aromatic system. We consider here only systems with the O=C-NH- and $-N=C-NH_2$ groups within a base. Note that such moieties are characteristic of the natural nucleic acid base-pairing systems as well.

We choose six sets of two nitrogenous bases each (depicted in Fig. IV.1) whose H-bonded pairing properties are examined here using density functional theory. **Set I** consists of the two substituted azoles **A1** (imidazol-2-one) and **A2** (pyrazol-3-one), where both bases contain the O=C-NH moiety. **Set II** consists of the two substituted imidazoles **B1** (imidazol-2-one) and **B2** (2-aminoimidazole), where the base **B1** has the O=C-NH- moiety and **B2** has the -N=C-NH_2 moiety. **Set III** is comprised of the two substituted pyrimidines **C1** (4-aminopyrimid-6-one) and **C2** (pyrimid-2,4-dione) with the O=C-NH- and –N=C-NH_2 functionalities respectively. **Set IV** consists of the substituted pyrimidines **D1** (2-amino-pyrimid-5-one) and **D2** (pyrimid-2,5-dione), where **D1** has the -N=C-NH_2 moiety while **D2** has the O=C-NH- moiety. **Set V** consists of the two bicyclic base systems **E1** (pyrrolo[5,6-c]pyrid-2-one) and **E2** (pyrrolo[5,6-d]pyrid-2-one), both having the O=C-NH- moiety. **Set VI** is comprised of the two bicyclic bases **F1** (pyrrolo[5,6-d]pyrid-2-amino) and **F2** (pyrrolo[5,6-d]pyrid-2-one), where **F1** has the -N=C-NH_2 moiety while **F2** has the O=C-NH- moiety. Each base has a ring nitrogen atom through which it may be attached to the macromolecular backbone, and which does not participate in H-bonding. The would-be N-C bond connecting the base to polymeric backbone is indicated by a pointed arrow in Fig. IV.1 For these calculations, the backbone moiety is modeled simply by a methyl group bonded to the ring nitrogen atom of the base.

Each set of bases is tested for the ability of both constituent bases to dimerise through H-bonding to furnish two stable self-associative base pairs of closely similar configuration, as well as for the inability to pair among themselves one with the other within this configuration. These twin characteristics are essential for any of these Sets

to provide viable candidate pairs with the capacity to store information. The structure of the backbone for the macromolecule is worked out in Chapter Five, where a sugar phosphate backbone and a polyamide backbone are proposed and studied.

IV.2 Theoretical Methodology

The B3LYP density functional theory (DFT) model[25-27] was used with the 6-31G* basis set for all molecular species (solitary bases and base pairs). Full geometry optimization was carried out without any symmetry constraints using initial structures from PM3 SCF-MO calculations as input. The pairing energy E_p was calculated from the difference in total energies between the pair and the component bases. The basis set superimposition error (BSSE) energy correction was estimated by the Boys-Bernardi counterpoise method[28] involving re-calculation of the two monomers within the dimer using the basis set for the dimer. This works well for hydrogen-bonded base pairs since the monomers do not change much in geometry when incorporated into the dimer. The BSSE-compensated pairing energy is denoted here as E_p(cr).

The B3LYP/6-31G* geometries were further subjected to single point calculations at the MP2/6-311++G(d,p) level of theory for all the systems studied. The pairing energy thus obtained is given as E_p(MP2). In order to establish the isomorphism of the base pairs of Set III (selected as the optimal one here – see later), the pairing configurations and pairing energies were all fully re-calculated using the B3LYP/ 6-31++G(d,p) and the MP2/6-31G(d,p) levels of theory. All calculations were carried out using the GAUSSIAN 2003 program package.[29]

IV.2.1 Descriptors of pairing configuration and H-bond geometry

This study adopts the following descriptors (or markers) of pairing configuration, which correspond to those commonly used for DNA base pairs:

1. The distance R_{cc} between the carbons of the two methyl groups attached to the two component bases of the pair.
2. The angles θ_1 and θ_2 spanning the C1-N1-N2 and C2-N2-N1 atoms in the two bases of the pair, where C1 and C2 are the two methyl carbons, while N1 and N2 are the nitrogen termini of the would-be glycoside C-N bonds.
3. The dihedral φ which spans the C1-N1-N2-C2 atoms, serving as an indicator of co-planarity (or the lack of it) between the two bases within the pair.

A hydrogen bond within a base pair is represented as **X....H-Y** or **X-H....Y**, where **X** and **Y** are electronegative atoms belonging to two different component bases. Geometry around the hydrogen bonds was studied using the following descriptors:

1. The length R_{hb} of the actual hydrogen bond **X....H** or **H....Y**.
2. The length R_{xy} between the two electronegative atoms **X** and **Y**.
3. The hydrogen bond angle θ_{hb} of the moiety **X...H-Y** or **X-H...Y**.

Atom **X** belongs to the base on the left, while atom **Y** belongs to the base on the right, as they appear in the pairs of Figs. IV.2, 3 and 4. The H-atom is numbered as per the atom **X** or **Y** to which it is bonded covalently. If the H-bond involves an atom **X** on the left as the electron pair donor and an atom **Y** on the right as the proton donor, it is labeled as **X....H-Y**, where **X** and **H** are numbered as per convention, and. the converse for H-bonds a proton donor on the left and a lone pair donor on the right.

IV.3 Results and Discussion

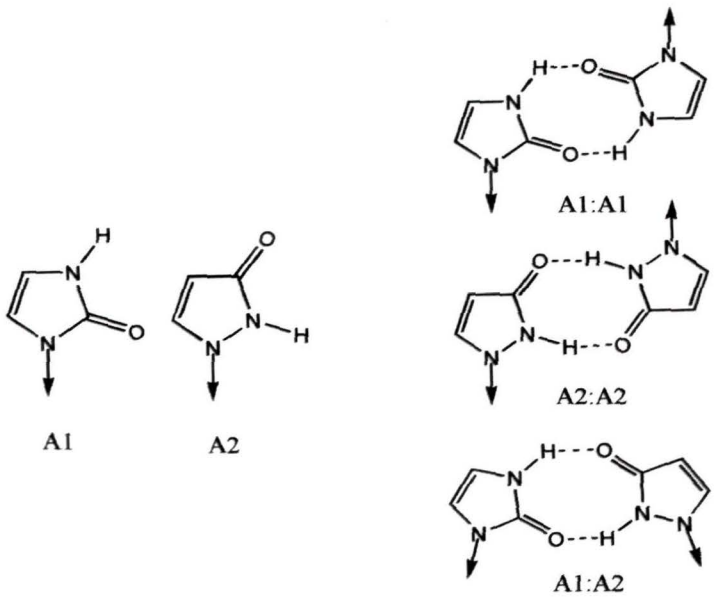
Table IV.1 presents B3LYP/6-31G* values of the pairing energy E_p, and the BSSE-compensated pairing energy E_p(cr) for each pair of the Sets I and II, giving the configurational data for the base pairs of Sets I and II using the markers R_{cc}, θ_1, θ_2 and φ described earlier, together with the pairing energy E_p(MP2) obtained from single point calculations at the MP2/6-311++G(d,p) level of theory. Table IV.2 gives data about the geometry of the H-bonds of each base pair for Sets I and II. Table IV.3 presents pairing energies and configuration data for Sets III and IV, while Table IV.4 gives the H-bond data for these two sets. Table IV.5 presents pairing energies and configuration data for Sets V and VI, and Table IV.6 presents their H-bond data.

The B3LYP/6-31G* BSSE energy is only a small fraction of the total energies of the bases and base pairs, ranging from 0.0029111 to 0.0069254 hartree (1.83 to 4.34 kcal/mol). However, it does appreciably alter the values of the pairing energies E_p. These pairing energies have rather large values (-10.63 to -22.53 kcal/mol). Pairing energies of comparable range (from -9.8 to -26.8 kcal/mol) have been obtained in an extensive study by Machado et al.[30] on 30 different DNA base pairs using their own linear-scaling DFT scheme, along with the use of standard Hartree-Fock and MP2 methods. The earlier work of Sponer et al.[31] also yielded similar ranges of values.

IV.3.1 Azole-azole pairs from Set I

Figure IV.1(a) portrays schematically the possible H-bonded pairs arising out of the azole bases of **Set I**, *viz.*, the self-associative pairs **A1:A1** and **A2:A2** along with the hetero-associative pair **A1:A2**. The **A1:A1** and **A2:A2** pairs are symmetrical dimers, while **A1:A2** is not so. Pairs **A1:A1** and **A2:A2** have their two would-be attachment

Set I (substituted azole-azole pairs)



Set II (substituted imidazole-imidazole pairs)

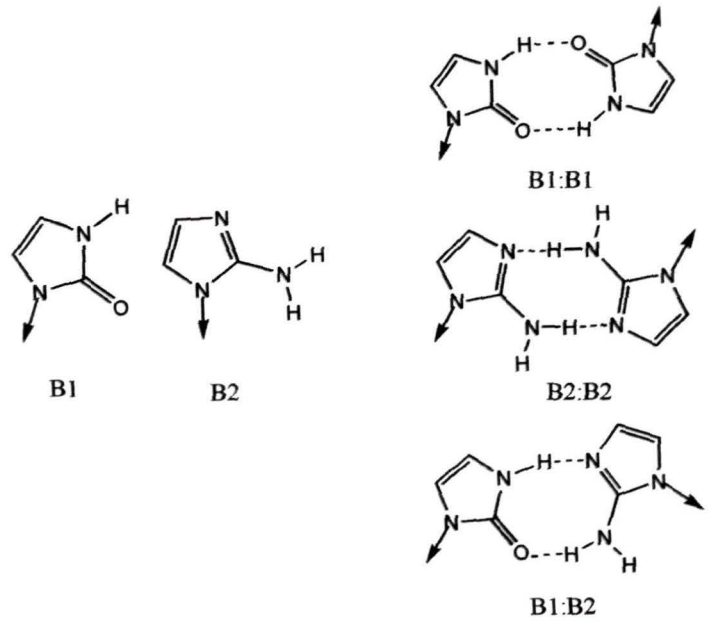


Fig. IV. 1 : Component bases and resulting base pairs built out of
(a) the azole-azole base-pairing **Set I**, and
(b) the imidazole-imidazole base-pairing **Set II**

Table IV.1 Uncorrected pairing energy E_p and the BSSE-corrected pairing energy E_p(cr) along with configurational data for the H-bonded base pairs of **Sets I** and **II** as obtained from B3LYP/6-31G* optimized geometries and energies. Given also are the pairing energies E_p(MP2) without BSSE correction as calculated by the MP2/6-311++G(d,p)// B3LYP/6-31G* strategy [a]

Base Pair	E_p	E_p(cr)	R_{cc}	θ_1	θ_2	φ	E_p(MP2)
Set I *Azole pairs*							
A1:A1	-20.96	-17.23	8.849	131.0	131.0	180.0	-20.48
A2:A2	-22.53	-18.14	7.875	117.8	117.6	180.0	-23.15
A1:A2	-21.54	-17.51	6.820	111.1	100.8	32.2	-21.47
Set II *Imidazole pairs*							
B1:B1	-20.96	-17.23	8.849	131.0	131.0	180.0	-20.48
B2:B2	-14.87	-11.57	9.103	130.9	130.9	178.3	-13.79
B1:B2	-17.96	-14.40	8.750	132.4	131.9	3.4	-17.09

[a] Pairing energies in kcal/mol; R_{cc} in angstrom; angles in degrees

Table IV.2. H-bond data[a] for the base pairs of **Sets I** and **II** (B3LYP/6-31G* geometries) along with the BSSE-corrected pairing energy $E_p(cr)$[b]

Base Pair	H-bond	R_{hb}	R_{xy}	θ_{hb}	$E_p(cr)$
Set I *Azole pairs*					
A1:A1	O2-H3	1.785	2.813	172.3	-17.23
	H3-O2	1.785	2.813	172.3	
A2:A2	H2-O3	1.718	2.759	173.1	-18.14
	O3-H2	1.717	2.758	172.9	
A1:A2	O2-H2	1.759	2.793	172.7	-17.51
	H3-O3	1.765	2.796	172.4	
Set II *Imidazole pairs*					
B1:B1	O2-H3	1.785	2.813	172.3	-17.23
	H3-O2	1.785	2.813	172.3	
B2:B2	N3-H2	1.982	3.012	175.7	-11.57
	H2-N3	1.982	3.012	175.7	
B1:B2	H3-N3	1.876	2.900	169.2	-14.40
	O2-H2	1.889	2.916	175.0	

[a] Bond distances in angstrom; angles in degrees

[b] Energies in kcal/mol

bonds *trans* to each other, while in pair **A1:A2** they are *cis*. These pairs exhaust all the possibilities for H-bonded pairing between monomers **A1** and **A2** which involve at least two H-bonds and also exclude from involvement in H-bonding the ring N-atom terminus of the would-be glycoside bonds. These pairs make use of H-bonds of the C=O....H-N type only, and each base possesses the same type of pairing functionality, namely, the O=C-NH moiety.

The B3LYP/6-31G* uncorrected pairing energy E_p and the BSSE-corrected pairing energy E_p(cr) have large values for all three pairs (Table IV.1). Values of E_p and E_p(cr) are respectively -20.96 and -17.23 kcal/mol for the **A1:A1** dimer, -22.53 and -18.14 kcal/mol for the **A2:A2** dimer, and -21.54 and -17.51 kcal/mol for the **A1:A2** pair. Use of the MP2/6-311++G(d,p) single point calculations on the B3LYP/6-31G* geometries leads to values of the BSSE-uncorrected pairing energy E_p(MP2) which compare favorably with the BSSE-uncorrected B3LYP/6-31G* optimized values, being -20.48 for the **A1:A1** dimer, -23.15 kcal/mol for the **A2:A2** dimer, and -21.47 kcal/mol for the **A1:A2** pair. These large values of the pairing energy are linked to:

(a) the presence in each pair of two H-bonds of the C=O...H-N type, rated as strong in the earlier PM3 studies,[17-19]

(b) relatively small values for the H-bond lengths R_{hb} and R_{xy} (Table IV.2), being respectively 1.785 and 2.813 Å for the **A1:A1** pair, about 2..758 and about 2.759 Å for the **A2:A2** pair, and 1.759 or 1.765 and 2.793 or 2.796 Å for the **A1:A2** pair,

(c) the near-linearity of the H-bonds, where values of the H-bond angle θ_{hb} given in Table II are 172.3° for the **A1:A1** pair, 172.9 or 173.1° for the **A2:A2** pair, and 172.4° or 172.7° for the **A1:A2** pair.

Table IV.1 also gives values of the four configurational markers R_{cc}, θ_1, θ_2 and φ for the base pairs of Set I, where the **A1:A1** and **A2:A2** pairs have D_{2h} symmetry, while the **A1:A2** pair is not symmetrical. This is evident in the values of the θ_1 and θ_2 markers, which are equal for the symmetrical **A1:A1** and **A2:A2** pairs but differ for the **A1:A2** pair. The values of 180.0° for the dihedral angle φ in the self-associative **A1:A1** and **A2:A2** pairs point to essential co-planarity of the two base moieties for these two pairs. This is not seen for the hetero-associative **A1:A2** pair, where φ equals -32.2°, indicating a propeller twist for this non-planar pair.

Values of the internuclear distance R_{cc} differ for all three pairs, being 8.849 Å for **A1:A1**, 7.875 Å for **A2:A2**, and 6.820 Å for **A1:A2**. Note also the dissimilarities in value of the θ_1 and θ_2 angles among the three base pairs. All this indicates three different pairing configurations for the three pairs. It is thereby apparent that since no two pairs have closely similar pairing configurations, the azole Set I does not qualify as a candidate base set which can furnish viable repeat units for the macromolecular H-bonded duplex.

IV.3.2 Imidazole-imidazole pairs from Set II

Figure IV.1(b) portrays schematically the H-bonded base pairs built out of the imidazole bases of **Set II**, *viz.*, the self-associative pairs **B1:B1** and **B2:B2,** and the hetero-associative pair **B1:B2** (where base **B1** of Set II is the same as **A1** of Set I). The **B1:B1** and **B2:B2** pairs are symmetrical dimers, while **B1:B2** is not so. The pairs **B1:B1** and **B2:B2** have their two would-be backbone attachment bonds *trans* to each other, while in pair **B1:B2** they are *cis*. These three pairs exhaust all possibilities for H-bonded pairing between monomers **B1** and **B2** which involve at least two H-bonds

and exclude the ring N-atom termini of the would-be glycoside bonds. These pairs make use of H-bonds of both the C=O....H-N and the C=N....H-N types; where **B1** has the O=C-NH moiety while **B2** the -N=C-NH_2 moiety.

The B3LYP/6-31G* pairing energies E_p and E_p(cr) for the three base pairs (Table IV.2) are quite large, where E_p values are -20.96, -14.87 and -17.96 kcal/mol, while values of E_p(cr) are -17.23, -11.57 and -14.40 kcal/mol for **B1:B1**, **B2:B2** and **B1:B2** respectively. Single point MP2/6-311++G(d,p)//B3LYP/6-31G* values of E_p(MP2) are close to the BSSE-uncorrected B3LYP/6-31G* optimized values, being -20.48 kcal/mol for the **B1:B1** dimer, -13.79 kcal/mol for the **B2:B2** dimer, and -17.09 kcal/mol for the **B1:B2** pair. These large values of pairing energy may be attributed to

(a) strong H-bonds of the C=O....H-N and C=N....H-N types,

(b) near-linearity of the H-bonds (Table II), (θ_{hb} ranging from 169.2 to 175.7°),

(c) short H-bond lengths (Table IV.2) which are, however, on the whole longer (R_{hb} from 1.785 to 1.982 Å and R_{xy} from 2.813 to 3.012 Å) than those for the Set I pairs.

The order of magnitude of pairing energy with respect to pair is **B1:B1** (two C=O....H-N H-bonds) > **B1:B2** (one C=O....H-N and one C=N....H-N H-bond) > **B2:B2** (two C=N....H-N H-bonds). This suggests that H-bonds of the C=O....H-N type are stronger than the C=N....H-N type, as noticed in other pairs discussed below.

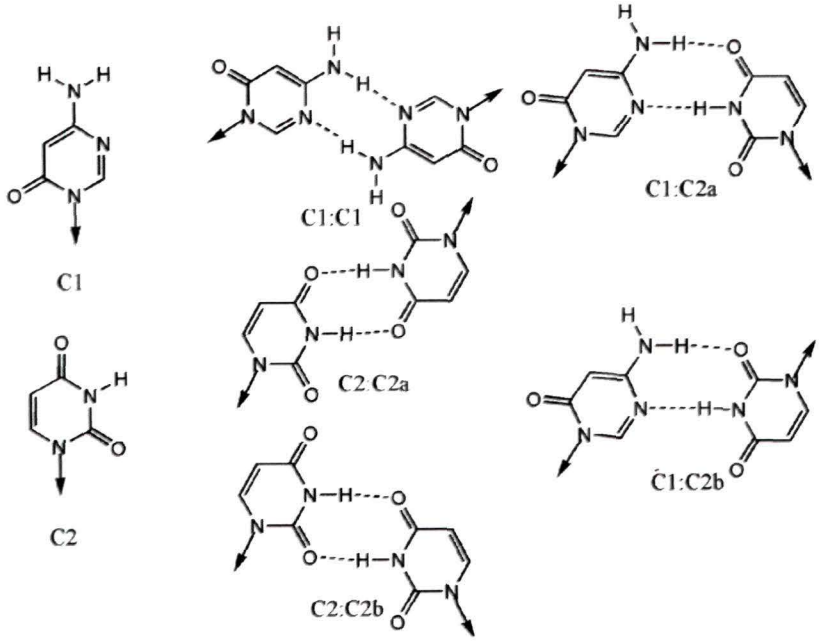
Table IV.1 gives B3LYP/6-31G* optimized values of the configuration markers for the pairs of Set II. Pairs **B1:B1** and **B2:B2** have D_{2h} symmetry (where the θ_1 and θ_2 markers are equal in value), while the **B1:B2** pair is not symmetrical. The self-associative pairs **B1:B1** and **B2:B2** are essentially planar, with the dihedral φ close to 180°. The **B1:B2** pair, although asymmetrical, is also planar, with φ close to zero.

Among all these pairs, the data points to some similarity of configuration for the self-associative pairs **B1:B1** and **B2:B2**. The θ_1 and θ_2 markers have closely similar values for these two pairs, ranging between 130.9 and 131.0°, pointing to reversibility of both these pairs. The dihedrals φ for **B1:B1** and **B2:B2** are very close to 180°. However, values of the inter-nuclear distance R_{cc} are 8.849 and 9.103 Å for **B1:B1** and **B2:B2** respectively. This difference of 0.254 Å may be regarded as detracting somewhat from true isomorphism of these pairs. In contrast, the hetero-associative pair **B1:B2** has a configuration markedly different from that of the self-associative pairs mainly in that the dihedral φ is 3.4° (not near 180°). Otherwise, the distance R_{cc} for **B1:B2** is 8.750 Å (not too different from that of **B1:B1**), while θ_1 and θ_2 equal 132.4 and 131.9° respectively, The φ value of 3.4° for **B1:B2** predicts that **B1** and **B2** cannot pair amongst themselves in a configuration similar to that of their respective self-associative dimers, but rather in a fashion reverse to them. The imidazole Set II only just barely qualifies as a candidate base set to furnish viable repeat units for the macromolecular duplex. That the monomers **B1** and **B2** of Set II yield pairs of at least borderline acceptability arises from the two *different* moieties O=C-NH and N=CH-NH_2 present, one in each base of this set.

IV.3.3 Pyrimidine-pyrimidine pairs from Set III

Figure IV.3(a) portrays five H-bonded pairs arising from the pyridimine bases **C1** and **C2** of **Set III**, where **C1** has the –N=C-NH_2 moiety and **C2** the O=C-NH moiety. The pairs include the self-associative dimers **C1:C1**, **C2:C2(a)** and **C2:C2(b)** along with the hetero-associative pairs **C1:C2(a)** and **C1:C2(b)**. These five pairs exhaust all the possibilities for H-bonded pairing between **C1** and **C2** that involve at least two H-

(a) **Set III** (substituted pyrimidine-pyrimidine pairs)



(b) **Set IV** (substituted pyrimidine-pyrimidine pairs)

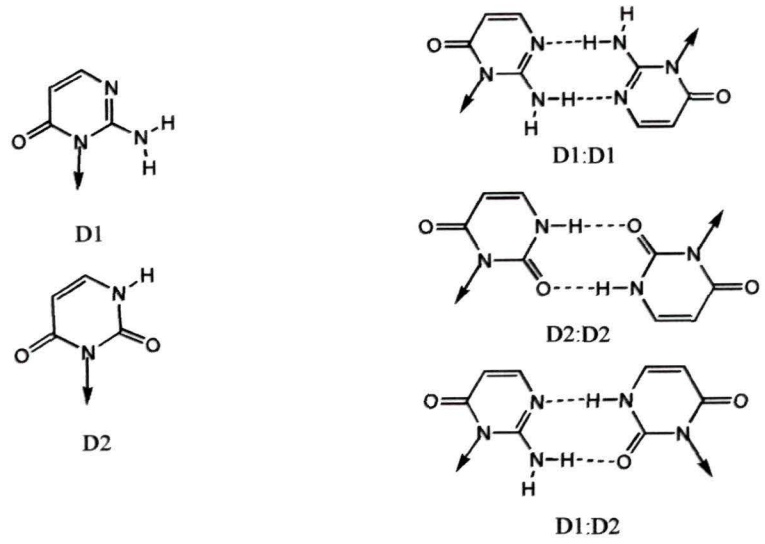


Fig. IV. 2 : Component bases and resulting base pairs built out of
(a) the pyrimidine-pyrimidine base-pairing **Set III,** and
(b) the pyrimidine-pyrimidine base-pairing **Set IV**

Table IV.3 Uncorrected pairing energy E_p and the BSSE-corrected pairing energy E_p(cr) along with configurational data for the H-bonded base pairs of **Sets III** and **IV** as obtained from B3LYP/6-31G* optimized geometries and energies. Given also are the pairing energies E_p(MP2) without BSSE correction as calculated by the MP2/6-311++G(d,p)// B3LYP/6-31G* strategy [a]

Base Pair	E_p	E_p(cr)	R_{cc}	θ_1	θ_2	φ	E_p(MP2)
Set III *Pyrimidine pairs*							
C1:C1	-10.63	-8.07	11.169	151.0	151.0	179.8	-9.91
C2:C2(a)	-14.43	-10.82	11.181	153.3	153.3	179.8	-12.81
C2:C2(b)	-13.47	-9.10	8.623	96.7	138.1	0.2	-12.50
C1:C2(a)	-15.37	-11.26	7.956	112.7	115.7	0.2	-13.46
C1:C2(b)	-14.88	-10.79	9.523	131.2	139.6	179.5	-13.29
Set IV *Pyrimidine pairs*							
D1:D1	-15.05	-11.56	8.900	119.9	119.9	180.0	-13.88
D2:D2	-19.73	-16.32	8.621	118.0	117.9	179.8	-18.69
D1:D2	-17.23	-14.08	8.391	119.4	118.8	3.8	-16.42

[a] Pairing energies in kcal/mol; R_{cc} in angstrom; angles in degrees

Table IV.4 H-bond data[a] for the base pairs of **Sets III** and **IV** (B3LYP/6-31G* geometries) along with the BSSE-corrected pairing energy $E_p(cr)$[b]

Base Pair	H-bond	R_{hb}	R_{xy}	θ_{hb}	$E_p(cr)$
Set III *Pyrimidine pairs*					
C1:C1	H4-N3	2.045	3.067	176.7	-8.07
	N3-H4	2.044	3.066	176.8	
C2:C2(a)	O4-H3	1.838	2.866	172.7	-10.82
	H3-O4	1.837	2.865	172.2	
C2:C2(b)	O4-H3	1.856	2.882	172.0	-9.98
	H3-O2	1.862	2.885	171.5	
C1:C2(a)	H4-O4	1.979	2.999	177.8	-11.26
	N3-H3	1.822	2.864	179.4	
C1:C2(b)	H4-O2	2.043	3.060	176.2	-10.79
	N3-H3	1.809	2.851	178.3	
Set IV *Pyrimidine pairs*					
D1:D1	N3-H2	1.933	2.965	178.3	-11.56
	H2-N3	1.934	2.965	178.3	
D2:D2	H3-O2	1.795	2.827	178.4	-16.32
	O2-H3	1.799	2.831	178.5	
D1:D2	H3-N3	1.883	2.922	178.9	-14.08
	O2-H2	1.842	2.869	179.0	

[a] H-bond distances in angstrom; angles in degrees

[b] Pairing energies in kcal/mol

bonds and exclude from H-bonding the ring N-atom termini of the would-be glycoside bonds. The would-be glycoside bonds are *trans* to each other for **C1:C1**, **C2:C2(a)** and **C1:C2(b)**, but point in the same direction for **C2:C2(b)** and **C1:C2(a)**.

The B3LYP/6-31G* uncorrected and BSSE-corrected pairing energies E_p and E_p(cr) for these base pairs are given in Table IV.3. Values of E_p range from -10.63 to -15.37 kcal/mol, while E_p(cr) ranges between -8.07 and -11.26 kcal/mol. The single point MP2/6-311++G(d,p)//B3LYP/6-31G* values of E_p(MP2) compare fairly well with the BSSE-uncorrected B3LYP/6-31G* optimized values, ranging from –9.91 to –13.46 kcal/mol. These fairly large values point to appreciable stability of these pyrimidine-pyrimidine pairs, although they are not as stable as the azole-azole pairs of **Set I** nor the imidazole pairs of **Set II**. These large pairing energies may be attributed to factors similar to those invoked for the pairs of Set I and Set II, namely, H-bonds of the C=O...H-N type, fairly short H-bond lengths, and near linearity of all the H-bonds, as may be seen from the data of Table IV.4. The observation that the pairing energies for the Set III pairs are smaller compared to those for the Set I and Set II pairs may also be linked to the appreciably longer H-bond lengths for the Set III pairs (ranging between 1.809 and 2.045 Å). We may say that the pyrimidine pairs of Set III are less tightly packed than the azole pairs of Set I and Set II, which results in generally larger pairing energies for the Set I and Set II pairs.

B3LYP/6-31G* optimized values of the configuration markers R_{cc}, θ_1, θ_2 and φ for the base pairs of Set III are given in Table IV.3. It is immediately obvious from this data that the self-associative pairs **C1:C1** and **C2:C2(a)** are highly similar to each other in configuration. They both have very similar values of the internuclear distance

R_{cc} (11.169 Å for **C1:C1** and 11.181 Å for **C2:C2(a)**), of the angles θ_1 and θ_2 (151.0° for **C1:C1** and 153.3° for **C2:C2(a)**), and of the dihedral φ (-179.8° for both pairs, indicating planarity). These values point to an essential isomorphism for the two pairs, and also indicate that both pairs are reversible, having D_{2h} symmetry.

C2 may dimerise in another way due to the presence of another C=O group adjacent to the endocyclic N-H group, yielding the **C2:C2(b)** pair with a configuration quite markedly different from that of **C1:C1** and **C2:C2(a)**. The R_{cc} distance is much shorter (8.623 Å), while the θ_1 and θ_2 values of 96.7 and 138.1° respectively diverge appreciably from each other and from the values in the **C1:C1** and **C2:C2(a)** pairs, indicating non-reversibility of the **C2:C2(b)** pair. Most of all, the near zero value of the dihedral φ in this pair points to a pairing configuration reverse to the other two.

The bases **C1** and **C2** may also pair in a hetero-associative manner to yield the two pairs **C1:C2(a)** and **C1:C2(b)**. These pairs differ markedly from each other in configuration, and also do not approach the configuration of the isomorphic pairs **C1:C1** and **C2:C2(a)**. While **C1:C2(a)** and **C1:C2(b)** are both basically planar, their backbone attachment bonds point in two different ways (φ equals 0.2 and 179.5° respectively). Their R_{cc} values differ as well, being 7.956 and 9.523 Å respectively. The θ_1 and θ_2 markers also differ appreciably in value among them-selves, ranging between 112.7 and 139.6°.

The data thus predicts that the pyrimidine base Set III provides good candidate pairs to act as viable repeat units for a self-associative information-bearing macromolecular duplex. Apart from the ability of this set to furnish two isomorphic dimeric pairs, **C1:C1** and **C2:C2(a)**, the bases **C1** and **C2** cannot pair among themselves within the

configuration emerging out of the isomorphic pairs, nor does the other dimer of **C2** fall within this desired configuration. This suitability of the pyrmidine Set III arises from the presence of the two *different* moieties O=C-NH and N=CH-NH_2, one in each pyrmidine base of this set.

IV.3.4 Pyrimidine-pyrimidine pairs from Set IV

Figure IV.3(b) depicts the three H-bonded pairs possible for the pyrimidine bases **D1** and **D2** of **Set IV**, *viz.*, the self-associative pairs **D1:D1** and **D2:D2** and the hetero-associative pair **D1:D2**. The base pairs of this Set are bonded by H-bonds of both the C=N...H-N and the C=O....H-N types. The **D1:D1** and **D2:D2** pairs are symmetrical dimers, while **D1:D2** is asymmetrical. Pairs **D1:D1** and **D2:D2** have their would-be backbone attachment bonds *trans* to each other, while in the pair **D1:D2** they are *cis*. Note that these three pairs exhaust all possibilities for H-bonded pairing between the monomers **D1** and **D2** within the context of this study.

The B3LYP/6-31G* values of the pairing energies E_p and E_p(cr) (Table IV.3) are generally larger for Set IV than for the other pyrimidine Set III. Here, the E_p values are -15.05, -19.73 and -17.23 kcal/mol, while the E_p(cr) values are -11.56, -16.32 and -14.08 kcal/mol for **D1:D1**, **D2:D2** and **D1:D2** respectively. The single point MP2/6-311++G(d,p)// B3LYP/6-31G* values of E_p(MP2) compare quite well with the BSSE-uncorrected B3LYP/6-31G* optimized values (-13.88, -18.69 and -16.42 kcal/mol for **D1:D1**, **D2:D2** and **D1:D2** respectively). The H-bonds are of the C=N...H-N and C=O...H-N types, and may be regarded as strong, being also fairly short (often shorter than those for Set III) and very linear (see Table IV.4). However, these Set IV pairs are generally less tightly packed than the azole pairs of Sets I and II.

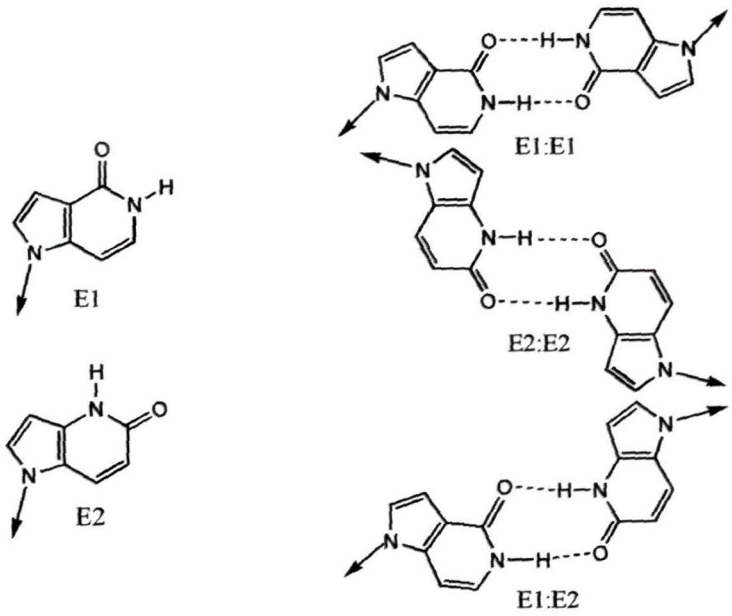
The B3LYP/6-31G* configuration data for the Set IV pairs (Table IV.3) reveal similarities in configuration between the two self-associative pairs **D1:D1** and **D2:D2** in that their θ_1 and θ_2 angles are all similar in value (117.9 to 119.9°) while the dihedral φ is close to 180°. These two pairs are thus planar and reversible. However, their R_{cc} values differ somewhat, being 8.900 and 8.621 Å for **D1:D1** and **D2:D2** respectively. This fairly significant difference of 0.279 Å detracts from true isomorphism for these pairs. Like the azole pairs **B1:B1** and **B2:B2** of Set II, these pairs may be described as only approaching isomorphism, and whether they are acceptable or not would depend on the context. In contrast, the hetero-associative pair **D1:D2** is of a configuration basically reverse to that of **D1:D1** and **D2:D2**, where the φ value approaches zero. The R_{cc} distance of 8.391 Å here is also shorter than that for the two self-associative pairs.

IV.3.5 Fused ring base pairs from Set V

Set V comprises the fused ring bases **E1** and **E2** which form the self-associative pairs **E1:E1** and **E2:E2** and the hetero-associative pair **E1:E2** of Figure IV.4(a), exhausting all the possibilities for H-bonding amongst bases **E1** and **E2** subject to the constraints of this study. Pairs **E1:E1** and **E2:E2** are planar with D_{2h} symmetry, while the **E1:E2** pair is asymmetrical. The base pairs of this Set are bonded together by H-bonds of only the C=O....H-N type, since bases **E1** and **E2** both possess the O=C-NH moiety

All pairs have large B3LYP/6-31G* values of the pairing energies (Table IV.5), where E_p is –21.19, -21.15 and -21.17 kcal/mol, while E_p(cr) equals –17.09, -16.80 and –16.84 kcal/mol for **E1:E1**, **E2:E2** and **E1:E2** respectively. Single point MP2/ 6-311++G(d,p)// B3LYP/6-31G* values of E_p(MP2) compare well with the BSSE-

(e) **Set V** (substituted bicyclic-bicyclic pairs)



(f) **Set VI** (substituted bicyclic-bicyclic pairs)

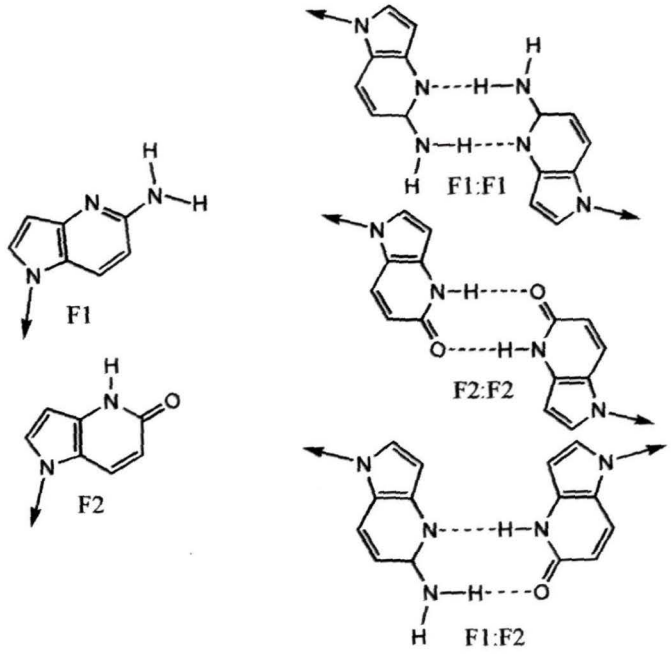


Fig.IV.3 : Component bases and resulting base pairs built out of
(e) the bicyclic-bicyclic base-pairing **Set V,** and
(f) the bicyclic-bicyclic base-pairing **Set VI**

Table IV.5 Uncorrected pairing energy E_p and the BSSE-corrected pairing energy E_p(cr) along with configurational data for the H-bonded base pairs of **Sets V** and **VI** as obtained from B3LYP/6-31G* optimized geometries and energies. Given also are the pairing energies E_p(MP2) without BSSE correction as calculated by the MP2/6-311++G(d,p)// B3LYP/6-31G* strategy [a]

Base Pair	E_p	E_p(cr)	R_{cc}	θ_1	θ_2	φ	E_p(MP2)
Set V *Fused ring pairs*							
E1:E1	-21.19	-17.09	13.653	145.6	145.6	179.3	-20.02
E2:E2	-21.15	-16.80	13.379	155.3	155.3	179.4	-20.71
E1:E2	-21.17	-16.84	13.310	161.3	171.7	0.9	-20.38
Set VI *Fused ring pairs*							
F1:F1	-11.18	-7.96	13.344	153.4	153.4	-121.8	-10.61
F2:F2	-21.15	-16.70	13.378	155.3	155.3	-178.7	-20.71
F1:F2	-15.58	-11.51	11.045	162.3	163.3	-39.8	-16.30

[a] Bond lengths in angstrom; bond angles and dihedrals in degrees; energies in kcal/mol

Table IV.6 H-bond data[a] for the various base pairs of **Sets V** and **VI** (B3LYP/6-31G* geometries) along with the BSSE-corrected pairing energy $E_p(cr)$[b]

Base Pair	H-bond	R_{hb}	R_{xy}	θ_{hb}	$E_p(cr)$
Set V *Fused ring pairs*					
E1:E1	O4-H5	1.777	2.814	178.6	-17.09
	H5-O4	1.777	2.814	178.7	
E2:E2	H4-O4	1.767	2.807	179.1	-16.80
	O5-H4	1.768	2.807	179.1	
E1:E2	O4-H4	1.778	2.816	179.5	-16.84
	H5-O5	1.768	2.806	178.1	
Set VI *Fused ring pairs*					
F1:F1	N4-H5	2.050	3.075	176.2	-7.96
	H5-N4	2.050	3.075	176.2	
F2:F2	H4-O5	1.768	2.808	178.8	-16.70
	O5-H4	1.768	2.808	179.0	
F1:F2	N4-H4	1.933	2.969	176.7	-11.51
	H5-O5	1.882	2.909	178.6	

[a] Bond lengths in angstrom; bond angles in degrees
[b] Pairing energy in kcal/mol

uncorrected B3LYP/6-31G* optimized values, which are –20.02, -20.71 and –20.38 kcal/mol for **E1:E1**, **E2:E2** and **E1:E2** respectively. These pairing energies are larger than those for the pyrimidine pairs of Sets III and IV, often exceeding those of Sets I and II, which may be linked to near-linearity of the H-bonds (Table IV.6) and their relatively small lengths (R_{hb} values from 1.767 to 1.778 Å).

The B3LYP/6-31G* configurations of the self-associative pairs **E1:E1** and **E2:E2** are both more or less planar, the dihedral φ for both pairs being close to 180°. The R_{cc} distances are 13.653 and 13.379 Å respectively, and it is debatable whether 0.274 Å difference is large enough to render Set V unsuitable. The θ_1 and θ_2 angles differ too, being 145.6° and 155.3° respectively for **E1:E1** and **E2:E2**, making the hexagonal H-bond region of **E1:E1** turn by about 20° from that of **E2:E2**. The hetero-associative pair **E1:E2** is planar (φ close to zero), while the R_{cc} distance is quite similar to that for the other two pairs. However, the φ dihedral makes the **E1:E2** configuration reverse to that of the other two pairs. We deem the pairs **E1:E1** and **E2:E2** as significantly different in configuration from each other (and reverse to **E1:E2**), thus marking the fused ring Set V as being not fully suitable for constructing viable repeat units here.

IV.3.6 Fused ring base pairs from Set VI

Figure IV.4(b) portrays the H-bonded base pairs **F1:F1**, **F2:F2** and **F1:F2** built up from the bicyclic fused ring bases **F1** and **F2** of Set VI. The symmetrical dimers **F1:F1** and **F2:F2** have their N-C would-be backbone attachment bonds pointing in opposite directions, while the hetero-associative pair **F1:F2** has these bonds pointing in the same direction. The pairs of this Set utilize H-bonds of both the C=O....H-N and the C=N....H-N types.

The B3LYP/6-31G* values of pairing energies E_p (Table IV.5) are –11.18, -21.15 and –15.58 kcal/mol for **F1:F1**, **F2:F2** and **F1:F2** respectively, while the BSSE-corrected pairing energies E_p(cr) are –7.96. –16.70 and –11.51 kcal/mol for **F1:F1**, **F2:F2** and **F1:F2** respectively. Single point MP2/6-311++G(d,p)/ B3LYP/6-31G* values of E_p(MP2) compare well with the BSSE-uncorrected B3LYP/6-31G* optimized values, being -10.61, -20.71 and -16.30 kcal/mol for the **F1:F1**, **F2:F2** and **F1:F2** pairs respectively. Pairing energies depend on the H-bond types present (Table IV.6), where **F1:F1** with two H-bonds of the C=O....H-N type is more strongly paired than **F2:F2** with 2 H-bonds of the C=N....H-N type, while the pair **F1:F2** with one H-bond of each type has an intermediate value of the pairing energy.

The B3LYP/6-31G* configurations of the self-associative pairs **F1:F1** and **F2:F2** are largely alike except for the dihedral φ. The R_{cc} distances are very close in value to each other (13.344 and 13.378 Å for **F1:F1** and **F2:F2** respectively). The θ_1 and θ_2 angles are very similar in value (153.4 or 155.3°) for both pairs. The dihedral φ is close to 180° for the co-planar pair **F2:F2**, but 121.8° for **F1:F1**, indicating a non-planar configuration for this pair. This non-planarity of **F1:F1** arises from the pyramidal H-atoms of the 4-amino groups in the bases **F1**. This departure from planarity makes it difficult to say whether the bases **F1** and **F2** of Set VI could constitute a suitable set for the construction of good repeat units for the macro-molecular duplex. The hetero-associative pair **F1:F2**, however, has a markedly different configuration from the other two pairs, being distinguished by a shorter R_{cc} value of 11.045 Å, θ_1 and θ_2 values of 162.3 and 163.3° respectively, and most of all by the dihedral φ value of 39.8° which predicts **F1:F2** to be non-planar.

IV.3.7 Establishment of isomorphism for Set III pairs

The B3LYP/6-31G* results reported above for the pairs **C1:C1** and **C2:C2(a)** arising from the chosen Set III bases (see Table IV.3) were tested with results using other *ab initio* strategies, *viz.*, the B3LYP/6-31++G(d,p) and MP2/6-31G(d,p) methods for the sake of comparison. These two strategies may be deemed as more rigorous and reliable than the B3LYP/6-31G* method used here. Results for the pairing energy E_p and configuration data as calculated by the two latter methods are given in Table IV.7 and compared with the B3LYP/6-31G* results. It is apparent that all these three sets of configurational data correspond closely to each other. Although the values of the pairing energies are smaller when calculated by the latter two methods, the three configuration data sets are almost identical. The R_{cc} distances are consistent within 0.021 to 0.050 Å, while the angles θ_1 and θ_2 are consistent within 0.2 to 0.6°. Values of the dihedral φ are all very close to 180°. It emerges that the configurations of the pairs are all highly similar to one another for all the three methods. More importantly, the essential isomorphism of the two pairs remains invariant of the calculation strategy used, which justifies the use of the B3LYP/6-31G* method for all the systems studied here.

IV.3.8 Comparison with other results

The range of values for the interaction energies obtained here for these novel base pairs have been compared above with values obtained using the HF/6-31G**, MP2 and DFT regimes for the whole set of 30 DNA base pairs, and the ranges are very comparable indeed. Our B3LYP/6-31G* results may also be compared with the results of a previous similar study on hetero-associative pairs,[19] as well as with earlier

Table IV.7 Pairing energies and configurational data[a] for the **C1:C1** and **C2:C2(a)** pairs as obtained from B3LYP/6-31++G(d,p) and MP2/6-31G(d,p) optimized geometries and total energies and compared with the B3LYP/6-31G* results

Base Pair	E_p	R_{cc}	θ_1	θ_2	φ
B3LYP/6-31++G(d,p)					
C1:C1	-8.16	11.190	150.8	150.8	-179.8
C2:C2(a)	-7.60	11.234	153.9	153.9	179.9
MP2/6-31G(d,p)					
C1:C1	-6.02	11.081	150.5	150.8	172.3
C2:C2(a)	-5.71	11.141	153.1	153.1	-179.5
B3LYP/6-31G*					
C1:C1	-10.63	11.169	151.0	151.0	179.8
C2:C2(a)	-14.43	11.181	153.3	153.3	179.8

[a] Bond lengths in angstrom; bond angles and dihedrals in degrees; energies in kcal/mol

PM3 SCF-MO work done on analogous systems.[17,18] We note that while the hetero-associative pairs have their would-be N-glycoside bonds either pointing in the same direction or else away from each other, all the viable self-associative pairs of this study have their N-glycoside bonds pointing only in the same direction. Furthermore, the pairing energies E_p are much larger in the B3LYP/6-31G* method than in the PM3 method. The B3LYP values for E_p range from -10.63 to -22.53 kcal/mol, while the PM3 values range from only -0.68 to -5.11 kcal/mol. Pyramidalisation of the exo-cyclic amino groups, especially in solitary bases, is reproduced well in the B3LYP studies, but not by the PM3 method. Other higher level calculations testify to the same. In view of the much greater sophistication and accuracy of the *ab initio* B3LYP /6-31G* strategy over the semi-empirical PM3 SCF-MO method, the results arising from the former are more to be trusted in this context, falling in line with the results of other higher level calculations by various workers.

IV.3.9 Information encoded

The two H-bonded base pairs arising from Set III imply an informational content of two bits per pair. We may envisage a situation where the self-associative duplex encodes information on a three-to-one basis as occurs in nature between DNA and proteins. Having three base pairs per codon allows for a total of 2^3 or 8 encrypted species to be encoded. In contrast to this situation for a self-associative duplex, a hetero-associative duplex with the same number of distinct pairs (namely, two) would encode for 64 encrypted species. A hetero-associative duplex like DNA is thus much better suited to encode information, with more information content per given length than a corresponding self-associative duplex.

IV.4 Conclusions

1. Facility of H-bonded pairing among the bases studied is linked to the number and type of H-bonds involved, as well as to the geometry of the H-bonds.
2. All the self-associative pairs studied here are of the reverse type, having their would-be glycoside bonds *trans* to each other, unlike the hetero-associative pairs studied earlier.
3. Existence of the same type of difunctional H-bonding moiety in both bases of a set does not allow for isomorphism in the resultant self-associative pairs.
4. Only one set (**Set III**) comprised of the pyrimidine bases **C1** and **C2** is predicted as suitable for furnishing viable repeat units for an information-bearing macroduplex.
5. Such a self-associative duplex would be much less efficient in information-bearing capacity than a corresponding hetero-associative duplex.

References

1. Watson, J. D; Crick, F. H. C. *Nature*, 1953, **171**, 737.

2. Sponer, J; Jurecka, P; Hobza, P. In *Computational Studies of DNA and RNA*, Sponer, J; Lankas, F. (eds.), Springer, Dordrecht, 2006, pp.343-388.

3. Sponer, J; Hobza, P; Leszczynski, J. In *Computational Molecular Biology*, Leszczynski, L. (ed), Elsevier Science, 1999; Chapter 3.

4. Leszczynski, J (ed). Computational Molecular Biology, Elsevier Science, 1999.

5. Sponer, J; Hobza, P. *Collec Czechoslovak Chem Comm*, 2003, **68**, 2231.

6. Dabkowska, I; Jurecka, P; Hobza, P. *J Chem Phys*, 2005, **122**, 204322.

7. Sponer, J; Jurecka, P; Hobza, P. *J Am Chem Soc*, 2004, **126**, 10142.

8. Sponer, J; Leszczynski, J; Hobza, P. *Biopolymers*, 2001, **61**, 3.

9. Lind, M. C; Bera, P. P; Richardson, N. A; Wheeler, S. E; Schaefer III, H. F. *Proc Natl Acad Sci USA*, 2006, **103**, 7554.

10. Bera, P. P; Schaefer III, H. F. *Proc Natl Acad Sci USA*, 2005, **102**, 6698.

11. Richardson, N.A; Wesolowski, S.S; Schaefer III, H. F. *J Am Chem Soc*, 2002, **124**, 10163.

12. Venkateswarlu, D; Lyngdoh, R. H. D. *J Chem Soc Perkin Trans 2*, 1995, 839.

13. Venkateswarlu, D; Lyngdoh, R. H. D; Bansal, M. *J Chem Soc Perkin Trans 2*, 1997, **2**, 621.

14. Schlonvogg, I; Pitsch, S; Lesueur, C; Eschenmoser, A. *Helvet Chim Acta*, 1996, **79**, 2316.

15. Wipro, H; Kudick, R. A; Krishnamurthy, R; Eschenmoser, A. *Helvet Chim Acta*, 2001, **9**, 2411.

16. Jaun, B; Eschenmoser, A. *Helvet Chim Acta*, 2003, **84**, 1778.

17. Buam, D. M. L; Lyngdoh, R. H. D. *J Mol Struct (THEOCHEM)*, 2000, **505**, 149.

18. Buam, D. M. L; Lyngdoh, R. H. D. *Indian J Chem*, 2002, **41B**, 2346.

19. Neihsial, S; Lyngdoh, R. H. D. *J Mol Struct* (THEOCHEM), 2007, **806**, 213.

20. Guerra, C. F; Bickelhaupt, F. M; Snijders, J. G; Baerends, E. J. *J Am Chem Soc*, 2000, **122**, 4117.

21. Crick, F. H. C; Watson, J. D. *Proc Roy Soc (London) Ser A*, 1954, **223**, 80.

22. Seeman, N. C; Rosenberg, J. M; Suddath, F. L; Kim, J. J. P; Rich, A. *J Mol Biol*, 1976, **104**, 109.

23. Rosenberg, J. M; Seeman, N. C; Day, R. O; Rich, R. A. *J Mol Biol*, 1976, **104**, 145

24. Pauling, L; Corey, R. B. *Archiv Biochem Biophys*, 1956, **65**,164.

25. Becke, A. D. *Phys Rev B*, 1998, **38**, 3093.

26. Becke, A. D. *J Chem Phys*, 1993, **98**, 5648.

27. Lee, C; Yang, W; Parr, R. G. *Phys Rev B*, 1998, **37**, 785.

28. Boys, S. F; Bernardi, F. *Mol Phys*, 1985, **19**, 553.

29. Frisch, M. J; Trucks, G W; Schlegel, H.B; Scuseria, G.E; Robb, M.A; Cheeseman, J. R; Montgomery Jr, J. A; Vreven, T; Kudin, K. N; Burant, J. C;Millam, J. M; Iyengar, S. S; Tomasi, J; Barone, V; Mennucci, B; Cossi, M; Scalmani, G; Rega, N; Petersson, G. A; Nakatsuji, H; Hada, M; Ehara, M; Toyota, K; Fukuda, R; Hasega-wa, J; Ishida, M; Nakajima, T; Honda, Y; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V; Adamo, C; Jaramillo, J; Gomperts, R; Stratmann, R. E; Yazyev, O; Austin, A. J; Cammi, R; Pomelli, C; Ochterksi, J. W; Ayala, P. Y; Morokuma, K; Voth, G. A; Salvador, P; Dannenberg,

J. J; Zakrzewski, V. G; Dapprich, S; Daniels, A. D; Strain, M. C; Farkas, O; Malick, D. K; Rabuck, A. D; Raghavachari, K; Foresman, J. B; Ortiz, J. V; Cui, Q; Baboul, A. G; Clifford, S; Cioslowski, J; Stefanov, B. B; Liu, G; Liashenko, A; Piskorz, P; Komaromi, I; Martin, R. L; Fox, D. J; Keith, T; Al-Laham, M. A; Peng, C. Y; Nanayakkara, A; Challacombe, M; Gill, P. M. W; Johnson, B; Chen, W; Wong, M. W; Gonzalez, C; Pople, J. A. Gaussian 03 Revision A.1, Gaussian, Inc., Pittsburgh PA, 2003.

An electron sitting in a prison asked a second electron cellmate, "What are you in for?" To which the latter replied, "For attempting a forbidden transition."

A tidy laboratory means a lazy chemist.

CHAPTER FIVE

TOWARDS H-BONDED DUPLEXES WITH PYRANOSE PHOSPHATE AND POLYAMIDE BACKBONES

V.1 Chosen DNA Base Mimic Set

The stage is now set for the design and theoretical construction of the complete repeat units for the novel information-bearing macromolecular duplexes which are the goal of the studies embodied in this Dissertation. By repeat unit here is meant not the solitary base or base pair, but the whole structure including backbone which is repeated indefinitely along the duplex. For the case of DNA, the full repeat unit is the nucleotide (2'-dexoribonucleoside 3',5'-diphosphate). Such single-strand repeat units are comprised of derivatives of the bases contained within the chosen DNA base mimic set, where a polymeric backbone is used to attach the bases.

The choice of the DNA base mimic set for the study of this Chapter arises out of the studies of Chapter Three on hetero-associative base pairs arising out of various sets of nitrogenous bases including substituted pyrimidines, pyridines and pyrazines. Out of the 10 hetero-associative base pairs studied, choice was made of the two complementary base pairs arising out of a DNA mimic base set of four substituted pyrimidines **A1**, **A2**, **A3** and **A4**, where **A1** is pyrimid-2,4-dione, **A2** is 4,6-diaminopyrimid-2-one, **A3** is 4-aminopyrimid-2,6-dione, and **A4** is 4-aminopyrimid-2-one (see Figs. III.1 and III.2). These four bases can undergo complementary H-bonded base-pairing to yield only the *two unique pairs* **A1:A2** and **A3:A4** within the specified configurational constraint for pairing, and *no other* pairs at all.

V.2 Pyranose Phosphate and Polyamide Backbones

Since H-bonded base pairs cannot be correctly aligned in a specified sequence without a backbone to connect them together, DNA employs a 2'-deoxyribose-3',5'-diphosphate polymeric backbone for this purpose. The backbone is necessary for the information-bearing property of the DNA macroduplex since a properly defined sequence of bases or base pairs cannot exist without it. For our synthetic H-bonded macromolecular duplexes built up from the chosen DNA base mimic set, we propose *two* different structures for the backbone in the macromolecular duplex – a *sugar phosphate* backbone and a *polyamide* backbone.

V.2.1 Sugar phosphate backbone

The first type of backbone proposed is related to the original prototype present in DNA, being of the sugar phosphate type. We propose a 2'-deoxy-β-allopyranose-3',6'-diphosphate moiety as the repeat unit for the backbone, where the base is attached at the 5'-carbon of the sugar. The complete repeat unit including base is thus a *pyranonucleotide*, where the five-membered sugar ring of DNA is replaced by the six-membered allopyranose ring here, and the 4 DNA bases are replaced by the bases of the novel mimic set devised here in Chapter Three. Here, the phosphate groups are left unionized, since this would lead to difficulties in performing DFT calculations on the dianion species resulting from H-bonded pairing of the two units, for the reasons outlined in Chapter Two Section II.4.

Since the mimic set consists of 4 bases, 4 different repeat units can be built up. These four repeat units (nucleotides) are built up from the four bases of the mimic set (**A1**, **A2**, **A3** and **A4**). The four nucleotides thus emerge as: *(a)* (pyrimid-2,4-dione)-

2'-deoxy-β-allopyranose 3',6'-diphosphate, abbreviated as Py**A1**, *(b)* (4,6-diamino-pyrimid-2-one)-2'-deoxy-β-allopyranose 3',6'-diphosphate or Py**A2**, *(c)* (4-amino-pyrimid-2,6-dione)-2'-deoxy-β-allo-pyranose 3',6'-diphosphate or Py**A3**, and lastly *(d)* (4-aminopyrimid-2-one)-2'-deoxy-β-allo-pyranose 3',6'-diphosphate or Py**A4**. Note that **A4** of Py**A4** is actually the nucleic acid base cytosine.

The four novel nucleotides Py**A1**, Py**A2**, Py**A3** and Py**A4** are portrayed in Fig. V.1, where each pyranonucleotide is drawn as a schematic valence-bond structure. Note the pyranose ring is in the standard chair conformation, where the base moiety is bonded to the C1'-atom and occupies an equatorial position. As such, the plane of the base would be expected to be approximately perpendicular to the plane bisecting the six-membered ring through the C1' and C4' atoms, and the N1, C1' and C4' atoms may be expected to be roughly collinear. The C3' phosphate group occupies an axial position, while the C6' phosphate group is bonded to the exocyclic and equatorial CH_2 group. The phosphate group conformations are aligned such as to avoid steric crowding as much as possible, being directed away from the pyranose ring.

Note that the repeat units are represented here as *dinucleotides*, (not mono-nucleotides – see below) with *two* phosphate groups attached to the C3' and C6' atoms of the pyranose. This is done to make the environment of the central part (sugar plus base) more similar to the actual situation in the macromolecular single stranded polymer. In such a situation, phosphate group moieties are present on both the C3' and C6' ends of each pyranose ring unit along the chain, except for those situated at the two extremities of the entire polymer (which may be called as the 3' and the 6' termini, analogous to the 3' and 5' ends of DNA single strands).

Schematic Representation

Pyranonucleotide repeats units

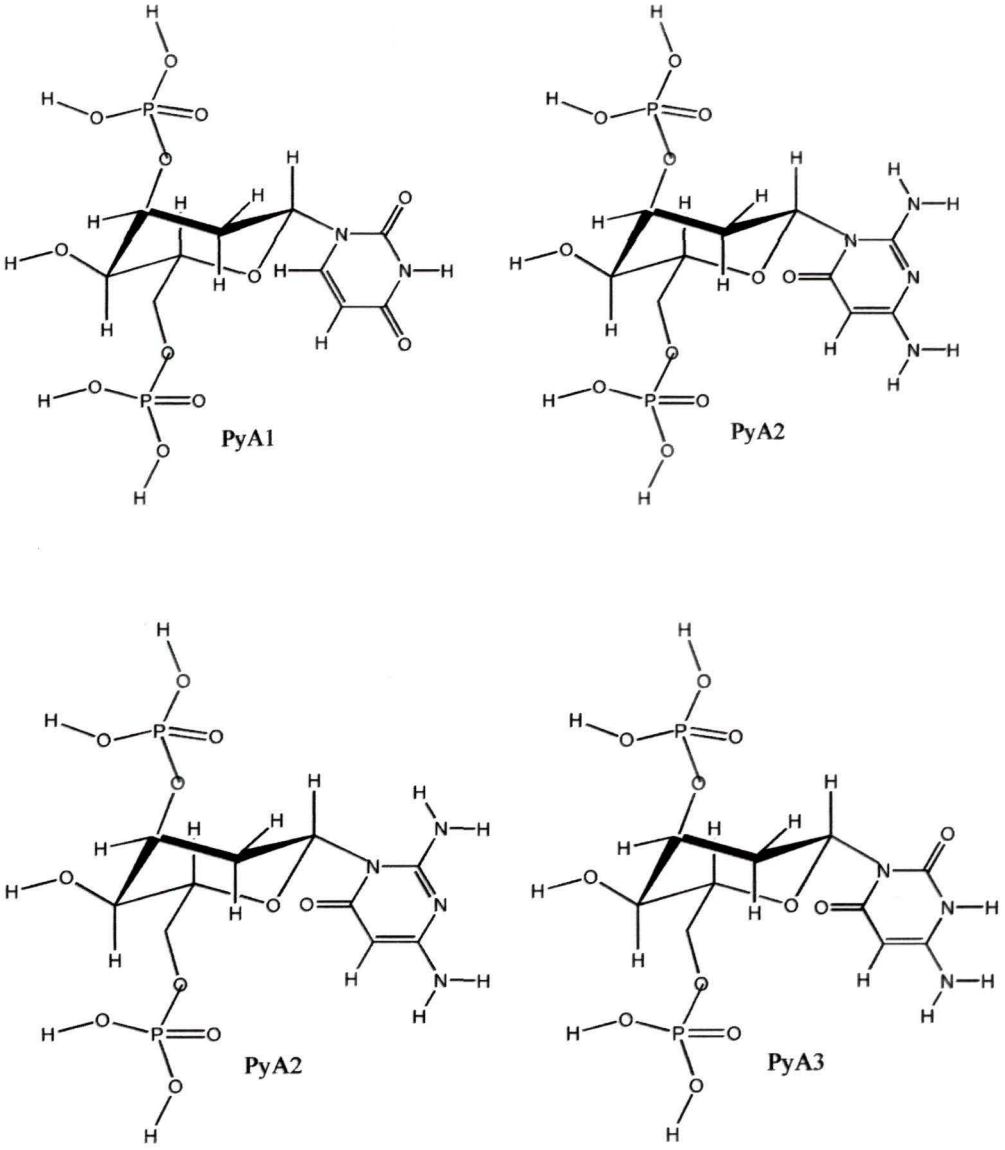


Fig. V. 1 : Schematic representation of the pyranonucleotide repeat units **PyA1, PyA2, PyA3** and **PyA4**

In fact, the actual repeat unit should be considered as a *mononucleotide*, with a single phosphate group at either the 3' and or the 6' end, but not at both. This is the case with DNA and RNA, where the repeat units are actually mononucleotides. Now, for our putative macro-molecular single strand, if the repeat unit is conceived of as a 3'-mononucleotide, then the 6' atom of the sugar ring should bear a hydroxyl group, and vice versa if the 6'-mononucleotide is taken as the repeat unit. If it is 3'-mononucleotides that are taken as the repeat units, the macropolymer strand may be built up successive formation of new O-P bonds between the 6'-hydroxy oxygen atom of the incoming 3'-nucleotide and the phosphorus atom of the 3'-phosphate group belonging to the previously attached nucleotide moiety.

This choice of a pyranose phosphate backbone takes the cue from the pioneer experimental work of Eschenmoser's group on various sugar phosphate backbones, where different sugars had been worked out to yield oligomeric structures for the H-bonded polymeric duplex [1,2,3].

V.2.2 Polyamide backbone

We now take a drastic step away from the known world of nature, making a radical break from previous ideas on primary information-bearing macromolecules. The second type of backbone proposed here may be called a *polyamide* backbone because the primary structure of the single-stranded macropolymer containing it takes the form *poly*(-CH_2-CH**B**-CO-NH-) or (-CH_2-CH**B**-CO-NH-)$_n$ with **B** as the base moiety. This structure has some resemblance to a polypeptide or protein which has the form *poly*(-CH**A**-CO-NH-) or (-CH**A**-CO-NH-)$_n$, where **A** is an amino acid residue. Now, there is a problem if we adopt the corresponding structure *poly*(-CH**B**-CO-NH-) or

(-CH**B**-CO-NH-)$_n$ as the single-stranded component for our H-bonded double-stranded information-bearing macromolecular duplex (where **B** is an H-bonding base moiety). This is because successive or adjacent base moieties in the single-stranded *poly*(-CH**B**-CO-NH-) chain would be situated on *opposite* sides of the chain if the chain adopts the facile zig-zag conformation. So all the base moieties **B** within the chain will not be able to pair through H-bonding on one side; only half of them will be able to do so. This will result in a waste of base moieties, since only half of them will contribute to the information content of the H-bonded macroduplex.

However, if we adopt the structure *poly*(-CH_2-CH**B**-CO-NH-) or (-CH_2-CH**B**-CO-NH-)$_n$ for our single-stranded macropolymer, successive or adjacent base moieties **B** within the chain will all be situated on the *same* side of the chain when the chain assumes the usual zig-zag conformation. This means all the bases in one strand will be able to participate in H-bonding with the opposite strand, and none of them will be wasted as far as informational content is concerned. For such a single-stranded polymer, it is possible to envisage repeat units of the form HO-CH_2-CH**B**-CO-NH_2 where covalent bonding between one repeat unit and the next occurs by formation of a new N-C bond through elimination of water and creation of a dimer as shown below:

HO-CH_2-CH**B**-CO-NH_2 + HO-CH_2-CH**B**-CO-NH_2 $\rightarrow$

HO-CH_2-CH**B**-CO-NH--CH_2-CH**B**-CO-NH_2 + H_2O

It may be expected that an appropriate dehydrating agent could be used to facilitate this process, which (*in vitro* at least) would ordinarily be a simple SN2 reaction. For instance, DCC (dicyclohexylcarbodiimide) is used for creating peptide bonds between

amino acids in the synthesis of polypeptides. The above condensation process can be repeated to make a single-stranded polymer of desired length, size and base sequence.

The species HO-CH_2-CH**B**-CO-NH_2 is the most fundamental repeat unit for a macro-polymer of this type, and may be termed as a β-hydroxy-α-(pyrimid-1-yl)propionamide in general, which may be abbreviated here as *pyrimidylamide*. Here, the base **B** may be any of the four pyrimidine bases of the chosen DNA base mimic set. Ultimately, the polymer will have two distinct ends – the OH end and the NH_2 end - which may be compared with the 3'-phosphate and 5'-phosphate ends of natural DNA. The base sequence of the single-stranded macropolymer may be read from either end, depending on which convention is taken as normative. We propose here that the sequence be read from the OH end to the NH_2 end

If the backbone is considered *without* the bases, it takes the form (-CH_2-CHOH-CO-NH-)$_n$ or *poly*(-CH_2-CHOH-CO-NH-). The OH group on the second carbon atom of the repeat moiety -CH_2-CHOH-CO-NH- may be removed by an SN2 attack of the 1-nitrogen atom of the pyrimidine bases of the mimic set in order to attach the base to the polyamide backbone. If the macropolymer is conceptualized in this manner, the bases may be successively affixed to the already synthesized bare polymeric backbone by a series of such SN2 reactions. This, however, is not the way in which DNA is synthesized in nature or in vitro. In nature, the DNA chain is built up by successive polymerization of incoming *nucleotides* (not bases) as the chain grows, and there is no pre-existing bare sugar-phosphate backbone to which solitary bases may affix themselves one after the other. So likewise here, the polyamide backbone in itself may not be expected to exist simply on its own prior to formation of the base-

attached single stranded polymer. The single stranded polymer may rather be built up by the successive attachment of pyrimidylamide units, one by one, through successive formation of N-C bonds. The new N-C bond forms through SN2 attack of the amide nitrogen atom in the incoming repeat unit on to the C2 carbon atom of the previously attached repeat unit, along with departure of the C2-hydroxy group as water.

The smallest repeat unit of this single-stranded polymer is thus the HO-CH_2-CH**B**-CO-NH_2 species. Fig. V.2 depicts the four different pyrimidylamide repeat units built up from the four bases of the chosen mimic set. These are *(a)* β-hydroxy-α-(pyrimid-2,4-dione-1-yl)-propion-amide or Pa**A1**, *(b)* β-hydroxy-α-(4,6-diaminopyrimid-2-one-1-yl)propionamide or Pa**A2**, *(c)* β-hydroxy-α-(4-aminopyrimid-2,6-dione-1-yl)propionamide or Pa**A3**, and *(d)* β-hydroxy-α-(4-aminopyrimid-2-one-1-yl)propionamide or Pa**A4**. These portrayals are not the fully optimized structures calculated here, though.

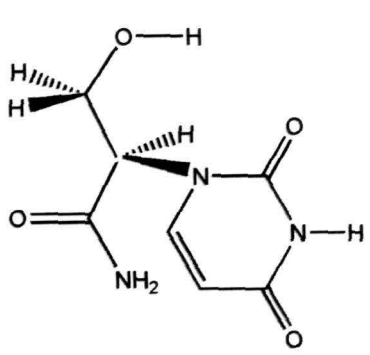
V.2.3 H-bonded repeat unit pairs

Having defined the complete repeat units for the single-stranded polymers with the sugar phosphate and polyamide backbones, the stage is set for H-bonded pairing between such units to construct double-stranded units as described below:

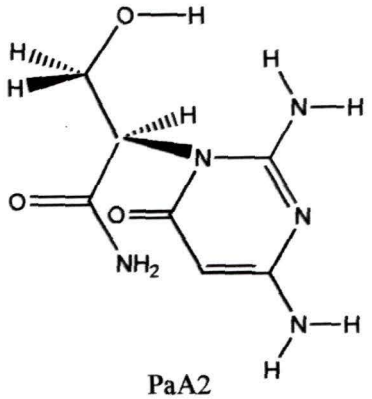
Sugar phosphate species. The four units designated as Py**A1**, Py**A2**, Py**A3** and Py**A4** (see above) may pair among themselves through H-bonding between the pyrimidine moieties to yield only the two unique pairs Py**A1**:A2**Py** and Py**A3**:**A4**Py. These two unique H-bonded nucleotide pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py are schematically portrayed in Fig. V.3. Here the pyranose phosphate moiety Py for the second nucleotide in each pair is written *after* the base moiety to signify that the sugar

Schematic Representation

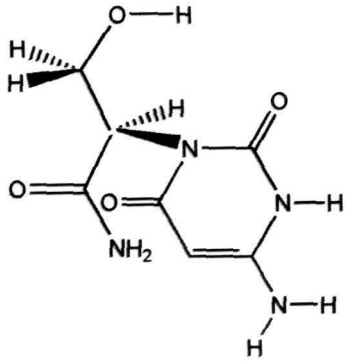
Pyrimidylamide repeat units



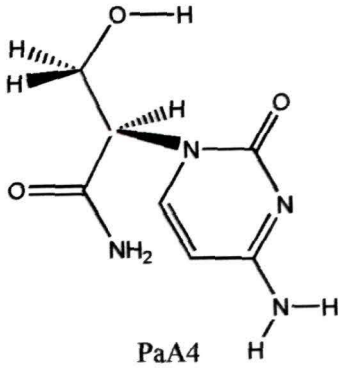
PaA1



PaA2



PaA3

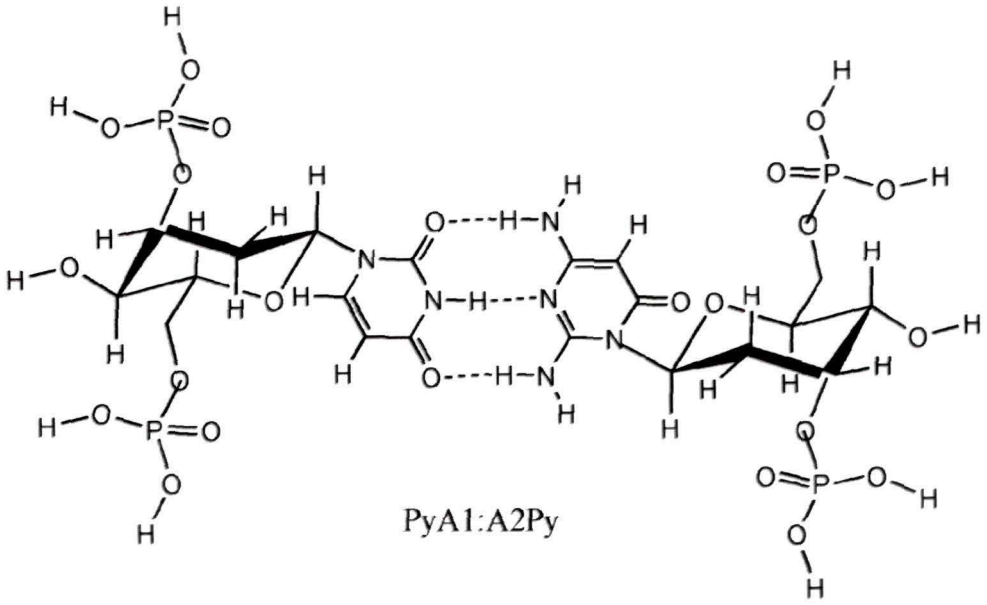


PaA4

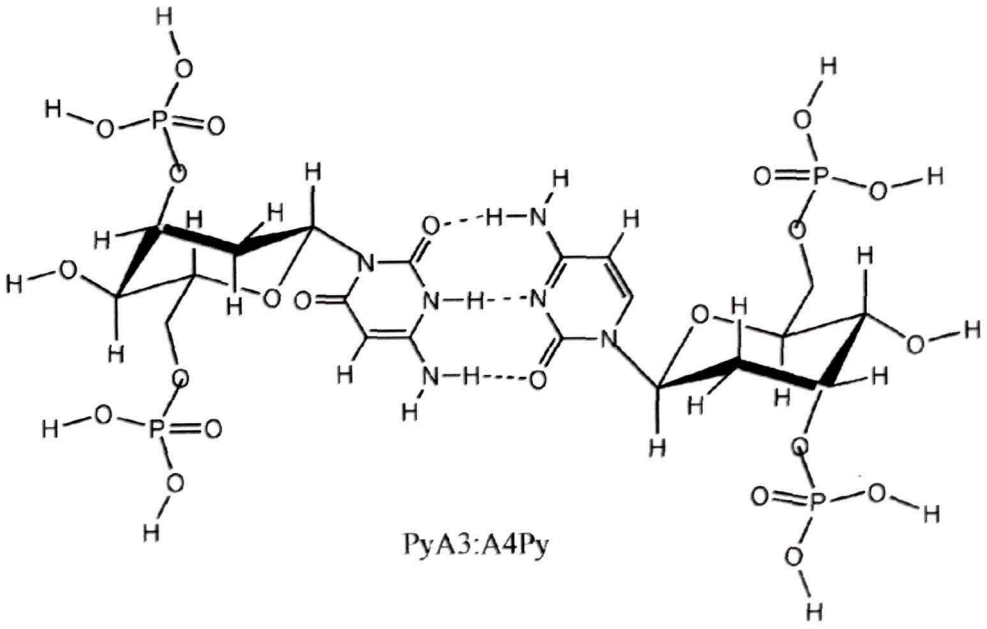
Fig. V. 2 : Schematic representation of the pyrimidylamide repeat units **PaA1, PaA2, PaA3** and **PaA4**

Schematic Representation

Pyranonucleotide Pairs



PyA1:A2Py



PyA3:A4Py

Fig. V. 3 : Schematic representation of the two pyranonucleotide H-bonded pairs **PyA1:A2Py** and **PyA3:A4Py**

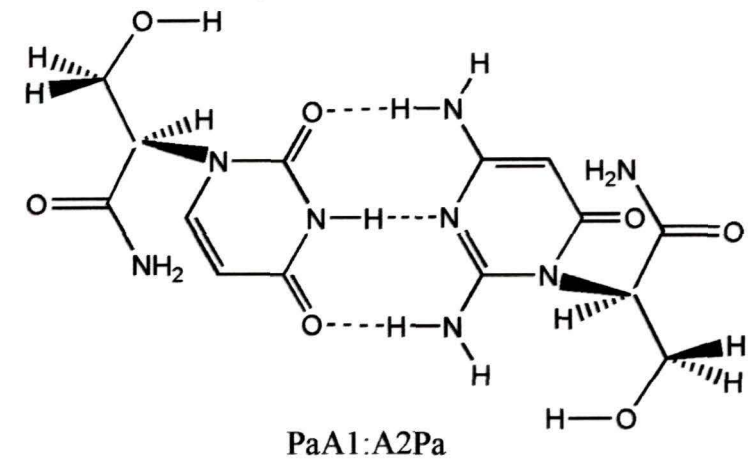
phosphate part does not take part in H-bonding. From Fig. V.3, it is clear that we have adopted an *antiparallel* pairing scheme for these two pairs, which would become more immediately obvious if two polynucleotide *chains* were involved in H-bonded pairing instead of just two single nucleotides. The middle point of each nucleotide pair ((midway between the two N3 atoms of the two base moieties) acts as a sort of centre of inversion, especially for the various atoms and groups involved in the sugar phosphate moieties. This means that, in the polymeric duplex, one polynucleotide single strand would run from the C3' end to the C6' end, and the other strand runs from the C6' end to the C3' end. This antiparallel H-bonded pairing pattern is found in DNA and in the doubly-stranded regions of RNA, as well as in the triplet pairs involved in codon-anticodon pairing between messenger RNA and transfer RNA.

Polyamide species. The fundamental units designated as Pa**A1**, Pa**A2**, Pa**A3** and Pa**A4** (see above) can pair among themselves through H-bonding between the pyrimidine moieties to yield only the two unique pairs Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa, with the amido moiety Pa for the second unit in each pair written after the base moiety. These two unique H-bonded pairs Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa are schematically portrayed in Fig. V.4, which makes it clear that, here too, an antiparallel type of arrangement is adopted for pairing between the two sides. This would become more apparent when oligomeric strands and not single units are involved in the H-bonded duplex, similar to the situation in natural DNA.

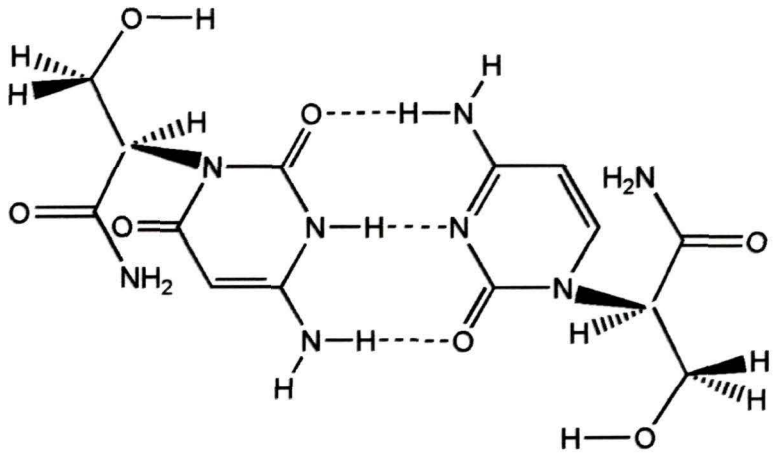
Isomorphism. The anticipation here is that, as was the case for the *solitary base* pairs, the two H-bonded pairs involving *complete repeat units* in each case would also be isomorphic, having closely similar alignments or configurations. The H-bonded

Schematic Representation

Pyrimidylamide Pairs



PaA1:A2Pa



PaA3:A4Pa

Fig. V. 4 : Schematic representation of the two pyrimidylamide H-bonded pairs **PaA1:A2Pa** and **PaA3:A4Pa**

pairing configuration, in fact, is expected to be more or less the same for the solitary base pairs, for the pyranose nucleotide pairs and for the pyrimidylamide pairs as well. If this is borne out by the calculations of this Chapter, it would serve to establish that the pairing configuration is largely independent of the backbone (whether sugar phosphate or polyamide) or the lack of it.

V.3 Methodology

The B3LYP density functional theory model[4,5,6] was used with the 6-31G(d,p) basis set for all molecular species (single units and H-bonded pairs) with full optimization of geometry. Preliminary PM3 SCF-MO calculations provided starting geometries for the B3LYP/6-31G(d,p) optimization. Species treated thus incorporated (a) the four nucleotides Py**A1**, Py**A2**, Py**A3** and Py**A4** along with their H-bonded pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py, and (b) the four pyrimidylamides Pa**A1**, Pa**A2**, Pa**A3** and Pa**A4** along with their H-bonded pairs Pa**A1**:**A2**Py and Pa**A3**:**A4**Pa. All these species were treated as neutral singlet ground states, with no ionization of the monomeric units, so that the H-bonded pair is also a neutral singlet ground state.

A higher level study was also made of the solitary bases **A1**, **A2**, **A3** and **A4** along with their H-bonded base pairs **A1**:**A2** and **A3**:**A4** using the MP2 method with the 6-311++G(d,p) basis set and full optimization of molecular geometry. This is to augment the B3LYP calculations done on these simple systems as presented in Chapter Three, and to provide more accurate estimates of their pairing energies and pairing configurations. Data on these systems may also be compared with corresponding data for the more complex, back-boned systems dealt with in this Chapter.

For all these systems, their status as true minima on the potential energy surface was confirmed by vibrational frequency analyses[7,8] which yielded only positive eigenvalues for the force constant matrix in each case. The extra calculations involved in vibrational frequency analysis were also implemented to (a) evaluate the zero point vibrational energy (ZPVE) correction to the total energy (appropriately scaled), and also (b) to express energy quantities in terms of the Gibb's free energies, which include the entropy term ΔS as calculated using the partition functions corresponding to various degrees of freedom. All calculations were carried out using the GAUSSIAN 2003 suite of DFT and MP2 programs[9].

V.3.1 Estimates of H-bonded pairing facility

Unlike in Chapters Three and Four, pairing facility for the H-bonded pairs studied here is estimated not only in terms of the simple pairing energy E_p as calculated from the total electronic energies taken as such. Nor is any use made of the BSSE-corrected energies, as was done for the base pairing energies of Chapter Four. Pairing facility for each H-bonded system was determined in terms of *(a)* the simple pairing energy E_p (an enthalpy term) as calculated from the difference in total energies between the pair and the sum of the component units each calculated in isolation, *(b)* the ZPVE-corrected pairing energy E_p(ZP), using a scaling factor of 0.9821 prescribed for the 6-31G(d,p) basis set and a scaling factor of 0.9897 for the 6-311++G(d,p) basis set, and *(c)* the free energy change ΔG_p for pairing between the component units in each case. The free energy change is, in fact, the real thermo-dynamic criterion for pairing stability, being directly related to the equilibrium constant for the reaction. Each of these three types of pairing energy was calculated from the appropriate energy of the

H-bonded pair minus the sum of the appropriate energies of the two units of the pair (each calculated independently in isolation).

V.3.2 Descriptors of H-bonded pairing configuration

For each of the H-bonded pairs between the nucleotide units and between the pyrimidyl-amide units, this study adopted the following descriptors of pairing configuration, which correspond to those commonly used for DNA and RNA base pairs and also used in Chapters Three and Four for the various H-bonded base pairs:

1. The distance R_{cc} between two carbon atoms of the backbone which are bonded to the N1-atoms of the base moieties, including the distance between *(a)* the two C1'-carbon atoms of the two β-allopyranose rings of the pyranose nucleotide units within the Py**A1**:**A2**Py and Py**A3**:**A4**Py pairs, and *(b)* the two α-carbon atoms Cα of the amido moieties of the pyrimidylamide units within the Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa pairs.
2. The angles θ_1 and θ_2 spanning *(a)* the C1'-N1-N1 and C1'-N1-N1 atoms in case of the pyranonucleotide systems, and *(b)* the Cα-N1-N1 and Cα-N1-N1 atoms for the pyrimidylamide systems; here N1 and N1 are the nitrogen termini of the respective two base moieties in the pair which are bonded to the backbones.
3. The dihedral angle φ which spans the C1'-N1-N1-C1' atoms for the pyranonucleotide systems and the Cα-N1-N1-Cα atoms for the pyrimidylamide systems, serving as an indicator of co-planarity (or lack of it) between the two base moieties within the pair.

For the *solitary* base pairs **A1**:**A2** and **A3**:**A4**, the descriptors of configuration are similar to the above, except that the C1'-atoms of the nucleotide pairs and the Cα-atoms of the pyrimidylamide pairs are simply replaced by the carbon atoms C of the methyl groups attached to the N1-atoms of the base moieties within the pairs. This

follows the system used for the solitary base pairs of Chapters Three and Four. The symbols for the configurational markers thus remain the same, *viz.*, R_{cc}, θ_1, θ_2 and φ.

V.3.3 H-bond geometry

A hydrogen bond within a base pair is represented as **X...H-Y** or **X-H...Y**, where **X** and **Y** are electronegative atoms belonging to two different component base moieties. The geometry around the hydrogen bonds was studied using the following determinants:

1. The length R_{hb} of the actual hydrogen bond **X...H** or **H...Y**.
2. The total length R_{xy} between the two electronegative atoms **X** and **Y**.
3. The hydrogen bond angle θ_{hb} of the moiety **X...H-Y** or **X-H...Y**.

The convention here is that the atom **X** belongs to the base on the left, while atom **Y** belongs to the base on the right, each atom being numbered as per the conventions of heterocyclic chemistry. The hydrogen atom in the middle is numbered as per the atom **X** or **Y** to which it is bonded covalently.

V.4 Energetic and Structural Aspects

This Section presents the results and discusses the findings related to pairing facility, pairing configuration and H-bond geometry of (a) the two *solitary* base pairs **A1**:**A2** and **A3**:**A4**, and (b) the two *pyranonucleotide* pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py, and (c) the two *pyrimidyl-amide* pairs Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa. Following this, a survey is made of the *charge transfers* occurring during H-bonding where the charge distributions of the various species are calculated by the various methods described in Chapter Two Section II.10. This emphasis on charge here arises from the observation

that the stabilization of H-bonds is due primarily to the electrostatic or columbic component of the net H-bonded interaction energy.

V.4.1 Solitary base pairs

The 4 component bases of the chosen DNA base mimic set have been portrayed in Fig. III.1 of Chapter Three, along with the unique and isomorphic base pairs **A1:A2** and **A3:A4** that they form (Fig. III.2). Table V.1 presents the MP2/6-311++G(d,p) optimized values of the pairing energy E_p, for the solitary base pairs **A1:A2** and **A3:A4**. Table V.1 further presents the optimized values of the three markers of pairing configuration for these base pairs, *viz.*, the R_{cc}, θ_1, θ_2 and φ descriptors. Table V.2 then gives data about the geometry of the three H-bonds present in each of these two base pairs.

Pairing facility. The data of Table V.1 indicates that H-bonded pairing is more facile for the **A3:A4** pair than for the **A1:A2** pair, where the E_p value for **A3:A4** (-26.95 kcal/mol) is noticeably larger than that for **A1:A2** (-18.86 kcal/mol). This observation based on MP2 results well reproduces the B3LYP findings of Chapter Three (Table III.1). Although these MP2 values for E_p are smaller than the corresponding B3LYP values of Chapter Three, they are presumably more accurate. Since both pairs have the same number of H-bonds and the same number of each type, the difference in H-bonding facility between the two pairs may be difficult to rationalize on the basis of H-bonding alone.

Pairing configuration. The MP2/6-311++G(d,p) optimized values of the three markers of pairing configuration (Table V.1) for **A1:A2** and **A3:A4** show that these solitary base pairs may indeed be described as *isomorphic*, as was also indicated

Table V.1 Various estimates of pairing facility and values of configurational markers for the solitary base pairs **A1:A2** and **A3:A4** as obtained from MP2/6-311++G(d,p) optimized geometries[a]

Base pair	E_p	R_{cc}	θ_1	θ_2	φ
A1:A2	-18.86	9.625	136.6	135.9	-156.7
A3:A4	-26.95	9.581	135.8	135.2	-161.4

[a] Interatomic distances in angstrom; bond angles and dihedrals in degrees; pairing energies in kcal/mol

Table V.2 H-bond geometry data[a] for the simple base pairs **A1:A2** and **A3:A4** as calculated from the fully optimized MP2/6-311++G(d,p) geometries

Pair	*H-bond*	R_{hb}	R_{ab}	θ_{hb}
A1:A2	O2....H4-N4	1.984	2.999	176.6
	N3-H3....N3	1.837	2.883	177.9
	O4....H2-O2	1.897	2.919	178.5
A3:A4	O2....H4-N4	1.814	2.841	176.3
	N3-H3....N3	1.864	2.897	176.4
	N4-H4....O2	1.920	2.940	176.9

[a] Bond lengths in angstrom; bond angles in degrees

earlier by the results of lower level B3LYP calculations on these base pairs in Chapter Three (Table III.1). The R_{cc} values differ by only 0.044 Å, where the value for **A1:A2** is slightly larger than that for **A3:A4**. This could mean that the bases **A1** and **A2** within the **A1:A2** pair are somewhat less compactly held together than the bases **A3** and **A4** in the **A3:A4** pair. The H-bond lengths (discussed below) also point to this. This observation is reflected in the prediction that the **A3:A4** pair has a larger pairing energy than the **A1:A2** pair (see above).

The θ_1 and θ_2 angles for both pairs all fall within the narrow range of 135.2 to 136.7°, much the same as the range obtained (135.7 to 137.8°) in the B3LYP calculations of Chapter Three, here again indicating that both pairs are interchangeable like the DNA base pairs. Regarding the dihedral φ, the MP2 results point to some departure from co-planarity of the two bases in each pair. The φ values of -156.7 and -161.4° obtained here for **A1:A2** and **A3:A4** respectively diverge further from 180° than do the corresponding φ values of 173.5 and -171.3° obtained in the previous B3LYP calculations. Nevertheless, these MP2 values are still quite close to each other, differing by only 3.7°. The overall MP2 data on pairing configuration thus indicates essential *isomorphism* for the two base pairs **A1:A2** and **A3:A4**, as indicated already by the B3LYP calculations of Chapter Three. It is thereby inferred that the isomorphism of these two pairs is independent of the theoretical method used.

H-bond geometries. The MP2/6-311++G(d,p) values of the R_{hb}, R_{xy} and θ_{hb} markers for the H-bond geometries (Table V.2) yield almost linear H-bonds with fairly typical lengths for the H-bonds of both pairs (R_{hb} ranging from 1.814 to 1.984 Å and R_{xy} from 2.841 to 2.999 Å). These values show good accordance with the B3LYP results of

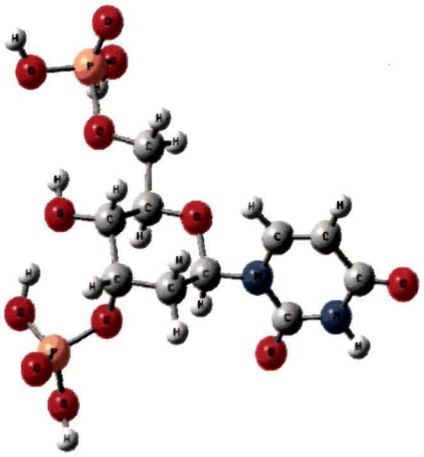
Chapter III (Table III.2). The O2...H4 H-bond of the **A3:A4** pair (1.814 Å) is much shorter than the corresponding O2...H4 bond of the **A1:A2** pair (1.984 Å), as evident from the R_{hb} values, indicating that **A3:A4** is somewhat more compactly aligned than **A1:A2**. This greater compactness for **A3:A4** is also apparent from its shorter R_{cc} value (see above). This may also be linked to the larger pairing energy for the **A3:A4** pair, where compactness of pairing leads to a larger pairing energy. This correlation of compactness of pairing configuration and H-bond geometry with pairing energy may also be noted for many of the base pairs of Chapters Three and Four.

V.4.2 Pyranonucleotide pairs

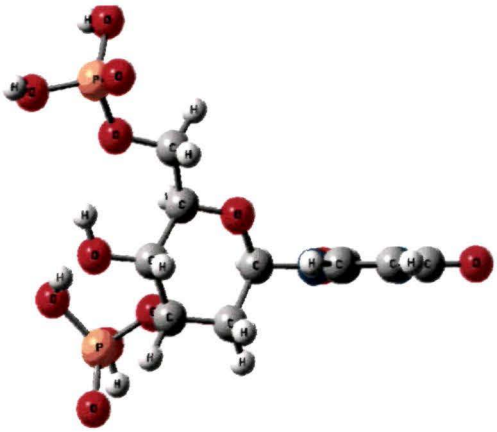
Figs. V.5 and V.6 depict the fully optimized three-dimensional structures for the four pyranonucleotide repeat units Py**A1**, Py**A2**, Py**A3** and Py**A4** as obtained by the B3LYP/6-31G(d,p) method with frontal and side views illustrated using the GAUSSVIEW program. In all cases, the base moieties maintain their essential planarity, and the plane of the base is approximately perpendicular to the plane bisecting the pyranose ring through the C1' and C4' atoms. All four nucleotides have basically the same structure and conformational characteristics. For each pyranonucleotide, it is possible to choose P-O bonds in the two phosphate groups (the C3' and C6' groups) which point in the same direction so as to ensure a kind of conformational periodicity of structure in the proposed polymeric single strand. Such periodicity is characteristic of DNA too, so that DNA has even been described as a "periodic crystal", where periodicity runs along the one-dimensional axis of the double helix.

Figs. V.7 and V.8 portray the fully optimized three-dimensional structures for the two H-bonded pyranonucleotide pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py respectively,

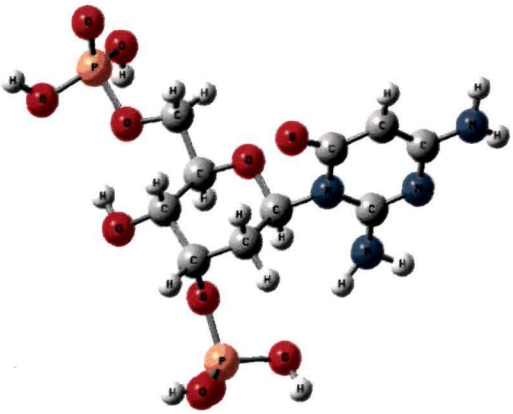
Optimised Gaussview



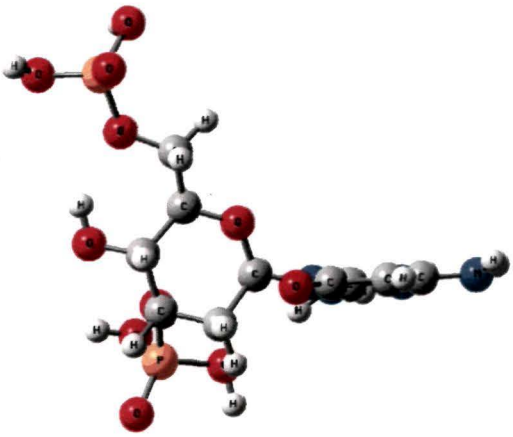
PyA1 (*frontal*)



PyA1 (*side*)



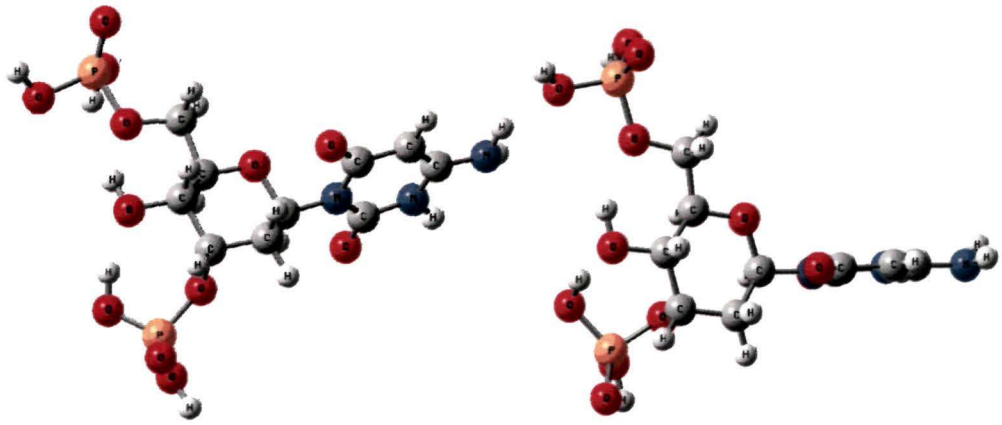
PyA2 (*frontal*)



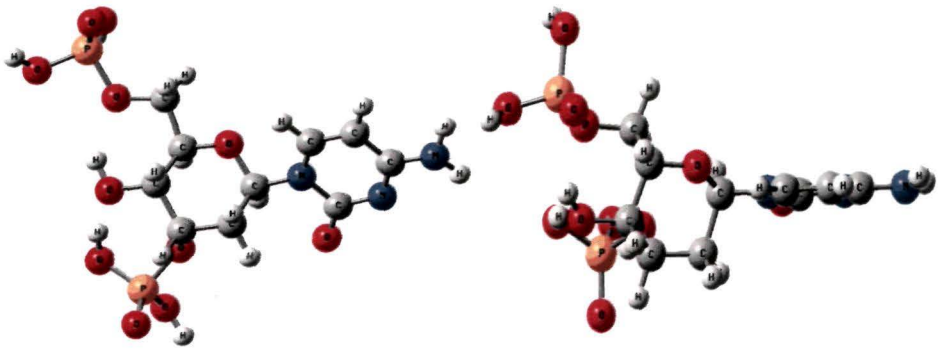
PyA2 (*side*)

Fig. V. 5 : The three dimensional optimised geometries of the pyranonucleotides **PyA1** and **PyA2** (frontal and side views)

Optimised Gaussview



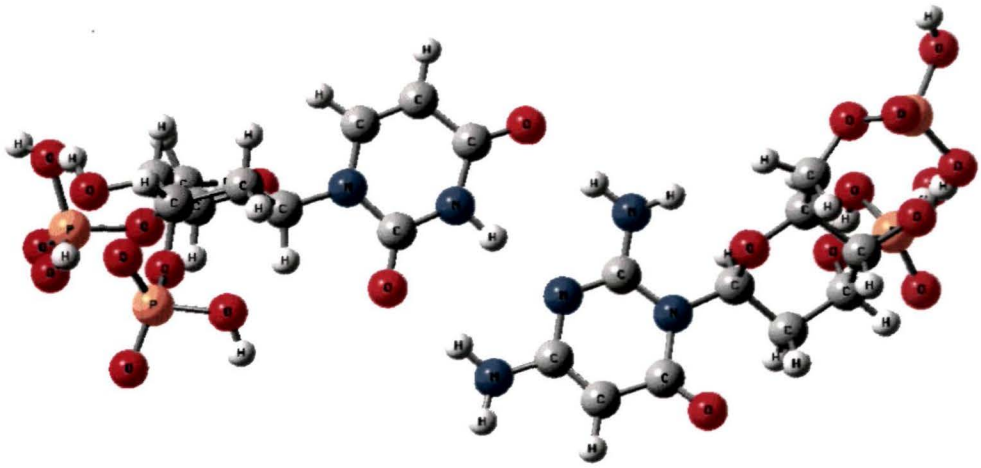
PyA3 (*frontal*) **PyA3** (*side*)



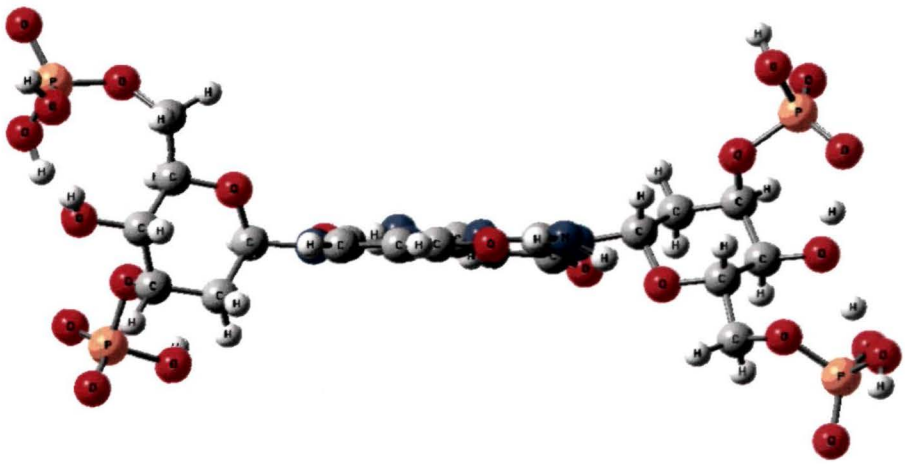
PyA4 (*frontal*) **PyA4** (*side*)

Fig. V. 6 : The three dimensional optimised geometries of the pyranonucleotides **PyA3** and **PyA4** (frontal and side views)

Optimised Gaussview



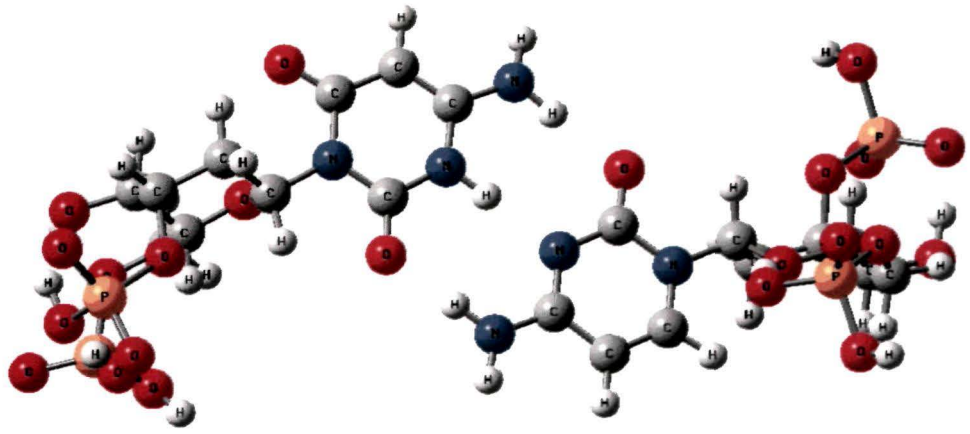
PyA1:A1Py (*frontal*)



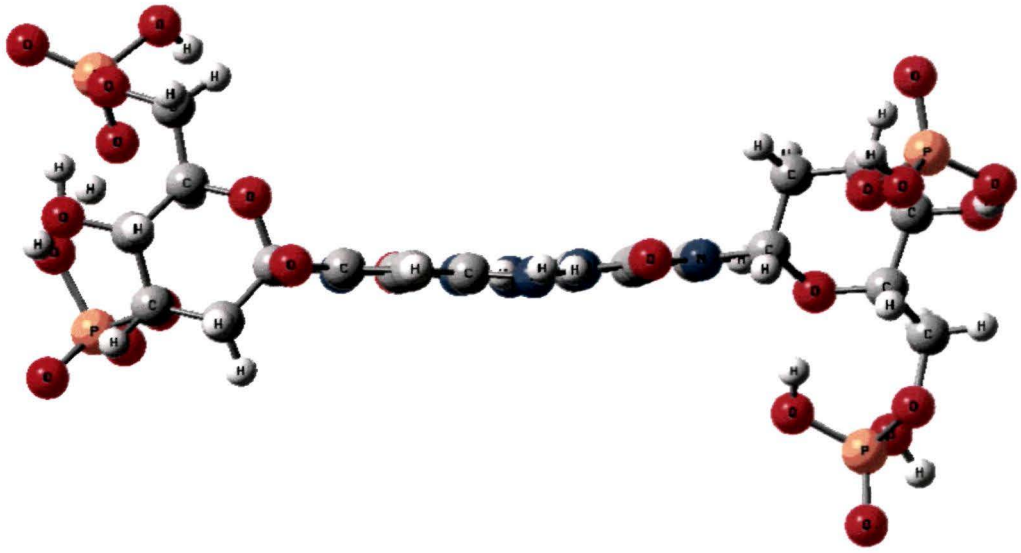
PyA1:A1Py (*side*)

Fig. V. 7 : The three dimensional optimised geometries of the pyranonucleotide H-bonded pair **PyA1:A2Py** (frontal and side views)

Optimised Gaussview



PyA3:A4Py (*frontal*)



PyA3:A4Py (*side*)

Fig. V. 8 : The three dimensional optimised geometries of the pyranonucleotide H-bonded pair **PyA3:A4Py** (frontal and side views)

where frontal and side views are included. These Figures indicate that, for each nucleotide pair, the two base components are essentially co-planar. For both the pairs, there is a kind of centre of inversion (only approximate) situated midway between the two N3 atoms of the component bases. This element of symmetry applies especially to the atoms of the pyranose ring and to the exocyclic phosphorus atoms. However, the individual OH groups and oxygen atoms of the phosphate groups do not always follow this centre of inversion, but this is only a matter of appropriate choice of the relevant dihedral angles. In any case, the symmetry around the central P atom of each phosphate group would be augmented if all the hydroxyl groups were fully ionized.

Table V.3 presents B3LYP/6-31G(d,p) values for the various estimates of pairing energy [the simple pairing energy E_p, the ZPVE-corrected pairing energy E_p(ZP) and the Gibb's free energy of pairing ΔG_p] for the pyranonucleotide pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py. Table IV.3 also presents optimized values of the markers of pairing configuration for these pyrano-nucleotide pairs, *viz.*, the R_{cc}, θ_1, θ_2 and φ descriptors. Table V.4 presents data regarding the geometry of the three H-bonds present in each of these two nucleotide pairs.

Pairing facility. The optimized B3LYP/6-31G(d,p) values of the simple pairing energy E_p for the pyranonucleotide pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py are -28.23 and -27.69 kcal/mol respectively (Table V.3), which are larger than the corresponding B3LYP values for the free base pairs **A1:A2** and **A3:A4** without backbone (Chapter Three, Table III.1). They are also larger than the corresponding MP2 values of E_p calculated for the solitary base pairs (Table V.1). Apparently, net H-bonded pairing facility is greater for the pyranonucleotide pairs than for the free base pairs, which

Table V.3 Estimates of pairing facility of the pyranonucleotide pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py along with values of their configurational markers as obtained from B3LYP/6-31G* optimized geometries[a]

Pair	E_p	E_p(ZP)	ΔG_p	R_{cc}	θ_1	θ_2	φ
Py**A1**:**A2**Py	-28.23	-26.47	-11.70	9.759	136.3	136.6	162.6
Py**A3**:**A4**Py	-27.69	-26.83	-14.28	9.656	135.2	135.0	-179.1

[a] Pairing energies in kcal/mol; interatomic distances in angstrom; bond angles and dihedrals in degrees

Table V.4 H-bond geometry data[a] for the pyranonucleotide pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py, for the pyridylamide pairs Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa (as obtained from B3LYP/6-31G* optimised geometries)

Pair	H-bond	R_{hb}	R_{ab}	θ_{hb}	R_{hb}	R_{ab}	θ_{hb}
		Pyranonucleotide pairs			*Pyrimidylamide pairs*		
A1:A2	O2...H4	1.965	2.978	177.2	1.971	2.986	176.8
	H3...N3	1.923	2.970	178.7	1.908	2.953	178.9
	O4...H2	1.917	2.935	175.8	1.909	2.931	176.9
A3:A4	O2...H4	1.818	2.845	177.7	1.823	2.853	178.1
	H3...N3	1.892	2.927	177.4	1.916	2.950	178.4
	H4...O2	1.860	2.884	179.0	1.891	2.914	177.3

[a] Bond lengths in angstrom; bond angles in degrees

points to a stabilizing effect of the pyranose phosphate backbone moiety. As was the case for the solitary base pairs, the values of E_p(ZP) are somewhat smaller than the corresponding E_p values. The corresponding values of the free energy of pairing ΔG_p are smaller still (-11.70 and -14.28 kcal/mol respectively), but serve to indicate that H-bonded pairing facility would be greater for Py**A3:A4**Py than for Py**A1:A2**Py. This parallels the result obtained from the previously discussed B3LYP and MP2 calculations for the solitary pairs without backbone, *viz.*, that base pairing would be more facile for the **A3:A4** pair than for the **A1:A2** pair. However, the simple pairing energies E_p for the nucleotide pairs do not reflect this trend, while the E_p(ZP) values are almost equal (-26.47 and -26.83 kcal/mol).

Pairing configuration. Values of the configuration descriptors for the pairs Py**A1**:**A2**Py and Py**A2**:**A3**Py are given in Table V.3. These values indicate firstly that pairing configuration does not change appreciably when passing from the solitary base pairs to the nucleotide pairs. The R_{cc} values for the nucleotide pairs are increased slightly in comparison with those for the solitary pairs, and the smaller value for the Py**A3:A4**Py pair indicates greater compactness of pairing for this pair than for the Py**A1:A2**Py pair. The θ_1 and θ_2 angles have almost the same range as that for the solitary pairs, while the φ values also do not change appreciably, and also indicate essential co-planarity for both the H-bonded systems. We may thus conclude that the basic H-bonded pairing configuration does not change appreciably upon passing from the solitary pairs without backbone to the pyranonucleotide pairs themselves.

More significantly, the configuration markers for the nucleotide pairs indicate that the two pairs Py**A1**:**A2**Py and Py**A2**:**A3**Py are basically *isomorphic* with each other,

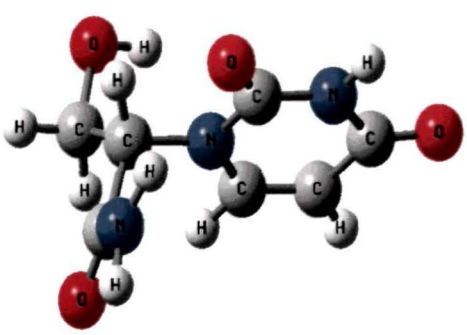
as may be seen by comparing the values for the two pairs. Note, however, that the pair Py**A1**:**A2**Py has somewhat lesser base co-planarity than the pair Py**A3**:**A4**Py, which is reflected in the lesser base co-planarity of the solitary **A1**:**A2** pair over the **A3**:**A4** pair. This isomorphism of H-bonded pairing is thus characteristic for all the unique pairing systems (solitary pairs and nucleotide pairs) derived from the mimic base set consisting of the bases **A1**, **A2**, **A3** and **A4**. This finding is important since it means that isomorphic pairing systems may be successfully designed on the basis of the solitary base pairs alone without reference to a backbone.

H-bond geometries. The H-bond data (Table V.4) points to H-bonds of the two nucleotide pairs that are quite linear (like for the solitary pairs), having lengths of broadly the same range as for the solitary base pairs. The O2...H4 H-bond of the Py**A3**:**A4**Py pair is the shortest H-bond within these two pairs, while the O2...H4 bond of the Py**A1**:**A2**Py pair is the longest. These trends appear in the solitary base pairs as well (Table V.1). These differences are reflected in the smaller R_{cc} value for the more compact Py**A3**:**A4**Py pair.

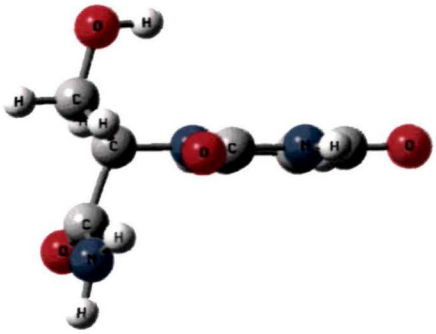
V.4.3 Pyrimidylamide pairs

Figs. V.9 and V.10 portray the optimized three-dimensional structures for the four pyrimidylamide repeat units Pa**A1**, Pa**A2**, Pa**A3** and Pa**A4** obtained using the B3LYP /6-31G(d,p) method. Frontal and side views are illustrated using the GAUSSVIEW program. For each of the pyrimidylamide species, the base moieties may be seen to be essentially planar. The plane of the base is more or less perpendicular to the plane containing the carbon, oxygen and nitrogen atoms of the β-hydroxypropionamide moiety. All the four nucleotides are found to possess basically the same structural and

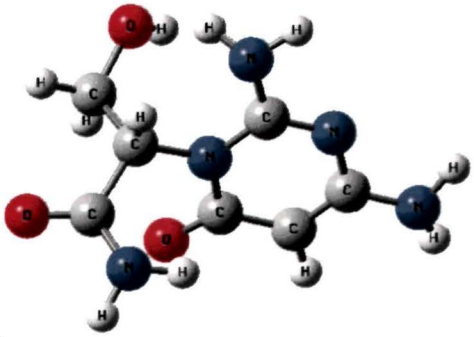
Optimised Gaussview



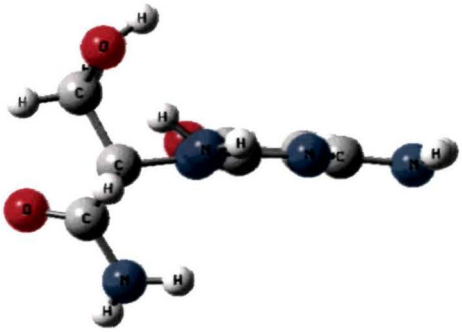
PaA1 (*frontal*)



PaA1 (*side*)



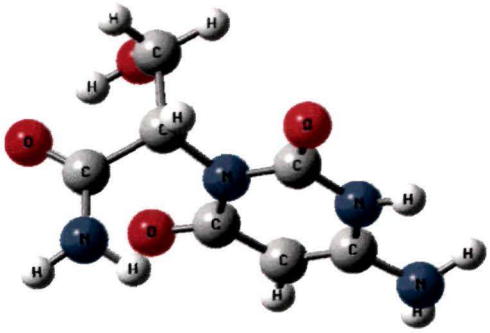
PaA2 (*frontal*)



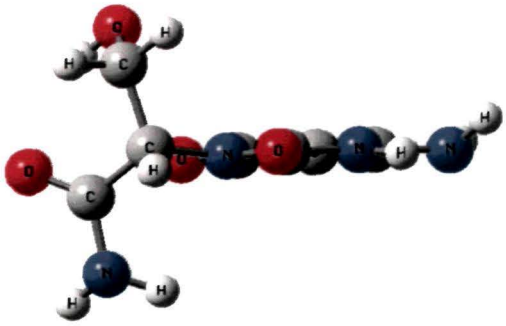
PaA2 (*side*)

Fig. V. 9 : The three dimensional optimised geometries of the pyrimidylamide units **PaA1** and **PaA2** (frontal and side views)

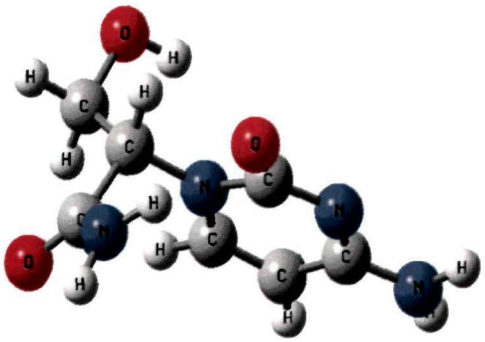
Optimised Gaussview



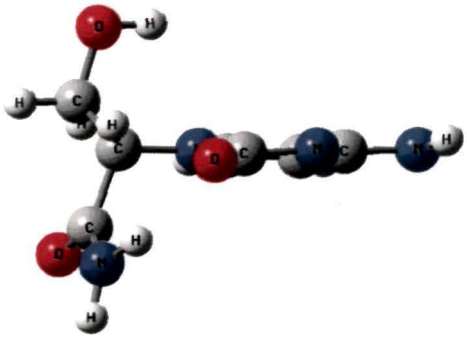
PaA3 (*frontal*)



PaA3 (*side*)



PaA4 (*frontal*)



PaA4 (*side*

Fig. V. 10 : The three dimensional optimised geometries of the pyrimidylamide units **PaA3** and **PaA4** (frontal and side views)

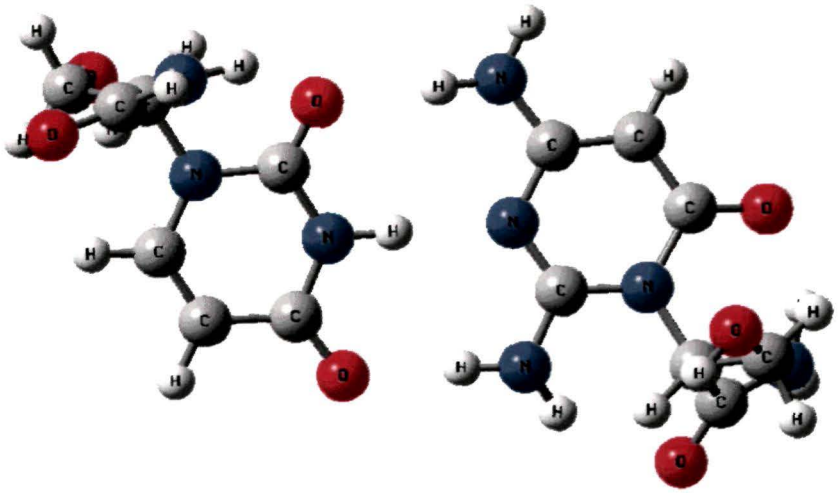
conformational characteristics. Note that this is also correspondingly true of the DNA nucleotides dA, dT, dG and dC, which are all conformationally similar.

Figs. V.11 and V.12 portray the fully optimized three-dimensional structures for the two H-bonded pyrimidylamide pairs Pa**A1**:**A2**Pa and Pa**A2**:**A3**Pa obtained using the GAUSSVIEW program. The frontal and side views indicate that the base pair moieties within each pairing system are basically co-planar with each other. For each pyrimidylamide pair, the carbon, nitrogen and oxygen atoms of the propionamide moiety fall more or less in one plane, which is approximately perpendicular to the plane containing the H-bonded bases. The antiparallel arrangement for pairing is quite evident, and a sort of centre of inversion (only approximate) exists midway between the N3 atoms of the two bases involved in each H-bonded pair.

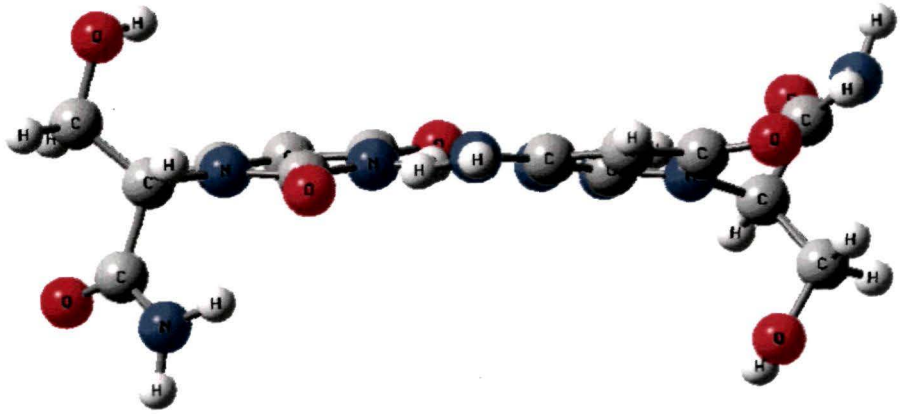
Table V.5 presents B3LYP/6-31G(d,p) values for the various pairing energy indices [the simple pairing energy E_p, the ZPVE-corrected pairing energy E_p(ZP) and the Gibb's free energy of pairing ΔG_p] for the pyrimidylamide pairs Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa. Table V.5 also presents optimized values of the markers of pairing configuration (R_{cc}, θ_1, θ_2 and φ) for these two pyrimidylamide pairs. Table V.4 gives data about the geometry of the H-bonds present in the two pyrimidylamide pairs.

Pairing facility. The data of Table V.5 indicates that facility of H-bonded pairing between the monomer units in each pyrimidylamide pair is quite appreciable, but still noticeably less than for the pyranonucleotide pairs (Table V.3). The E_p values for Pa**A1**:**A2**Pa and Pa**A2**:**A3**Pa are respectively -19.36 and -26.62 kcal/mol, which are smaller than the corresponding E_p values for the nucleotide pairs Py**A1**:**A2**Py and Py**A2**:**A3**Py, especially the former. This trend is also followed by the E_p(ZP) and ΔG_p

Optimised Gaussview



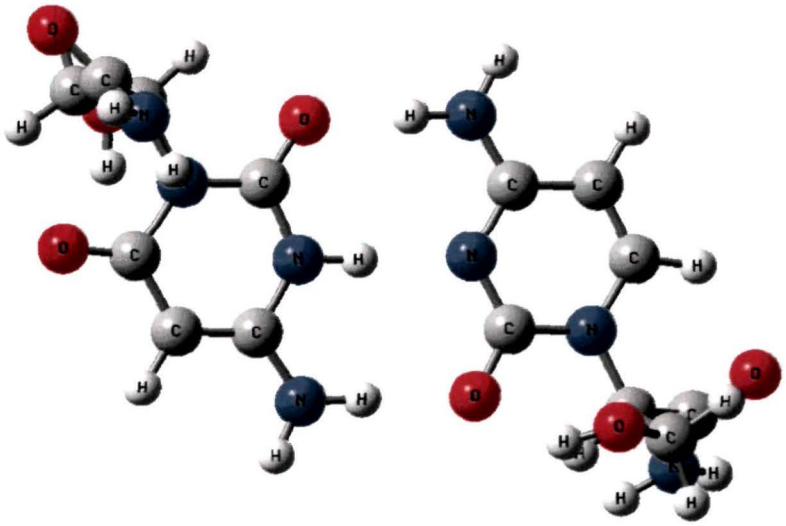
PaA1:A2Pa (*frontal*)



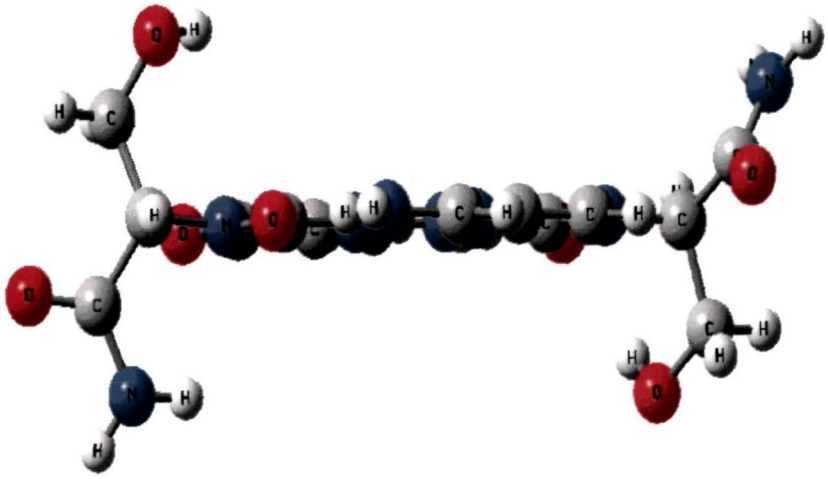
PaA1:A2Pa (*side*)

Fig. V. 11 : The three dimensional optimised geometries of the pyrimidylamide H-bonded pair **PaA1:A2Pa** (frontal and side views)

Optimised Gaussview



PaA3:A4Pa (*frontal*)



PaA3:A4Pa (*side*)

Fig. V. 12 : The three dimensional optimised geometries of the pyrimidylamide H-bonded pair **PaA3:A4Pa** (frontal and side views)

Table V.5 Estimates of pairing facility of the pyrimidylamide pairs Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa along with values of their configurational markers as obtained from B3LYP/6-31G* optimized geometries[a]

Pair	E_p	E_p(ZP)	ΔG_p	R_{cc}	θ_1	θ_2	φ
Pa**A1**:**A2**Pa	-19.36	-18.07	-6.64	9.736	136.9	134.5	168.1
Pa**A3**:**A4**Pa	-26.62	-25.32	-13.84	9.714	136.2	135.0	-176.9

[a] Pairing energies in kcal/mol; interatomic distances in angstrom; bond angles and dihedrals in degrees

indices. As for the nucleotide pairs, the order of absolute magnitude for the pairing energies in each case emerges as: E_p > E_p(ZP) > ΔG_p., with the free energies being markedly smaller in value. Compared to the solitary pairs without backbone (Table V.1), the E_p values for the pyrimidylamide pairs here predict that H-bonded pairing facility does not change very much on passing from the solitary pair to the pyrmidylamide pair in each case. This was not observed to be the case for the pyranonucleotide pairs, where addition of the sugar phosphate moiety appreciably augments H-bonded pairing facility. It may also be observed that all the different pairing energy indices predict that the Pa**A3**:**A4**Pa pair is noticeably more stable than the Pa**A1**:**A2**Pa pair, which trend is also reflected in the greater stability of the free **A3**:**A4** pair over the free **A1**:**A2** pair (Table V.1).

Pairing configuration. Optimized values of the configuration descriptors for the pyrimidyl-amide pairs Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa (Table V.5) indicate first of all that the H-bonded pairing configuration remains consistent to an appreciable extent on passing from the solitary base pairs to the pyrimidylamide pairs. The R_{cc} values for the pyrimidylamide pairs are only marginally larger than those for the solitary pairs, and the slightly smaller R_{cc} value for the Py**A3**:**A4**Py pair indicates somewhat more compact pairing for this pair than for the other pair Py**A1**:**A2**Py. The θ_1 and θ_2 angles have almost the same range as that for the solitary pairs. The φ values also do not change appreciably, indicating essential co-planarity for both these H-bonded pairs. We may thus conclude that the basic pairing configuration does not change appreciably upon passing from the solitary pairs (without backbone) to the pyrimidyl-amide pairs. This is what was noted for the two pyranonucleotide pairs Pa**A1**:**A2**Pa

and Pa**A3**:**A4**Pa as well (Table V.5). The H-bonded pairing configuration of these systems is thus largely *independent* of the presence or type of backbone incorporated.

Furthermore, values of the configuration markers for the pyrimidylamide pairs indicate that the two pairs Pa**A1**:**A2**Pa and Pa**A2**:**A3**Pa are essentially *isomorphic* with each other, as may be seen by comparing the values for the two pairs. The R_{cc} values differ by only 0.022 Å, while the θ_1 and θ_2 angles fall within the narrow range of 134.5 to 136.9°. However, the pair Pa**A1**:**A2**Pa has somewhat lesser base co-planarity than the pair Pa**A3**:**A4**Pa; this is reflected also in the lesser base co-planarity of the solitary **A1**:**A2** pair over the **A3**:**A4** pair, as well as in the lesser base co-planarity of the nucleotide pair Py**A1**:**A2**Py over the Py**A2**:**A3**Py pair.

We may now conclude that, in fact, *all* these H-bonded pairing systems are more or less isomorphic with one another, regardless of whether they are solitary base pairs, pyranonucleotide pairs or pyrimidylamide pairs. Based upon the B3LYP/6-31G(d,p) optimized configurations for all these three types of H-bonded pairs, we define the value ranges for the configuration markers in general as follows. The R_{cc} marker has a value range from 9.581 to 9.759 Å (a narrow range of only 0.178 Å). The θ_1 and θ_2 angles range between 134.5 and 136.9° (a range of only 2.4°). While the dihedral φ ranges from -156.7 to 179.1°, this does not detract very much from the desired feature of essential co-planarity for all these systems. Isomorphism of H-bonded pairing is thus characteristic for *all* the unique pairing systems (solitary pairs, pyranonucleotide pairs and pyrimidylamide pairs) derived from the one mimic base set comprised of the four bases **A1**, **A2**, **A3** and **A4**. This observation thus highlights that viable isomorphic H-bonded pairing systems may be successfully designed on the basis of

the free base pairs alone without reference to a backbone. Addition of the backbone (sugar phosphate or polyamide) does not nullify the essential isomorphic nature.

H-bond geometries. The H-bond geometry data for the pyrimidylamide pairs (Table V.4) does not reveal anything significantly different from the data for the other H-bonded systems (solitary base pairs or pyranonucleotide pairs). All the H-bonds are quite linear, as evinced from the θ_{hb} angle values being all very close to 180°. The lengths R_{hb} of the actual H-bonds **X...H** or **H...Y** range from 1.823 to 1.971 Å, the shortest being the O2...H4 H-bond of the Pa**A3**:**A4**Pa pair, and the longest being the O2...H4 H-bond of the Pa**A1**:**A2**Pa pair. We thus observe that the trends concerning H-bond lengths, as evident in the solitary base pairs (Table V.1), are reproduced in the pyrimidylamide and pyranonucleotide pairs as well. This serves to further corroborate our expectation that H-bonded pairing configuration as seen in the solitary base pairs should be closely reproduced in the more complex repeat unit systems which are covalently bonded to the backbone moieties (sugar phosphate or polyamide).

V.5 Charge Distribution and Transfer during H-Bonding

Chapter Four had discussed briefly some aspects of the manner in which charge transfer takes place during H-bonding in the various self-associative base pairs studied therein. This kind of treatment is important since the stabilizing interactions involved in H-bonding in general are believed to be primarily electrostatic. This Section approaches this aspect by firstly deriving the charge distribution around the H-bonding region of the various species treated in this Chapter, namely, the pyrano-nucleotides and their pairs along with the pyrimidylamides and their pairs. The charge

distribution is expressed here in terms of the *point charges* situated on the atoms directly involved in H-bonding, *viz.*, the atoms contained within the **X...H** or **H...Y** moieties of each H-bond. Atomic charge derivation is carried out here using *(a)* the *Mulliken* charges obtained through the standard and well-known Mulliken population analysis, and *(b)* charges derived from the *molecular electrostatic potential* (MESP or simply ESP) for which three different strategies are employed here.

The main aims of the studies embodied in this Section may be delineated as follows:

1. To calculate and compare the atomic charges on the H-bonding atoms of the various species involved, employing a variety of models and computational strategies.
2. To take note of the manner in which the atomic charges within the monomer units undergo change when H-bonding occurs creating the pair.
3. To calculate the net charge transfer occurring from one monomer unit to the other when H-bonding occurs leading to the pair, and to compare the dipole moments for the monomer species and their pairs.

Atomic charge derivation is conducted only for the atoms most directly involved in the H-bonds **X...H-Y** or **X-H...Y**, *viz.*, the lone pair donor atoms (**X** or **Y**) and the proton **H** involved. The standard Mulliken-type atomic charges for these atoms are calculated based upon the wave function obtained by the B3LYP/6-31G* method on the fully optimized geometries, using the formula involving the density matrix.

Use is also made of atomic charges derived from the molecular electrostatic potential (MEP) generated from the wave function itself, for which single-point

calculations at the B3LYP level employing an extended basis set (the 6-311++G(d,p) basis set) were conducted on the optimized geometries already obtained from the B3LYP/6-31G* calculations. The theory underlying calculation of molecular electrostatic potentials has been discussed in Chapter Two. For derivation of atomic charges from the MEP, three strategies were used, designated in the GAUSSIAN 03 program suite as the CHELP[10], CHELPG[11] and MK[12,13] options (see Chapter Two).

From the data of Tables V.6 to V.9, the partial charges on the relevant atoms display a broad range of values. Mulliken charges for the H-atoms range from 0.292 to 0.425 a.u., larger in the H-bonded pairs than in the solitary units. Mulliken nitrogen charges range from -0.651 to -0.807 a.u., while oxygen charges range from -0.496 to -0.602 a.u. Negative charges greater than unity (calculated by the CHELP and MK methods) appear for exocyclic nitrogens within the H-bonded pairs.

V.5.1 Pyranonucleotide systems

The monomer units here are the four solitary pyranonucleotides Py**A1**, Py**A2**, Py**A3** and Py**A4** (Fig. V.1), while the unique H-bonded pairs are the Py**A1**:**A2**Py and Py**A3**:**A4**Py pairs (Fig. V.2). The atoms directly involved in the three H-bonds of each pair are *(a)* the O2, N3, H3 and O4 atoms of Py**A1** along with the H4, N4, N3, N2 and H2 atoms of Py**A2**, which constitute the O2...H4-O4, N3-H3...N3 and O4...H2-N2 H-bonds of the Py**A1**:**A2**Py pair; and *(b)* the O2, N3, H3, N4 and H4 atoms of Py**A3** along with the H4, N4, N3 and O2 atoms of Py**A4**, which constitute the O2...H4-O4, N3-H3...N3 and N4-H4...O2 hydrogen bonds of the H-bonded Py**A3**:**A4**Py pair.

Table V.6 presents the various estimates of atomic charge on the above-mentioned atoms as they exist in the non-H-bonded pyranonucleotide units Py**A1**, Py**A2**, Py**A3**

Table V.6 Comparison of various atomic charge estimates* for the H-bonding atoms of the four unpaired pyranonucleotides Py**A1**, Py**A2**, Py**A3** and Py**A4**.

Nucleo-tide	Atom	B3LYP/6-31G* Mulliken	ESP-derived B3LYP/6-311++G(d,p) single point CHELP	CHELPG	MK
Py**A1**	O2	-0.503	-0.562	-0.594	-0.581
	H3	0.292	0.342	0.377	0.374
	N3	-0.617	-0.661	-0.688	-0.654
	O4	-0.496	-0.658	-0.603	-0.589
Py**A2**	H4	0.273	0.288	0.370	0.376
	N4	-0.642	-0.708	-0.867	-0.871
	N3	-0.592	-0.622	-0.784	-0.756
	H2	0.282	0.307	0.380	0.392
	N2	-0.658	-0.684	-0.843	-0.846
Py**A3**	O2	-0.511	-0.591	-0.609	-0.585
	H3	0.283	0.337	0.400	0.399
	N3	-0.644	-0.522	-0.686	-0.628
	H4	0.274	0.265	0.355	0.358
	N4	-0.649	-0.712	-0.828	-0.823
Py**A4**	H4	0.287	0.325	0.403	0.409
	N4	-0.628	-0.857	-0.928	-0.926
	O2	0.530	-0.637	-0.645	-0.638
	N3	-0.571	-0.745	-0.823	-0.828

* All charge values in atomic units

and Py**A4**. Each set of atomic point charges is calculated by the various methods mentioned earlier, which include the B3LYP/6-31G* Mulliken charges, and the ESP-derived charges using the single-point B3LYP/6-311++G(d,p) method with the CHELP, CHELPG and MK options. One can note that the electronegative oxygen and nitrogen atoms are all negatively charged as expected, while the hydrogen atoms are all positively charged. In general, the ESP-derived charges have larger magnitudes than the Mulliken charges, which means that the region around the H-bonding atoms is predicted as more polar by the ESP-derived methods than by the Mulliken method (except for the N3-atom of Py**A3**). For each hydrogen atom, the charge generally increases in magnitude in the order Mulliken < CHELP < (CHELPG or MK) with respect to the method, and this trend is a good extent followed by the electronegative atoms as well (except the N3 atoms of Py**A1** and Py**A3**).

The B3LYP/6-31G* Mulliken charges are more negative for the nitrogen atoms (whether ring or exocyclic) than the oxygens. Although this trend is not maintained invariantly by the CHELP and MK versions of the ESP-derived charges, the CHELPG charges point to this trend in consistent fashion, and give the order N(exocyclic) > N(ring) > O for magnitude of the charges. Furthermore, any nitrogen attached covalently to a hydrogen is more negatively charged than the ring N3-atoms of Py**A2** and Py**A4** which have no attached hydrogens.

Table V.7 presents the various estimates of atomic charge on the atoms directly involved in H-bonding as they occur in the H-bonded nucleotide pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py. Each set of atomic point charges Q as they occur in these pairs is

Table V.7 Comparison of atomic charge estimates for the H-bonding atoms of the two pyranonucleotide pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py with the corresponding changes ΔQ occurring upon H-bonding

Pair and Monomer	Atom	B3LYP/6-31G* Mulliken		ESP-derived B3LYP/6-311++G(d,p) single point CHELP		CHELPG		MK	
		Q	ΔQ	Q	ΔQ	Q	ΔQ	Q	ΔQ
*Py***A1**:**A2***Py Pair*									
Py**A1** part	O2	-0.577	-0.074	-0.560	0.002	-0.722	-0.128	-0.677	-0.096
	H3	0.340	0.048	0.314	-0.028	0.320	-0.057	0.252	-0.122
	N3	-0.651	-0.034	-0.578	0.083	-0.655	0.033	-0.539	0.115
	O4	-0.514	-0.018	-0.667	-0.009	-0.627	-0.024	-0.605	-0.016
Py**A2** part	H4	0.305	0.032	0.407	0.119	0.449	0.079	0.428	0.052
	N4	-0.672	-0.030	-0.862	-0.154	-0.921	-0.054	-0.890	-0.019
	N3	-0.659	-0.067	-0.673	-0.051	-0.666	0.118	-0.462	0.294
	H2	0.312	0.030	0.470	0.163	0.495	0.115	0.488	0.096
	N2	-0.694	-0.036	-0.824	-0.140	-0.951	-0.108	-0.904	-0.058
*Py***A3**:**A4***Py Pair*									
Py**A3** part	O2	-0.573	-0.062	-0.612	-0.021	-0.734	-0.125	-0.687	-0.102
	H3	0.332	0.049	0.353	0.016	0.436	0.036	0.380	-0.019
	N3	-0.677	-0.033	-0.606	-0.084	-0.777	-0.091	-0.623	0.005
	H4	0.329	0.055	0.464	0.199	0.533	0.178	0.499	0.141
	N4	-0.693	-0.044	-0.912	-0.200	-1.076	-0.248	-1.030	-0.207
Py**A4** part	H4	0.342	0.055	0.456	0.131	0.592	0.189	0.581	0.172
	N4	-0.652	-0.024	-0.917	-0.060	-1.148	-0.220	-1.134	-0.208
	N3	-0.654	-0.083	-0.673	0.072	-0.859	-0.036	-0.764	0.064
	O2	-0.585	-0.055	-0.680	-0.043	-0.731	-0.086	-0.677	-0.039

* All charge values in atomic units

calculated by the methods mentioned earlier, listing the B3LYP/6-31G* Mulliken charges, as well as the ESP-derived charges using the single-point B3LYP/6-311++G(d,p) method with the CHELP, CHELPG and MK options. The change that occurs for each atomic charge after H-bonding takes place is given by:

$$\Delta Q(\text{atom}) = Q(\text{atom in nucleotide pair}) - Q(\text{atom in solitary nucleotide}).$$

The data of Table V.7 indicates that the electronegative atoms are all negatively charged while the hydrogens are all positive. Unlike for the unpaired nucleotides, the MEP-derived atomic charges for the atoms within the H-bonded pairs are not invariantly larger than the Mulliken charges, though this is always true for the exocyclic N2 and N4 atoms and their bonded hydrogens. These exocyclic nitrogen atoms also bear the maximum negative charge among all the atoms studied here, an observation markedly seen in all the MEP-derived atomic charge values. The MEP-derived data set also predicts maximum positive charges for the H-atoms bonded to these exocyclic nitrogens. This means that the bond dipoles across these covalent N-H bonds are appreciable, and the magnitude of the bond dipoles are expected to be greater in the Py**A3**:**A4**Py pair than in the Py**A1**:**A2**Py pair since the charge disparity between the N and H atoms is more marked for the former than for the latter.

The differences in charge ΔQ between the atoms of the solitary units and the corresponding ones of the nucleotide pair do not show a uniform trend. The Mulliken charge data set gives consistently negative values of ΔQ for the electronegative oxygen and nitrogen atoms and consistently positive values of ΔQ for the hydrogen atoms. This consistency is not reflected, though, in the values of ΔQ calculated by the ESP-derived methods CHELP, CHELPG and MK. In fact, the H3 atoms attached to

the ring N3 atoms give negative values for ΔQ when calculated by these MEP-derived methods, while the ring N3 atoms give positive ΔQ values. The exocyclic N2 and N4 atoms do, however, show consistency in the negative values of ΔQ that they provide, where the ΔQ values for their attached H-atoms are consistently positive.

V.5.2 Pyrimidylamide systems

The monomer units here are the four solitary pyrimidylamides Pa**A1**, Pa**A2**, Pa**A3** and Pa**A4** (Fig. V.3), while the two unique H-bonded pairs are the Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa pairs (Fig. V.4). Since the fundamental constituent base pairs here are the same as those for the pyranonucleotide systems, the atoms involved in H-bonding are all the same. The atoms directly involved in the three H-bonds of each pair are thus *(a)* the O2, N3, H3 and O4 atoms of Pa**A1** along with the H4, N4, N3, N2 and H2 atoms of Pa**A2**, which constitute the O2...H4-N4, N3-H3...N3 and O4...H2-N2 H-bonds of the Pa**A1**:**A2**Pa pair; and *(b)* the O2, N3, H3, N4 and H4 atoms of Pa**A3** along with the H4, N4, N3 and O2 atoms of Pa**A4**, which constitute the O2...H4-N4, N3-H3...N3 and N4-H4...O2 hydrogen bonds of the Pa**A3**:**A4**Pa pair.

Table V.8 presents the various estimates of atomic charge on the relevant atoms in the non-H-bonded pyridylamide units Pa**A1**, Pa**A2**, Pa**A3** and Pa**A4**, including the B3LYP/6-31G* Mulliken charges, and the ESP-derived charges using the single-point B3LYP/6-311++G(d,p) method (the CHELP, CHELPG and MK options). As expected, the electronegative oxygen and nitrogen atoms are all negatively charged, while the hydrogen atoms are all positively charged. In general, the MEP-derived charges as calculated by the CHELPG and MK methods have larger magnitudes than the Mulliken charges, while the CHELP method often predicts smaller charge values

Table V.8 Comparison of various atomic charge estimates* for the H-bonding atoms of the four unpaired pyrimidylamides Pa**A1**, Pa**A2**, Pa**A3** and Pa**A4**.

Pyrimidyl-amide	Atom	B3LYP/6-31G* Mulliken	ESP-derived B3LYP/6-311++G(d,p) single point CHELP	CHELPG	MK
Pa**A1**	O2	-0.528	-0.551	-0.600	-0.590
	H3	0.360	0.314	0.374	0.376
	N3	-0.699	-0.559	-0.667	-0.646
	O4	-0.494	-0.616	-0.594	-0.584
Pa**A2**	H4	0.340	0.283	0.352	0.355
	N4	-0.776	-0.711	-0.794	-0.794
	H2	0.344	0.289	0.353	0.371
	N2	-0.778	-0.689	-0.796	-0.809
	N3	-0.580	-0.606	-0.746	-0.708
Pa**A3**	O2	-0.518	-0.581	-0.606	-0.577
	H3	0.350	0.342	0.401	0.392
	N3	-0.729	-0.642	-0.699	-0.590
	H4	0.338	0.305	0.361	0.367
	N4	-0.780	-0.677	-0.832	-0.831
Pa**A4**	H4	0.355	0.320	0.405	0.410
	N4	-0.767	-0.738	-0.913	-0.910
	O2	-0.542	-0.564	-0.647	-0.655
	N3	-0.561	-0.706	-0.795	-0.810

* All charges in atomic units

than the Mulliken method. For each H-atom, the charge generally increases in magnitude in the order CHELP < Mulliken < CHELPG < MK with respect to the method, a trend not followed by the electronegative atoms. The B3LYP/6-31G* Mulliken charges are more negative for the nitrogen atoms (ring or exocyclic) than for the oxygen atoms. Although this trend is not maintained invariantly by the CHELP and MK versions of the MEP-derived charges, the CHELPG charges point to this same trend in consistent fashion, giving the order N(exocyclic) > N(ring) > O for magnitude of the charges.

Table V.9 presents the various estimates of atomic charge on the atoms directly involved in H-bonding as they occur in the H-bonded pairs Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa, incorporating the B3LYP/6-31G* Mulliken charges and MEP-derived charges using the single-point B3LYP/6-311++G(d,p) strategy with the CHELP, CHELPG and MK options. Included also are the changes occurring for each atomic charge after H-bonding takes place (given as ΔQ).

The data of Table V.9 again indicates (as expected) that the electronegative atoms are all negatively charged while the hydrogens are all positive. The CHELPG and MK charges for all the atoms are almost always larger than those calculated by CHELP method. Generally, it is the exocyclic nitrogen atoms that bear the maximum negative charge among all the atoms studied here, an observation markedly seen in the CHELPG and MK data sets. The general order of magnitude observed for the electronegative atomic charges emerges as N(exocyclic) > N(ring) > O. The bond dipoles across the exocyclic N-H bonds are expected to be larger in the Pa**A3**:**A4**Pa pair than in the Pa**A1**:**A2**Pa pair since the charge disparity between the N and H atoms is more marked for the former than for the latter.

Table V.9 Comparison of atomic charge estimates for the H-bonding atoms of the two pyrimidylamide pairs Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa with the corresponding changes ΔQ occurring upon H-bonding

Pair and Monomer	Atom	B3LYP/6-31G* Mulliken		ESP-derived B3LYP/6-311++G(d,p) single point CHELP		CHELPG		MK	
		Q	ΔQ	Q	ΔQ	Q	ΔQ	Q	ΔQ
*Pa***A1***:***A2***Pa Pair*									
Pa**A1** part	O2	-0.556	-0.028	-0.485	0.066	-0.639	-0.039	-0.623	-0.033
	H3	0.419	0.059	0.328	0.014	0.389	0.015	0.346	-0.030
	N3	-0.741	-0.042	-0.499	0.060	-0.691	-0.024	-0.591	0.055
	O4	-0.524	-0.030	-0.548	0.068	-0.639	-0.045	-0.616	-0.032
Pa**A2** part	H4	0.390	0.050	0.293	0.010	0.461	0.109	0.449	0.094
	N4	-0.835	-0.059	-0.748	-0.037	-0.969	-0.175	-0.944	-0.150
	N3	-0.657	-0.077	-0.678	-0.072	-0.761	-0.015	-0.641	0.067
	H2	0.400	0.056	0.353	0.064	0.457	0.104	0.459	0.088
	N2	-0.833	-0.055	-0.705	-0.016	-0.868	-0.072	-0.852	-0.043
*Pa***A3***:***A4***Pa Pair*									
Pa**A3** part	O2	-0.561	-0.043	-0.563	0.018	-0.693	-0.087	-0.651	-0.074
	H3	0.414	0.064	0.296	-0.046	0.467	0.066	0.420	0.028
	N3	-0.774	-0.045	-0.538	0.104	-0.782	-0.083	-0.608	-0.018
	H4	0.408	0.070	0.292	-0.013	0.454	0.093	0.425	0.058
	N4	-0.849	-0.069	-0.678	-0.001	-0.966	-0.134	-0.920	-0.089
Pa**A4** part	H4	0.425	0.070	0.378	0.058	0.581	0.176	0.581	0.171
	N4	-0.807	-0.040	-0.799	-0.061	-1.128	-0.215	-1.130	-0.220
	N3	-0.643	-0.082	-0.621	0.085	-0.866	-0.071	-0.816	-0.006
	O2	-0.602	-0.060	-0.540	0.024	-0.672	-0.025	-0.643	0.012

* All charges in atomic units

The differences in charge ΔQ between the atoms of the solitary units and the corresponding ones of the nucleotide pair do not always show a uniform trend. Here again, the Mulliken charge data set gives consistently negative values of ΔQ for the electronegative oxygen and nitrogen atoms and consistently positive values of ΔQ for the hydrogen atoms, a trend not consistently reflected in the values of ΔQ calculated by the ESP-derived CHELP method. The CHELPG and MK charges do, however, follow the trend shown by the Mulliken data set for the exocyclic N2 and N4 atoms and the H-atoms attached to them.

V.5.3 Analysis of charge transfers

Charge transfer is important for H-bonding since much of the attractive force involved in H-bonding is attributed to electrostatic interactions. An interesting facet here is the manner in which charge redistribution and transfer occurs following the H-bonding process. This is analyzed for the pyranonucleotide and pyrimidylamide pairs studied here by referring to the types of H-bonds involved, the direction in which they point, and the magnitudes of changes in atomic charges that occur.

The H-bonded pairs studied here basically involve two different kinds of H-bonds, *viz.*, the O...H-N (or N-H...O) type and the N-H...N type of H-bonds. The O...H-N and N-H...O bonds include a carbonyl oxygen and an exocyclic nitrogen (as an amino group), while the N-H....N bonds involve two ring nitrogens. All the H-bonded pairs studied here each have two H-bonds of the first type and one of the second type. However, some differences exist in the directionality of the H-bonds. Pairs which incorporate the **A1:A2** base pair moiety (*viz.*, the Py**A1**:**A2**Py and Pa**A1**:**A2**Pa pairs) have two O...H-N bonds (with both oxygens on **A1** and both amino groups on **A2**)

both pointing in the same direction. In contrast, pairs which incorporate the **A3:A4** base pair moiety (*viz.*, the Py**A3**:**A4**Py and Pa**A3**:**A4**Pa pairs) have one O...H-N bond (with an oxygen on **A3** and an amino group on **A4**) and one N-H...O bond (with an oxygen on **A4** and an amino group on **A3**). Here, the two H-bonds point in opposite directions. Both types of pairs, though, have one N-H...N bond (with a ring N-H group on **A1/A3** and a ring N-atom on **A2/A4**), pointing in the same direction. In order to conduct this analysis of charge redistribution, we use the CHELPG data set of MEP-derived charges among the four data sets available (see Tables V.7 and V.9).

Table V.10 presents values of the *net charge transfer* associated with H-bonding for the pyranonucleotide pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py and the pyrimidylamide pairs Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa, calculated by *(a)* the B3LYP/6-31G* Mulliken method, and *(b)* the MEP-derived strategies CHELP, CHELPG and MK. Table.V.10 presents values of the dipole moments for the species associated with each system - the solitary monomers (the 1st and 2nd units) and the whole H-bonded pairing system itself.

Pyranonucleotide pairs. From the data of Table V.7 for the nucleotide pairs, one observes appreciable changes in atomic charge ΔQ for the atoms involved in the O2...H4-N4 H-bond between the **A1** and **A2** moieties of the Py**A1**:**A2**Py pair. The CHELPG charge differences for this bond (ΔQ = -0.128, 0.079 and -0.054 a.u. for the O2, H4 and N4 atoms respectively) may be compared with those for the O4...H2-N2 H-bond (ΔQ = -0.024, 0.115 and -0.108 a.u. for the O4, H2 and N2 atoms respectively). Charge difference values for these two O...H-N H-bonds indicate some degree of comparability with respect to the overall change in charge distribution. Negative charge accumulates more on the oxygen atom for the O2...H4-N4 H-bond, but on the

Table V.10 Variously estimated values of the net charge transfer occurring during H-bonded pairing and values of the dipole moments for solitary monomers and resultant H-bonded pairs (the pyranonucleotide and the pyrimidylamide cases)

Pair	Amount of net charge transferred				Direction of transfer	Dipole moments		
	Mulliken	CHELP	CHELPG	MK		*1st unit*	*2nd unit*	*Pair*
Pyranonucleotide pairs								
Py**A1**:**A2**Py	-0.0136	-0.0724	-0.0870	-0.1031	Py**A2** to Py**A1**	4.49	5.84	12.91
Py**A3**:**A4**Py	-0.0839	-0.0617	-0.0679	-0.0883	Py**A4** to Py**A3**	8.38	7.58	14.60
Pyrimidylamide pairs								
Pa**A1**:**A2**Pa	-0.0186	-0.0353	-0.0397	-0.0490	Pa**A2** to Pa**A1**	1.07	6.04	4.31
Pa**A3**:**A4**Pa	-0.0782	-0.0699	-0.0743	-0.0905	Pa**A4** to Pa**A3**	8.20	5.58	12.39

[a] Charges in atomic units; dipole moments on debye

nitrogen atom for the O4...H2-N2 H-bond. The atomic charges differences on the central H-bond N3-H3...N3 show an unusual reversal of the expected trend, where the ΔQ values are negative for the H3 atom and positive for the nitrogen atoms.

Appreciable changes in atomic charges are also noted for the O2...H4-N4 and N4-H4...O2 H-bonds of the Py**A3**:**A4**Py pair. The CHELPG charge differences for the O2...H4-N4 bond (ΔQ = -0.125, 0.189 and -0.220 a.u. for the O2, H4 and N4 atoms respectively) are of about the same order as those for the N4-H4...O2 H-bond (ΔQ = -0.248, 0.178 and -0.086 a.u. for the N4, H4 and O2 atoms respectively), where it is the two N4 atoms on Py**A3** and Py**A4** that are associated with the largest charge differences. The ΔQ values for these two H-bonds are larger than those for the central N3-H3...N3 H-bond (ΔQ = -0.091, 0.036 and -0.036 a.u. for the N3, H3 and N3 atoms respectively), indicating larger charge transfer for the O2...H4-N4 and N4-H4...O2 H-bonds than for the central N3-H3...N3 H-bond. Charge differences for the O2...H4-N4 and N4-H4...O2 H-bonds of the Py**A3**:**A4**Py pair are also larger than those for the O2...H4-N4 and O4...H2-N2 H-bonds of the Py**A1**:**A2**Py pair just discussed above.

All these observations point towards the conclusion that the Py**A3**:**A4**Py pair would be more strongly H-bonded than the Py**A1**:**A2**Py pair, as is corroborated by the larger pairing energy for the former expressed as the E_p(ZP) and ΔG_p terms of Table V.3. Table V.10 also shows that the net charge transfer calculated by the B3LYP/6-31G* Mulliken method is smaller for the Py**A1**:**A2**Py pair (-0.0136 a.u.) than for the Py**A3**:**A4**Py pair (-0.0839 a.u.). This finding is, however, not reflected in the charge transfers calculated by the CHELP, CHELPG and MK methods, which all invariably point to greater charge transfer for the Py**A1**:**A2**Py pair than for the Py**A3**:**A4**Py pair.

The Mulliken-based dipole moments (Table V.10) indicate that, for both the pairs Py**A1**:**A2**Py and Py**A3**:**A4**Py, H-bonding leads to greater net charge separation in the H-bonded pair than is present in the solitary units, where the dipole moment for each pair is larger than for the constituent units. The data regarding the Mulliken-based dipole moments also points to a larger dipole moment for the Py**A3**:**A4**Py pair (14.60 D) than for the Py**A1**:**A2**Py pair (12.91 D), which may be taken to mean that the net charge transfer is larger for the former than for the latter pair. This Mulliken-based charge data fits well with the pairing energies E_p(ZP) and ΔG_p of Table V.3.

All the four methods of charge derivation (Mulliken, CHELP, CHELPG and MK) predict the same results regarding the direction of net charge transfer from one unit to the other during H-bonding in each pair. For the Py**A1**:**A2**Py pair, the direction is predicted as Py**A2** to Py**A1** since the atom associated the most negative CHELPG charge increment ΔQ (*viz.*, the O2 atom) is situated on the Py**A2** unit, which also bears the H3 atom associated with a negative ΔQ value. The net charge transfer in the Py**A3**:**A4**Py pair is predicted in the direction Py**A4** to Py**A3** since the O2, N3 and N4 atoms of the Py**A3** unit within this pair each bear negative CHELPG charge increments ΔQ which are larger than those for the N4, N3 and O2 atoms of the Py**A4** unit; furthermore, the H4 atom of the Py**A4** unit has a larger positive ΔQ value than does the H3 atom of the Py**A3** unit.

Pyrimidylamide pairs. From the data of Table V.9 for the pyrimidylamide pairs Pa**A1**:**A2**Pa and Pa**A3**:**A4**Pa, one observes that for the first pair Pa**A1**:**A2**Pa, differences in atomic charge ΔQ are more marked for the atoms involved in the O2...H4-N4 and the O4...H2-N2 H-bonds than in the N3-H3...N H-bond. The

CHELPG charge differences for this bond (ΔQ = -0.039, 0.109 and -0.175 a.u. for the O2, H4 and N4 atoms respectively) may be compared with those for the O4...H2-N2 H-bond (ΔQ = -0.045, 0.104 and -0.072 a.u. for atoms O4, H2 and N2 respectively). Charge difference values for these two O...H-N H-bonds are somewhat comparable. Unlike the Py**A1**:**A2**Py pair, negative charge accumulates more on the nitrogen atoms for both the O2...H4-N4 and O4...H2-N2 H-bonds. The atomic charges on the central N3-H3...N3 H-bond show no reversal of the expected trend, where the ΔQ values are positive for the H3 atom and negative for the two nitrogen atoms, as may be expected.

Appreciable changes in atomic charges are also noticed for the O2...H4-N4 and the N4-H4...O2 H-bonds of the Pa**A3**:**A4**Pa pair. The CHELPG charge differences for the O2...H4-N4 bond (ΔQ = -0.087, 0.176 and -0.215 a.u. for the O2, H4 and N4 atoms respectively) are about the same order as for the N4-H4...O2 H-bond (ΔQ = -0.134, 0.093 and -0.025 a.u. for the N4, H4 and O2 atoms respectively), where the two nitrogen atoms on Pa**A3** and Pa**A4** are associated with the largest charge differences. The ΔQ values for these two H-bonds are generally larger than those for the central N3-H3...N3 H-bond (where ΔQ = -0.083, 0.066 and -0.071 a.u. for the N3, H3 and N3 atoms respectively), indicating that larger charge transfer occurs for the O2...H4-N4 and N4-H4...O2 H-bonds than for the central N3-H3...N3 H-bond. Charge differences for the O2...H4-N4 and N4-H4...O2 H-bonds of the Py**A3**:**A4**Py pair are also larger than those for the O2...H4-N4 and O4...H2-N2 H-bonds of the pyranonucleotide Py**A1**:**A2**Py pair just discussed above.

All these observations point towards the conclusion that the Pa**A3**:**A4**Pa pair would be more strongly H-bonded than the Pa**A1**:**A2**Pa pair, as is clearly and consistently

corroborated by the larger pairing energy for the former, expressed as the E_p, E_p(ZP) and ΔG_p terms of Table V.3. Table V.10 also shows that the net charge transfer as estimated by the B3LYP/6-31G* Mulliken method is smaller for the Pa**A1**:**A2**Pa pair (-0.0186 a.u.) than for the Pa**A3**:**A4**Pa pair (-0.0782 a.u.). Unlike for the nucleotide pairs discussed above, this is well reflected in the net charge transfer calculated by the CHELP, CHELPG and MK methods; all invariably point to greater charge transfer for the Pa**A3**:**A4**Pa pair than for the Pa**A1**:**A2**Pa pair. The Mulliken dipole moments (Table V.10) indicate that, for the Pa**A3**:**A4**Pa pair (but not the Pa**A1**:**A2**Pa pair), H-bonding leads to greater net charge separation in the H-bonded pair than in the solitary units, where the dipole moment for Pa**A3**:**A4**Pa is larger than for the constituent units Pa**A3** and Pa**A4**. The data regarding the Mulliken-based dipole moments also points to a larger dipole moment for the Pa**A3**:**A4**Pa pair (12.39 D) than for the Pa**A1**:**A2**Pa pair (4.31 D), which may be taken to mean that the net charge transfer is larger for the former than for the latter pair. This Mulliken-based charge data thus fits in well with the H-bonded pairing energies E_p, E_p(ZP) and ΔG_p of Table V.3.

The four methods thus predict the same direction of net charge transfer from one unit to the other during H-bonding for the Pa**A3**:**A4**Pa pair, but not always for the Pa**A1**:**A2**Pa pair. For the Pa**A3**:**A4**Pa pair, the direction is Pa**A4** to Pa**A3** since the O2 and N3 atoms of Pa**A3** have larger negative ΔQ values than the O2 and N3 atoms of Pa**A4,** while the H4 atom of Pa**A4** has the largest positive ΔQ value of all these atoms. However, for the PaA1:A2Pa pair, while the CHELPG and MK methods predict the direction of net charge transfer as Pa**A2** to Pa**A1**, the CHELP methods predicts the opposite direction. Taking the CHELPG and MK estimates as more reliable, the

direction Pa**A2** to Pa**A1** is considered as the better choice, but this is not fully corroborated by the various CHELPG ΔQ values of the relevant atoms. The large positive ΔQ values for the H4 and H2 atoms do, however, point towards this trend.

V.6 Towards the Polymeric Duplexes

The studies embodied within this Chapter deal only with the H-bonded monomer units of the actual macromolecular polymeric duplexes envisaged in this Dissertation. These are the two pyranonucleotide pairs PyA1:A2Py and PyA3:A4Py, and the two pyrimidylamide pairs PaA1:A2Pa and PaA3:A4Pa, each of them being a monomer of the H-bonded duplex.The grand aim of constructing the entire H-bonded duplexes is limited by the lack of appropriate means to conduct such an investigation. Clearly, the current quantum chemical theories in vogue, as contained in the GAUSSIAN 03 package, are not a practicable means for this, since the macromolecular duplexes (even if treated as oligomers of modest size) are far too large for the DFT and post-Hartree-Fock schemes to handle. Here, the recently developed semi-empirical PM6 SCF-MO method of the MOPAC 2009 package might seem more promising. Of course, the ideal instrument to tackle macromolecules of this type would be the classical potential strategies of the CHARm, AMBER and similar packages, which have force fields quite suitable and adequate to handle large systems of this type.

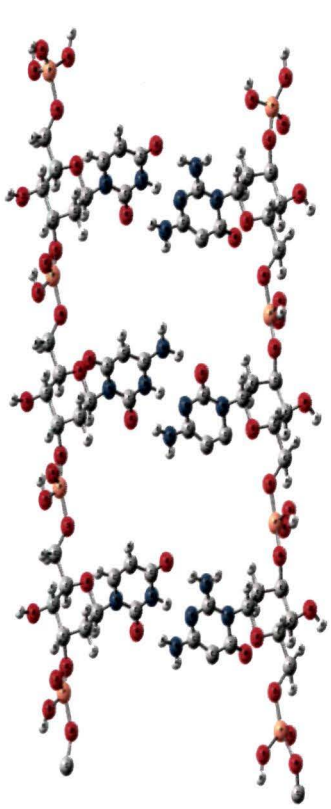
As we can see, the macromolecular duplexes envisaged here are of two types – the sugar phosphate type and the polyamide type. No attempts have been made here yet to picture the actual three-dimensional situation for these duplexes, since this would have to involve proper consideration of the various conformations that arise. The eventual result of all this is not yet possible to imagine. Even for DNA, it is not

possible to arrive at the three-dimensional double helical structure from consideration of only the mononucleotide pairs in themselves.

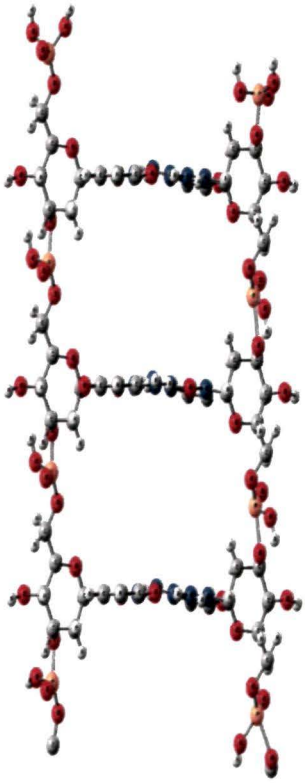
In the absence of adequate information to propose realistic three-dimensional structures for these artificial polymeric duplexes, one has hypothetically constructed a "ladder" kind of supramolecular structure, since there is no way to predict the relevant dihedrals at this stage. It is expected that this flat kind of "ladder" structure would be turned and twisted, but the precise mode of this is not possible to predict as of now.

Pyranonucleotide duplex. Fig. V.13 portrays the ladder structure for a trimeric H-bonded pyranonucleotide duplex, giving *(a)* the frontal (actually somewhat lateral) view showing the central H-bonded pairs, *(b)* the first side view, where only the base pair plane is seen from the side, and (c) the second side view, where the whole ladder structure is seen from the side. The trimer is seen to be evenly-spaced and it is expected that the actual pyranonucleotide duplex (even when it is realistically turned and twisted) would retain this evenly-spaced *periodicity* (which is also an important characteristic of the DNA duplex itself). This periodicity actually is a consequence of the *isomorphism* of the base pair units. The plane containing the base pairs is approximately perpendicular to the axes containing the two pyranose phosphate backbones. Due to lack of proper consideration of the actual conformations and the relevant dihedrals, the twisted three-dimensional conformation (or *secondary* structure) is not evident. Whether it would form an *even* (symmetrical) helix, a *double* helix, or perhaps *no* helix at all, is not possible to state. The studies of Eschenmoyer's group on synthetic p-RNA duplexes have, in fact, revealed *no helix* at all. Fig. V.14 goes on to depict a more extended *oligomeric* ladder-type structure for the pyranonucleotide H-

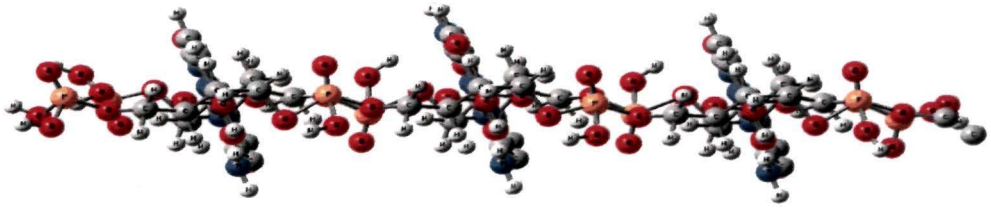
Ladder form of pyranonucleotides



3-mer (*frontal*)



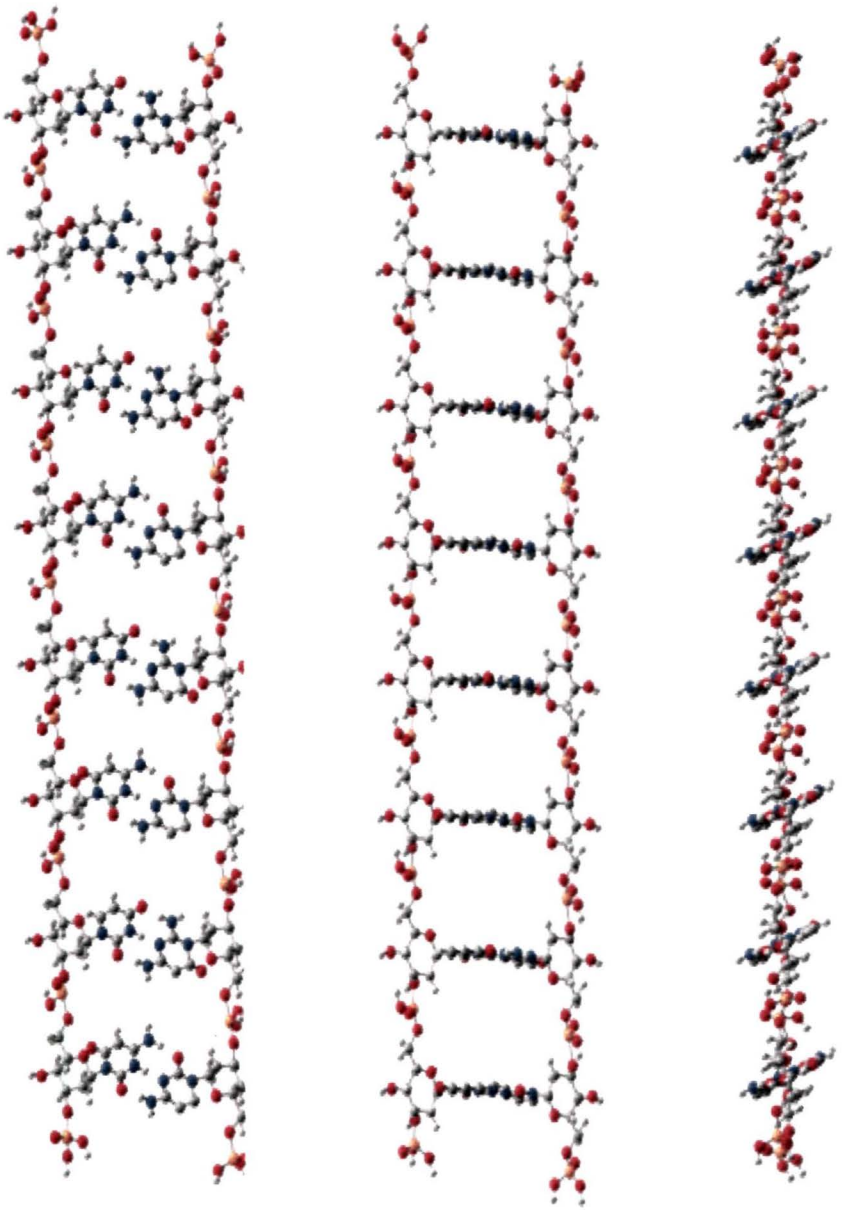
3-mer (*side*)



3-mer (*side/horizontal*)

Fig. V. 13 : Ladder structure for trimeric pyranonucleotide H-bonded duplex (frontal and side views)

Ladder form of polypyranonucleotides



8-mer (*frontal*) **8-mer (*side*)** **8-mer (*side*)**

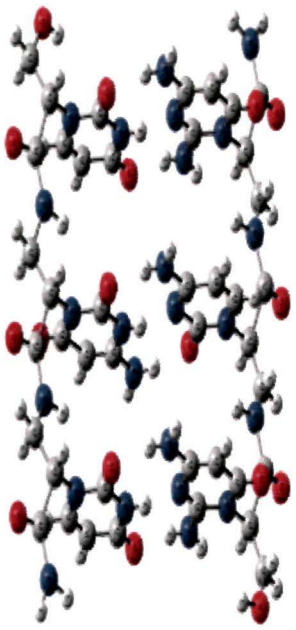
Fig. V. 14 : Ladder structure for oligomeric (8-mer) pyranonucleotide H-bonded duplex (frontal and side views)

bonded duplex, giving the frontal and the two side views, where the evenly-spaced periodicity becomes even more evidently visible.

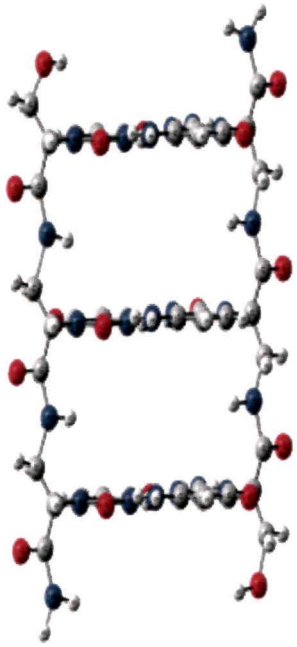
Pyrimidylamide duplex. Fig. V.15 portrays the "ladder" structure for a trimeric H-bonded pyrimidylamde duplex, depicting *(a)* the frontal view showing the central H-bonded pairs, *(b)* the first side view, where the base pair plane is seen from the side, and (c) the second side view, where the whole ladder structure is seen from the side. Like for the pyranonucleotide pairs, the trimer is evenly-spaced. This periodicity of structure is expected to be retained even when the duplex is turned and twisted). The plane containing the base pairs is approximately perpendicular to the axes containing the two amide backbones. Since the actual conformations and dihedrals are not incurporated here, the twisted three-dimensional conformation (or *secondary* structure) is not evident. Whether it would form a symmetrical helix, a *double* helix, or *no* helix at all, is not possible to state. Fig. V.16 depicts an extended *oligomeric* ladder structure for the pyrimidylamde duplex, giving the frontal and the two side views, where the periodicity is even more evidently visible.

Stacking structure. For both the pyranonucleotide and the pyrimidylamde H-bonded pairing systems, it is seen that the co-planar base pair moiety is more or less *perpendicular* to the axis contained in the backbone moiety. This indicates that the polymeric duplex in both cases would contain the successive base pair moieties all more or less *parallel* to one another. This means that, like for DNA, the adjacent base pairs would be able to be stabilized by *stacking interactions.* This stacking arrangement is also expected to be periodic, like it is in DNA. However, all these expectations require the

Ladder form of polyamide units



3-mer (*frontal*)

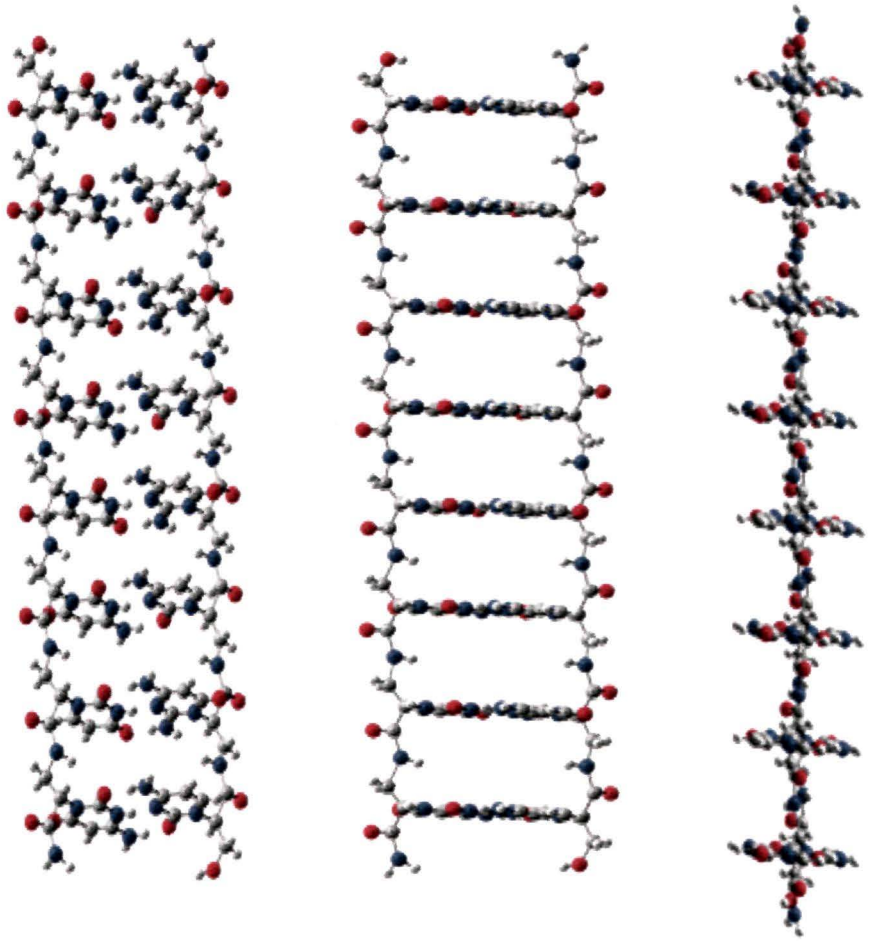


3-mer (*side*)

3-mer (*side/horizontal*)

Fig. V. 15 : Ladder structure for trimeric pyrimidylamide H-bonded pair duplex (frontal and side views)

Ladder form of polyamides units



8-mer (*frontal*) **8-mer (*side*)** **8-mer (*side*)**

Fig. V. 16 : Ladder structure for oligomeric (8-mer) pyrimidylamide H-bonded duplex (frontal and side views)

use of classical potential methods in order to ascertain the precise nature of these stacking interactions and the periodicity associated with the polymeric duplexes.

V.6 Conclusions

1. The backbones proposed, *viz.*, the *pyranonucleotide* and the *polyamide* backbones, have been tested by theoretical calculations and have been both found suitable.
2. Although these backbones are quite different structurally, they both lead to repeat units all of *similar structure and conformation* within each backbone type.
3. *Isomorphism* of H-bonded pairing is retained within the paired units for both types of back-boned pairs (nucleotide and amide).
4. H-bonded pairing *configuration* of the base pair moieties is *retained* throughout (whether solitary base pairs, pyrnonucleotide pairs or pyrimidylamide pairs).
5. The pyrimidylamide pairs are *more stable* thermodynamically (and more compactly bound) than the pyranonucleotide pairs, which predicts that the polyamide duplex would be more stable and compact than the pyranonucleotide duplex.
6. Although the three-dimensional secondary structure of these duplexes is not yet possible to predict fully, it is expected that they would form *periodic* structures with *evenly spaced* distribution of the repeat units along the duplex axis, together with a *stacking arrangement* of the base pairs which stabilizes the duplex..
7. *Charge transfer* for the three types of pairs (solitary base pairs, pyranonucleotide pairs and pyrimidylamide pairs) is generally predicted to occur in the directions **A2** → **A1** and **A3** → **A4** for pairs containing the **A1:A2** and **A3:A4** moieties.

References

1. Schlonvogg, I; Pitsch, S; Lesueur, C; Eschenmoser, A. *Helvet Chim Acta*, 1996, **79**, 2316.

2. Wipro, H; Kudick, R. A; Krishnamurthy, R; Eschenmoser, A. *Helvet Chim Acta*, 2001, **9**, 2411.

3. Jaun, B; Eschenmoser, A. *Helvet Chim Acta*, 2003, **84**, 1778.

4. A.D. Becke, *Phys Rev B*, 38, 3093, 1998.

5. A.D. Becke, *J Chem Phys*, 98, 5648, 1993.

6. C. Lee, W. Yang, R.G. Parr, *Phys Rev B*, 37 ,785, 1998.

7. Hehre, W. J; Radom, R; Schleyer, P. v. R, Pople, J. A. Ab Initio Molecular Orbital Theory, Wiley, New York, 1986, 253.

8. Curtiss, L. A; Raghavachari, K; Redfern, P. C; Rassolov, V, Pople, L. A. 1998, *J Chem Phys*, 1998, **109**, 7764

9. Frisch, M. J; Trucks, G. W; Schlegel, H. B; Scuseria, G. E; Robb, M. A; Cheeseman, J. R; Montgomery Jr, J. A; Vreven, T; Kudin, K. N; Burant, J. C; Millam, J. M; Iyengar, S. S; Tomasi, J; Barone, V; Mennucci, B; Cossi, M; Scalmani, G;Rega, N; Petersson, G. A; Nakatsuji, H; Hada, M; Ehara, M; Toyota, K; Fukuda, R; Hasegawa, J; Ishida, M; Nakajima, T; Honda, Y; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V; Adamo, C; Jaramillo, J; Gomperts, R; Stratmann, R. E; Yazyev, O; Austin, A. J; Cammi, R; Pomelli, C; Ochterksi, J. W; Ayala, P. Y; Morokuma, K; Voth, G. A; Salvador, P; Dannenberg, J. J; Zakrzewski, V. G; Dapprich, S; Daniels, A. D; Strain, M. C; Farkas, O; Malick, D. K; Rabuck, A. D; Raghavachari, K;

Foresman, J. B; Ortiz, J. V; Cui, Q; Baboul, A. G; Clifford, S; Cioslowski, J; Stefanov, B. B; Liu, G; Liashenko, A; Piskorz, P; Komaromi, I; Martin, R. L; Fox, D. J; Keith, T; Al-Laham, M. A; Peng, C. Y; Nanayakkara, A; Challacombe, M; Gill, P. M. W; Johnson, B; Chen, W; Wong, M. W; Gonzalez, C; Pople, J. A. Gaussian 03 Revision A.1, Gaussian, Inc., Pittsburgh PA, 2003.

10. Chirlian, L. E; Francl, M. M. *J Comp Chem* 1987, **8**, 894

11. Breneman, C. M; Wiberg, K. B. *J Comp Chem* 1990, **11**, 361.

12. Besler, B. H; Merz, K. M; Kollman, P. A. . *J Comp Chem* 1990, 11, 431.

13. Singh, U. C; Kollman, P. A. . *J Comp Chem* 1984, **5**, 129

A neutron goes into a bar and asks the bartender, "How much for a beer"? The bartender replies, "For you, no charge".

Two atoms run into each other. One atom says, "I think I lost an electron"

The second atom asks, "Are you sure"?

The first atom replies, "I'm positive"

CHAPTER SIX

LOOKING AHEAD !

V.1 Brief Summary of Work Accomplished

Let us now survey briefly the achievements of the work embodied within this Thesis, as recounted below as per the aims and objectives of the Dissertation, aiming towards the innovative design of novel and totally synthetic information-bearing H-bonded macromolecular duplexes:

. *(a)* ***First*** of all, out of a large number of candidate Sets of DNA base mimics (hetero-associative and self-associative), only *one* was chosen as the best ***DNA base mimic Set*** which satisfied all the criteria prescribed. This chosen Set included the *pyrmidine* bases **A1**, **A2**, **A3** and **A4** which successfully provided only the *two isomorphic base pairs* **A1:A2** and **A3:A4**. DFT calculations at the B3LYP level along with single-point MP2 calculations have established the optimized structure and stability (pairing energies) for these two pairs. This was accomplished in Chapters Three and Four.

(b) ***Secondly***, two types of ***backbones*** were tested out on these unique pairs, which are based on a *pyranonucleotide* and a *polyamide* structure. These led to the two pairs **PyA1:A2Py** and **PyA3:A4Py** built up from the pyranonucleotide repeat units; and to the two pairs **PaA1:A2Pa** and **PaA3:A4Pa** built up from the pyrimidylamide repeat units. Each set of two pairs was found to be *stable* and *isomorphic* (having the same pairing configuration), so that design of the backbone structure was very successfully worked out. In fact, it was found that the solitary pairs, the pyranonucleotide pairs and the pyrimidylamide pairs were *all* more or less isomorphic to one another.

(c) ***Thirdly***, the structure of the backboned base pairs was projected hypothetically to give some idea of the *polymeric H-bonded macroduplexes* themselves. It is predicted that the structures would be *evenly spaced* between the monomeric repeat units, and that the whole macromolecular structure would be *periodic* along the backbone axis, with successive base pairs parallel to each other to maximize *stacking* interactions.

VI.2 Conception of Further Work

This Dissertation has quite successfully wrought designs for the fundamental repeat units of two distinct types of information-bearing H-bonded duplex macromolecules. However, the precise structure of the actual polymeric duplex needs to be properly established through further computational work.

Actual *quantum chemical* calculation (using GAUSSIAN 03) can at most be capable of application only to the *dimeric* (or perhaps, the trimeric) H-bonded duplex. This would, of course, be almost prohibitively expensive, but may soon be possible here with the installation of the powerful PARAM computer system and of the parallel-computing version of the GAUSSIAN 03 software. Such accurate and very compute-intensive calculations would shed much light on the geometrical and conformational structure of the H-bonded macromolecular duplex and its charge distribution.

A much more practicable strategy would be to resort to the very well-established *classical potential* methods of the AMBER and CHARMm type of program software. Such software requires good initial estimates of the charge distribution (atomic point charges, bond dipoles etc.) as well as the geometrical structure of the repeat units in isolation or as the H-bonded pairs. The repeat units must also include the monomeric

backbone structures as well. Fortunately, this work of charge derivation and structure determination has already been accomplished through the calculations described in Chapter Five itself. This information serves as input data for calculations on the larger polymeric duplex systems using the AMBER 09 package recently installed here.

The actual three-dimensional (or secondary) structure of the polymeric duplex may be studied by AMBER one step at a time by building up first the trimer, then the hexamer, and finally the decameric or dodecameric forms of the macromolecule, beginning with the monomer or dimer structures obtained through accurate quantum chemical means. This successive building up (with energy minimization, geometry optimization and molecular dynamics simulation) eventually leads to a reliable and trustworthy model for the overall three-dimensional structure and conformation of the two types of information-bearing macromolecular H-bonded duplexes designed here. It would then be confirmed whether the duplex structure is ladder-type, helical or otherwise, which the present calculations of this Thesis cannot yet conclude.

A variety of base sequences may be worked out for the oligomeric ctructures treated in this fashion. The effects of base sequence upon the overall secondary structure are, of course, expected to be minor, since the fundamental base pair repeat units are isomorphic. Base sequence effects upon DNA structure have been well studied using experimental and theoretical tools of investigation

It is expected that such polymeric structures as calculated out here for these artificial systems would display all the desired features of periodicity, even-spaced distribution and base stacking interactions such as are characteristic of the DNA macromolecule of nature itself. All these features contribute to the stability of the macroduplex.

What follows next could be the grand attempt to actually *synthesize* the H-bonded duplexes themselves in the laboratory. This is beyond the means of this research group, and would have to be taken up through collaboration with other research groups which are engaged in synthetic organic chemistry and polymer chemistry. The first step is to synthesize the pyrimidine bases **A1**, **A2**, **A3** and **A4** and their back-boned counterparts. The next step would be to construct H-bonded duplex sections as per the methods employed by Eschenmoser and his group. Upon achieving complete synthesis of these H-bonded duplexes in oligomeric form, their structural characteristics may be experimentally determined and compared with the results of theoretical calculations. A good correspondence between experiment and theory here could constitute quite an important landmark in the world of chemistry today

The H-bonded macroduplexes designed here have four base units which form two unique base pairs (like in DNA). Depending upon the size of the code words desired, ***novel informational codes*** may be built up with singlet, doublet, triplet or quadruplet combinations of units to form code words. For our set of four basic informational units forming two unique pairs, if the code words are three letters long (as in DNA and mRNA), a total of 64 coding words results. A four-letter code word dictionary would comprise 254 coding words, and so on. In the context of application to the world of *information technology* at the nano level, special methods would have to be devised for the easy synthesis and replication of such macromolecular duplexes, in order to allow for the efficient storage, retrieval and copying of information. Such possibilities may be one day be successfully exploited as a result of the pioneer work carried out here and embodied in this Dissertation.

Curriculum Vitae

Siamkhanthang Neihsial

Date of birth	: 08-07-1976
Place of birth	: D.Phailian, Lamka, Churachandpur-795 158, Manipur, India
Nationality	: Indian
Marital Status	: Single
Present affiliation	: PhD Research Scholar, Department of Chemistry, North-Eastern Hill University, Shillong 793 022, India

Education and Degrees

Ph.D 2009: (PhD Thesis submitted)
North-Eastern Hill University, Shillong, India (Computational Chemistry) Thesis "*Quantum Chemical Studies on Hydrogen-bonded Base Pair and Backbone Components of Novel Information-bearing MacromolecularDuplexes*".
PhD Supervisor Prof. R. H. Duncan Lyngdoh

M.Sc 2003 : North-Eastern Hill University, Shillong, India (Project work in Physical Chemistry)

B.Sc 1999 : St. Edmunds College, Shillong, India. Majoring in Chemistry with Physics and Mathematics as allied subjects, securing 10th rank.

Fellowships/Awards

1. JRF/SRF Fellowship granted by the University Grant Commission, New Delhi under the programme Rajiv Gandhi National Fellowship Schemes for SC/ST. Candidates(RGNFS), 1st April 2005-1st Nov 2009.
2. Manipur Post Matric Scholarship, BSc, MSc, Ph.D 1st year
3. NEHU Post Graduate Meritorious Scholarship, 1st semester MSc

Professional Experience

1. Teaching Experience

1. Mathematics and Science tutor for high school and higher secondary students at Residency Hostel, Shillong (2001-2003).
2. Serve as an assistant teacher at St Albert's Residential School, Shillong. Taught Mathematics and Science. (2003)
3. Taught Younger Research Lab mates Linux, Quantum Chemistry, MOPAC, Gaussians (2004-2008)

2. Programming Experience

Linux, C, C++, FORTRAN

3. Research in Physical Chemistry:

Published one paper during MSc IVth Semester

4. Scientific Software

MOPAC, PC GAMESS, GAUSSIANS

Language Known

Read and write English, speak Hindi

Research Interests

Quantum chemical calculations (DFT and MP2) on nucleic acids components and design of novel information-bearing macromolecular duplexes.

Publications

(a) Full paper in Refereed Scientific Journals

1. Kinetics and Mechanism of the Oxidation of Substituted Benzoic acids by Quinolinium Dichromate
   Suante, H. Siamkhanthang, N. Lalnundanga, Mahanti, M. K.
   Oxidation Communications, 28, **2005**, 99-107

2. Novel H-bonded base pairs as potential repeat units for information-bearing

macromolecular duplexes: A B3LYP/6-31G* search
Siamkhanthang Neihsial, R.H. Duncan Lyngdoh *
J Molec Struct (THEOCHEM) 806, **2007,** 213-221

3. Novel H-bonded base dimers as repeat units for information-bearing self-associative duplexes: A B3LYP/6-31G* search
Siamkhanthang Neihsial, R.H. Duncan Lyngdoh *
J Comput Chem 29, **2008**, 1788-1797

4. Novel Information-bearing Macromolecular Duplexes Built from Pyranose Phosphate and Polyamide Backbones
Siamkhanthang Neihsial and R. H. Duncan Lyngdoh*
J Phys Chem A .(communicated)

(b) Abstracts in proceeding

1.Published abstract entitled " Novel H-bonded Base Pairs as Potential Repeat Units for Information-bearing Macromolecular Duplexes: A B3LYP/6-31G* Search" in the proceeding at the national Symposium on "Advances in Chemistry and Environmental Impact (ACE-2006)" , 2^{nd} -3^{rd} November, 2006. Department of Chemistry, North-Eastern Hill University, Shillong.

2. Published abstract entitled ""Novel H-bonded Base Dimers as Repeats Units for Information-bearing Self-associative Duplexes: A B3LYP/6-31G* Study" in the proceeding at the 2^{nd} Mid-Year Symposium of the Chemical Research Society of India", 21^{st} July, 2007, Department of Chemistry, Indian Institute of Technology, Guwahati.

3. Abtstract published in the proceeding at the 96^{th} Indian Science Congress, 3^{rd} -7^{th} Jan, 2009, North-Eastern Hill University, Shillong, entitled "Novel Information-bearing Macromolecular Duplexes Built from Pyranose Phosphate and Polyamide Backbones".

Presentations/Attended at Scientific Conferences/workshops

1. "Workshop on Molecular Modelling", 1^{st}-4^{th} December, 2005. Centre for Biotechnology, Anna University, Chennai. Got a participation certificate.

2. National Symposium on "Advances in Chemistry and Environmental Impact (ACE-2006)" , 2^{nd} -3^{rd} November, 2006. Department of Chemistry, North-Eastern Hill University, Shillong. Presented poster entitled "Novel H-bonded Base Pairs as Potential Repeat Units for Information-bearing Macromolecular Duplexes: A B3LYP/6-31G* Search".

3. "2^{nd} Mid-Year Symposium of the Chemical Research Society of India", 21^{st} July, 2007, Department of Chemistry, Indian Institute of Technology, Guwahati. Presented a poster entitled "Novel H-bonded Base Dimers as Repeats Units for Information-

bearing Self-associative Duplexes: A B3LYP/6-31G* Study". Got a participation certificate.

4. School on "Numerical Quantum Many-body Methods in Physics and Chemistry", 29th Oct- 4th Nov, 2007, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Jakkur Campus, Bangalore. Got a participation certificate.

5. 96th Indian Science Congress, 3rd -7th Jan, 2009, North-Eastern Hill University, Shillong. Presented a poster entitled "Novel Information-bearing Macromolecular Duplexes Built from Pyranose Phosphate and Polyamide Backbones". Got a participation certificate as well as a poster presentation certificate.

6. Training course on "In Silico Approach to Genome Analysis", 5th-11th feb, 2009, Bioinformatics Centre, North-Eastern Hill University, Shillong. Got a participation certificate.

7. Workshop on " High Performance Computing ", 3rd-6th Mar 2009, Computer Centre, NEHU, Shillong and Centre for Development of Advanced Computing (CDAC), Pune. Got a participation certificate.

Honours and Awards (Academic)

1. Awarded 10th rank in the BSc result by North-Eastern Hill University.

Membership

-Indian Science Congress Association (ISCA): Student Member during 2006-2007, Life Member since 2008

-World Association of Theoretically Oriented Chemist (WATOC) : Student Member since 2007