รายงานฉบับสมบูรณ์

โครงการ: การพัฒนา universal base โดยมี peptide nucleic acid เป็นองค์ประกอบหลัก เพื่อนำมาใช้ในการวินิจฉัยโรค

โดย ดร. อุทัย วิชัย และคณะ

กรกฎาคม 2554

สัญญาเลขที่ MRG4780184

รายงานฉบับสมบูรณ์

โครงการ: การพัฒนา universal base โดยมี peptide nucleic acid เป็นองค์ประกอบหลัก เพื่อนำมาใช้ในการวินิจฉัยโรค

คณะผู้วิจัย	สังกัด
1. ดร.อุทัย วิชัย	มหาวิทยาลัยนเรศวร
2. Prof. Stephen A. Woski	The University of Alabama

สนับสนุนโดยทบวงมหาวิทยาลัย และสำนักงานกองทุนสนับสนุนการวิจัย

(ความเห็นในรายงานนี้เป็นของผู้วิจัย ทบวงฯ และสกว. ไม่จำเป็นต้องเห็นด้วยเสมอไป)

ACKNOWLEDGEMENT

I would like to deeply thank my mentor, Prof. Dr. Stephen A. Woski for his support and guidance throughout the completion of my research. Many thanks are extended to Assistant Professor Dr.Yuthana Tantirungrotechai who supports for computational methodology and theoretical knowledge of DNA simulations and computational calculation in this investigation.

I would like to thank for the financial support from TRF (MRG 4780184) and Faculty of Science, Naresuan University.

Uthai Wichai

ABSTRACT

The pentadecamer geometries of DNA:DNA duplexes (5'-TAGTTGT**X**GCGTACA-3':3'-ATCAACA**Y**CGCATGT-5') containing carbazole derivatives (carbazole (CBZ), 3,6-dicyanocarbazole (DCC), 3,6-dicyanocarbazole (DNC), and 3-nitro-6-cyanocarbazole (NCC)) at the middle position (X) of DNA strand were simulated using AMBER 10. The missing parameters of carbazole derivatives were obtained by the RESP procedure. On simulation, the study systems were treated in explicit water under constant temperature and pressure for 6 ns. The equilibrium trajectories of most systems were found in the range of 3 to 4 ns. In the case of CBZ opposite natural nucleobases occurred the highest mobility, this was probably due to non-polar behavior. From simulated geometries, carbazole derivatives displayed unsuitable arrangement within modified DNA duplexes that could be discussed in terms of electrostatic interaction, steric volume, and polarization.

The potential universal bases of carbazole derivatives were evaluated with calculation of interaction energies between carbazole derivatives and neighboring nucleobases by using the PM6 and DFTB methodologies. These methods predicted the universality and suggested that DNC was the most universality. Furthermore, this investigation found that the interaction ability of carbazole derivatives was also related to their dipole moment.

Keywords Molecular dynamics simulation, Universal base, Carbazole derivative, Stacking interaction

CHAPTER I

INTRODUCTION

Rationale for the study

Currently, universal bases, a chemically modified nucleobase, are attracted much attention due to the equally hybridization with any four natural bases. Particularly, they are widely used as potential tools for many applications that depend on the selective pairing against natural nucleobases. For examples, polymerase chain reaction (PCR), hybridization probes, and therapeutic applications [1-9]. Frequently, the indecisive information occurs from ambiguous bases during DNA sequencing which are an account of the degeneracy in the genetic code and the presence of a polymorphism. Consequently, universal bases are useful for their applications to minimize base-discrimination of ambiguous bases that were detected in the target sequence. In all of their applications, the potential universal base should be indiscriminately paired with each of the natural bases and can be incorporated without significantly destabilizing of the double helical structure.

Prevalently, only two universal bases, 3-nitropyrrole (**1**, Figure 1) and 5-nitroindole (**2**, Figure 1), have been approved for commercial and it was shown that **2** is the best candidate as a universal base thus far. For example, **2** has been successfully used as hybridization probe for the detection of single nucleotide polymorphism [10]. Although this analogue has a great potential in the universal behavior, its incorporation still destabilized the duplexes.

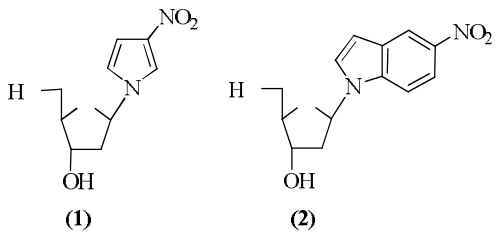


Figure 1 Chemical structure of 3-nitropyrrole and 5-nitroindole monomer

Thus far, an alternative solution to design novel universal base was investigated in order to improve the universal base properties. The design of an ideal universal base *via* hydrogen bonding matching-position and base stacking approaches has been utilized and the latter seemed to receive the much attention for the novel base design. Generally, the design of universal base is to concentrate on base stacking as the major dominant for stabilizing effect of duplex structure. There are three different forces that assist to stacking interaction including; *van der Waals* interactions, desolvation, and electrostatic effects. Additionally, dipole moment is also another important factor that contributes to binding property and selectivity of universal base.

Previously, 3-nitrocarbazole (MNC) and 3,6-dinitrcarbazole (DNC) have been incorporated in PNA system and were investigated for their universal base properties compared with 5-nitroindole (NI) [11]. These carbazole derivatives were chosen because of theirs large surface area and the presence of the high electronegativity group provided the large dipole vector with both inward and outward direction to the aromatic moiety. Consequently, these properties should affect the enhancement of stacking ability. As the results, it was found that DNC has the highest average T_m (61.6 °C) and NI has the lowest ΔT_m (0.9 °C), while ΔT_m of DNC was 2.9 °C. These indicated that DNC has higher binding affinity and showed lower non-discrimination in its base pairing properties when compared with NI.

Therefore, NI, MNC, and DNC were also performed in the context of DNA:DNA for their future applications [12]. From this investigation, DNC still maintained the highest average T_m (50 °C), this revealed that DNC was more stable than NI (average T_m was 47 °C) and MNC (average T_m was 48.5 °C). These results displayed that introducing the nitro group strongly polarized the carbazole moiety and resulted in better stacking interactions. In addition, it was found that MNC have lower ΔT_m (1.0 °C) than DNC and NI (ΔT_m = 1.8 °C). Therefore, these could be summarized that DNC exhibited higher stability during incorporation step than the others, however its universality was poor. This was probably a cause from the steric volume of nitro group.

In order to improve the universal base properties of this analogue, reduction of the steric volume of nucleobase analogues by replacement of the nitro group with other groups having lower steric volume and maintaining the high dipole moment is our target.

Under this study, modified carbazoles were designed as universal bases. In order to achieve the effective design for universal bases, surrounding interactions between modified base and natural nucleobases must be clarified. However, these interactions such as stacking interaction is still difficult to obtained from experiments, computational chemistry can be use to investigate their systems in order to primarily screen the potential universal base of modified bases. Here, the dipole moment of each modified carbazoles was initially computed by using the Guassiun03 program. Next, the target modified carbazoles were chosen by considering from their calculated dipole moments. In the next stage, molecular dynamics (MD) simulation was performed using the AMBER software package to investigate the influence of the chosen carbazole derivatives that affect to the overall stability of the DNA duplex structure. Finally, the interaction energies of modified carbazoles were calculated from their MD production runs by employing the PM6 and DFTB semi-empirical model.

Purpose of the study

To examine the universality of carbazole derivatives by consideration from computed dipole moments and interaction energies.

Significance of the study

The information from this study will be an important data to explain the mechanism of stacking interaction of modified carbazoles within DNA duplexes that affect to the stability of overall duplex structure. Additionally, these data will lead to the efficient design of the novel universal bases for further synthesis and application in experiment.

Scope of the study

The dipole moment of carbazole derivatives including carbazole, 3-cyano-carbazole, 3,6-dicyanocarbazole, 3-nitrocarbazole, 3,6-dinitrocarbazole, and 3-nitro-6-cyanocarbazole were firstly computed, the second was MD simulation study of the pentadecamer DNA duplexes containing modified carbazoles and the last step was calculation of the interaction energies of carbazole derivatives from simulated structural geometries of modified pentadecamer DNA duplex.

CHAPTER II

LITERATURE REVIEW

The design of the ideal universal base requires the appropriate research methodology that can experimentally observe interactions between designed bases and neighboring nucleobases within DNA duplex. Therefore, the information and background regarding DNA and factors that contribute to stabilize the duplex formation are necessary. In this chapter, characteristic of DNA, significant factors to DNA duplex stability, aromatic stacking interaction, universal bases and their applications, and method overviews will be discussed.

Deoxyribonucleic acid (DNA)

Deoxyribonucleic acid (DNA) is a macromolecule that contains genetic information and is transmitted from one generation to the next. Its primary structure is a polymer of 2′-deoxyribonucleotide repeating units and each unit joins to neighbor nucleotides through phosphodiester linkage from its 5′-hydroxyl group to the 3′-hydroxyl group. In addition, the monomer unit of DNA consist of three components; heterocyclic nitrogenous bases, pentafuranose sugars, and phosphate groups. In the nucleotides, the bases are either purines (adenine (A) and guanine (G)) or pyrimidines (thymine (T) and cytosine (C)) that are linked to the C-1′ of the sugar *via* *N*-glycosidic bond in the β configuration (see Figure 2).

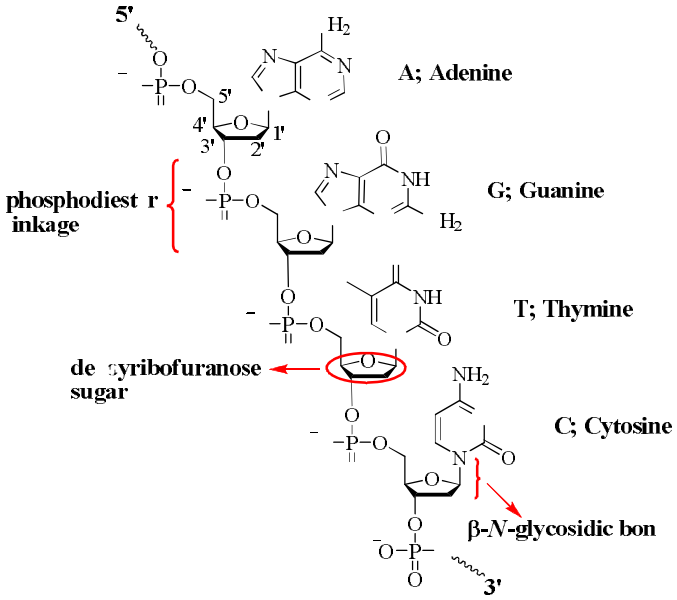


Figure 2 Structure of DNA-single strand

Watson and Crick proposed the model of a double helix (the secondary structure of DNA) in 1953 (Figure 3), where adenine pairs specifically with thymine and guanine specifically with cytosine *via* hydrogen bonding and they are called “Watson-Crick base-pairing” (Figure 4) [13].

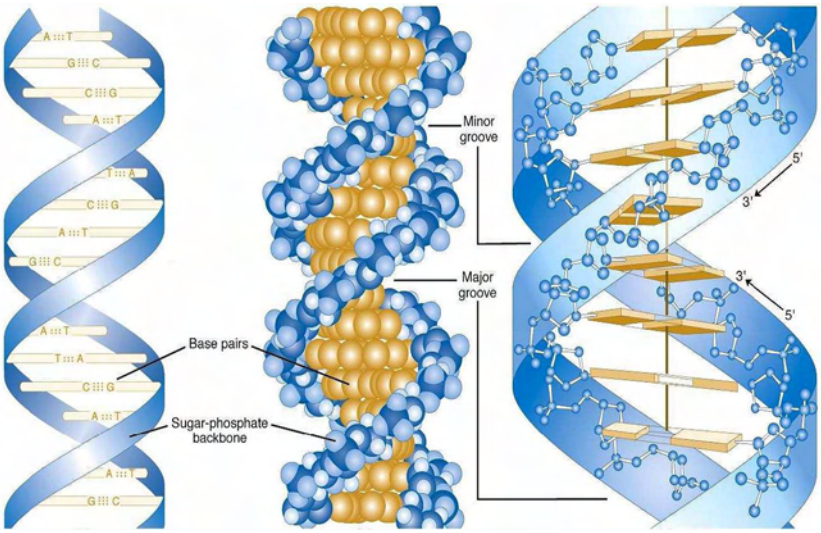


Figure 3 DNA double helix formation

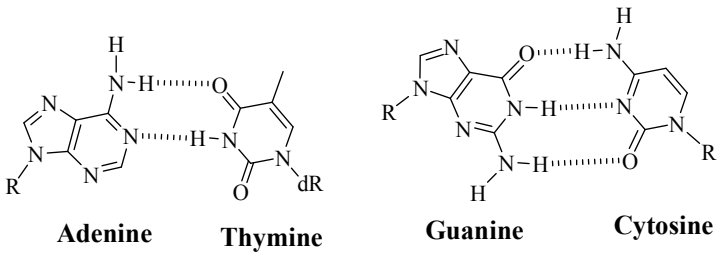


Figure 4 Hydrogen bonded (Watson-Crick base pairing) A:T and G:C base pairs in DNA

Additionally, the nucleobases form the double helix structure by inter-molecular hydrogen bonding between complementary bases in the interior and sugar phosphate backbone extends along the outside in anti-parallel manner (Figure 3) Several distinct classes of helical DNA structures have been identified [14-16] depending upon their environment such as relative humidity, types of salt, ionic strength, and the solvent. DNA is now known to exist in three conformers which are denoted as A-DNA, B-DNA, and Z-DNA (Figure 5).

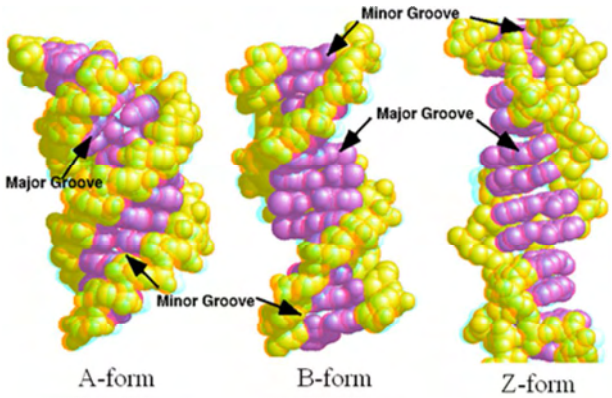


Figure 5 Double helical DNA: A-form, B-form and Z-form

B-DNA have 10 bases per turn with little tilting of the bases to enhance the base stacking whereas A-type helix has 11 bases per turn. The major groove of B-DNA is

wide and the minor groove is narrow, both are of moderate depth and are well solvated by water molecules. Furthermore, the B form is sufficiently flexible to permit a conformational response in the backbone. An important difference between the A- and B-form duplexes is the sugar puckering mode (Figure 6). In an A-type duplex, the sugar is a C(3′)-*endo* conformation, whiles in a B-type duplex is a C(2′)-*endo* conformation [17].

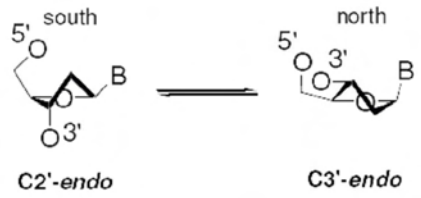


Figure 6 The sugar puckering modes

Factors to double helix stability

The DNA double helix is much more rigid than its component single strands. However, the negatively charged phosphate groups that have a lowest effect on one another are all located on the outer surface. Therefore, the repulsive electrostatic interaction from these charged groups are partially neutralized by interaction with cations such as Mg^{2+}, basic polyamines such as putrecine and spermidine, and the positively charged side chains of chromosomal proteins. Mainly, the structure of DNA duplex is governed by a balance of non-covalent forces in aqueous solution [18]. In fact, the core of the double helix is the base pair holding together by the specific hydrogen bonding. Although hydrogen bonding is a significant factor in DNA bases pairing specificity, there is also suggested that the stacking between the planes of adjacent DNA bases primarily contributes to the overall stabilization of the double helix [19]. There were several major non-covalent interactions that result in stacking energy including *van der Waals* interactions, electrostatic effects and desolvation. Furthermore, dipole moment was also another important factor that contributed to stacking interaction and increased binding property and selectivity of universal bases.

Aromatic stacking interaction

Generally, aromatic stacking interaction (or π-π stacking interaction), and also called “phenyl stacking”, is non-covalent interaction between aromatic moieties. Overall, π-π interaction is caused by intermolecular overlapping of p-orbital in π-conjugated systems. These interactions are made up from a combination of many forces that are discussed in this following.

1. A theoretical model for aromatic interaction

There are several important factors that contribute to the stacking interaction including *van der Waals* interactions, electrostatic interactions, induction energy, charge transfer and desolvation. However, consideration of relative effect of each factor on the interaction is important in order to understand aromatic stacking interaction.

1.1 *Van der Waals* interactions

Van der Waals interactions arise from overlapping between the large planar surfaces of the stacked aromatic bases, and make a favorable energetic contribution to the stacked helix. In most case of bases, they have tend to stack in an offset configuration rather than direct π-π overlap, while *van der Waals* interactions is unlikely the main contributor to stacking interaction because *van der Waals* area is small.

1.2 Electrostatic interactions

Electrostatic interaction is the interaction between the static molecular charge distributions. Several geometries have been proposed on the basic of the electrostatic component of the interaction by Hunter and Saunders in 1990 [20] (Figure 7).

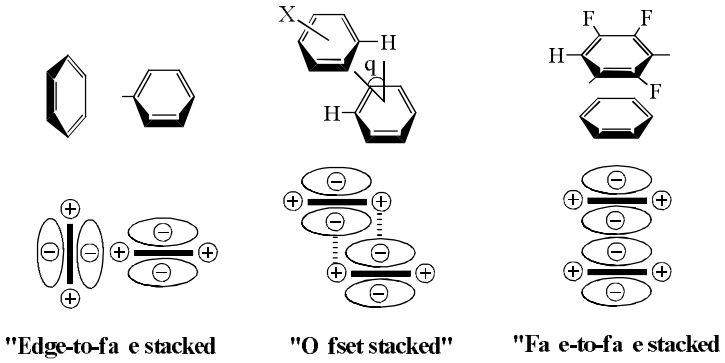


Figure 7 Basic types of stacking patterns; edge-to-face, offset and face-to-face stacked

The π-electron density above and below the planes of the aromatic units is the featured aromatic interaction of this model. The aromatic ring is portrayed as a positively charged σ-framework that is sandwiched between two regions of negatively charged π-electron density. In Figure 7, the edge-to-face geometry (or called T-shaped) is considered as a C-H-π interaction which is commonly observed between aromatic residues in proteins. The offset stacked orientation (Figure 7) is also commonly found in proteins and in DNA. The *van der Waals*, in the case of the offset stacked, does not depend on surface area of ring. Commonly, this orientation would be appeared when the electron density on the face of one or both ring was reduced. A third geometry is the face-to-face stacked orientation (Figure 7). This is commonly observed with donor-acceptor pair and compounds that have opposite quadrupole moments. For an excellent example of this type of aromatic interaction is the benzene-perfluorobenzene interaction. REF

Generally, this proposed model indicated that the geometries of π-π interactions were controlled by electrostatic interaction and *van der Waals* interaction. According to the proposed model of Hunter and Saunders, there was implied that the electron-donor-acceptor concept might be mislead when used for π-π interactions. On the other hand, since the properties of atoms at the point of intermolecular contact rather

than the overall molecular properties, therefore this might be important contribution for the net favorable abilities of π-π interactions.

Cozzi and Siegel showed the evidence that the electrostatics was the importance influence on aromatic stacking [21-23]. They created a model that consisted of substituted 1,8-diarylnaphtalene molecules (Figure 8) in chloroform for investigating the influence of substituted group (X) on aromatic stacking interaction by using dynamic NMR studies. From the study, the activation energy for the isomerism was provided to evaluate the strength of the aromatic interaction between the stacked phenyl rings in the ground state, as shown in Table 1.

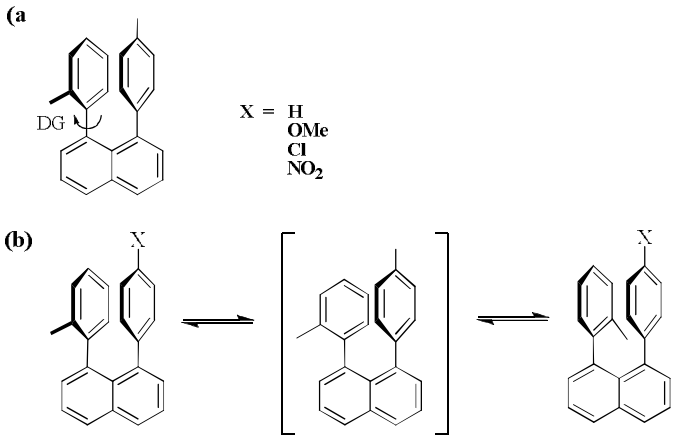


Figure 8 The model for used to investigate the substituent effects on aromatic stacking interactions, (a) is the model of substituted 1,8-diarylnaphtalenes and (b) is the dynamic equilibrium model

Table 1 The activation energies of barriers to rotation substituted 1,8-diaryl-naphthalene molecules

Substituents	$\Delta G^{\ddagger}$/kJ mol^{-1}
OMe	58.2
Me	60.2
H	61.5
Cl	64.9
CO_2Me	70.7
NO_2	72.4

From this study, it was found that an electron withdrawing substituent (e.g. -NO_2) decreased the electron density therefore; decreased the π-electron repulsion, whereas an electron donating substituent (e.g. -OMe) increased the electron density. This led to an increase the π-repulsion and thereby to a less favorable interaction due to electronic repulsion (Figure 9).

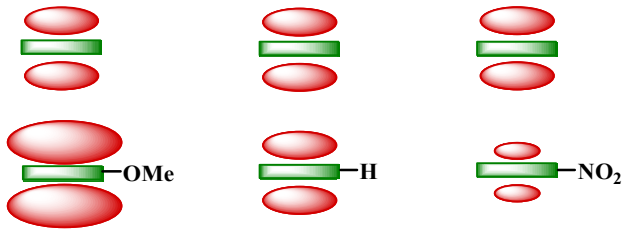


Figure 9 Schematic representation of the effect of substituent on stacking interactions

Addition to non-polarized π-systems by either a substituent or a heteroatom, the dominant interaction was π-electron repulsion, therefore an electron deficient π-system could be stabilized the interaction because the repulsion was decreased.

1.3 Induction

Induction energy is the interaction between the static molecular charge distribution of one molecule and the induced charge distribution of the other. There is a few experimental suggested that induction effects were important in aromatic interactions [24].

1.4 Desolvation

Two molecules which form a complex in solution must be desolvated before complexation can occur. Additionally, the solvent may be completed for recognition site, and also destabilize the complex. Alternatively in polar solvent, consequence of solvophobic effects can stabilize the complex. Therefore, the flat π-electron surface of aromatic molecules are non-polar, so that solvophobic forces are favorable for stacking. There was study of solvation effects on aromatic interactions in host-guest systems of pyrene and cyclophane by Diederich and coworkers (Figure 10) [25]. The stabilities of the complexes and the solvent polarity had linear relationship by description of the empirical $E_T(30)$ parameter (Figure 10). Diederich's model displayed that solvent properties appeared to be the most important for determination of the strength of non polar host–guest complexation. In polar solvent, stronger binding can be occurred. Due to high cohesive interactions solvents could interact more strongly with "like" bulk solvent than with the non polar surfaces of the host and guest molecules. When complexation took place, free energy was gained with the release of surface-solvating molecules to bulk solvent. Therefore, Diederich concluded that water was the best solvent for non polar binding.

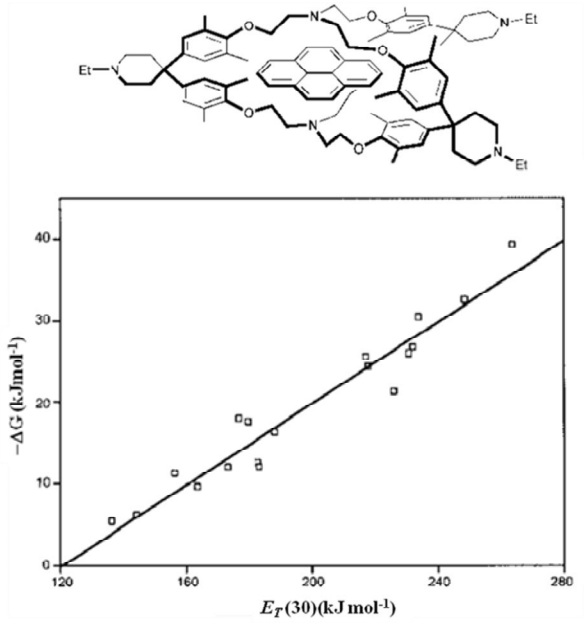


Figure 10 Relationship between the free energy of cyclophane·pyrene complexation of, ΔG (kJ mol^{-1}) and the solvent polarity, $E_T(30)$ (kJ mol^{-1})

1.5 Charge-transfer interactions

Charge-transfer (CT) is stabilization due to the mixing of the ground state (AB) with an excited charge-separate state (A^+B^-). Although CT bands are common in aromatic complexes, this is rarely observed. Theoretical calculations suggested that these effects made a very small contribution to the stability of the ground state of molecular complexes [24].

2. Base stacking in DNA

Base stacking is the non-covalent interaction that is actually a complex interaction. Base stacking can be explained in a very simple way by comparison base stacking with coins that stack on top of each other [26]. Of course, this picture is very simplified and high resolution structures shows that base-base stacking interaction are usually associated with an offset rather than a face-to-face geometry [20], and this offset between two aromatic planes in a stacked structure is 3.4 Å, this is the *van der Waals* distance, where π-repulsion is minimized, and this distance corresponds to the rise of

the B-DNA helix per base pair. The twist of one base pair relative to its π-stacked neighbor about the helix axis was calculated that it was 30°– 40° and it was shown that this twist would be optimized the π-π interaction [20]. Several non-covalent forces should be considered and all forces are likely played an important role such as electrostatic forces, which includes *van der Waals* forces and solvation effect undoubtedly influence the stacking stability [27]. Among of all the forces, the electrostatic forces have been investigated most intensively because electrostatic effects are considerably easier to computationally model than those in both *van der Waals* forces and solution effects.

In the 1996, Kool and coworkers addressed base stacking of natural and non-natural bases so called “dangling end” experiments [28, 29]. The nucleoside analogues of interest (Figure 11) were placed at the ends of a base-paired duplex where it remained unpaired, meaning it has no hydrogen bonded partner (Figure 12). The resulting stabilization of the duplex was measured by thermal denaturation experiments and was compared with the stability of the duplex lacking the end-dangling nucleotide. The energetic of stacking was also measured separately from hydrogen bonding.

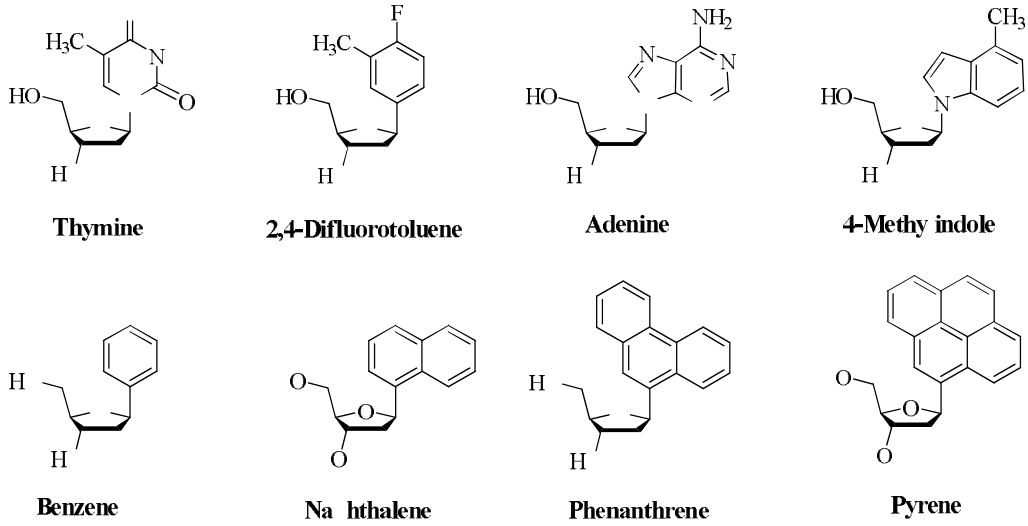


Figure 11 Structure and DNA sequences of “dangling end” experiments

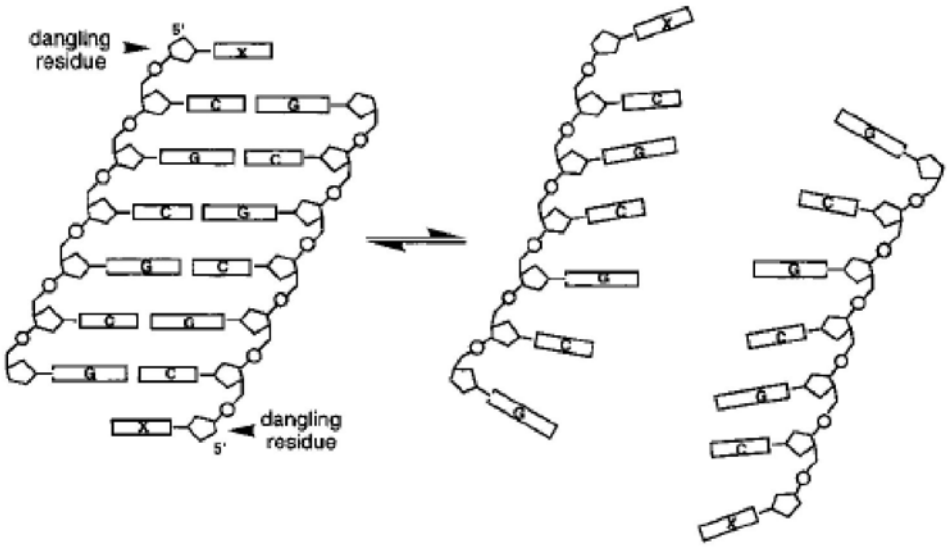


Figure 12 Illustration of dangling ends in a duplex (left) and when melted into two single strands (right)

Kool and coworker summarized and interpreted that the purines stacked more strongly than the pyrimidines, presumably due to their larger surface area and greater polarizability. The difluorotoluene (F), a non-polar base shape mimic of thymine (T), stacked considerably more strongly than T, even though they have the same size, shape and surface area. These results suggested that surface area alone was not a good predictor of stacking affinity. It seemed likely that the increased hydrophobicity of non-polar analogues favorably contributed to enhance their stacking. An alternative explanation was the difference in polarizability which would influence the *van der Waals* forces.

To further investigate the effect of size and hydrophobicity on stacking interactions, non-natural nucleosides such as benzene, naphthalene, phenanthrene and pyrene were placed on such a “dangling end” position. The overall order of stacking ability of the tested molecules was found to be thymine < benzene < adenine < F $\leq$ naphthalene ~ phenanthrene $\leq$ methylindole < pyrene, in the given context. Solvophobic effects would appear to be the most important factor in determining this order. When molecules of the same size (T and F) were compared, the less polar structure stacked more strongly than that of higher polarity. When non-polar molecules of different size

were compared, it was found that the propensity for stacking appeared to correlate with the surface area excluded from water.

In addition to most recent study, Kool and coworkers tried to correlate the stacking energies of natural and non-natural nucleosides with hydrophobicity, dipole moment, polarizability, and surface area by using “dangling end” experiments and the obtained values are depicted in Figure 13. In the following, they used a measurement of solvent driven effect (log *P*), overlapping surface area, dipole moment, and polarizability reflecting the importance of electrostatic effect on base stacking.

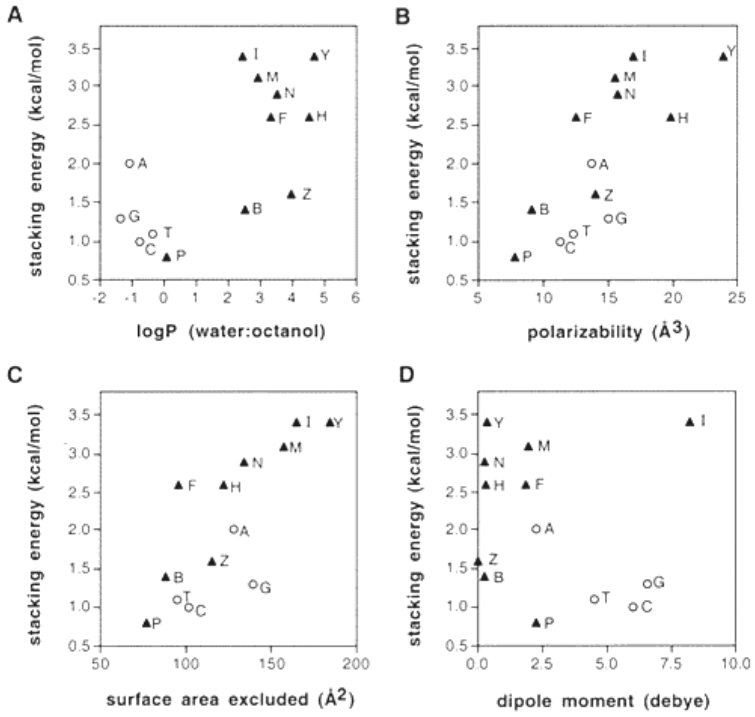


Figure 13 Relationship between the stacking free energies and calculated physical properties of the DNA bases and aromatic analogues, abbreviations are as follows: A (adenine), B (benzene), C (cytosine), F (difluorotoluene), G (guanine), H (phenanthrene), I (5-nitroindole), M (4-methylindol), N (naphthalene), P (pyrrole), T (thymine), Y (pyrene), and Z (trimethylbenzene)

According to these obtained results, Kool and coworker concluded that the different forces may contribute to the aromatic stacking in aqueous solution in the context of DNA structure including:

1. Electrostatic effects are devised into three kinds of interactions:

1.1 dipole-induced-dipole interaction, which is this interaction do not appear to be a dominant factor, at least not for the majority of the studied species.

1.2 the permanent interaction between partial charges (dipole-dipole), which can stabilize or destabilize depending on the orientation and magnitude of local bond polarization for each stacking participant.

1.3 the quadrupolar effects, which has been suggested to be important when aromatic species are adjacent in face-to-face orientation.

2. Dispersion forces

Dispersion can be contributed to stacking of aromatic systems in aqueous solution. Such attractive forces depend on overlap surface area (geometry dependent) and on the polarizability of two molecules. However, the magnitude of these forces seems to be rather insignificant when compared with other factors.

3. Solvation effects

These effects contribute quite strongly to stacking of the “dangling ends”. Solvent-driven effects depend on the relative energies of solvation of bases in the stacked or unstacked arrangement as well as on the amount of surface area which is desolvated upon stacking in the helix. This leads to cause the system to minimize its contact surface with water. From this study, it was found that the pyrene-DNA interaction was shown to be more significantly sensitive to ethanol than the interactions of the natural bases (Figure 14).

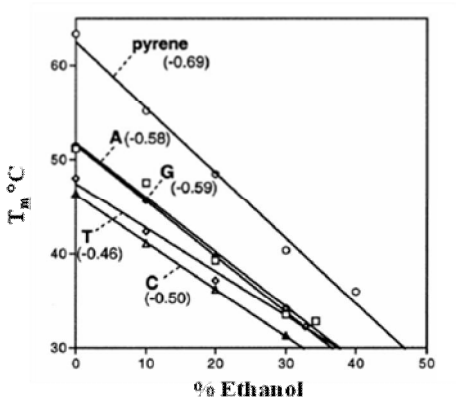


Figure 14 The effect of added ethanol on the T_m for the four natural base and pyrene

Generally, it was concluded that hydrophobic effects were larger stabilizing the stacking than other effects (electrostatic effects, dispersion forces).

Recently, Wichai [30] investigated the effect of dipole moment on stability of phenyl derivatives within DNA:PNA duplex structure. From this study, it was found that adding electron-withdrawing group, such as nitro group, strongly polarized aromatic moiety and improved base stacking interaction, as follows in Figure 15. Additionally, this received result could be reconfirmed previous observations that the dipole moment strongly affects on overall duplex stability.

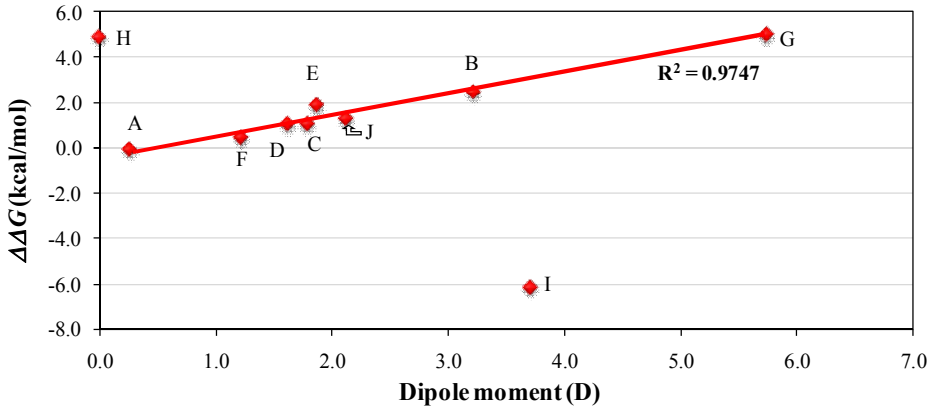


Figure 15 Effect of residue dipole moment on stacking energies ($\Delta\Delta G$) of modified phenyl (A = Phenyl, B = 4-Acetophenone, C = 4-Bromophenyl, D = 4-Chlorophenyl, E = 4-Fluorophenyl, F = 4-Methoxyphenyl, G = 4-Nitrophenyl, H = 4-Tolyl, I = 4-Trifluoromethylphenyl, J = Pentafluorophenyl)

Currently, no theoretical study exists which fully agree with the observed experimental data for natural or non-natural DNA bases, further studies are still required in order to gain future insight into how the dispersive, electrostatic and solvation forces cooperate to stabilize the stacking of the DNA bases.

Universal bases and their applications

There are many applications of biomolecular technique and medicinal chemistry that rely on the selective pairing of the natural DNA nucleobases. Accordingly, pairing between the standard DNA bases strongly favors A:T and G:C aliment. During DNA sequencing of these applications, the problem from ambiguous base position occurs frequently. Consequently, the potential solution to this problem is applied a universal base that shows non-discrimination in hybridization against all of both natural DNA and RNA bases [31-33]. The development of universal bases for incorporation into DNA has been designed based on two basic strategies including the first aims to maintain a Watson-Crick-like network of hydrogen bonding between base pairs, while the second strategy is mimicking the shape and size of natural nucleobases by using hydrophobic

groups to enhance base stacking interaction. Generally, universal bases are characterized as hydrophobic base analogues, non-hydrogen bonding. These mispairing of universal bases within a DNA duplex and loss of solvation are energetically unfavorable causing duplex structure become unstable. Therefore, overall duplex stability is compensated between the stacking and hydrophobic interaction.

Although there are many other nucleobase analogues that has been described in previous researches, the two universal bases that have received most attention are 3-nitropyrrole (**1**) and 5-nitroindole (**2**), as shown in Figure 16. Based on their universality, these bases and their derivatives has been incorporated into DNA to apply in many applications, namely they are widely use as potent tools such as modified primer for PCR, hybridization probes for DNA analysis, and diagnostic tools for gene therapy.

The first universal base receiving prominent attention was 3-nitropyrrole (**1**, Figure 16), described by Bergstrom and co-worker [34-36]. The design of this analogue was aimed to maximize base stacking interaction. The presence of the nitro group enhanced stacking by polarization of the π-aromatic system of the pyrrole ring. This analogue behaved indiscriminately in base pairing with the four natural bases; however it caused to destabilize in duplexes [34]. Then, a related analogue to **1,** 5-nitroindole (**2**), (Figure 15) was described by Loakes and co-worker [37]. Interestingly, 5-nitroindole was shown to be more stabilizing than 3-nitropyrrole when positioned opposite the four natural bases [37].

In addition, the 4-nitroindole (**3**) and 6-nitroindole (**4**) derivatives [38] (Figure 16) have been studied in melting experiments [37]. However, these derivatives were more destabilizing than 5-nitroindole. When contiguous substitutions of 4-**,** 5- and 6-nitroindole were incorporated into a duplex, the T_m value remained constant, while the stacking enthalpy (ΔH) increased. Conversely, 3-nitroindole did not increase in stacking enthalpy with contiguous substitution and the T_m value decreased with each additional substitution. This indicated that 5-nitroindole was better stacking within the duplex than 3-nitropyrrole.

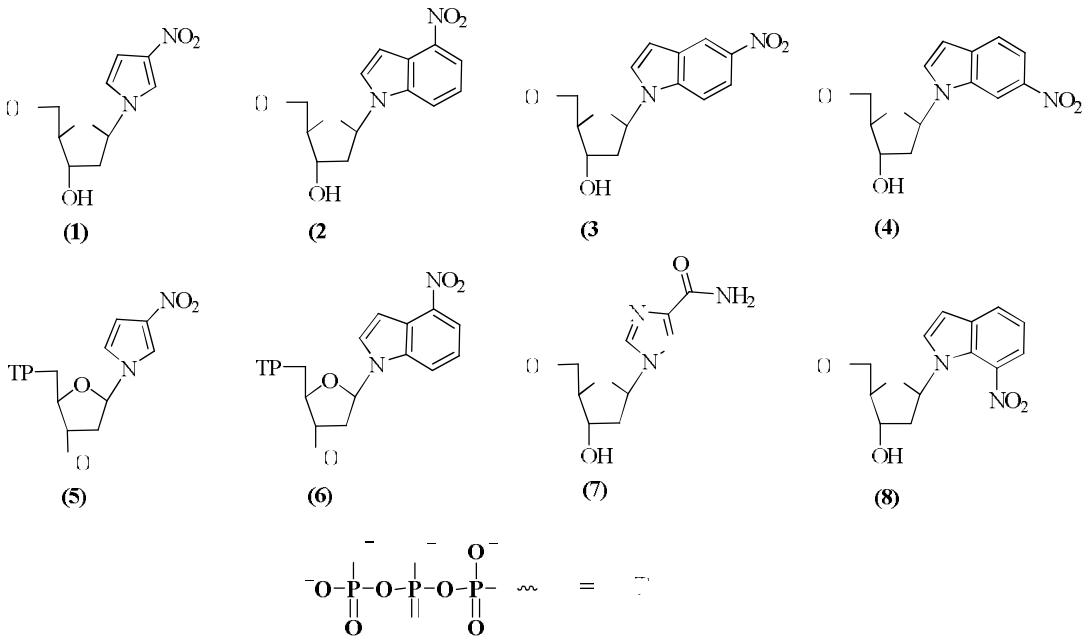


Figure 16 Chemical structures of nucleoside analogues containing universal base

Luo and coworker has been used 3-nitropyrrole to enhance mismatch discrimination to isolate and characterize mutant ligases [39]. In addition, Bergner and coworker similarly employed 3-nitropyrrole for the detection of single nucleotide polymorphisms (SNPs). They observed that mismatch stability was increased in the presence of 3-nitropyrrole, and suggested that could be used as a hybridization probe for high-throughput genotyping of SNPs [40-44]. Additionally, 5-nitropyrole was successfully used to analyze the structure of DNA hairpins [45] and used as hybridization probes for the detection of single nucleotide polymorphisms [24].

Recently, Peterson *et al* has been investigated 3-nitropyrrole and the corresponding triphosphate (**5**; Figure 15) as a lethal mutagen against poliovirus [7]. This application is called “lethal mutagenesis” that is increasing the viral mutation rate beyond a biologically-tolerable threshold, resulting in reduction and extinction of potential virus (or viral fitness). They found that 3-nitropyrrole did not exhibited any significant antiviral effect against poliovirus because it was not accepted by poliovirus RNA dependent RNA polymerase (RdRP). Later on, Peterson and coworker have also been investigated 5-nitroindole and its triphosphate (**6;** Figure 16) to function as lethal mutagens in poliovirus

[9]. They reported that 5-nitroindole triphosphate (**6**) could be universally incorporated against each natural RNA bases by the RdRP from poliovirus and it inhibited this enzyme more potently than the triphosphate metabolite of the commercial available antiviral drug, ribavirin (**7;** Figure 16) which were also universal base [47].

Furthermore, a relate isomer to 5-nitroindole, 7-nitroindole (**8**) having photochemically cleavable property and frequently employing for DNA analysis was introduced as DNA fragmentation tools for studying the 2′-deoxyribonolactone abasic site lesion [48]. On photolysis, 7-nitroindole is underwent a radical process to produce excited nitro group, as trigger, and then the trigger induced an intramolecular H1′ abstraction leading quantitatively to the highly labile deoxyribonolactone moiety to cleave of the DNA strand, as shown in Figure 17. After that, Berthet and coworker have also been studied the hybridization properties and enzymatic replication of 7-nitro isomer [49]. They reported that 7-nitroindole displayed hybridization and enzymatic recognition properties.

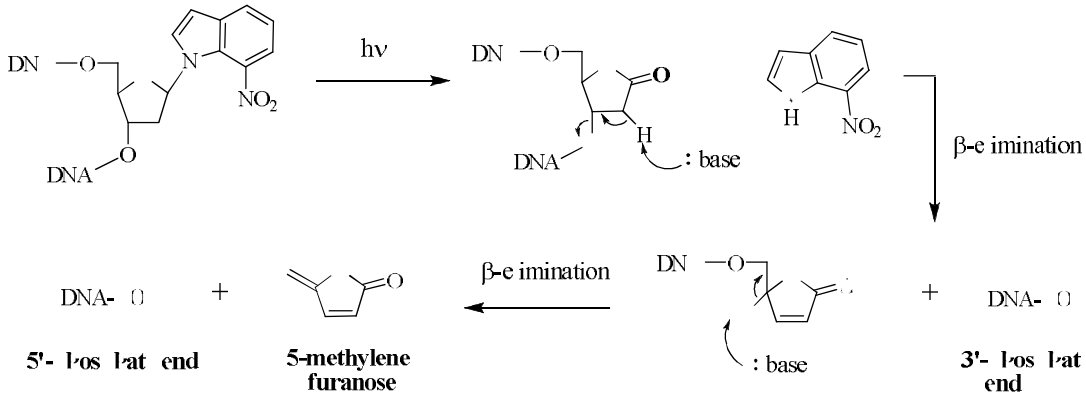


Figure 17 Photolysis of 7-nitroindole (8) results in the formation of the labile deoxyribonolactone abasic lesion to produced the DNA backbone cleavage

Method overviews

1. Molecular dynamics (MD) simulation

Molecular dynamics (MD) simulations are important computational tools for investigation of dynamic properties and thermodynamic properties in order to understand the macromolecular stability [50], conformational and allosteric properties [51], the role of dynamics in enzyme activity [52, 53], molecular recognition and the properties of complexes [52, 54] ion and small molecule transport [55, 56], protein association [57], protein folding, [58, 59] and protein hydration [60]. Therefore, they could provide for various studies including molecular design (drug design [61] and protein design [62]) and structure determination, refinement, and prediction (X-ray [63], NMR [64] and modeling [65]). Furthermore, they also can be applied in numerous fields such as molecular biology, biophysics, biotechnology, enzymology and pharmaceutical chemistry, and often more easily than real experiments.

Under this study, MD simulation was used to preliminarily investigate the molecular dynamics properties, structural properties and simulated structure determination of the study systems containing carbazole derivatives modified DNA (15 mers) duplex by employing the AMBER 10 package.

1.1 AMBER

1.1.1 Introduction

AMBER is the programs for molecular dynamics simulations in order to study time-dependent evolution of systems, explicitly model solvents, and treat systems too large to investigate with electronic structure method, especially on biopolymers such as proteins [66] and nucleic acids [67]. AMBER, stands for Assisted Model Building with Energy Refinement, is used for a set of force fields, Parm99 [68] which is modification of the older Parm98 [69] to the second-generation AMBER force field [70]. The general AMBER force fields model is given in this following equation (1).

$$E_{total} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 +$$

$$\sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\emptyset - \gamma)] + \sum_{i<j}^{atoms} [\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}] \qquad (1).$$

The total energy (E_{total}) is the sum of several components that related to bond lengths, bond angles, dihedral angles, and non-bond interactions, where bond lengths, K_r is a force constant altering the stiffness of bond stretching/contraction in a particular bond type, $r - r_{eq}$ represents the deviation of actual bond length r from equilibrium bond length r_{eq}, bond angles, $K\theta$ is similarly a constant altering the energy cost for an angle θ in a certain bonding arrangement to deviate from the equilibrium bond angle θ_{eq}, dihedrals, V_n is a constant giving the energy barrier to rotation for each of the n terms, γ is the phase angle, and Ø is the dihedral angle under evaluation. Finally, non-bond interactions, these interactions are *van der Waals* and electrostatic interactions that contribute to the total energy for pairs of atoms, where $\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}$ is a Lennard-Jones potential representing the *van der Waals* interaction, where A and B are constants determined by atom types of the atoms *i, j* involved, and R_{ij} is the distance between the atoms, $\frac{B_{ij}}{R_{ij}^{6}}$ represents the attractive dispersion force between atoms, which is dependent to the inverse 6th power on distance, $\frac{A_{ij}}{R_{ij}^{12}}$ represents the repulsive force between atoms based on their interacting electron clouds; its inverse 12th power dependence on distance is justified on grounds of computational efficiency rather than theoretical purity, and electrostatic interactions are represented by $\frac{q_i q_j}{\varepsilon R_{ij}}$, where q_i and q_j are the charges on atoms *i* and *j*, R_{ij} is the distance between the atoms, and $\mathcal{E}$ is the effective dielectric function.

AMBER is not a single program, but is rather a collection of codes that are designed to work together. There are three main steps that comprise of system preparation, simulation, and trajectory analysis. The process of MD simulation includes parameterization step, model study building, preparation of force field system, MD simulation, and MD trajectory analysis.

1.1.2 Parameterization step

The parameters that contain the possibility of bound lengths, angles, and dihedrals are potentially required for MD simulation in order to describe all possible dynamics and conformational properties of study systems. Carbazole residues whose parameter values are not found in the standard database AMBER, Therefore, each missing parameter must be defined prior simulations which can be used the restrained electrostatic potentials (RESP) procedure in AMBER [71, 72] by building the force fields around the concept of pair-wise charges. The structure's topology and atom types were defined on LEAP module in AMBER. The LEAP is the main preparation program for construction of biopolymer from the component residues, solvates the system, and prepares lists of force field terms and their associated parameters. There are two files that received from the refinement of structure's topology and atom types including coordinate (*inpcrd*) file that comprises just the Cartesian coordinates of all atom in the system and parameter topology (*prmtop*) file that contains all other information need to compute energies and forces such as atom names, atom mass, force field parameter, lists of bonds, angles, and dihedral [73, 74].

1.1.3 Preparation of the model study and the force field system for MD simulations

The initial duplex structure of the model was generated by using NUCGEN and produced the result format as the PDB file that contains the atoms named according to the AMBER conventions. The next step is modifying the initial duplex with parameterized carbazole derivatives by using the Sirius visualization program.

Then the input files was prepared, *prmtop* and *inpcrd* files, for MD simulation. The DNA (15 mer) duplex structure containing carbazole derivatives, model study, were taken from the PDB files into Xleap. Sodium counterions (Na^+) were randomly added to neutralize the phosphate backbone, and then solvated this system with explicit water model, TIP3P model, in a 10 Å of a truncated octahedral box.

1.1.4 Molecular dynamics simulation

Prior to MD simulation, the starting models were minimized in order to remove any *van der Waals* that existed due to this interaction might otherwise lead to local structural distortion and result in an unstable simulation. In the first stage of

minimization, the DNA molecule was fixed in its energy minimized position and both the water molecules, and sodium ions were allowed to readjust to the DNA molecule. Then the last minimization was carried out by allowing all atoms in the system to relax. After minimization steps, equilibration step was started from heating the system by assigning the initial velocities at low temperature and repeated until the desired temperature was reached. During this step, the DNA molecule was fixed and let the water move to adjust to the present of DNA. Once the desired temperature was achieved, the constraints on the DNA could be removed and the simulation in explicit water system continues. During simulation, several properties were monitored, particularly the structure, the pressure, the temperature, and the energy. The equilibration was simulated until these properties became stable with respect to the time. If the temperature increased or decreased significantly, the velocities could be rescaled to give the correct temperature. Classically, it was important to accurately simulate the experimental conditions to be replicated. Usually, the canonical-ensemble (NVT conditions) that was the collection of all systems whose thermodynamic state was characterized by fixed number of atoms (N), fixed volume (V), and fixed temperature (T). Another condition, the isobaric-isothermal ensemble (NPT conditions), NPT was characterized by fixed number of atoms (N), fixed pressure (P), and fixed temperature (T). The temperature was observed to fluctuate during a simulation at constant energy because of the spontaneous interconversion of the kinetic and potential components of the total energy.

The atomic velocities can be applied to evaluate the instantaneous temperature by using equation 2, as follows:

$$3k_B T = \frac{1}{N}\sum_{i=1}^{N} m_i v_i v_j \qquad (2)$$

Where k_B is Boltzmann's constant, m_i is the mass, and υ_i and υ_j are the velocity of atom i and j, respectively, and N is the total number of atoms. If there are to rescale or otherwise modifying the atomic velocities for the temperature to remain practically constant during the course of a simulation, it is worth mentioning that equation 2 must be corrected when constraint algorithms. This case can be used SHAKE for constraining

the bond of hydrogen atoms to fixed lengths [75]. The SHAKE algorithm, otherwise known as the constrained Verlet method, is a straightforward modification of Verlet algorithm.

Accordingly, a constant pressure during a simulation can be maintained by allowed the fluctuation to adjust the dimensional of the periodic box and rescaled the atomic positions.

1.1.6 Analysis of trajectory

During simulations, coordinates and velocities of the system were saved as MD trajectory data for the analysis. In AMBER package, *ptraj* program is available tool for analyzing this trajectory data, which read in AMBER parameter, topology files (*prmtop*), and allows a user to interactively inspect the information within the data such as list of bonds, angles, dihedrals, and other parameters. These properties can be graphically displayed, where one of the axis (x) correspond to time and the other to the quantity of interest, such as energy, RMSD, average structure, distance, etc.

2. Calculation of interaction energy

Interaction energy (ΔE) of DNA duplex formation is non-covalent interactions that contain stacking interaction, hydrogen bonding, electrostatic interaction, and *van der Waals* interaction. Non-covalent interaction energy can be calculated with the perturbation and variation theories. On perturbation theory, the interaction energy is directly calculated from the sum of various energies terms such as electrostatic, dispersion, *van der Waals*, and exchange repulsion. The variation method is indirectly determined the interaction energy by calculating the difference of the energy of molecular cluster (supersystem) and the energies of the isolated subsystems (subsystems). The perturbation method is well suited for determining interaction between rigid systems, but its utility is limited for system of many intramolecular and intermolecular degrees that have to be simultaneously optimized. Generally, the variation method is exclusively used [76] for the evaluation of non-interaction energies of larger systems and can be computed following the equation 3, where $E(R \cdots T)$ is total

energy of molecular cluster $R \cdots T$ formation, and *E(R)* and *E(T)* are energies of subsystems *R* and *T*.

$$\Delta E = E(R \cdots T) - [E(R) + E(T)] \quad (3)$$

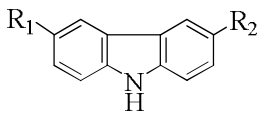
Therefore, the stacking energies of the carbazole derivatives modified DNA duplex systems will be evaluated by computing based on the variation theory.

The stabilization of non-covalent complexes is due to the favorable interaction energy. It means that the total energy of the complex is lower than the energies of its separated subsystems ($\Delta E < 0$) which occur systematically in the case of a complex is formed *in vacuo*. Contrarily, if a complex is formed in the water environment, the binding can be now realized not only favorable non-covalent interaction energy but also because of favorable entropy. In this case, the total energy is higher than energy of its separated system ($\Delta E > 0$) [77].

CHAPTER III

RESEARCH METHODOLOGY

Calculation of dipole moments



Carbazole (CBZ); $R_1 = H$, $R_2 = H$
3-Cyanocarbazole (M C); $R_1 = H$, $R_2 = CN$
3,6-Dicyanocarbazole (D C); $R_1 = CN$, $R_2 = CN$
3-Nitrocarbazole (MNC); $R_1 = H$, $R_2 = NO_2$
3,6-Dinitrocarbazole (NC); $R_1 = NO_2$, $R_2 = NO_2$
3-Nitro-6-cyanocarbazole (N C); $R_1 = CN$, $R_2 = NO_2$

Figure 18 Chemical structures of carbazole derivatives

The geometries of Carbazole (CBZ), 3-cyanocarbazole (MCC), 3,6-dicyano-carbazole (DCC), 3-nitrocarbazole (MNC), 3,6-dinitrocarbazole (DNC), and 3-nitro-6-cyanocarbazole (NCC) were optimized by using Gaussian03 program with B3LYP/6-31G(d) basis set before dipole moment calculations. Then, all dipole moment calculations were accomplished by using Gaussian03 at B3LYP/6-311 +G(2d,p) level.

Molecular dynamics simulations

The MD simulations stage of carbazole derivatives modified-DNA duplexes in the presence of sodium counter-ions and water molecules were carried out according to simplified flow diagram (see Figure 19).

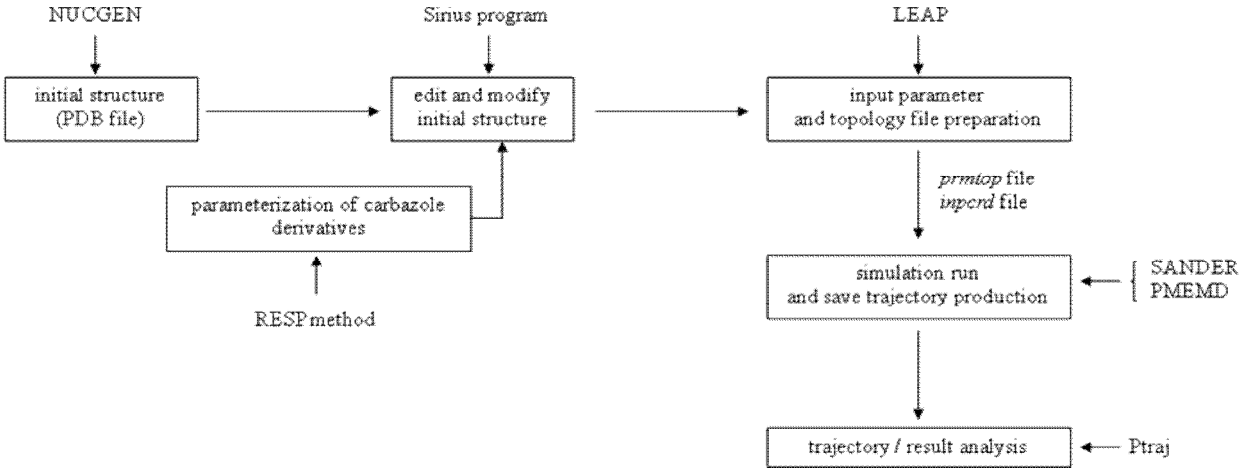
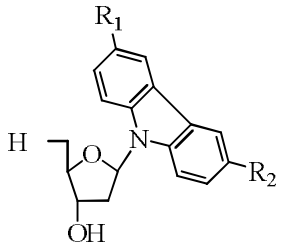


Figure 19 Simplified flow diagram for the MD simulation of study systems

1. Parameterization for modified carbazole residues



CBZ ucleoside analogue; $R_1 = H$, $R_2 = H$
D C ucleoside analogue; $R_1 = CN$, $R_2 = C$
NC ucleoside analogue; $R_1 = NO_2$, $R_2 = NO_2$
N C ucleoside analogue; $R_1 = CN$, $R_2 = NO_2$

Figure 20 Nucleoside analogues of carbazole derivatives

Conformational geometries of the nucleoside analogues containing carbazole derivatives (CBZ, DNC, DCC, and NCC; Figure 18) were optimized by using Gaussian03 program with HF/6-31G* basis set. Then, these optimized analogues were calculated an electrostatic potential (ESP) at a large number of grid point around each of each molecules with Gaussian03 at HF/6-31G* level [77, 78]. The obtained ESP data of each

nucleoside analogues was converted to the restrained electrostatic potentials (RESP) format. Afterward, the converted ESP points were fitted the atomic charges by using the RESP procedure [77, 78]. In addition, only the atomic charges of C1′ and H1′ atoms, and the modified carbazole residues were fitted. The atomic charges on the deoxyribose moiety were kept at the AMBER force field database [77, 78]. Finally, the topology and atom types of modified carbazole residues were parameterized into FF99SB force field of AMBER 10 using GAFF atom types [73].

2. Preparation of the model study for MD simulation and experimental force field

Each initial duplex structure that depicted in Figure 21 was generated to the 15 base-pairs with a canonical B-DNA conformation by using the NUCGEN module of AMBER software.

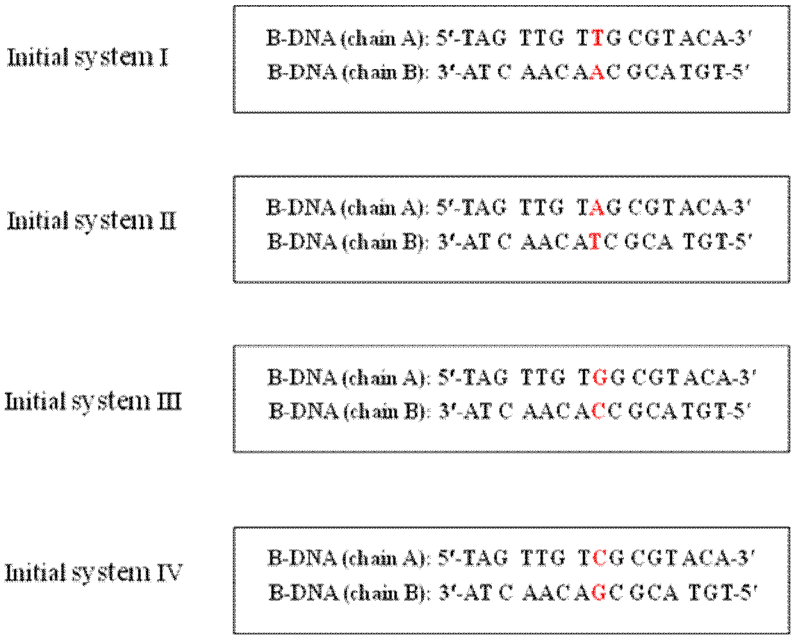


Figure 21 The initial model study

After that, each initial duplex structures were modified (see Figure 22 for all modifications) by replacement the middle position (X = T, A, G and C) of chain A on

each systems with the four nucleoside analogues containing carbazole derivatives (Figure 20) using superimposing method with the Sirius visualization program.

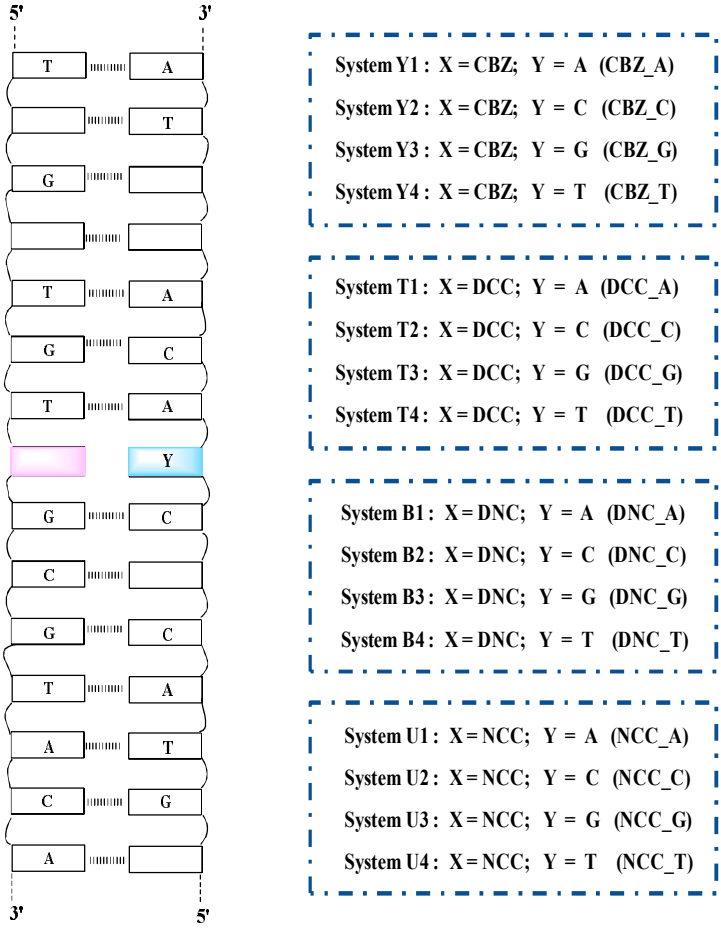


Figure 22 Model studies for molecular dynamics simulations

After that, each modified DNA duplex was neutralized with sodium counterions (28 atoms; one Na^+ for one phosphate group) and explicitly solvated by a 10 Å water shell in all direction of a truncated octahedral box (TIP3PBOX) by employing the Xleap module in AMBER program.

4. Molecular dynamics simulations

All molecular dynamics simulations were performed using the SANDER and PMEMD module in AMBER program and were started from system minimization of solvent and sodium ions with a 500 kcal/mol restraint force on the DNA molecule for 3000 cycles at 0 K. Then, another minimization was carried out by allowing relaxation of the DNA molecule with 3000 cycles of running. After complete minimization, equilibrations were performed by heating from 0 K to 300 K over a period of 30 ps for 30000 steps under constant volume condition and restraint weak force (10 kcal/mol) on the DNA molecule by allowed residue 8 to relax. Then, the last simulation was performed for 6 ns at NTP conditions that including constant temperature (300 K) and pressure (1 atm) without restraint on the DNA molecule. The hydrogen atoms were constrained using the SHAKE and 2 fs time step.

5. Analysis of trajectory

A final 5 ns of each trajectory was analyzed by using the *ptraj* procedure in AMBER program. In analysis, RMSD, B factor, average energy structure, and distance between modified carbazole residues and neighboring nucleobases were calculated.

Additionally, the distance between carbazole derivatives and neighboring nucleobases were analyzed in three cases. The first case was the distance between thymine (T; residue 7) and carbazole derivatives (CBZ, DCC, DNC, and NCC; residue 8). The second case was calculation of distance between carbazole derivatives (residue 8) and guanine (G; residue 9). Finally, the third case was calculation of the distance between carbazole derivatives (residue 8) and their opposite nuclobases (A, C, G, and T; residues 23). In addition, the distance between residue 7 (T) with 8 (CBZ), residue 8 (CBZ) with 9 (G), and the distance between residue 8 with residues 23 were calculated from C1′ atom on the deoxyribose moiety of each residue [79]

Calculation of interaction energies

Each of interaction energies of carbazole derivatives within the DNA duplex structure were evaluated by sampling their structural geometries every 50 snapshots (or every 25 ps) from the trajectories and started from the first snapshot. All sampling structures were fragmented to produce three clusters containing system 1 (S1), system 2 (S2), and system 3 (S3) (Figure 23). System 1, only the three consequential base-pairs were extracted and the others were stripped. System 2, the three consequential base-pairs were extracted and removed the residue 8 (X). System 3, only the residue 8 was extracted and the others molecules were stripped. The hydrogen atoms were added to the N1 or N9 position of the base residues after removal of the sugar phosphate backbone in each system. Then, the interaction energies (ΔE) of carbazole derivatives within the DNA duplex structure were calculated by employing the PM6 and the DFTB semi-empirical models in the Gaussian09 program as follows:

$$\Delta E = ES1 - (ES2 + ES3)$$

where *ES*1, *ES*2, and *ES*3 represent the energies of the system 1, system 2, and system 3, respectively.

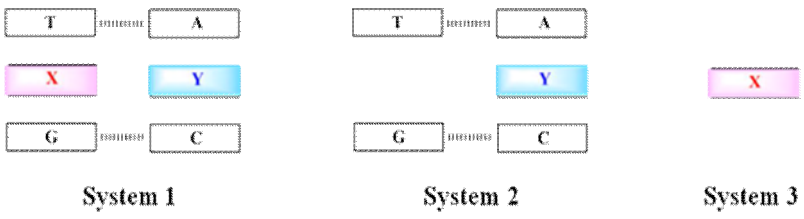


Figure 23 The duplex fragments for the stacking calculations

CHAPTER IV

RESULTS AND DISCUSSION

Dipole moment of carbazole derivatives

The B3LYP/6-311 +G(2d,p) basis set in the Gaussiun03 program was applied to calculate the dipole moment values of carbazole derivatives and the obtained values are summarized in Table 2.

Table 2 Calculated dipole moment of carbazole derivatives

Carbazole derivatives	Dipole moment/D
CBZ	1.89
MCC	7.24
DCC	9.12
MNC	7.61
DNC	9.42
NCC	9.24

As the results, DNC (3,6-dinitrocarbazole), DCC (3,6-dicyanocarbazole), and NCC (3-ntro-6-cyanocarbazole) exhibited the high dipole moment when compared with the others (Table 2). These results confirmed that the presence of the high electro negativity group (-NO_2 and -CN groups) would affect the enhancement of the dipole moment. Therefore, these molecules were chosen to estimate their universal base properties. The substituted group on the carbazole molecule, -NO_2 and -CN groups, are high electro negativity group that should provide the large dipole vector with both inward and outward direction to the aromatic moiety. Therefore, this should affect the enhancement of stacking interaction of the π-aromatic system of the carbazole derivatives residues. As the presented in Figure 24, DNC and DCC had the dipoles in a same direction, and the dipole vector of NCC was non-symmetrical. From previous study, *van der Waals* volumes of base candidates were also calculated [80=ref STT pay]. It was found that the replacement of nitro group with cyano group would be

reduced *van der Waals* volume (DNC = 225.28 Å^3, DCC = 208.15 Å^3 and NCC = 216.61 Å^3). However, cyano-substituted carbazole still maintained high dipole moment value when compared with the dinitro-substituted carbazole.

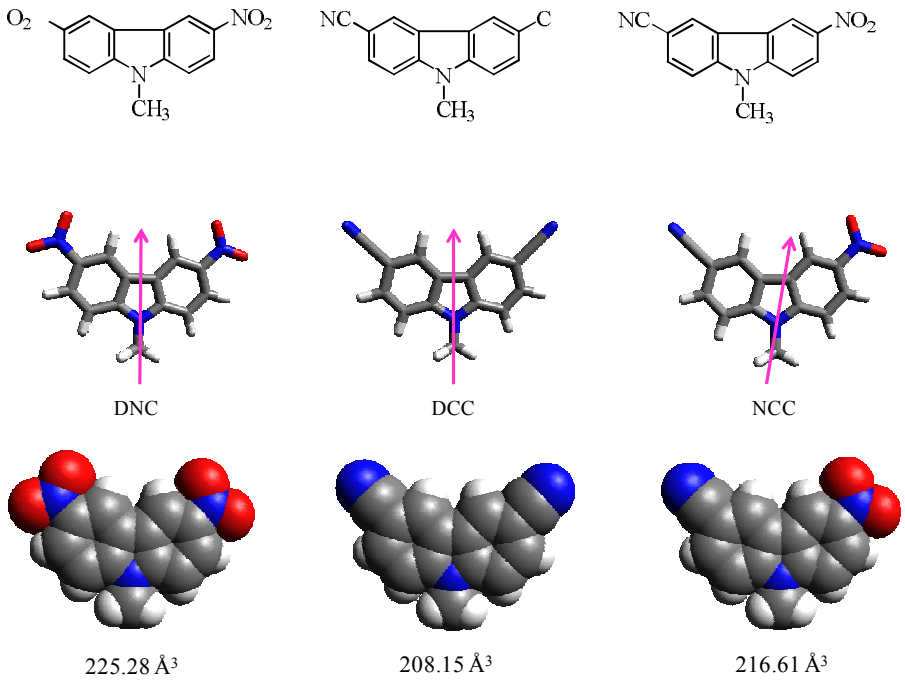


Figure 24 Dipole vectors and *van der Walls* spheres of DNC, DCC, and NCC

Overall properties of the MD simulation system

The overall properties of system that includes the potential, kinetic and total energies (= potential + kinetic), temperature, and density were analyzed at last 5 ns of the simulation. Firstly, all energies data were plotted with time along the 5830 ps (5 ns). As the results, the variations of all energies with the time showed slightly fluctuation, this revealed that all simulations reached the equilibration as well. The others properties (average temperature, density, and energy) are summarized in the Table 3. Temperature and density quite stabled during simulation, while energy of the system largely fluctuated. This might due to temperature changed during simulation.

Table 3 Overall properties of the MD simulation system

B-DNA: 5′-TAG T TG TXG CGT ACA-3′

B-DNA: 3′-ATC AAC AYC GCA TGT-5′

System	X	Y	Properties		
			Avg. Temperature (K)	Avg. Density (g/cm^3)	Avg. Energy (kcal/mol)
Y1	CBZ	A	300.05 ± 1.89	1.02	-68440.96 ± 18.84
Y2	CBZ	C	300.06 ± 1.91	1.02	-68648.29 ± 20.07
Y3	CBZ	G	299.97 ± 1.90	1.02	-69130.10 ± 19.49
Y4	CBZ	T	300.10 ± 1.90	1.02	-68571.90 ± 25.72
T1	DCC	A	299.97 ± 1.90	1.02	-68417.30 ± 17.19
T2	DCC	C	299.94 ± 1.93	1.02	-66798.21 ± 24.92
T3	DCC	G	300.01 ± 1.89	1.02	-68991.85 ± 18.64
T4	DCC	T	299.95 ± 1.92	1.02	-67381.59 ± 25.61
B1	DNC	A	300.04 ± 1.88	1.02	-68393.48 ± 20.68
B2	DNC	C	299.88 ± 1.89	1.02	-69041.58 ± 22.20
B3	DNC	G	299.92 ± 1.91	1.02	-67057.58 ± 23.00
B4	DNC	T	299.96 ± 1.89	1.02	-68872.78 ± 12.29
U1	NCC	A	299.97 ± 1.92	1.02	-67294.59 ± 23.90
U2	NCC	C	299.92 ± 1.88	1.02	-67619.07 ± 19.83
U3	NCC	G	300.03 ± 1.90	1.02	-69138.91 ± 23.24
U4	NCC	T	299.95 ± 1.91	1.02	-68330.71 ± 17.59

Stability of the DNA duplex structure during MD simulation

The deviation of the structural geometry of the double helix from the starting structure during simulation directly implied the stability of simulated structures. In this analysis, the stability of the simulated structure was confirmed by calculation of the root mean square deviation (RMSD) that is depended on the simulation time. The calculated

RMSD values in all systems were not constant and were fluctuated between 0 to 5.4 Å, as shown in Figure 25, 26, and 27. It was found that a period of time around 1230 ps (md12) to 3030 ps (md20) were more constant than the others periods. Consequently, md12 to md20 of simulated geometries of all system study were selected for the next calculation of interaction energies because their structural geometries were stable rather than other simulation products.

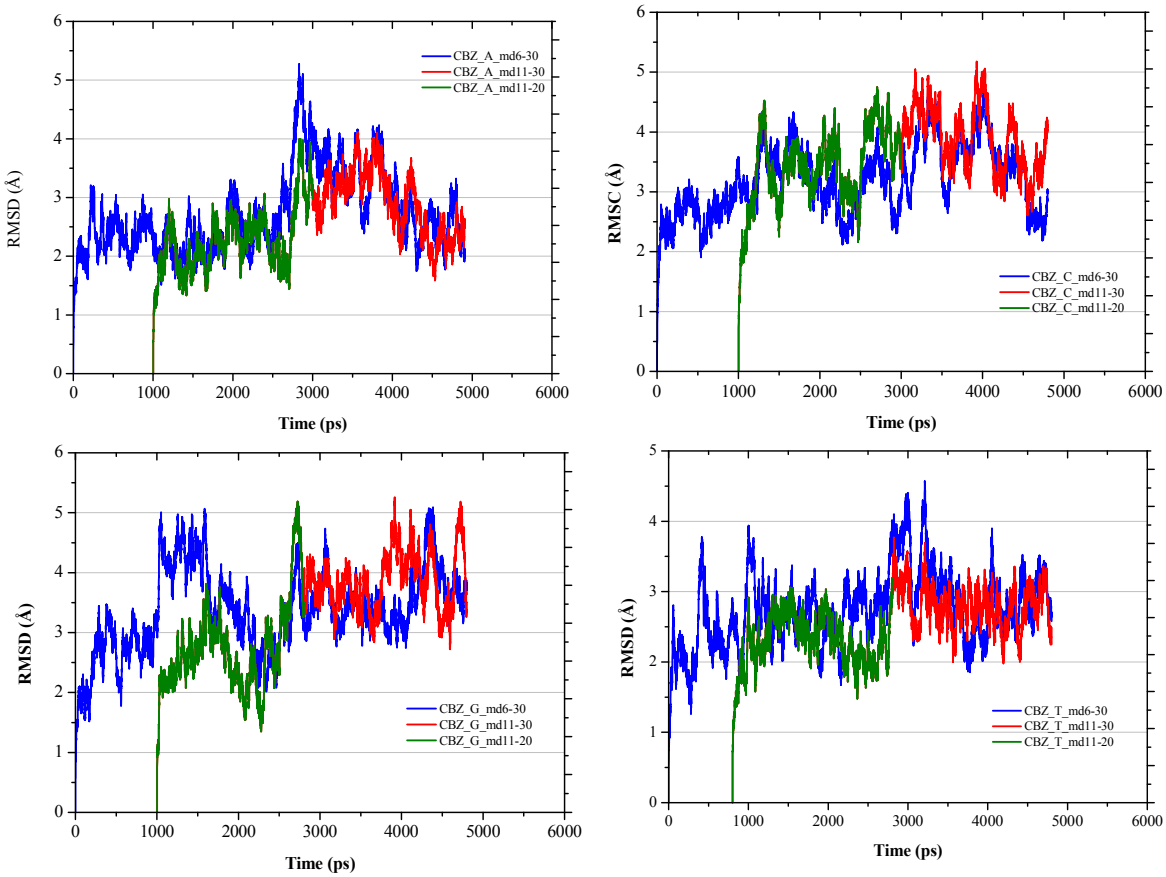


Figure 25 The RMSD of the simulate structure from CBZ_A, CBZ_C, CBZ_G, and CBZ_T systems where the blue line, green line and red line represents at last 5 ns, 2 ns - 4 ns, and 2 ns – 6 ns production runs

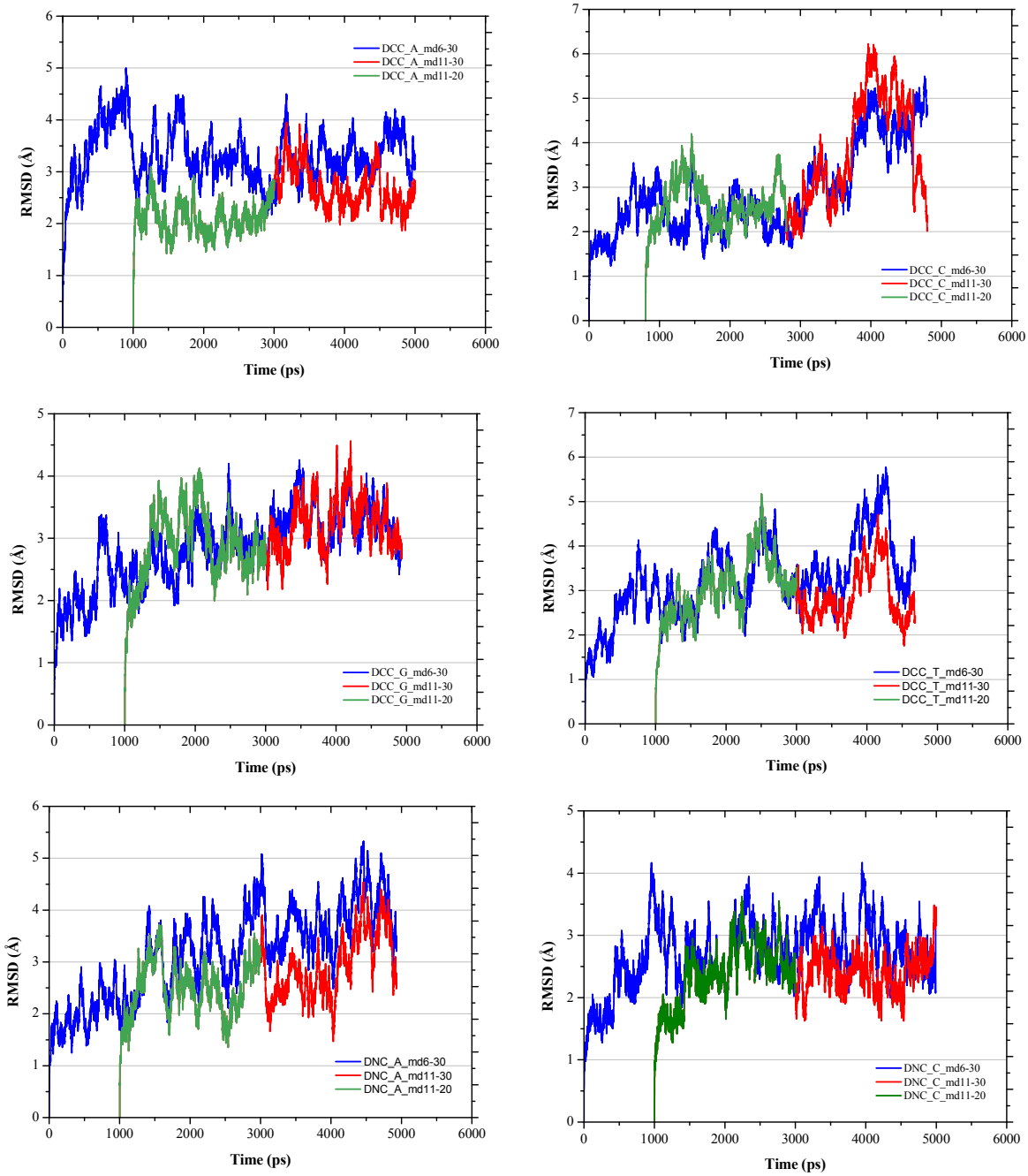


Figure 26 The RMSD of the simulate structure from DCC-A, DCC-C, DCC-G, DCC-T, DNC-A, and DNC-C systems where the blue line, green line and red line represents at last 5 ns, 2 ns - 4 ns, and 2 ns – 6 ns production runs, respectively

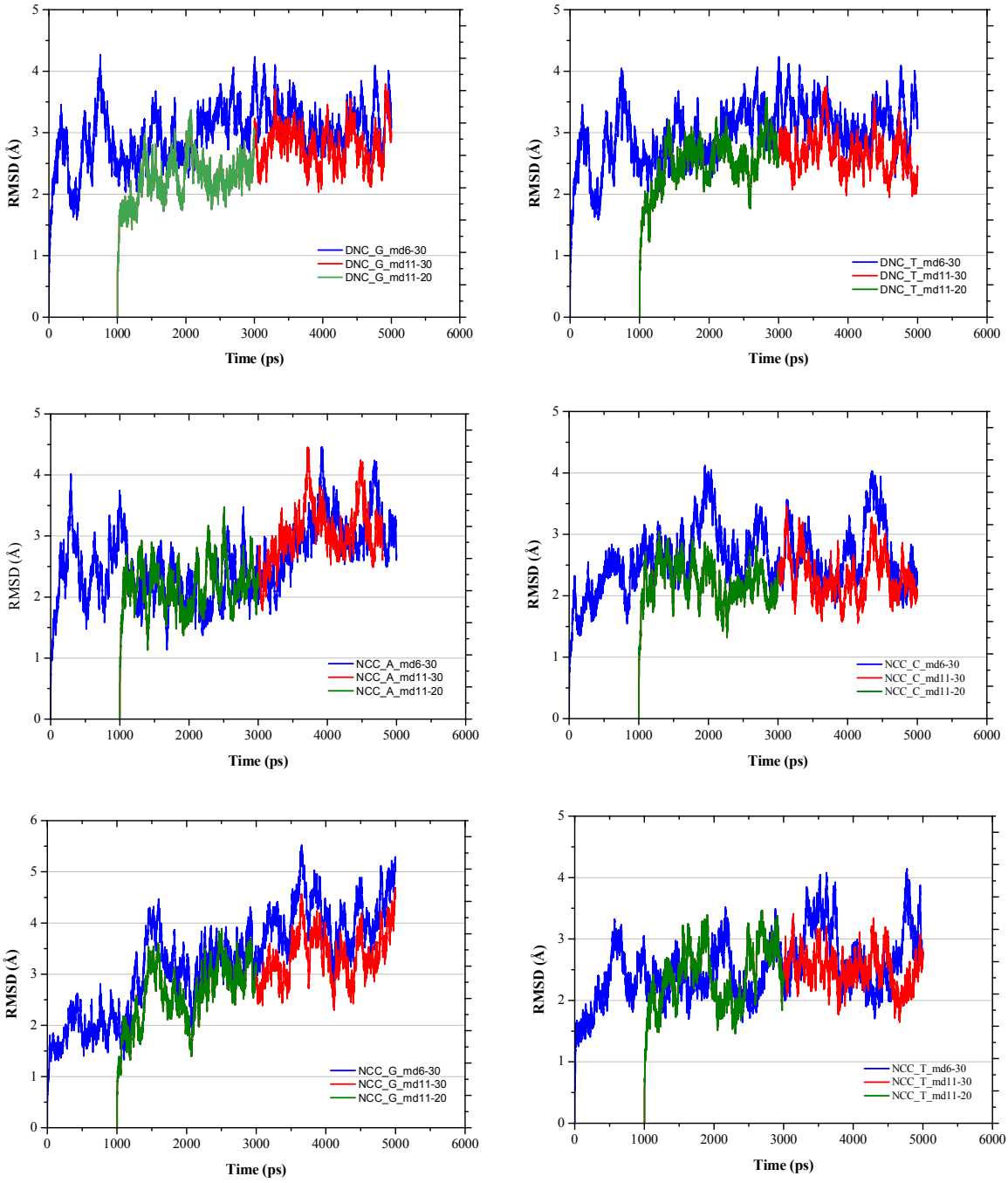


Figure 27 The RMSD of the simulate structure from DNC-G, DNC-T, NNC-A, NNC-C, NNC-G, and NNC-T systems where the blue line, green line and red line represents at last 5 ns, 2 ns - 4 ns, and 2 ns – 6 ns production runs, respectively

Properties of simulated structure

The structural properties were analyzed over selected MD production runs (md12-md20) in order to consider the dynamics disorder of simulated structure from calculation of B-factors, the average structure, and the distance between modified carbazole residues and neighboring nucleobases.

1. Analysis of B-factors

The B-factors of the simulation products were calculated in order to measure the mobility of base residues within the modified DNA duplex structure. From the computed B-factor of molecular base fluctuation in all modified 15 mer DNA systems, it was found that the B-factors decreased as CBZ (carbazole) > DNC (3,6-dinitrocarbazole) > DCC (3,6-dicyanocarbazole) > NCC (3-ntro-6-cyano-carbazole), as shown in Figure 28 and 29.

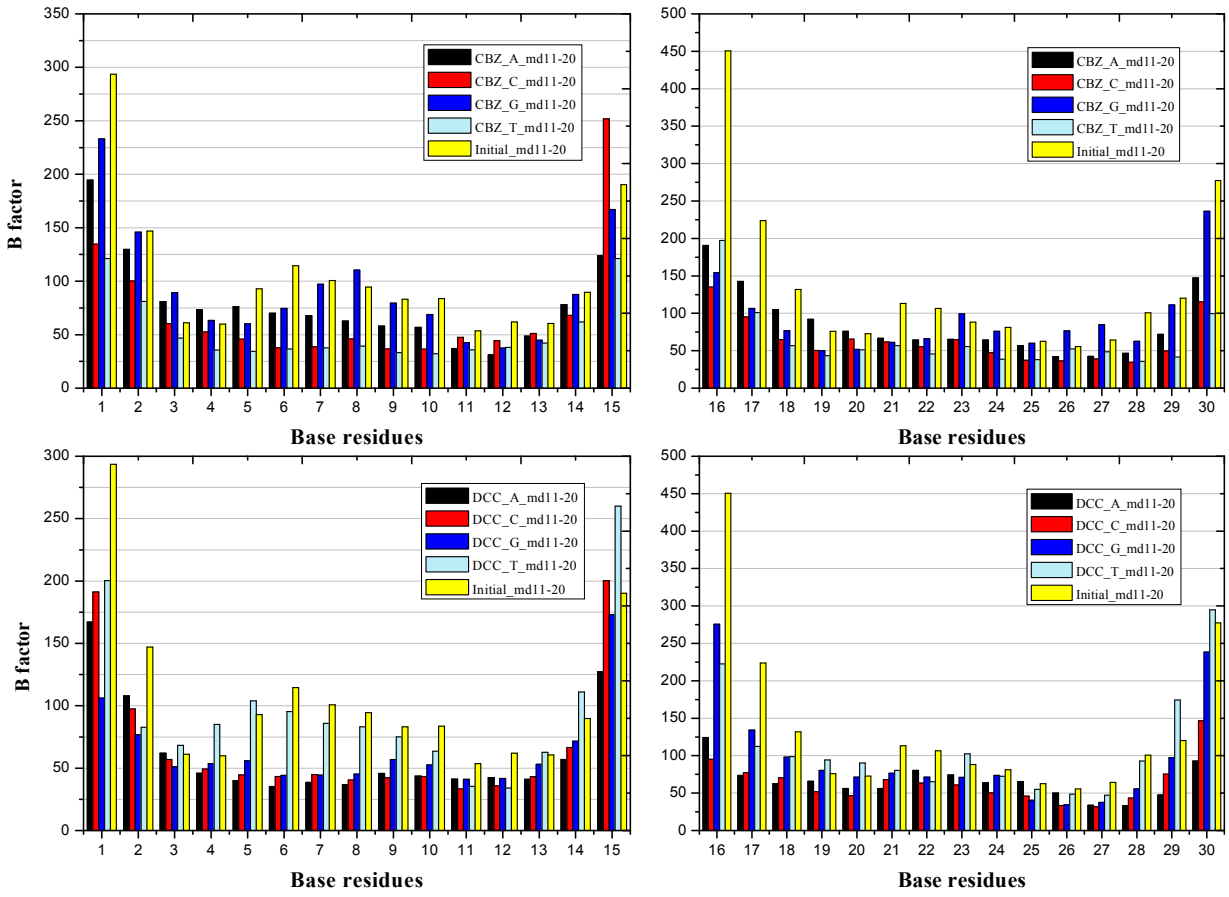


Figure 28 The B-factor of the simulated structures from 2-4 ns (md11-md20)

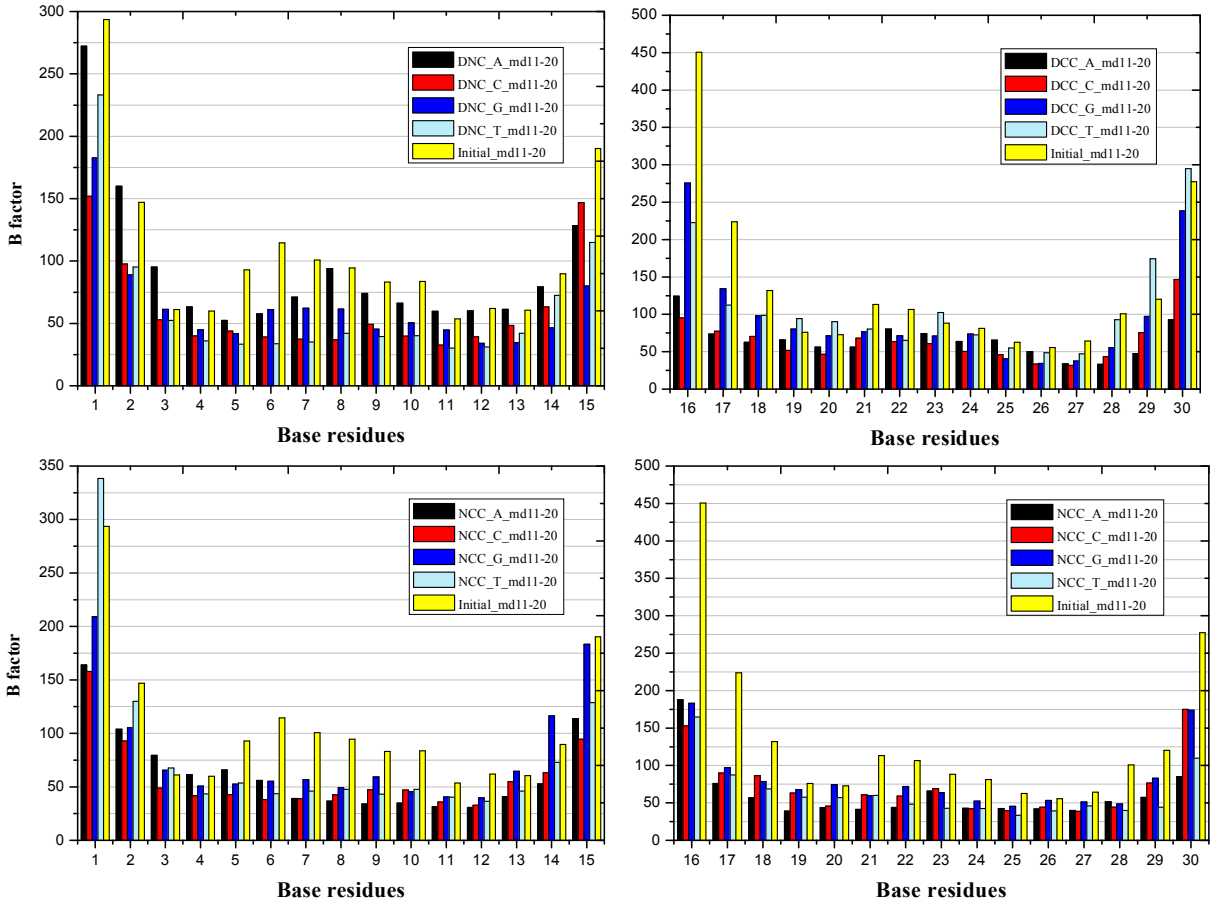


Figure 29 The B-factor of the simulated structures from 2-4 ns (md11-md20)

As the overall results, all system study had lower B-factor than initial system because carbazole derivatives are larger molecule and more non-polar than the natural nucleobases. In addition to residue 8, CBZ showed the highest B-factors. This implied that CBZ residue had the highest mobility because CBZ is more non-polar and smaller molecule than DNC, DCC, and NCC.

2. Average structure

The average structural geometries of central three consecutive base pairs of the modified-15 mer DNA duplexes containing carbazole derivatives are showed in Figure 30-35. Figure 30 and 31 is the average geometries of the systems containing

CBZ at middle base position within DNA duplexes. The substitution of carbazole derivatives in double helical DNA caused minimal structural distortion of duplex structure due to their large molecule. In the cases of CBZ opposite A, CBZ forced A and resulted that A was outwardly flipped from the DNA duplex. This was due to non polar property of CBZ. On CBZ against G, CBZ and G possessed the overlap between them that this was an unsuitable orientation for base pairing within the DNA duplex. In addition, this might result from the effect of the large molecules of CBZ and G and intermolecular force from partial negative charge of oxygen (ketone group) on G molecule that electrostatically attracted to hydrogen atom on CBZ molecule. In the part of CBZ against C and T, CBZ can be hybridized to C and T in similar pattern with natural nucleobases. This probably was due to both pyrimidine bases are small molecules and have the partial electronegative atoms such as oxygen and nitrogen.

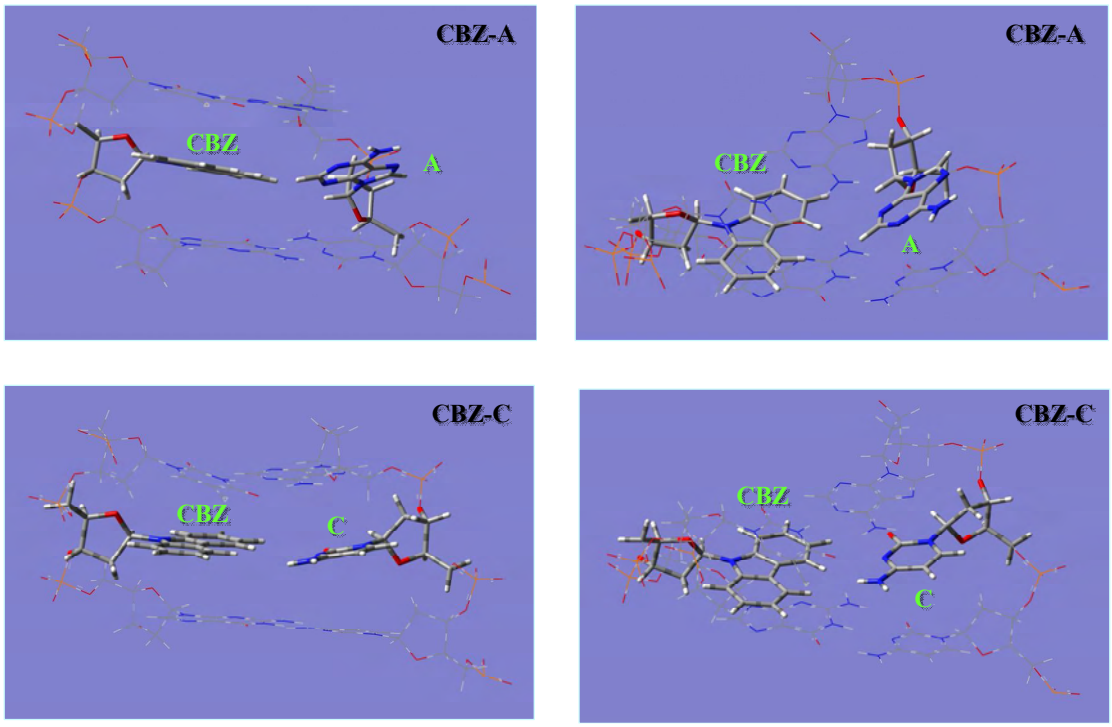


Figure 30 The average simulated structure of CBZ against A and C from central three consecutive base pairs of system study

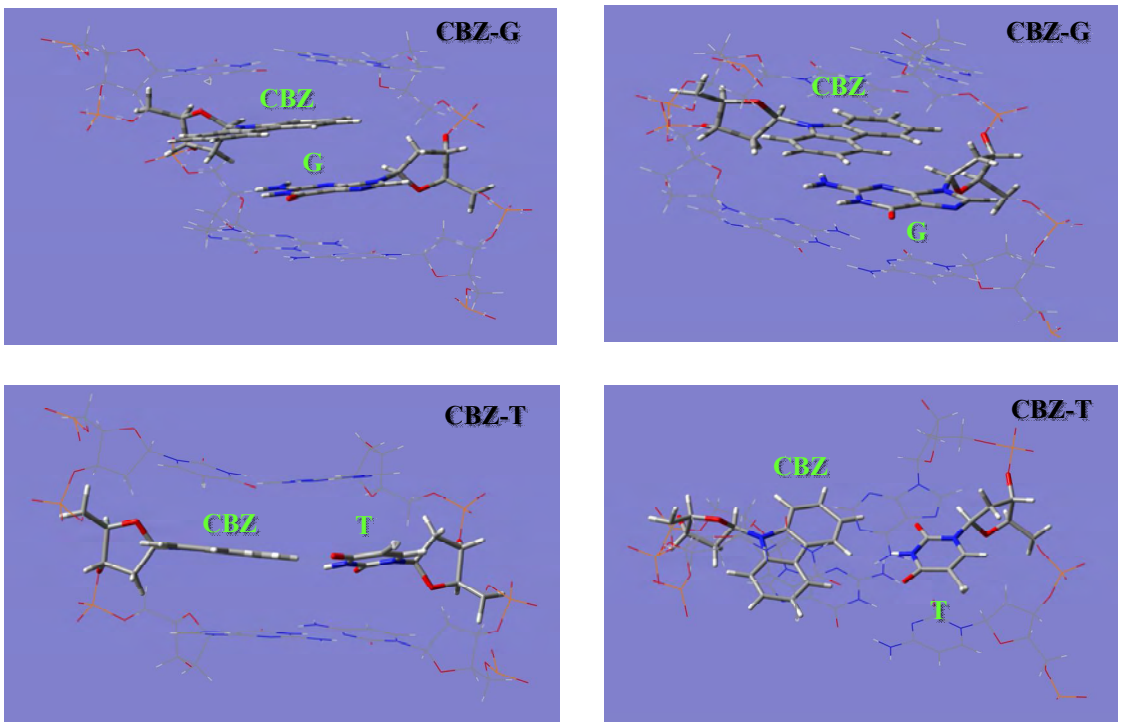


Figure 31 The average simulated structure of CBZ against G and T from central three consecutive base pairs of system study

In the systems of DCC modified DNA duplexes, their average geometries are exhibited in Figure 32. In the case of DCC against purine bases (A and G), they showed non-symmetry overlap. These might due to the result of the repulsion from the electrostatic that are similar arrangement of the dipole direction between DCC and purine bases. Because the dipole-dipole interaction can be exhibited either stabilizing or destabilizing effect and depend on the relative arrangement of the dipole direction [81, 82]. As the part of DCC opposite pyrimidine bases (C and T), hybridization between DCC with C and DCC with T have repulsion from each partial similar charge on each molecules.

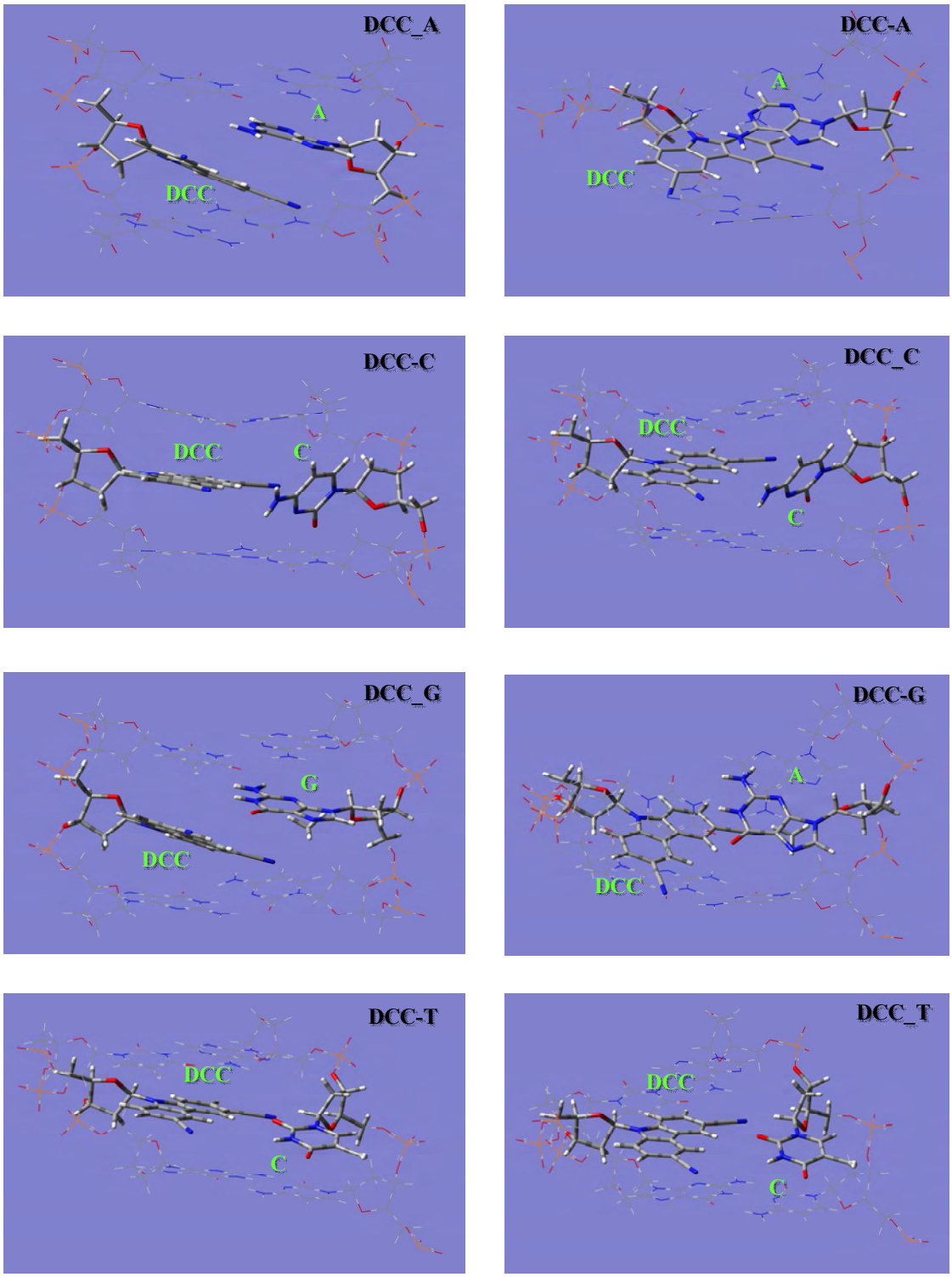


Figure 32 The average simulated structure of DCC against A, C, G, and T from central three consecutive base pairs of system study

Figure 33, 34, and 35 illustrate average geometries of DNC and NCC. Although, DNC and NCC residue were exhibited minimal fluctuation, average geometries of two derivatives were still not ideal pairing in against all of natural nucleobases. Additionally, the electrostatic effect and steric hindrance of the substituent group on carbazole moiety should be major affect to base pair between carbazole derivatives and natural nucleobases. In the case of DNC against nucleobases, the nitro group of DNC point towards the opposite strand, whereas another nitro group was located in the major groove. The nitro group did not form hydrogen bond in any of the structures because nitro group was relatively poor hydrogen bonding acceptors [83].

Especially, all modified DNA duplexes that were simulated in solution still maintained in B-form (the conformation of sugar is a C(2′)-*endo*).

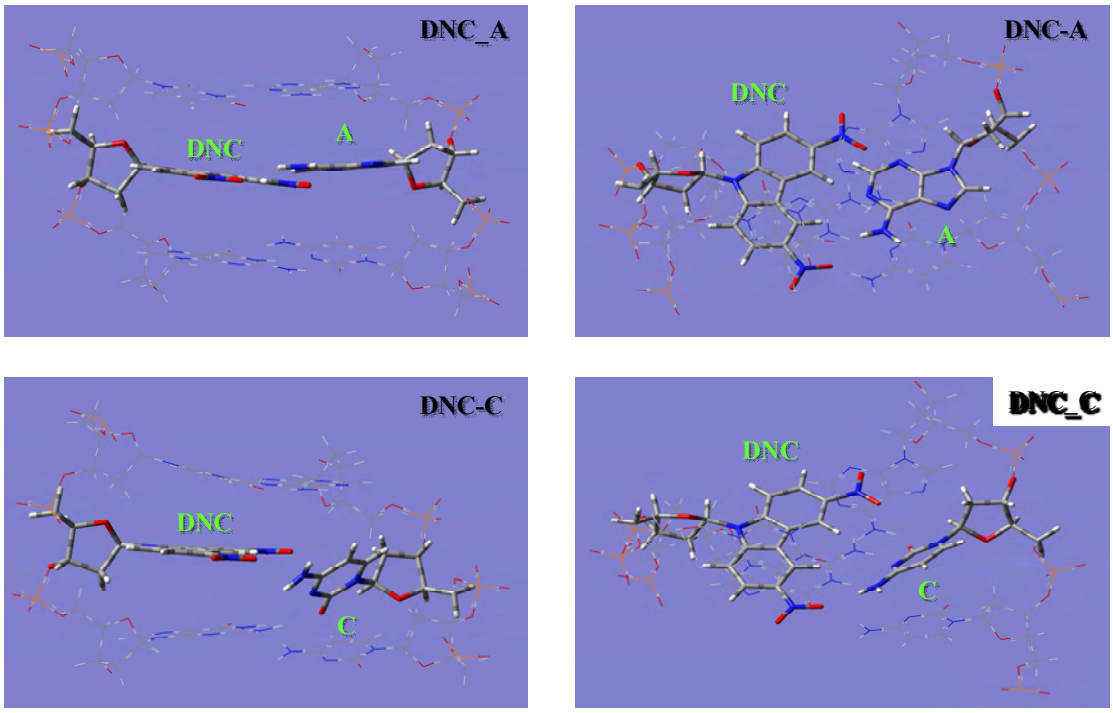


Figure 33 The average simulated structure of DNC against A and C from central three consecutive base pairs of system study

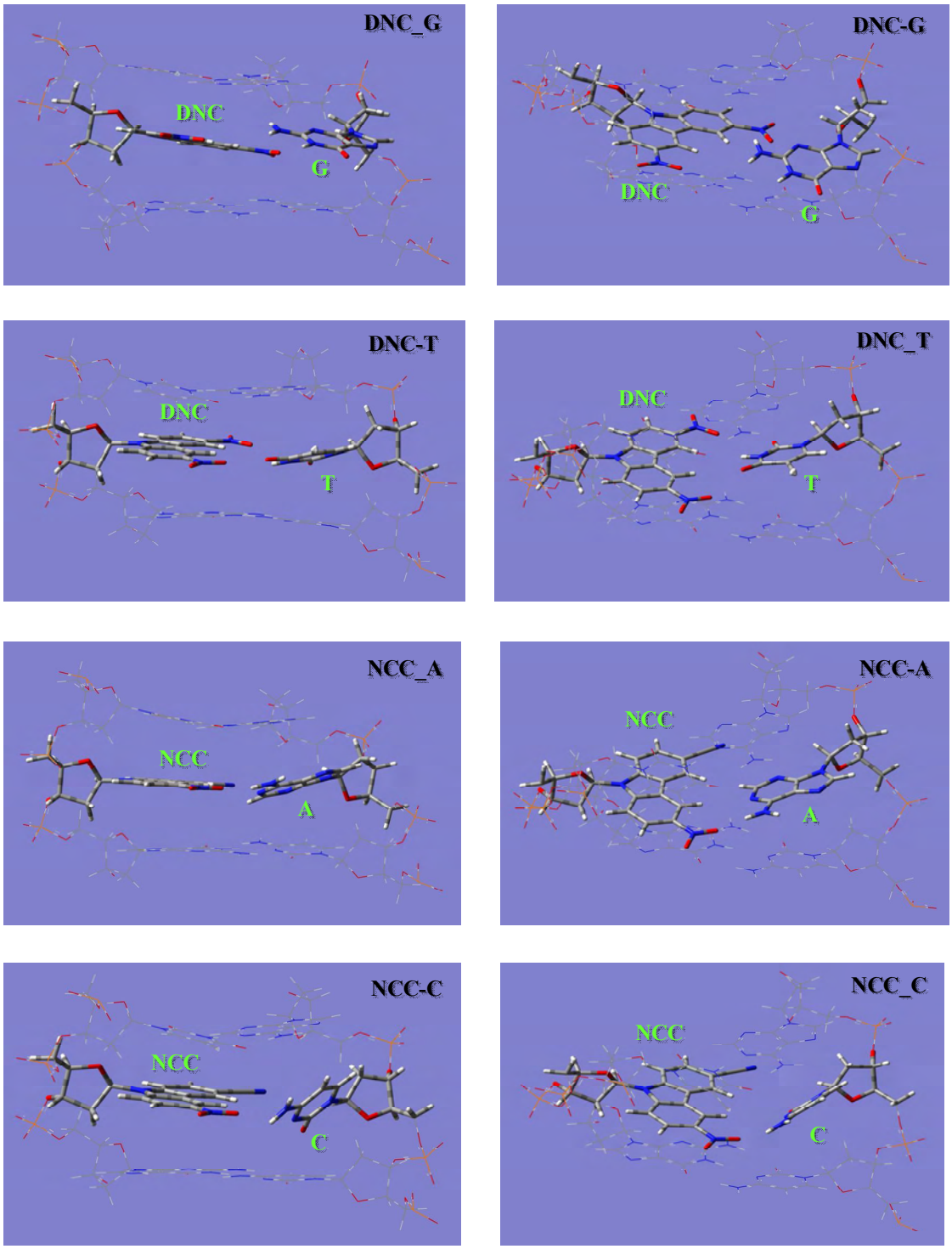


Figure 34 The average simulated structure of DNC against G and T and NCC against A and C from central three consecutive base pairs of system study

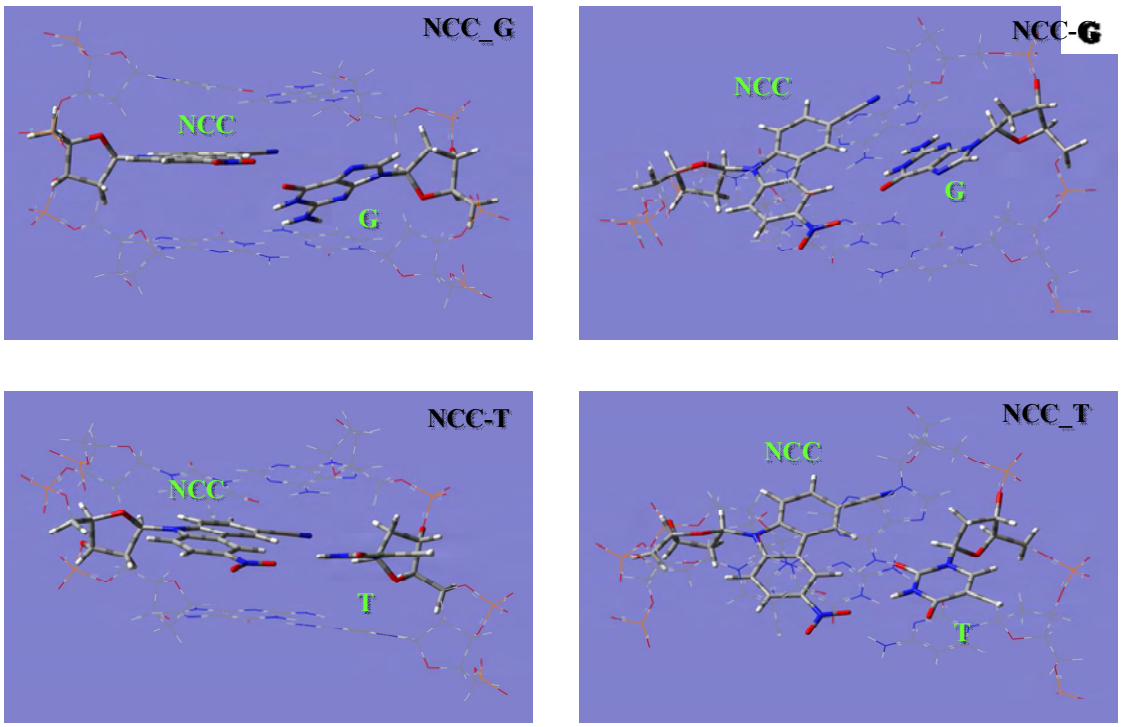


Figure 35 The average simulated structure of NCC against A and C from central three consecutive base pairs of system study

3. Analysis of the distance between carbazole derivatives and their neighboring nucleobases

The distance between modified carbazole (residue 8) and its adjacent nucleobases such as residue 7 (T), residue 9 (G), and its opposite nucleobases containing A, C, G, and T (residue 23) were computed to investigate the effect of carbazole derivatives on the duplex structures. All calculated distances were present in Table 4. Due to carbazole derivatives are hydrophobic and large molecules, and the effect from electrostatic of substituent groups on their molecules causes the calculated distance were deviated from C1′ atom on the deoxyribose moiety of each base residue.

Table 4 The distances between modified carbazole and neighboring nucleobases

B-DNA: 5′-TAG TTG TXG CGT ACA-3′

B-DNA: 3′-ATC AACAYC GCA TGT-5′

System	X (Res.8)	Y (Res.23)	Distances (Å)		
			Res.7(T) - Res. 8	Res.8 - Res.9(G)	Res.8 - Res.23
Y1	CBZ	A	4.99 ± 0.49	4.50 ± 0.31	9.84 ± 0.54
Y2	CBZ	C	5.20 ± 0.51	4.41 ± 0.36	12.03 ± 0.60
Y3	CBZ	G	4.25 ± 0.39	6.86 ± 0.43	8.00 ± 0.42
Y4	CBZ	T	5.18 ± 0.44	4.25 ± 0.27	11.53 ± 0.69
T1	DCC	A	6.28 ± 0.56	3.97 ± 0.25	9.35 ± 0.77
T2	DCC	C	4.76 ± 0.46	4.01 ± 0.24	11.27 ± 0.64
T3	DCC	G	6.20 ± 0.57	4.13 ± 0.23	11.47 ± 1.05
T4	DCC	T	5.38 ± 0.45	4.48 ± 0.51	10.18 ± 0.43
B1	DNC	A	5.74 ± 0.49	4.18 ± 0.29	11.60 ± 1.34
B2	DNC	C	5.19 ± 0.35	4.02 ± 0.24	11.11 ± 0.57
B3	DNC	G	5.02 ± 0.41	4.23 ± 0.30	10.91 ± 1.04
B4	DNC	T	5.55 ± 0.50	4.13 ± 0.26	12.71 ± 0.30
U1	NCC	A	5.23 ± 0.45	4.08 ± 0.24	12.51 ± 0.30
U2	NCC	C	5.29 ± 0.67	3.96 ± 0.28	11.10 ± 0.92
U3	NCC	G	6.10 ± 0.40	4.32 ± 0.36	12.80 ± 0.42
U4	NCC	T	5.63 ± 0.55	4.15 ± 0.25	10.14 ± 0.36
Initial	T	A	4.34 ± 0.36	4.80 ± 0.45	10.62 ± 0.36

Calculation of interaction energies

The interaction energies of each molecular cluster containing carbazole derivative modified 15 mer DNA duplex were computed on MD simulated geometries by sampling every 25 ps. The PM6 and DFTB method were used to calculate interaction energy. The obtained interaction energies were summarized to average values that presented in Table 5. The average energy that received from PM6 and DFTB method was also compared as showed in Figure 35. As the result, the standard deviation values of the average energies are mostly lower than 3 kcal/mol. Both PM6 and DFTB method have similar trend of the calculated interaction energies, whereas the values from the PM6 method was higher than the calculated energies from DFTB method. It was found that the interaction energy of DNC and NCC are higher than CBZ and DCC. These revealed that DNC and NCC can be incorporated into DNA duplex and more stable than CBZ and DCC. The interaction of NCC was favorable to pyrimidine bases (C and T) with the greatest interaction ability, specifically C, while DNC presented the similar manner to purine bases (A and G) and the strongest interaction was found when it opposite A. Interestingly, DNC exhibited the most universality due to it showed the most indiscrimination to hybridize with each natural nucleobase. This can considered form interaction energy that each of calculated energy in the case of DNC against DNA bases had lowest difference. Furthermore, the interaction abilities of carbazole derivative with DNA bases were also relate to their dipole moment (Table 5). These evidenced that large dipole moment would affect the enhancement of interaction of the π-aromatic system of the carbazole derivatives residues. Additionally, the presences of nitro groups were expected that they would assist in polarization of the carbazole aromatic system resulting the repulsion was reduced and enhanced the interactions with DNA bases [84].

Table 5 Calculated stacking energy of carbazole and its derivatives in 15 mer DNA duplex

5'- T**X**G-3'

3'- A**Y**C-5'

System	X	Y	Average ΔE (kcal/mol)		Dipole moment (D)
			PM6	DFTB	
Y1	CBZ	A	0.36 ± 0.74	0.36 ± 0.42	1.89
Y2		C	-2.17 ± 0.82	0.24 ± 0.64	
Y3		G	-4.48 ± 1.26	-0.78 ± 0.72	
Y4		T	-2.23 ± 0.70	0.16 ± 0.49	
T1	DCC	A	-7.16 ± 1.44	-2.89 ± 1.41	9.12
T2		C	-7.67 ± 2.35	-2.50 ± 1.65	
T3		G	-6.35 ± 1.21	-2.32 ± 0.93	
T4		T	-8.23 ± 1.24	-3.88 ± 0.94	
B1	DNC	A	-11.00 ± 1.92	-6.32 ± 1.48	9.42
B2		C	-9.99 ± 1.54	-5.24 ± 1.34	
B3		G	-10.57 ± 2.32	-5.64 ± 1.91	
B4		T	-10.22 ± 2.09	-6.26 ± 1.99	
U1	NCC	A	-9.83 ± 1.35	-5.40 ± 1.91	9.24
U2		C	-12.26 ± 2.75	-7.02 ± 2.10	
U3		G	-7.94 ± 1.73	-4.97 ± 1.80	
U4		T	-10.80 ± 2.34	-6.97 ± 1.98	
Initial	T	A	-8.67 *	-6.49 *	

* Data from Assoc. Prof. Dr. Yuthana Tantirungrotechai, National Nanotechnology Center (NANOTEC), National Science and Technology Development Agency, Thailand

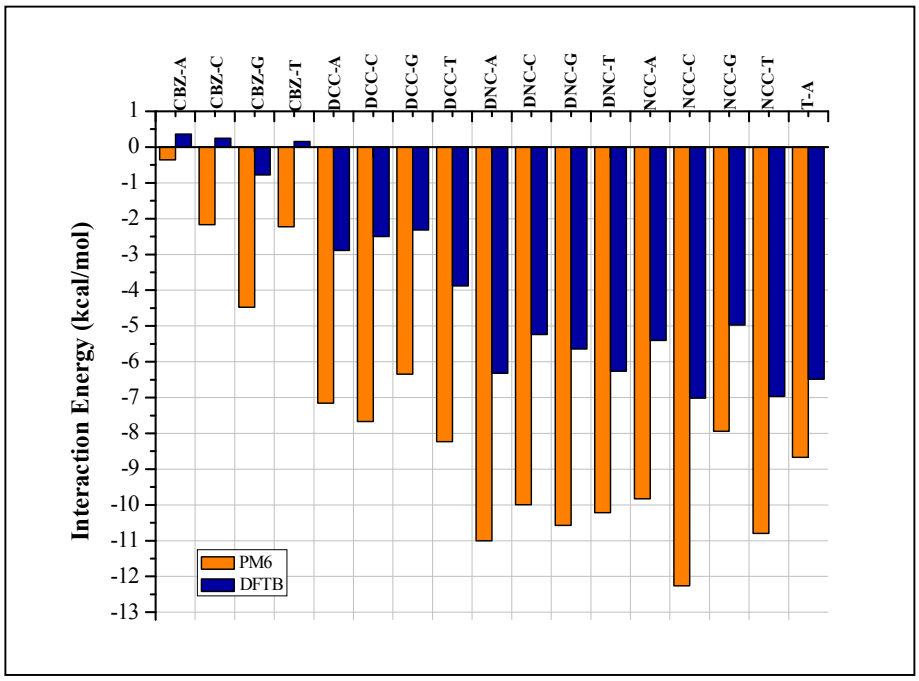


Figure 35 Comparison of average interaction energies between PM6 and DFTB methood

CHAPTER V

CONCLUSIONS

Carbazole derivatives that contained 3-cyanocarbazole, 3,6-dicyanocarbazole, 3-nitrocarbazole, 3,6-dinitrocarbazole, and 3-nitro-6-cyanocarbazole were designed as universal bases. The B3LYP/6-311+G(2d,p) on Gaussian03 was used to calculate their dipole moments. 3,6-Dicyanocarbazole, 3,6-dinitro-carbazole, and 3-nitro-6-cyano-carbazole were chosen to serve as universal bases because they showed large dipole moment. Then, molecular dynamics simulations were applied for modified pentadecamer DNA duplex systems in explicit waters at constant temperature and pressure (NPT condition) for 6 nanosecond by using AMBER 10 package. In each systems, pentadecamer DNA:DNA duplex (5'-TAGTTGT**X**GCGTACA-3':3'-ATC AAC A**Y**C GCATGT-5') were modified by substitution the middle position (X) of the DNA strand with carbazole, 3,6-dicyanocarbazole, 3,6-dinitro-carbazole, and 3-nitro-6-cyanocarbazole and variegated their opposite position (Y) with A, C, G, and T, respectively. At last 5 ns of trajectory were stable and the energies were well-conserved. The selected trajectories product (started from 2.2 nanosecond to 4 nanosecond) showed well equilibrium. From B-factor analysis, there are large molecular fluctuation from carbazole derivatives, particularly carbarzole and 3,6-dicyanocarbazole. This might due to molecular size and non-polar behaviors of them. As simulated geometries, carbazole derivatives were displayed unsuitable arrangement within modified DNA duplexes that could be discussed in terms of electrostatic effect, steric volume, and polarization.

In the past of interaction energy, it was found that both PM6 and DFTB method have similar trend of the calculated interaction energies. The PM6 method gave higher value than the calculated energies from DFTB method. These method predicted that the interaction energy of DNC and NCC modified DNA duplexes had almost similar values. From calculated interaction energy, DNC preferentially hybridized with purine bases (A and G) and the strongest hybridization was found when it opposite A, while NCC preferentially hybridized to pyrimidine bases (C and T) with the greatest stacking ability, specifically C. Additionally, DNC showed the most indiscrimination to hybridize with each natural nucleobase therefore; it has the most universality. Furthermore, the interaction energies of carbazole derivatives were also related to their dipole moment values,

namely large dipole moment could affect the enhancement of stacking interaction of the π-aromatic system of the carbazole derivatives residues.