

REACTIVITY OF SUBSTITUTED SYSTEMS:
A COMPUTATIONAL STUDY

DISSERTATION
ZUR ERLANGUNG DES GRADES
DES DOKTORS DER NATURWISSENSCHAFTEN
DER NATURWISSENSCHAFTLICH-TECHNISCHEN FAKULTÄT
DER UNIVERSITÄT DES SAARLANDES

Jihène Jamil JERBI

Saarbrücken
2018

Tag des Kolloquiums: 20 Juni 2018

Dekan der Fakultät: *Prof. Dr. Guido Kickelbick*

Vorsitzender: *Prof. Dr. Gregor Jung*

Berichterstatter: *Prof. Dr. Michael Springborg*

Prof. Dr. Reinhold F. Fink

Akademischer Mitarbeiter: *Dr. Marcel Albrecht*

Declaration

I, Jihène Jamil JERBI, declare that this thesis titled: "Reactivity of substituted systems: A computational study" and the work presented in it are my own. I confirm that

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Saarbücken, April 20, 2018

For the one and the unique man in this world

*For the special man who gave me the greatest gift anyone could give another
person, the one who believed in me*

*For the one who taught me to be a strong woman so that one day I would
have the strength to bear his loss*

For the most handsome, kindest and greatest man ever

For my Jmayil

To the loving memory of my father who unfortunately didn't stay in this world long enough to see his daughter become a doctor

To my mother who had the arduous task of raising four crazy children by herself after the untimely death of my father

To Sabrina, Noussa and Yassinos

“There is no life without rain”

For you my dearest Professor Michael

“Science is beautiful when it makes simple explanations of phenomena or connections between different observations. Examples include the double helix in biology and the fundamental equations of physics.”

Stephen W. Hawking

Answer to question: What are the things you find most beautiful in science? Stephen Hawking: “There is no heaven; it’s a fairy story”, interview in the newspaper The Guardian (15 May 2011)

Acknowledgement

I would like to thankfully acknowledge my advisor *Prof Dr Michael Springborg* for giving me the greatest opportunity to perform research in his laboratory. For his guidance, help and the patience during this journey. He made me grow through this voyage to the scientist I am today. Thank you so much for the encouragement, support and many great words of advice. Without your amazing experience, the fieldwork would never have been carried out successfully. It is been an honor to work with such a kindhearted and caring person. May you find in this work the testimony of my respect and my gratitude.

My gratitude also goes to *Dr Valeri G. Grigoryan* and *Dr Yi Dong* for their love and support, for all what they did to make me feel at home especially after the death of my father.

I would like to thank my friends and colleagues: *Philipp Thiel*, *Dr Mohammed Molayem*, *Dr Stephan Kohaut*, *Nicolas Louis*, *Kai Huwig*, *Chencheng Fan* and *MeiJuan Zhou* for their kindness, their sympathy and for the good times we had spent together, for the lovely atmosphere at work and nice times shared. My sincere thanks also go to our Secretary *Silvia Nagel* for her help in facilitating the administrative affairs.

My deep and sincere thanks also go to Professor *José Pedro Cerón Car-*

rasco (Universidad Católica San Antonio de Murcia, Spain) for his help, his friendship and for the interest he gave to this work. May he find in this work the affirmation of my gratitude.

My gratitude goes also to my friends *Dr Sarra Manina*, *Ila Latifah*, *Dr Amany Al Awadhi*, *Dr Nesma El-Sayed*, *Dr Cheng Hua Bai*, and *Reema Dawabsha*. I thank them for all the beautiful moments together: Saarbrücken won't be the same without you.

Special thanks go to my best friend ever *MSc Rafik Gabsi* for his love and encouragement, for being with me in all stages of my life since we were 9 years old. For never leaving me and handling me when I was not a good person to be with or when I am under so much stress.

An exceptional acknowledgement goes to Professor *José Pedro Cerón Carrasco* (Universidad Católica San Antonio de Murcia, Spain) and Doctor *Javier Cerezo* (Universidad de Murcia, Spain) for examining this thesis and answering the huge number of emails I sent them without being bored or bothered. I really do appreciate your help.

Zusammenfassung

In einem Teil liefern die Verwendung von "konzeptionellen DFT-Prinzipien und die Übergangszustandstheorie Informationen über den Reaktionsmechanismus bei der Methylierung von Cytosin und dessen Demethylierung sowie zur Erklärung der Unterschiede in der Reaktivität verschiedener Moleküle.

Darüber hinaus wurde eine Studie zu BODIPY und seinen Derivaten durchgeführt um einen Einblick in die fluoreszierenden Eigenschaften des letzteren zu erhalten und zu untersuchen wie die Addition verschiedener funktioneller Gruppen die Klickreaktion beeinflusst.

Schlüsselwörter DNA, Methylierung, genetischer Fehler, DFT, Reaktivität, angeregter Zustand, Klick-Reaktion

Abstract

In the present work, two kinds of molecules were studied from the theoretical point of view using quantum chemistry methods under the framework of the density functional theory: the DNA nucleobase cytosine and its derivatives as well as the fluorescent BODIPY and some of its derivatives and the click reaction.

In one part, the use of “conceptual DFT” principles and the transition state theory provide information on the reaction mechanism involved in the methylation of cytosine and its demethylation and to rationalize the differences in the reactivity of different molecules. This makes it possible to identify some enzyme-like drugs against diseases caused by hypomethylation.

In addition, a study was made on BODIPY and its derivatives in order to shed light on the fluorescent properties of the latter and how the addition of different functional groups can affect the click reaction.

Keywords DNA, methylation, genetic error, DFT, reactivity, excited state, click reaction

Preface

This thesis is based on of the following papers:

- *J. Jerbi* and M. Springborg : Reactivity Descriptors for DNA bases and the Methylation of Cytosine. International Journal of Quantum Chemistry, 2017; e25538. Reprinted with permission of Wiley and Sons: Copy right 2017 Wiley
- *J. Jerbi*, M. Springborg, H. de Haan, and J.P.Cerón-Carrasco: S-Adenosyl-l-Methionine Analogs As Enhanced Methyl Donors: Towards Novel Epigenetic Regulators: Chemical Physics Letters 690 (2017) 74–81. Reprinted with permission from Elsevier
- *J. Jerbi*, M.Springborg: Computational study of the reactivity of cytosine derivatives. Journal of Computational Chemistry 2017, 38, 1049–1056. Reprinted with permission from Wiley
- M. Albrecht, A. Lippach, M. P. Exner, *J. Jerbi*, M. Springborg, N. Budisa, and G.Wenz: Site-specific conjugation of 8-ethynyl-BODIPY to a protein by [2 + 3] cycloaddition. Organic Biomolecular Chemistry, 2015, 13, 6728-6736. Reprinted with permission from Royal society of

Chemistry

- J.Jerbi, M. Albrecht and M.Springborg: Reactivity of BODIPY derivatives and click reaction (A working report on the last project during this PhD work and in preparation for submission)

Table of Contents

1	DNA AND ITS MODIFICATIONS	8
1.1	DNA methylation and DNA methylase	9
1.1.1	History	9
1.1.2	Representation and conservation within species	10
1.1.3	Reaction mechanism	11
1.2	DNA demethylation and DNA demethylase	12
1.3	Role of DNA methylation in some human pathologies	16
1.3.1	Cancer	16
1.3.2	Hypermethylation of tumor suppressor genes	16
1.3.3	Global hypomethylation of the genome	17
1.3.4	Other pathologies	18
1.3.5	Therapeutic hopes	18
2	BODIPY AND CLICK REACTION	20
2.1	Overview	21
2.2	Definition	21
2.3	Principle of Fluorescence	23
2.4	BODIPY	25

2.5	Why BODIPY	27
2.6	Click chemistry	30
2.6.1	Definition	30
2.6.2	DNA and click chemistry	31
3	THEORETICAL BACKGROUND	32
3.1	Overview	33
3.2	Schrödinger equation and Hamiltonian	33
3.3	Born-Oppenheimer approximation	34
3.4	DFT and Kohn Sham equations	35
3.5	Functional	36
3.5.1	Local Density Approximation LDA	36
3.5.2	Generalised Gradient Approximation: GGA	37
3.5.3	Hybrid functional	37
3.6	Conceptual DFT	39
3.6.1	Electronic Chemical Potential and Mulliken Electronegativity	40
3.6.2	HSAB Principle and Maximum Hardness Principle (MHP)	41
3.6.3	Chemical Hardness and Softness S	42
3.6.4	Fukui Function $f(r)$	43
3.6.5	Electrophilicity ω and Nucleophilicity Index	44
4	REACTIVITY OF BODIPY DERIVATIVES	47
4.1	BODIPY Derivatives	48
4.1.1	Introduction	48
4.1.2	Methods	50

4.1.3	Experimental details	50
4.1.4	Reactivity	52
4.2	Click reaction	63
4.2.1	Reaction with compound II	65
4.2.2	Reaction with compound III	67
4.3	Conclusion	69
	Bibliography	70

General introduction

This work was carried out in the “PTCS: Physical and Theoretical Chemistry Saarland” group in the Physical Chemistry department at Saarland University. The aim of this work was to understand in depth the reactivity of deoxyribonucleic acid DNA because of its important role in the conservation of genetic information in all living beings as well as BODIPY as a fluorescing molecule used to detect damaged cells in the human body.

During this thesis, we focused on two well-known reactions to identify simple tools for studying chemical reactivity. Thus, we studied the DNA methylation and demethylation which present the main part of this work. In addition, we studied the BODIPY as fluorescent and also studied some of its derivatives and how substitution affects a click reaction.

The large size of DNA molecule and its complex environment make it necessary to consider simplified models of the latter to understand its reactivity. Moreover, when adding long molecular chains to some of BODIPY’s positions, the system becomes very complex and also then a model is needed to get information about its reactivity. This can be studied experimentally,

using various methods of analytical chemistry, but it can also be simulated.

Among all the methods of quantum chemistry, those based on density functional theory (DFT) are used here thanks to the good compromise it offers between reliability and computational effort.

Unlike the *ab initio* quantum chemistry methods, DFT methods are not centered around the wave function, but on the electronic density. These methods, like all quantum chemistry methods, make it possible to calculate the energy and electronic structure of modeled molecular systems. From these two quantities, various indices of reactivity derived from qualitative models of the chemical reactivity can be evaluated, allowing to rationalize or even to predict the chemical reactivity.

We shall briefly describe the background for each part separately and how they are inter-linked. After providing definition and historical background of DNA methylation, we will discuss its importance in biology by describing the role held by this modification. We will present the mechanism and subsequently, we will present the demethylation process.

In Chapter 2, we will introduce the basics of fluorescence; the most used fluorophores and why BODIPY is the one used as a DNA bio-label. Moreover, we will present a brief description of the click chemistry.

In chapter 3, we will present the basis of the theory used in this work: Density Functional Theory and conceptual DFT.

The last chapter is a working report on the last project made during this PhD work: Reactivity of BODIPY derivatives and click reaction.

In the end, all the four papers produced in this thesis are listed.

Chapter 1

DNA AND ITS MODIFICATIONS

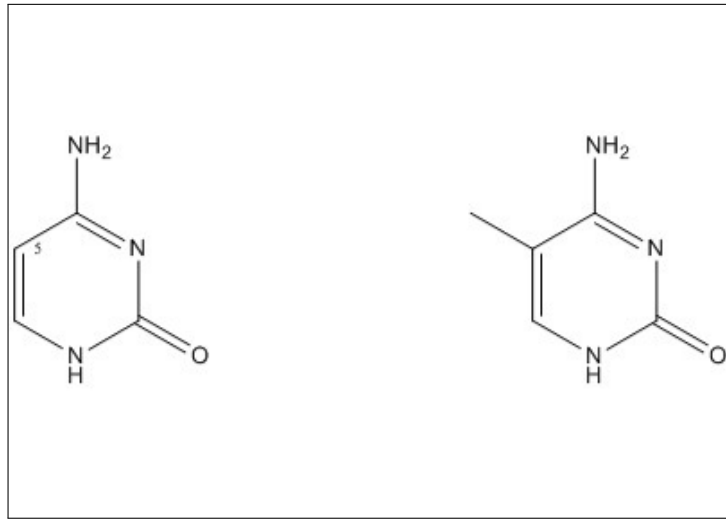


Figure 1.1: Schematic representation of Cytosine and 5-methyl-Cytosine.

1.1 DNA methylation and DNA methylase

DNA methylation is defined as a modification of one of the four nucleobases (guanine, cytosine, adenine, and thymine) of the deoxyribonucleic acid that occurs after the replication. This modification consists of adding a methyl group $-\text{CH}_3$ to the cytosine carbon position 5 replacing a hydrogen atom (see figure 1.1). It is possible that any of the four bases can be methylated. However, the cytosine one that frequently is modified in the mammals.

1.1.1 History

Fifty years ago, Rolin Hochtiss [1] discovered the methylated cytosine: 5-methyl-cytosine. It was only after twenty years that research on DNA methylation could be started thanks to Nirenberg and Matthaei and their discovery

of the genetic code [2]. In 1969, Kuhnlein et al. discovered the DNA methyltransferase in *E. coli* and mice [3]. Almost 25 years after the discovery of the methylated cytosine, Roy and Weissbach could isolate and characterize the first human DNA methyltransferase from HeLa extract [4]. Subsequently, Stein et al., in 1982, could confirm that the *in vitro* methylation of the gene *aprt0* (adeninephosphoribosyltransferase) previously transfected (the process by which purified nucleic acids are introduced into mammalian cells) into murine cells can lead to an inactive gene. These results suggested that DNA methylation has a role model in gene repression [5].

1.1.2 Representation and conservation within species

DNA methylation is a phenomenon that is largely conserved during evolution. It is known that this biological process is fundamental in bacteria, plants and animal cells since it helps to encode the information in a stable and reversible manner [6, 7]. In humans, the mutation rate is 10 to 50 times higher for 5-methylcytosines than for the other bases of DNA and this modification is the origin of a large proportion of mutations. This explains why CpG dinucleotides are underrepresented in the human genome [8, 9]. Within animal genomes, the rate methylation varies substantially [10]. It should be noted that in prokaryotes (unicell organism that lacks a cell nucleus) methylation plays a role not only as a defense against the introduction of foreign DNA (unmethylated, it will be degraded by enzymes restriction) but also a role of control of the fidelity of replication of the DNA [11, 12].

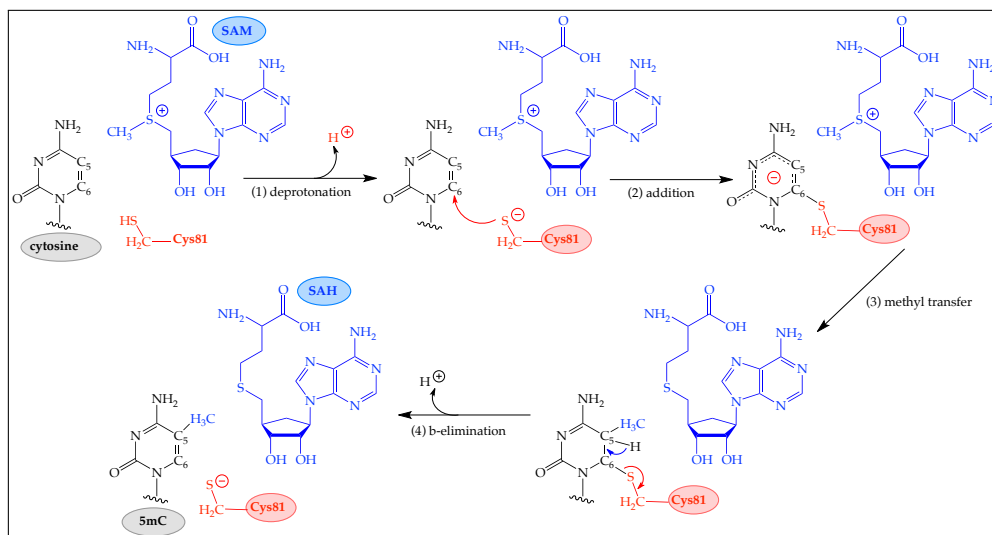


Figure 1.2: Mechanism of DNA methylation (Figure extracted from our recent work, [13]).

1.1.3 Reaction mechanism

The methylation mechanism involves an addition of a methyl group to a cytosine residue where the methyl group is transferred as a cation [14]. DNA methyltransferase (DNMT)s use S-adenosyl-L-methionine (SAM) as methyl donor in a three steps mechanism (see Figure 1.2):

- Since the parent cytosine is not nucleophile enough to spontaneously undergo the electrophilic attack by SAM, the process needs to be initiated by the addition of a cysteine residue located in the binding site (Cys81) to the C6 position of the flipped-out cytosine [15]
- The resulting activated-cytosine results in a partial-negatively charged C_5 that subsequently reacts with SAM by transferring of a positively charged methyl group towards DNA [14]

- The proton at C₅ position is released in the β -elimination step to a surrounding base, e.g., hydroxide anion. The reaction finally leads to the C₅-methylcytosine product (5mC) and the byproduct S-adenosyl-L-homocysteine (SAH) [16]. Although the complete picture of DNMTs activity is more complex than the simplified Fig. 2, [17] it is now well established that methyl transfer of step 3 is the rate-limiting reaction. [18]

1.2 DNA demethylation and DNA demethylase

The process of DNA demethylation can be both passive and active. Passive DNA demethylation occurs in dividing cells. Whereas the active process can occur in both dividing and non-dividing cells and this path involves enzymatic reactions to transform the 5mC back to the unmodified (natural) cytosine form [19, 20, 21, 22].

Since there is no well known direct mechanism in mammalian cells that can break the strong covalent bond connecting cytosine to the methyl group, the most logical pathway for demethylation appears to involve some different enzymes from methylation [23]. It is worth noting here that 5mC can be modified in two ways. The first mechanism is based on an initial deamination of the C₄ amino group by AID/APOBEC (activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme complex) after which 5mC is converted to Thymine. That process can eventually lead to a G/T mismatch

(mis-incorporation of Thymine instead of Cytosine in the base pair) which induces the base excision repair (BER) pathway to correct it as suggested by Rai et al., [24] and Bhutani et al. [25].

The second pathway studied in this thesis is the oxidation of 5mC which involves too the surveillance of BER to replace the residues with the unmodified cytosine.

First in the year 2009 that Tahiliani and coworkers [26] discovered the presence of 5-hydroxymethyl- cytosine (5hmC) in DNA sequences while searching for alternative pathways of active demethylation. The latter is an oxidation product of 5methyl-cytosine. Pfeifer et al., [27] suggested that this oxidation is the first step in the DNA demethylation pathway since the appearance of 5hmC reduces the level of 5mC at any nucleotide position [27].

Later in the same year, the group of Tahiliani discovered the so-called ten-eleven translocation (TET) enzymes: this family of demethylase was considered first as catalysts to the oxidation of the exocyclic methyl group of thymine. Iyer and coworkers, proved that oxidation of 5mC is facilitated by the TET enzyme supervision of the demethylation process. [26]

5hmC was found in cells during the development of embryo's brains and abundant in ES cells, primordial germ cells and fertilized oocytes [27]. Moreover, 5hmC was found *in vivo* in mammalian tissue and may play an important role in regulating DNA demethylation and gene expression [27]. This observation throws a light on the importance of 5hmC and then the demethylation process. The formation of 5mC was the initiation of more

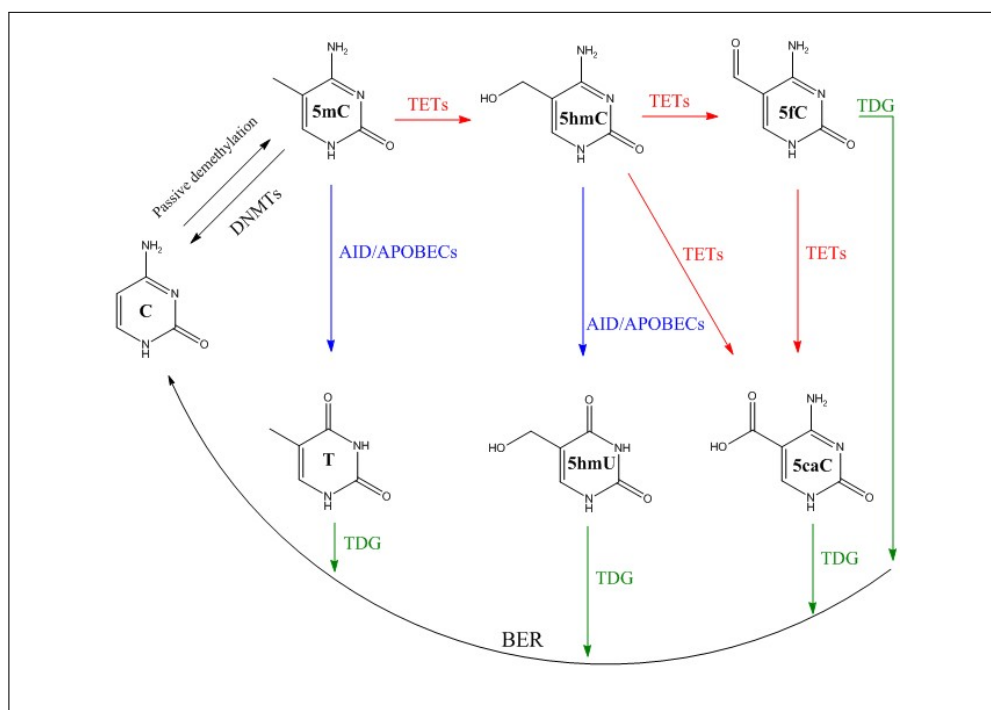


Figure 1.3: Paths for DNA demethylation (Figure inspired from reference 23).

useful molecules for the development of embryo.

After the formation of the sixth basis of DNA (5hmC), two new mechanisms can be proposed: The first, using AID/APOBEC as the one described above leads to 5-(hydroxymethyl)-uracil (5hmU). Under the supervision of thymine DNA glycosylase (TDG) and base excision mechanism (BER), 5hmU is transformed to cytosine [28]. In the second path, TET enzymes oxidize iteratively 5hmC to form 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) [29] (see figure 1.3).

TDG is essential for DNA demethylation and is required for normal cell development (the lack of this enzyme will lead to anomalies). It is associated with BER movement in all mentioned paths above to replace the modified cytosine residues with unmodified cytosine [30, 31]. According to a suggestion by Nabel and coworker in 2011 for some cells the active demethylation can be self-fixed in order to re-differentiate [32].

1.3 Role of DNA methylation in some human pathologies

1.3.1 Cancer

Day by day, we realize that methylation of DNA can be an alternative pathway to mutations in the development of cancers. Today, we know that methylation of DNA is involved in more than 65 % of cancers [33]. According to Jiang et al., DNA methylation was the first alteration observed in cancer cells [34]. Houshdaran and coworkers claim that the abnormal methylation can lead to hypermethylation or hypomethylation (defined in the next part) which is most likely to be the cause of most of the carcinogenesis [35].

During tumorigenesis (formation of Cancer), two types of unusual DNA methylation can be found: (1) hypermethylation of tumor suppressor genes (a gene that protects cells from Cancer). (2) hypomethylation of the genome leading to chromosomal instability or the expression of oncogenes (gene causing Cancer).

1.3.2 Hypermethylation of tumor suppressor genes

DNA hypermethylation can be defined as the regional methylation of CpG islands (region with large density of CG). The latter is found to be in tumor suppressor genes in cancer cells. Therefore, this aberrant methylation of the tumor suppressor genes can cause cancer.

An increase in the methylation of tumor suppressor genes leads to a decrease in their expression (existence) and a loss of their protective effect facing the

deregulation of cell development. Jones and Baylin [36] proved that there exists a growing number of hypermethylated genes associated with carcinogenesis, they concern promoters directing the transcription of genes involved in critical stages of cell growth [37].

The over-expression (excessive expression) of three DNMT was the first observation while studying human tumors [38, 39]. Indeed, a rise in the percentage of DNMT has been detected in different types of tumor [40, 41, 42]. The existence of DNMT 1 has been shown to be necessary and sufficient for oncoprotein-induced transformation [43]. DNMT3a seems to play an important role in the survival of tumor cells [44, 45].

One of the consequences of hypermethylation in a tumor suppressor gene is to switch the genes off. An assumption that comes directly to mind is that this increase could be responsible for the hypermethylation of promoters regularly observed in tumors as proposed by Feinberg and Rodenhiser [46, 38, 39]: over-expression of DNMTs is linked to cancer in humans.

Yet today, there is still no clear correlation between Cancer and the increase of DNMT level as this over-expression does not appear to exist in all cancers. It therefore seems that the increase in DNMT expression may be necessary but is not sufficient to cause the methylation defects observed in the tumors.

1.3.3 Global hypomethylation of the genome

It has long been known that the genetic material of malignant cells exhibit global hypomethylation [47, 48]: the decrease in methyl-CpG at the non-regulatory regions and structural elements [49, 50]. Feinberg and Tycko [46] claimed that in different hypomethylated tumor samples, changes in the cells

are correlated with tumor progression [46]. As consequence, cancer initiation is linked to a specific DNA hypomethylations.

Hypomethylation is found in breast cancer [51] and prostate cancer [52]. Hypomethylation also influences the expression of oncogene.

1.3.4 Other pathologies

The methylation of DNA is involved in other pathologies, too. An aspect common to this is mental retardation to varying degrees. The activity of methylation of DNA is, moreover, particularly high in neurons in spite of their state of terminal differentiation, when cell changes from one cell type to another, [53, 26]. This suggests that the DNA methylation/demethylation is particularly important for the development of the central nervous system.

1.3.5 Therapeutic hopes

Scientists worldwide are trying to find connections between methylation DNA alteration, gene expression and human diseases experimentally and computationally [12, 54]. The results of these studies can be the basis of the hope for treating patients with major diseases like cancer, as well as for understanding and preventing complications of treatment of damaged tissues.

The most well-known DNA methylation inhibitors used are the nucleoside analogues such as 5- azacytidine and 5-aza-deoxycytidine. These drugs differ from cytosine only in that they possess nitrogen replacing the carbon atom at the 5 position of the pyrimidine ring: nitrogen cannot accept methyl group, there follows a detachment of the methylating enzyme. Thus, incorporated into the DNA, after a few cycles of replication, many sites are demethylated.

Because of their mechanism of action, these analogues require the DNA replication and thus cell division [55, 56].

Several clinical studies have been conducted since the 1970s for the treatment of various types of cancer [57, 58]. Unfortunately, a reduction of the tumor is accompanied by many undesirable side effects such as haematological toxicity, gastrointestinal tract and phlebitis [59]. Today, teams therefore work on healthier programs to avoid the maximum adverse effects of these treatments by maintaining the demethylation activity for regression of cancers [60].

According to Cooney et al., reversibility of epigenetic changes is the key target in cancer therapy [61, 62]. Zebulainin, [1- (beta-D-ribofuranosyl) -1,2-dihydropyrimidin-2-one] is another nucleoside analogue which is considered among the Dnmt inhibitor drugs that changes gene activity. Thanks to its favorable pharmaco-kinetics, it seems a very good candidate for the demethylating treatment of tumors [63, 64, 56].

Other, more specific molecules such as antisense oligonucleotides directed against the DNMTs are being developed and are hoped to be able to reduce the harmful side effects of global demethylation.

Chapter 2

BODIPY AND CLICK REACTION

2.1 Overview

The use of fluorescence has developed significantly during the past 20 years, in the field of biological sciences. For instance, in biochemistry and biophysics, fluorescence spectroscopy and time-resolved fluorescence are considered as primary research tools for imaging.

In addition, fluorescence is used in biotechnology, flow cytometry, medical diagnostics, DNA sequencing, forensics, and genetic analysis. Moreover, the latter has been used dramatically in cellular and molecular imaging which may help to localize and measure the intercellular molecules [65]

The use of fluorescent probes coupled to an addressing sequence (RNA) or directly genetically encoded, allows the specific observation of each component of a biological cell.

2.2 Definition

In general, fluorescence provides an impressive tool for visualizing biological structures or processes. When a fluorescent probe is attached to a structure of interest one can have several observations. Contrary to phosphorescence, in fluorescence, in the excited orbital, the electron is paired to a second electron in the ground state by opposite spin [65] which makes the return to the ground state spin-allowed and happening fast by photon emission. And then the expected lifetime of fluorescence is about 10 ns (10×10^{-9} s) [65, 66, 67]

What distinguishes phosphorescent and fluorescent bodies is the duration of the intermediate phase: In the case of fluorescence, the emission almost

immediately follows the excitation whereas, in the case of phosphorescence, the emission can take place up to several days after the excitation.

Indeed, the use of fluorescent probes attached to a structure of interest will allow selective observation. In order to improve their properties (Blinking, photostability and brightness) a development of new fluorescent probes is needed.

Photostability [68] is defined as the average number of successive excitations from the ground state that a molecule can undergo before being degraded, which limits the illumination time of a compound during an experiment. This parameter is significant in the drug discovery field where it is needed to photo test the drug and drug product to make sure that exposing them to light won't cause photochemical degradation [69]. The latter can happen after successive excitation of the same probe which will keep it in a permanent non-fluorescent state.

Studies have notably been carried out on the synthesis of fluorescent organic nanoparticles [70] making it possible to overcome the problems associated with the use of conventional fluorescent molecules that photodegrade rapidly. Biocompatibility of the fluorescent probes [71] with the system characterizes the degree to which a probe can be placed in a biology medium without causing a cellular response or causing a malfunction or destruction of the medium.

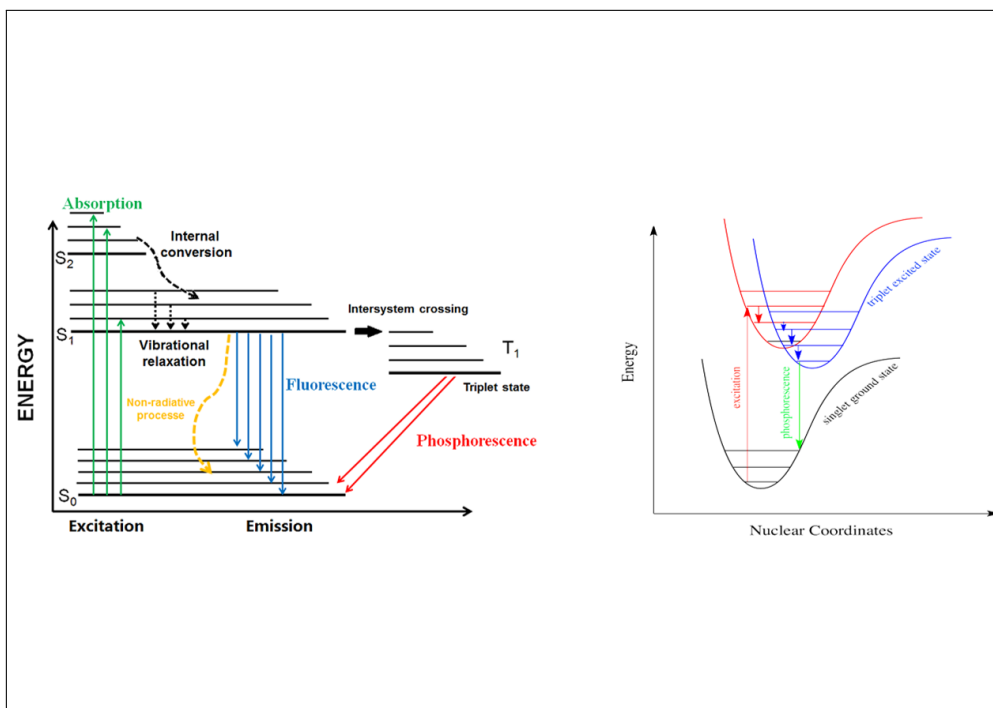


Figure 2.1: Jablonski diagram for a molecule after excitation. Adapted from reference [73, 74].

2.3 Principle of Fluorescence

After molecule absorbs a photon (light), it is excited from ground state (or any state n) to some excited state which results in passing from the ground state S_0 to an excited state S_n (with $n \geq 1$). Yet the excited state is energetically not stable and the absorbed energy will be released rapidly. Between the absorption and emission of light, a series of processes can occur and are usually illustrated with the help of a Jablonski diagram [72] (figure 2.1)

After the vertical excitation, the system is in a higher vibrational level and will rapidly relax to the lowest vibrational level of that electronic state. This

process is called “internal conversion and vibrational relaxations”. There are then two possible ways of de-excitation of the state S_1 with two kinetic constants k_r and k_{nr} , radiative and non-radiative respectively. Fluorescence represents the radiative pathway, transitions of internal conversion type (S_1 to S_0) and inter-system crossing type (S_1 to T_1) this is what can lead to phosphorescence. The latter is forbidden when relativistic effects can be ignored but can occur in reality if the molecule contains heavy atoms or is solvated. The process of emission from T_1 is actually the phosphorescence.

2.4 BODIPY

Till date, thousands of organic fluorophores have been discovered and synthesized. In the following section, we will present a brief overview of the well-known fluorophore families which comprise fluoresceins, rhodamines, cyanines and 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene (BoronDiPyrromethen or simply BODIPY) dyes.

As illustrated in Figure 2, the core structure in Bodipy dyes is a heteroaromatic polycyclic species composed of 2 pyrrole subunits, complexed by the nitrogen atoms to a difluoroborane group, BF_2 [75]. The boron atom of the latter adopts a tetrahedral geometry and therefore the two pyrrolic units to be coplanar. The official numbering (IUPAC) of the aromatic nucleus is presented in Figure 2.2.

In analogy to the current nomenclature of porphyrins and pyrroles, position 8 is often called meso position, positions 3 and 5 positions α , positions 2 and 6 positions β . Positions 1 and 7 will be called positions β' .

Treibes and Kreuzer [76] were the first to synthesize BODIPY dyes in 1968. However, these dyes received much more interest for biochemist and biologist as fluorescence producer only in the 90' s. Many studies were made on BODIPY in a wide diversity of applications [77]. As examples, we mention organic dyes for photovoltaic applications [78], drug delivery materials [79], agents for cellular imaging [80, 81], fluorescent [82], ion/molecule [83, 84] and pH probes [85], metal chelators [83, 86, 87], sensors for redox active molecules [83], photodynamic therapy [88, 89], molecular biolabeling [90, 91], chemical sensors [92, 93, 94], cell imaging [92, 95] dye-sensitized and bulk heterojunc-

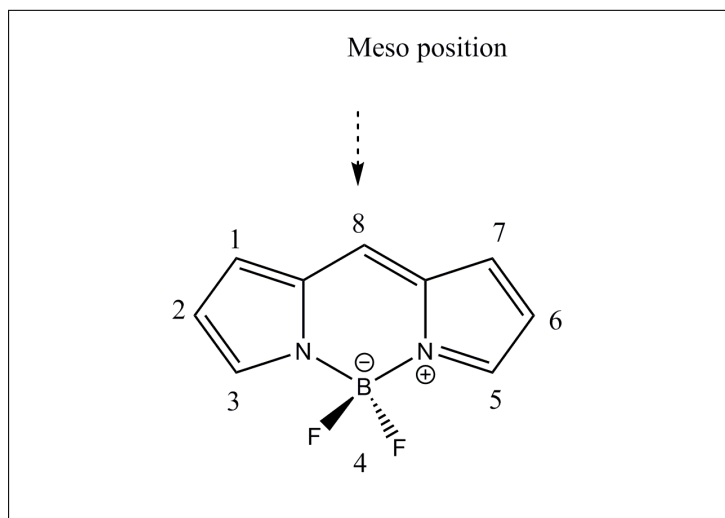


Figure 2.2: Labeling of the BODIPY molecule.

tion solar cells [96, 97, 98, 99], and two-photon absorption [100, 101].

These applications were only possible due to the exceptional properties of BODIPY such as high molar absorption coefficients and high fluorescence quantum yields, photochemical and chemical stability, negligible triplet-state formation, good solubility, excitation/emission wavelengths in the visible spectral region and fluorescence lifetimes in the nanosecond range, resistance to self-aggregation in solution, and narrow emission band widths with high peak intensities [82, 102, 103].

BODIPY can be considered one of the most promising successors (fluorescein, cyanine, and rhodamine) for *in vivo* fluorescent labels and it has found applications for low toxicities and has outstanding physical-chemical properties mentioned above. They have been applicable on toxins [104], lipids [105, 106], monosaccharides [107], and proteins [108, 109].

BODIPY derivatives with long linear alkyl chains on the periphery can min-

imize the aggregation problem associated with the *pi-pi* stacking, which is common for this compounds. In fact, BODIPYs using a single long alkyl chain have already been reported, and are also valuable fluorescent probes for applications in the study of fluidity changes in cell membranes [110] and of sol-to-gel transitions [111].

The small Stokes shift is the typical drawback for BODIPY [112, 113]. The relevance of this problem could be reduced by attaching an additional light absorber to BODIPY core [114, 115]. Thus, the hydrophobic nature of BODIPY is the real problem in biological applications since great water solubility is mandatory for fluorescence probes to be applicable in biological surroundings.

2.5 Why BODIPY

As reviewed by Kowada et al. [116], the benefit of small molecules compared to fluorescent proteins (defined earlier in this part) is their small size which makes their fictionalization easy. Moreover, their fluorescence near infrared is a plus to the list of properties compared to those of fluorescent proteins [117, 118].

A large number of fluorescent probes based on small-molecules have been developed [119] thanks to their short time in labeling target bio-molecules by combining either small molecules with bio-orthogonal ligations [120, 121] or peptide or protein tags [122, 123].

As mentioned above in this part, fluorescein, cyanine, rhodamine are among the most used fluorophores in bio-systems. Diverse small molecules based on

those dyes are commercially available because of their great properties.

However, while concentrating on the applications of those fluorophores on live-cell imaging, it is noticed that there are also many features that make them not good candidates and alternatives would be useful [116].

For example, the negatively charged fluorescein lacks cell-membrane permeability usually which leads to a reduction of its binding properties to cellular bio-molecules [116].

Rhodamine, on the other hand, can penetrate easily into cells even as a cationic dye. Unfortunately, it is also prone to either be localized in mitochondria, adsorb to proteins or lipids because of its cationic and hydrophobic properties [124]. This in turn can be problem in cellular imaging.

In addition, the low photostability of Cyanine makes it not suitable for cell imaging although it is very useful in tissues targets.

Yet, the outstanding criteria (listed above) of BODIPY and its vanishing total charge make it currently the best candidate as a fluorescent in cell imaging [125, 126]. In addition, BODIPY fluorescence is easy to be shaped by different methods [127, 128].

As described above, the application of BODIPY in biological imaging is not a trivial task because of few major drawbacks, including hydrophobic and lipophilicity, which makes it accumulate in sub-cellular membranes [116]. But as a neutral dye, new BODIPY based probes are needed to be developed since they will easily penetrate into the membrane.

In 2014, Hocek et al., [129] could synthesize new fluorescent nucleosides and nucleoside triphosphate analogues carrying an F-BODIPY fluorophore

using a short flexible nonconjugated tether. The synthesized BODIPY-labeled nucleoside triphosphates were easily integrated into DNA in primer extension, nicking enzyme amplification reaction, and polymerase chain reaction.

According to Hocek et al., BODIPY -based dyes belong to fluorescence family with exclusive properties to be used as fluorescent tags and the new discovered F- BODIPY can be detected in solutions and gels. Moreover, while incorporated into DNA the new F- BODIPY does not quench. As an application of the deoxyuridine derivatives fluorescent linked to a BODIPY fluophore tethered via long flexible linker, we mention the in cellulose fluorescence labeling of DNA and detection of apoptotic cells [130].

Figure 2.3 presents the new synthesized BODIPY based dye linked to one of the DNA bases Cytosine. A detailed description of the experimental results can be found in the work of Hocek et al. [129]

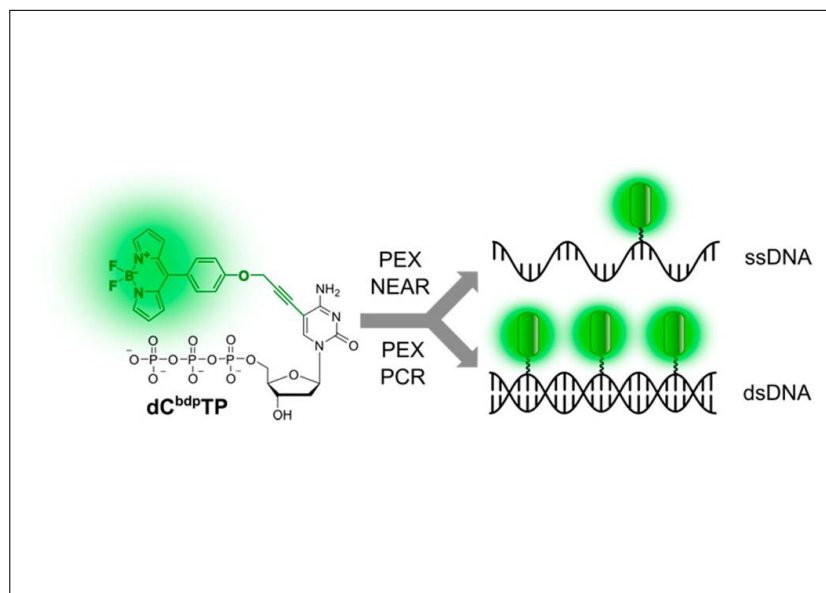


Figure 2.3: New synthesized Bodipy based dye linked to one of the DNA bases Cytosine (Adapted from reference [129].)

2.6 Click chemistry

2.6.1 Definition

In 2001, Sharpless and coworkers [131] connecting small molecular systems together in a fast and reliable way, under simple reaction conditions. The main goal is to mirror Nature's methods of creating molecular diversity [132] Moreover, the click reaction's selectivity and stability provide high sensitivity and low background signal. [133]

There are certain criteria that must be satisfied by a reaction before it can be called a "click" reaction [132]:

- High yielding and stereospecific
- Insensitivity against air and humidity
- TProduction of clean derivatives which can be removed by non-chromatographic methods
- Possible in favorable solvent or pure

As mentioned above, such reactions are fast since they have a great thermodynamic driving force (higher than 20 kcal/mol) which makes them highly selective for one product.

There exist many click reactions like alkene hydrothiolation reactions, epoxidations, aziridinations, dihydroxylations, and nucleophilic substitutions. However, the most well-known reaction which possess the specific criteria of the “click reaction” is copper-catalysed azide-alkyne cycloaddition which has high yielding and is regioselective and can be utilized in many applications [134]

2.6.2 DNA and click chemistry

Recently and in very short time, click chemistry has become an important tool in labeling DNA oligonucleotides. It can be used as biologically active moiety, fluorescent dye or in applications in synthesizing longer oligonucleotides [132].

Chapter 3

THEORETICAL BACKGROUND

3.1 Overview

In this third part of this bibliographic chapter, we will introduce the basic ideas of the Density Functional Theory (DFT) and the Conceptual DFT approach used in this thesis. An overview of the Schrödinger equation and the Hamiltonian are given as a first step, and then DFT is introduced briefly.

In the end of this part, we will present the conceptual DFT approach. This aims at identifying a correlation between electronic distribution and reactivity. In here we will introduce the theory of frontier orbitals, [135, 136] which derives from the theory of molecular orbitals and the conceptual density functional theory. [137, 138, 139]

3.2 Schrödinger equation and Hamiltonian

Quantum mechanics and theoretical chemistry are based on the Schrödinger equation. One can describe the state of a system with M atoms and N electrons by a wave function Ψ satisfying the stationary Schrödinger equation:

$$H\Psi = E\Psi \quad (3.1)$$

where H represents the Hamiltonian operator, Ψ is the wavefunction and E is the energy of the system. By using the atomic units $\hbar = e = m_e = 1$, the Hamiltonian can be written in the following form:

$$H = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (3.2)$$

where A, B represent the M nuclei and (i, j) denote the N electrons. Alternatively, the Hamiltonian can be written as:

$$H = T_N(R) + T_e(R) + V_{en}(R) + V_{nn}(R) + V_{ee}(R) \quad (3.3)$$

The first two terms of the equation show the kinetic energy of nuclei (T_N) and electrons (T_e), the other terms are Coulomb interaction for each pair of charged particles: V_{en} the attractive potential between electrons and nuclei, electron-electron repulsion term V_{ee} , and V_{nn} is the nuclear repulsion energy.

An exact solution of the Schrödinger equation is possible only for the case of the hydrogen atom and the one-electron hydrogen-like atoms. The presence of the term $\frac{1}{r_{ij}}$ in the multi-electron systems makes it impossible to find an exact solution but one has to apply approximations.

3.3 Born-Oppenheimer approximation

This is one of the first approximations used in chemistry for solving the Schrödinger equation: Since the electrons are much lighter than the nuclei, *i.e.* their movement is faster, the kinetic energy of the nuclei can be neglected as a first approximation. Consequently, the positions of the nuclei are considered fixed and their kinetic energy is neglected to allow the nuclear-nuclear repulsion term to be constant and then added later to the final result of the calculated electronic energy.

The total wave function can be written as a product of the nuclear $\Psi_{ni}(\vec{R})$ and electronic functions $\Psi_{ei}(\vec{R}, r)$ (the electronic wave functions that diago-

nalize the electronic Hamiltonian H_e)

$$H_e = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{Ai}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (3.4)$$

3.4 DFT and Kohn Sham equations

The principle concept of density functional theory (DFT) [140] is that the ground-state energy and all other ground-state properties are uniquely determined by the electron density of the ground state $\rho(r)$ as shown by Hohenberg and Kohn in 1964. They demonstrated that the energy is a functional of the electron density only $E[\rho]$ and can be written as follow:

$$E[\rho] = T[\rho] + E_{Ne}[\rho] + E_{ee}[\rho] \quad (3.5)$$

As observed by Kohn and Sham, the kinetic energy of the system in which the electrons are in interaction cannot be calculated exactly. The idea of the approach of Kohn and Sham is to replace the system studied by another equivalent one. This method allows the kinetic energy of the electrons to be separated into an exact calculable term T_s , and a calculated correction with an approximate functional. The energy then becomes:

$$E_{DFT}[\rho] = T_s[\rho] + E_{Ne}[\rho] + E_{ee}[\rho] = T_s[\rho] + E_{Ne}[\rho] + J[\rho] + E_{xc}[\rho] \quad (3.6)$$

T_s corresponds to the electron kinetic energy of the hypothetical system but ρ is the same as for the real system. $J[\rho]$ is the Coulomb interaction between electrons and E_{ex} is the exchange correlation energy, both they

form the potential energy of electron-electron interaction E_{ee} . The exchange-correlation energy can be written after equalizing the DFT and the exact energies as:

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (E_{xc}[\rho] - J[\rho]) \quad (3.7)$$

It is worth to mention that the exact E_{xc} is unknown but so rather small in a way that approximating it won't lead to dramatic errors. An advantage of this method is its lower computational requirements compared to other methods that can give results of the same level of precision for both geometry and thermochemistry problems. The theory of DFT is well described in several textbooks [140, 141, 142, 143, 144].

3.5 Functional

3.5.1 Local Density Approximation LDA

It is based on the approximation that the density can be approximated locally through a uniform electron gas. It is a simple model that gives good results for solids.

LDA, like any other methods, has its advantages and disadvantages. It can give accurate results when analyzing the molecular properties such as structures and vibrational frequencies but on the other hand it poorly characterizes the energetic details.

3.5.2 Generalised Gradient Approximation: GGA

Due to the problem of LDA with properly describing energetic properties, it was necessary to improve the results. To better reflect the non-homogeneity of the electron density, the gradient of the latter can be introduced into the description of the exchange-correlation effects. The resulting methods are called GGA functional. As the LDA approximation, exchange and correlation contributions are usually treated separately.

$$E_{xc}^{GGA} = E_x^{GGA} + E_c^{GGA} \quad (3.8)$$

In general it can be considered that GGA functional class allows obtaining better results than the LDA. In particular, they can describe relatively well systems involving hydrogen bonds. However, they poorly describe Van der Waals complexes.

3.5.3 Hybrid functional

In order to further improve the exchange-correlation energy functional, the way to build them has been modified to give rise to a new class of functionals that express the exchange energy as a fraction of the exact exchange energy (which can be obtained by the Hartree-Fock method provided when replacing the spin-orbitals HF by spin-orbitals KS), and a fraction of exchange and correlation energy from DFT, hence the name given to them of hybrid functionals. This class is widely used nowadays. B3LYP is a well-known example of a hybrid functional used nowadays. It is a mixture of LDA and

GGA functional and can be written as shown below:

$$E_{xc}^{B3LYP} = (1 - a)E_x^{LSDA} + aE_{xc}^{HF} + bE_x^{B88} + cE_c^{LYP} + (1 - c)E_c^{LSDA} \quad (3.9)$$

where $a = 0.20$, $b = 0.72$, and $c = 0.81$ are the three empirical parameters determined by fitting the predicted values to a set of atomization energies, ionization potentials, proton affinities, and total atomic energies. Functional LSDA is a functional derived from the functional LDA in which the spin is taken into account. For the open-shell systems case, the electronic density ρ is replaced by the spin electronic densities ρ_α and ρ_β such that $\rho = \rho_\alpha + \rho_\beta$

3.6 Conceptual DFT

The proficiency in enlightening and defining significant chemical concepts of molecular structure and reactivity was one of the outcomes of density functional theory. After the efforts made by Hohenberg and Sham to develop the quantum chemical equations for DFT and, in the late 1970s and early 1980s, Parr developed the “conceptual DFT” [145] defined as follow:

”Conceptual DFT is a DFT-subfield in which one tries to extract from the electron density relevant concepts and principles that make it possible to understand and predict the chemical behaviour of a molecule.”

In conceptual DFT, the ground state of N-electron system depends on the number of electron N and the external potential $\nu(r)$ which are determined by the density and we can accordingly write $E[\rho(r)] = E[N; \nu(r)]$: The responses of the system to changes of these two variables, independently or simultaneously can provide information about reactivity.

As described in the diagram in figure 3.1, the derivatives of $E[N; \nu(r)]$ with respect to N and $\nu(r)$ presents the first series of reactivity descriptors: the chemical potential (opposite of electronegativity), the hardness, the Fukui function $f(r)$ and the two variables linear-response function $X(r, r')$ (The latter is not used in this thesis).

Global reactivity descriptors are presented in the left side of the diagram; they give information about the overall susceptibility of a system to different types of reactions. However, in the right side of the diagram, local reactivity descriptors are defined. They are regioselectivity indicators that show where a molecule is most susceptible to the different type of reagents i.e., their values depend on the position where they are evaluated.

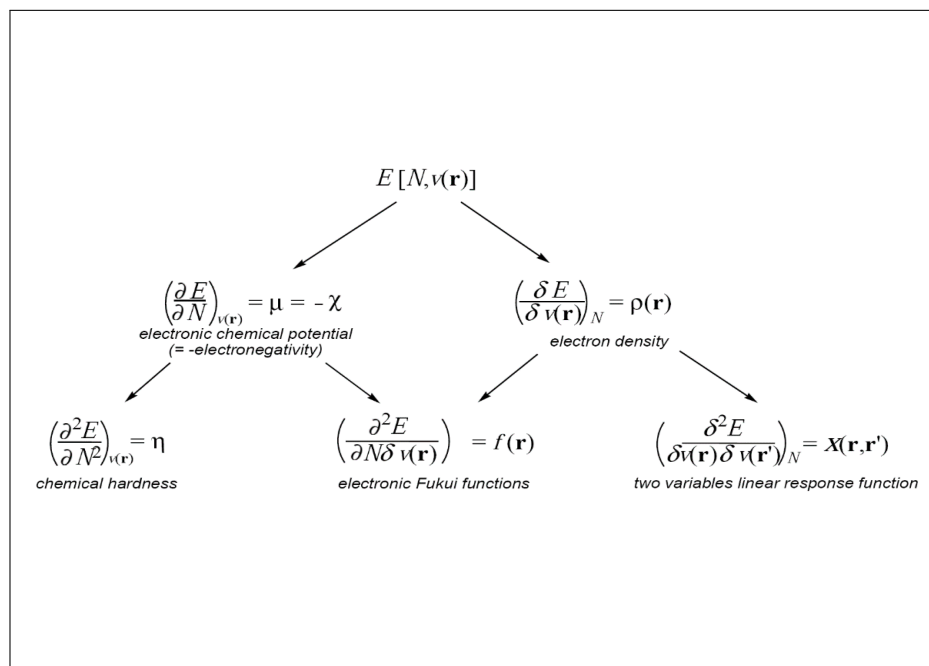


Figure 3.1: First and second derivatives of $E [N; \nu(r)]$ with respect to N and $\nu(r)$ (Figure extracted from reference [146].)

The calculations of the conceptual DFT reactivity indicators can be easily performed and even without complex computational demands.

3.6.1 Electronic Chemical Potential and Mulliken Electronegativity

The concept of electronic chemical potential describes the energy changes of a system when the number of electrons N changes at a fixed external potential

$\nu(r)$ (usually created by the nuclei) [147, 148]

$$\mu = \left(\frac{\partial E}{\partial N} \right)_{\nu(r)} \approx -\frac{(I + A)}{2} \quad (3.10)$$

Applying the finite approximation the chemical potential can be written as function of I and A, the ionization potential and the electron affinity of an atom or molecule, respectively. On the other hand, the chemical potential is identified as the negative of Mulliken electronegativity. This correlation gives an easy way to calculate the electronegativity of atoms and molecules [149, 150]:

$$\chi = -\mu = \left(\frac{\partial E}{\partial N} \right)_{\nu(r)} \approx \frac{(I + A)}{2} \quad (3.11)$$

3.6.2 HSAB Principle and Maximum Hardness Principle (MHP)

In 1963, Pearson [151, 152] proposed to classify Lewis acids and bases into hard and soft [151, 153] as four categories:

- Soft bases that group the highly polarizable nucleophiles
- Hard bases that bring together nucleophiles with little polarization
- Hard acids that group electrophiles of small size, very positively charged, and weakly polarizable
- Acids that group electrophiles of large size, weakly positively charged, and strongly polarizable

In an acid/base reaction "*Hard acids bind strongly to hard bases and soft acids bind strongly to soft bases*" which is the hard and soft acids and bases principle (HSAB) [153, 154, 155]. In addition to the HSAB principle, Pearson and Parr could found a relationship between energy and the variation of hardness [147] which leads them to announce the Maximum hardness principle (MHP):

"In any chemical process, the systems evolve toward an electronic state where its chemical hardness is a maximum. When combined with the minimum energy criterion that characterises stability, the MHP entails that, for a chemical process characterised by a negative (stabilising) energy change, the change in chemical hardness will always be positive." [146]

3.6.3 Chemical Hardness and Softness S

Besides, Parr in 1983 presented a quantitative expression for the hardness under the conceptual DFT framework. It can be defined as the changes of the electronic chemical potential of the system with respect to the electron number N at a fixed external potential $\nu(r)$ [147]:

$$\eta = \left(\frac{\partial \mu}{\partial N} \right)_{\nu(r)} = \left(\frac{\partial^2 E}{\partial N^2} \right)_{\nu(r)} \quad (3.12)$$

The chemical hardness can be seen as the resistance of a molecular system to electron transfer with the environment, whether gain or loss. It is therefore also a measure of the stability [156] of a molecule in the sense that the higher the resistance to charge transfer, the less this has the chance to occur. The chemical softness S was introduced too as a reactivity descriptor and defined

as the inverse of the chemical hardness:

$$s = \frac{1}{\eta} \quad (3.13)$$

3.6.4 Fukui Function $f(\mathbf{r})$

The frontier orbital theory (FOT) was developed in the 1950s by K. Fukui to explain the regioselectivity observed in reactions involving aromatic compounds. Since then it was extended first to all π systems, then to σ systems. The idea behind the FOT is that, during a reaction between a nucleophile and an electrophile, the charge transfer that takes place in the vicinity of the transition state involves primarily the electrons of the highest occupied molecular orbital (HOMO for Highest Occupied Molecular Orbital) of the nucleophile. The result must be that the electron density associated with these electrons, which he describes as frontier, must make it possible to explain reactivity and selectivity.

Later in 1984, Parr [157] introduced the $f(\mathbf{r})$ function named Fukui Function and defined as follow:

$$f(r) = \left(\frac{\partial \rho}{\partial N} \right)_{\nu(r)} \quad (3.14)$$

$f(\mathbf{r})$ is defined as the changes in electron density at a point \mathbf{r} with respect to the variation of the number of electrons N at a fixed external potential $\nu(r)$.

Depending on the type of attack experienced by the molecular system during a reaction, three Fukui functions are defined f^- for electrophilic, f^+ for nucleophilic and f^0 for radical reactions measured by the electron density

change upon removal, addition or as the average of the two respectively. Using the finite difference approximation, Fukui Functions could be related to the electron density of the molecule and can be written as [158]:

$$\begin{aligned} f^- &\approx \rho_N(r) - \rho_{N-1}(r) \\ f^+ &\approx \rho_{N+1}(r) - \rho_N(r) \\ f^0 &\approx \frac{\rho_{N+1}(r) - \rho_{N-1}(r)}{2} \approx \frac{f^+(r) + f^-(r)}{2} \end{aligned} \quad (3.15)$$

In other words, and as Parr postulated: *"the preferred direction for the approach of a reagent towards the other is the one for which the initial variation of the electronic chemical potential for a species is a maximum, and the one with the largest $f(r)$ is situated at the reaction site"* [157]. Mortier and Yang [158] proposed the atomic Fukui function:

$$\begin{aligned} f_k^- &= q_k(N) - q_k(N-1) \\ f_k^+ &= q_k(N+1) - q_k(N) \end{aligned} \quad (3.16)$$

$f_k^\pm(r)$ are the condensed Fukui function on atom k and $q_k(M)$ [$M = N-1, N, N+1$] are the Mulliken charge of the atom k in the molecule with $N-1, N$ and $N+1$ electrons.

3.6.5 Electrophilicity ω and Nucleophilicity Index

Electrophilicity index ω [159, 160] describes the energy stabilization of a molecule when it gains electron density ΔN from the surroundings: μ defines the tendency of an electrophile to gain more amount of density and η the resistance of the molecule to share electron density with the surroundings. So a good electrophile is characterized by a low hardness. Electrophilicity

can be written as:

$$\omega = \frac{\mu^2}{2\eta} \quad (3.17)$$

Many approaches have been proposed to study the nucleophilicity, may be the most distinguished one is the one proposed by Roy in 1998 [161] where he introduced the “*relative electrophilicity*” (S_k^+/s_k^-) and “*relative nucleophilicity*” (S_k^-/s_k^+) descriptors for the k^{th} atom of a molecule. The intention of these descriptors was to spot the reactive sites. More detailed information study is given in [161]. The electrophilic and nucleophilic local softness S_k^+ and S_k^- [161] are given by:

$$\begin{aligned} S_k^+ &= s.f_k^+ \\ S_k^- &= s.f_k^- \end{aligned} \quad (3.18)$$

Where S is the chemical softness and f^+ and f^- are the electrophilic and nucleophilic Fukui functions.

Chattaraj [162, 163] later in 2003 suggested ω are the electrophilic and nucleophilic Fukui functions. as a generalized philicity index to quantify the philicity properties in reactivity and regioselectivity studies. It is defined as a combination between the electrophilicity as proposed by Parr and the Fukui Function. Thus, three philicity index:

$$\begin{aligned} \omega_k^+ &= \omega.f_k^+ \\ \omega_k^- &= \omega.f_k^- \\ \omega_k^0 &= \omega.f_k^0 \end{aligned} \quad (3.19)$$

According to Chattaraj: ”A greater ω_k^+ value corresponds to a better

capability of accepting electron density, whereas a smaller value of ω_k^- of a system makes it a better donor of electron density.”

Chapter 4

REACTIVITY OF BODIPY DERIVATIVES AND CLICK REACTION

4.1 BODIPY Derivatives

4.1.1 Introduction

As described in the second part of the introduction of this thesis, BODIPY is one of the most convenient dyes for imaging and biological labelling because of its outstanding properties mainly being a neutral and charged molecule. However, because of its lack of solubility, the synthesis of new BODIPY-based dyes is desirable.

In the fourth paper of this thesis a joint theoretical/experimental study of novel 8-ethynyl-BODIPY derivatives is presented. For this, our colleagues used an alkyne-azide click chemistry to synthesize the needed molecule. Their results suggested that substituting an ethynyl group in the meso position of BODIPY dyes were catalyzed by Cu^+ and reacted with benzyl azide under physiological conditions.

The resulting blue shift during the formation of the triazole system allows us to further interpret the reaction and the optical changes: first, the protein used was conjugated successfully with the BODIPY fluorescent model; and second, we can distinguish easily between the free and coupled dyes.

Our part of the collaboration was to shed further light on those results with computational calculations adopting a TDDFT method. The presence of the blue shift was confirmed and justified by the fact that triazole was perpendicular orientated compared to the chromophore.

Fundamental conclusion of our previous work is that “derivative 3” as called

in paper four of 8-Ethynyl-BODIPY (where positions 6 and 2 were occupied by a Hydrogen and position 7 and 1 occupied by a methyl group) is an approved new fluorescent that can allow site-specific conjugation with and without copper catalysis.

Consequently, we decided to go for the 8-Ethynyl-BODIPY derivatives to have further reactivity information and then choose which of them will be suitable for the click reaction. TDDFT calculations will be a future step for this work to check the consistency of the blue shift seen before.

In this part, we will concentrate on analyzing the global and local reactivity of three 8-Ethynyl-BODIPY derivatives by adding different functional groups at different positions of the BODIPY core. The global reactivity descriptors as explained in details in the first and third paper of this thesis, can give us insights into the stability of the molecules and which of them will be suitable for the click reaction. Moreover, the local reactivity descriptors will allow us to identify at which position in the molecule exactly the reaction will take place.

Umezawa et al.[164, 165] proved that by incorporating different functional groups or substitutions to the BODIPY core the absorption and emission characteristics of the molecule will be modified and their fluorescence can be pushed to near infra-red region.

Equivalently, adding different donor/acceptor groups to the BODIPY core

can lead to changes in the energy gap between the HOMO and LUMO orbitals [166, 167, 168].

With the help of the information from the reactivity studies, we will proceed in our work with studying the click reaction by using the most stable derivatives and benzyl azide. As mentioned earlier in the introduction part, click chemistry is selective and stereospecific which means that usually in the end we will find exactly one product. However, the experimental group could identify both major and minor products. Our task was to apply theoretical tools to interpret those experimental measures by using the transition state theory as tool for our work.

Aiming to provide the most accurate results, the performed calculations were conducted not only in gas but also in aqueous phase. All the results were compared to derivative 3 of 8- Ethynyl-BODIPY synthesized in our recent work.

4.1.2 Methods

All the theory used during this work is explained in details in section one (first and third paper) of this thesis. Therefore, here we will focus on the outcome of the calculations.

4.1.3 Experimental details

The reaction investigated in this work is the copper-free [3+2]-azide-alkyne cycloaddition between an azide (benzyl azide) and alkyne-modified BOD-

IPY derivatives. Based on the reaction conditions (without toxic catalyst) and the optical properties (sharp emission bands and high quantum yield) this methodology is interesting for the site-specific labeling of various biomolecules, such as proteins or glycans.

The two main remaining questions are:

- Influence of substituents (electronic and steric effects) at the triple bond of alkyne BODIPY on the reactivity in the copper-free [3+2]-azide-alkyne cycloaddition
- Substitution pattern of the product (which regioisomer is formed?)

Kinetics of the product formation was investigated by using UV/Vis measurements. More specifically, the starting material undergoes a decreasing band of its absorption band at 540 nm (figure 4.1).

Rate constants of the model reaction of substituted BODIPYs with benzyl azide were estimated using the method and concentrations describe in our paper (fourth paper in this thesis [169]). Further investigations indicate a significant solvent dependence on the rate constants of the cycloaddition reaction. By using the method mentioned above the unsubstituted BODIPY reacts so fast ($3.56 \cdot 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ (H₂O/EtOH)) that the reaction progress could not be measured using our UV/Vis spectrometer. When TMS is added as functional group rate constant is $28.4 \cdot 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ (THF) and with phenyl the rate constant is $5.09 \cdot 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ (THF). So the reactivity (experimental) is H >TMS> Ph. Furthermore, the reaction of the TMS-modified BODIPY with various azides was performed to isolate the reaction products. In all cases only one single regioisomer could be isolated, though

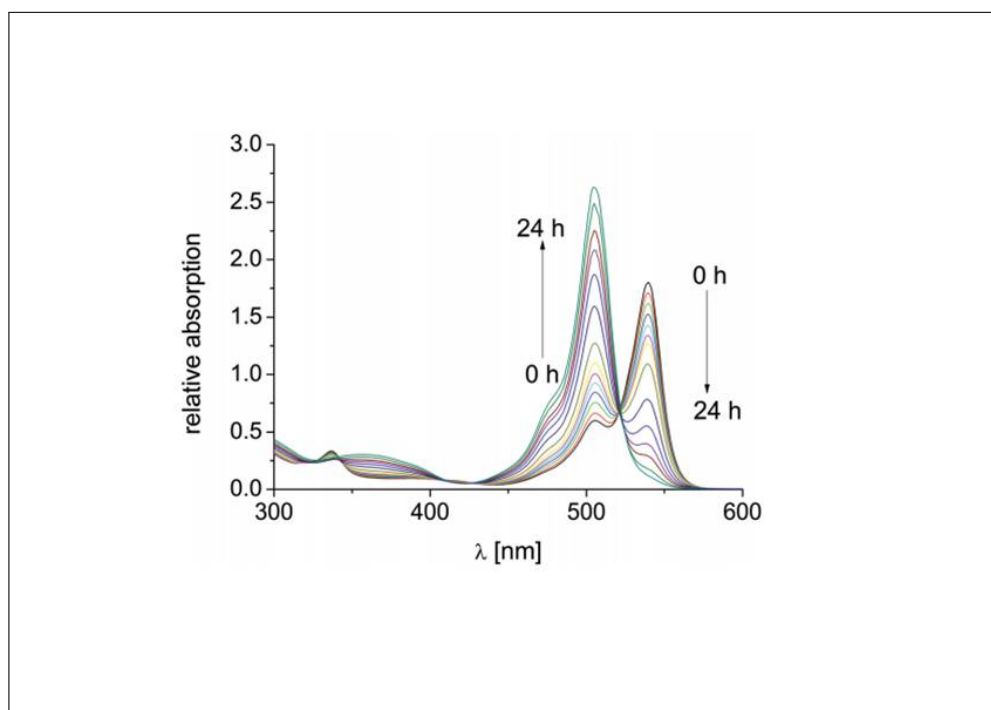


Figure 4.1: UV/vis spectra during the reaction of BODIPY 3 ($25 \mu\text{M}$) and benzyl azide (15mM) with BODIPY 5 under copper-free reaction conditions in water/ethanol (1:2 v/v) [169].

the available experimental data do not allow one to conclude which one is finally presented in the medium. We will perform special NMR experiments to investigate that.

4.1.4 Reactivity

Structure

It is relevant to mention that all calculations were performed with the Gaussian09 suite of programs [170] and GaussView 05 was used to visualize proper-

ties like HOMO/LUMO orbitals, densities, electrostatic potential and Fukui function. B3LYP [171, 172] method was used to optimize the structures with the standard 6-311G* basis set within the framework of density functional theory DFT [173, 174]. Moreover, any minimum in the potential energy surface is characterized by the absence of normal modes with imaginary frequencies, after optimization, confirming that all structures correspond to total energy minimum.

As a first step, we optimized the structures shown in figure 4.2 in both gas and aqueous phases. We studied neutral, cation and anion molecules for all species. We found that IV and I structures are planer, II is planer with the phenyl group perpendicular to the core due to the presence of methyl groups in both positions 1 and 7 (see figure 2.2, chapter 2 for numbering), and III is planer with the plane of the 3 methyl groups linked to Si perpendicular to the core's plan.

In Table 4.1, we present bond lengths using the atom numbering of figure 4.2, none of the substitutions dramatically affects the structure. All the results are in almost the same and it is necessary to indicate too that bond length 5-6 and 6-7 were identically the same this is why they were not shown in the table.

Since there is no previous works on these derivatives, our results can be compared with the work by Misra [175] only, where the authors investigated Aryl-substituted entities with the same structure as our compound number II but missing the ethynyl group in the meso position: the phenyl group is directly connected to the meso-carbon of the BODIPY core. The rotation of

Table 4.1: Calculated bond lengths (\AA) for each structure of molecules I to IV as obtained using DFT/B3LYP under 6-311G* basis set in gas phase.

	1 – 2	2 – 3	3 – 4	4 – 5
I	1.422	1.413	1.394	1.568
II	1.423	1.411	1.394	1.559
III	1.423	1.412	1.395	1.559
IV	1.423	1.408	1.394	1.556

the phenyl group is predicted with and without the presence of the ethynyl group. The phenyl group has to adjust its position with the presence of the density clouds of the two methyl groups in the positions 1 and 7.

As discussed by Tang and coworkers [176], linking an unsubstituted phenyl group to BODIPY core decreases the fluorescence quantum yield of the latter as compared to BODIPY alone. They added that attaching two methyl groups limits the motion of the phenyl ring and as a result leads to an increase in the fluorescence quantum yield of the molecule.

Global reactivity descriptors

According to our previous work [177, 178] polarizability can be used as a global descriptor for chemical reactivity. Therefore, we shall also study it here. A system with a low polarizability is considered as more stable.

Our calculated polarizabilities are presented in Table 4.2. As one can see, our calculations in gas phase hint that compound IV is the most stable structure which corresponds to the unsubstituted 8 – ethynyl- BODIPY by 213.47 \AA^3 followed by compound number III by 278.56 \AA^3 . Compounds I and II are the

least stable ones with values of 293.69 and 293.69 Å³ respectively. The same result is found, in solvent, except that compound II becomes more reactive than compound I.

The last row of table 4.2 shows the HOMO/LUMO gap of each molecule. According to the HSAB principle, when using Koopmanns' theorem in estimating ionization potential and electron affinity, a molecule with the largest gap value is more stable and can be considered as a hard molecule [177, 178]. Then, the values of the gap show that compounds II, III and IV have comparable values, i.e 2.72 (2.77), 2.76 (2.80) and 2.74 (2.79) eV respectively in gas phase (solvent phase) while compound I has a value of 3.65 (2.90) eV. All these accumulated findings suggest that is the most stable molecule among the studied systems.

The same remark can be applied to the results for the hardness where we obtain very similar values in both gas and solvent indicating that compounds II and III have very similar reactivity with values of 2.79 eV in gas and 1.27 and 1.25 eV respectively in solvent. Yet, these values suggest that these systems are more reactive than compound I and IV.

Moreover, as seen in Table 4.2, compound II is the worst electron acceptor with the smallest χ^+ of 5.40 eV in gas and 4.84 eV in solvent. On the other hand, it is the best electron donor with the smallest χ^- with a value of 2.62 eV and 3.57 eV in gas and solvent respectively which suggests that it is more reactive towards nucleophilic attack.

However, compound I may be considered as the best electron donor with values for χ^+ of 6.45 eV and 5.61 eV in gas and solvent which may be ascribed to the SO₃H groups at the position 1 and 7 of the BODIPY core. Compound

III and IV have similar properties as compound II.

In addition, when comparing electrophilicity of the investigated molecules, we can find that compound II is the less nucleophilic system with a value of 2.88 eV (6.97 eV) in gas phase (solvent phase).

Nonetheless, compound I is the most nucleophile one with a value of 4.55 eV (9.33 eV) in gas phase (solvent phase).

Local reactivity descriptors

When we change the functional group linked to the core of the BODIPY the charge distribution changes (see Table 4.3) and we will have different densities ascribed to the individual atoms, e.g., the values of f^+ for the Bromine atom is -0.026 and for the two Fluorine atoms it is -0.016 in the BODIPY (compound IV). Then, for compound III where we have additional SiMe_3 groups those two values change to -0.024 and -0.015 respectively, a change that is very similar to what we observe when we change the functional group attached to the phenyl ring. However, when adding the SO_3H group in the horizontal plan of the molecule (compound I) the values change to -0.02 and -0.014 in gas phase.

The most remarkable result is that, independently from functional groups added to the core and at which position, our results demonstrate that the distribution on the triple bond carbon number 13 (figure 4.4) is the highest for the three component of Fukui function. Indeed, it is the only positive values among all distributions on the other atoms. In contrast, when we added the solvent, the result changes for compound II as carbon 35 of the phenyl ring (see figure 4.5) has the highest values of Fukui function. Moreover, for

Table 4.2: Various global reactivity descriptors of the investigated molecules in eV (polarizability is in \AA^3) using DFT at the B3LYP/6-311G* level. (In each case, the upper line shows results for the gas phase and the lower line shows results for the aqueous phase).

	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>
χ	5.05	4.01	4.05	4.09
	4.95	4.21	4.22	4.24
χ^+	6.45	5.40	5.44	5.53
	5.61	4.84	4.85	4.88
χ^-	3.65	2.62	2.66	2.65
	4.29	3.57	3.59	3.60
ω	4.55	2.88	2.94	2.91
	9.33	6.97	7.10	7.07
η	2.80	2.79	2.79	2.88
	1.31	1.27	1.25	1.27
α	281.38	293.69 (^a 272.06)	278.56	213.47
	390.64	410.03	392.02	307.36
Gap	3.65	2.76	2.72	2.47
	2.90	2.80	2.77	2.79

^a: Taken from Reference [175]

compound III, the distribution becomes negative on the whole molecules but still the carbon 13 of the triple bond has the highest values (numbering is showed in figure 4.2).

When we analyze the plots of the Fukui function for radical attack, we can see that it is only compound II that actually has both positive and negative

Table 4.3: Calculated condensed Fukui function for the C≡C in the compounds investigated using DFT at the B3LYP/6-311G* level. (In each case, the upper line shows results for the gas phase and the lower line shows results for the aqueous phase).

	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>
f^+	0.018	0.015	0.004	0.018
	0.007	-0.002	-0.007	0.005
f^-	0.026	0.025	0.022	0.028
	0.011	0.005	0.005	0.010
f^0	0.022	0.020	0.013	0.023
	0.009	0.001	-0.001	0.008

charge distribution around the triple bond which will be helpful while put with benzyl azide in reaction. Since we need our dye to interact with benzyl azide to perform the click reaction, we conclude that compounds II and III are the most suitable derivatives for the aimed work as it has the characteristic of electron donor and acceptor at the same time, not so much electrophilic neither very nucleophilic. This is confirmed by the Fukui function calculations results, the gives insights into the ability of those two molecules to accept a radical attack in the ethynyl triple bond position.

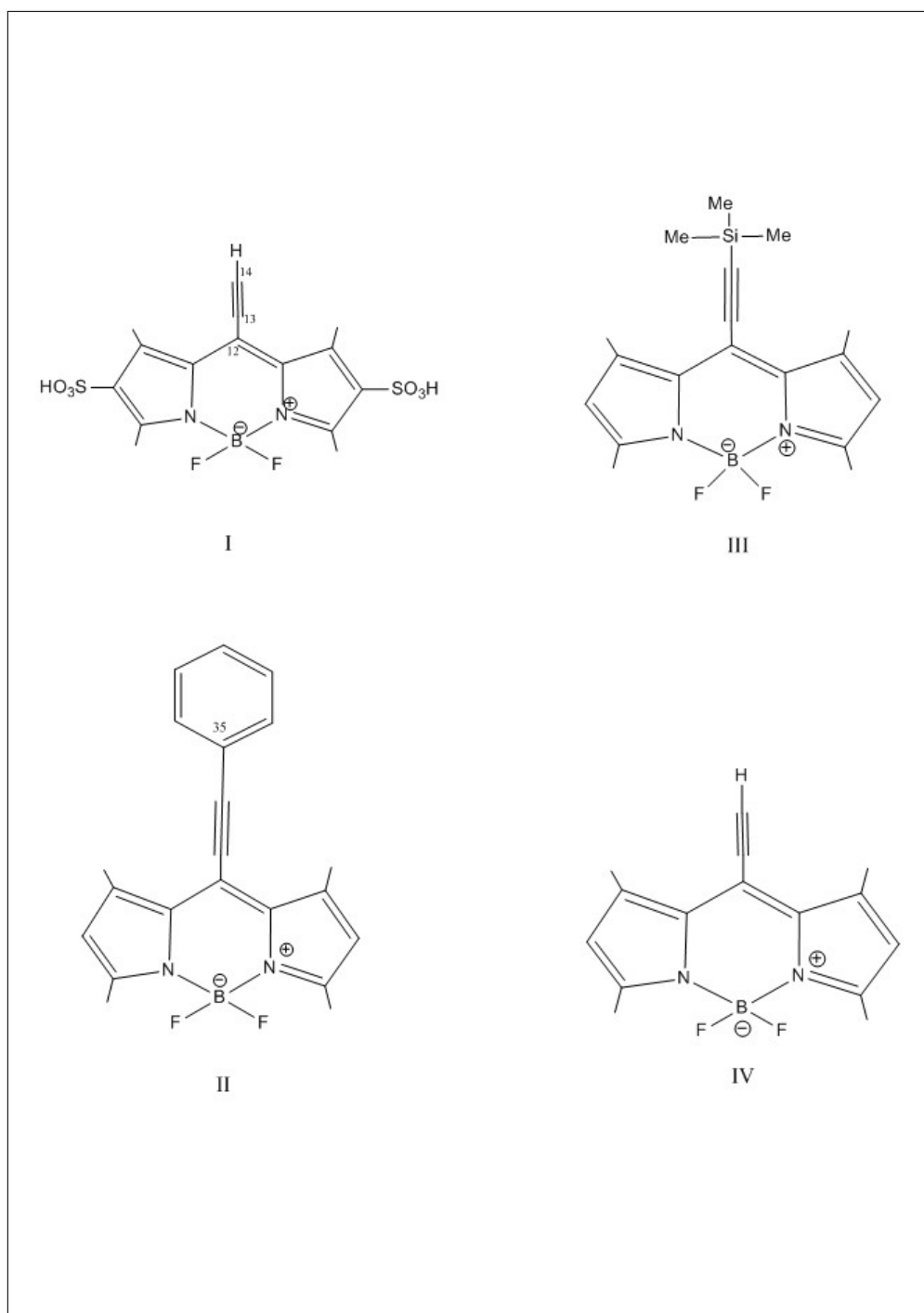


Figure 4.2: Schematic representation of the 8-Ethynyl-BODIPY derivatives studied.

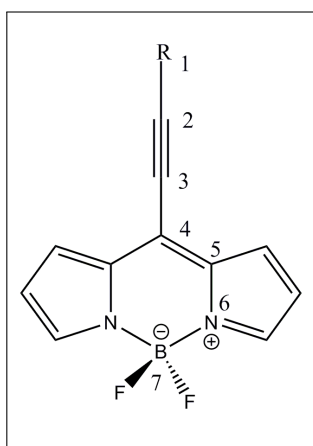


Figure 4.3: Schematic representation of the BODIPY dye under investigation (Numbering is used only as reference for this part of the work).

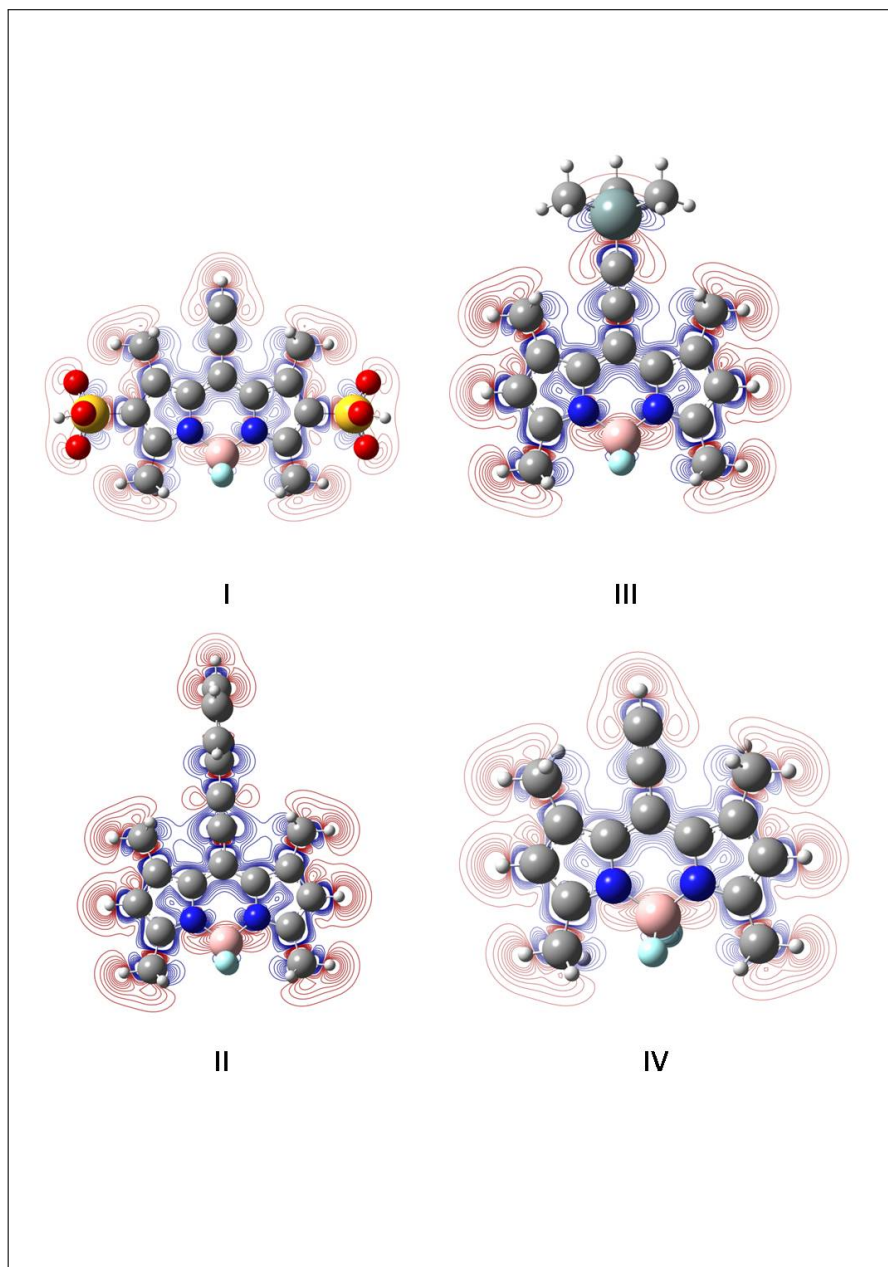


Figure 4.4: Fukui function for radical attack f^0 for the four compounds in gas phase Red color presents positive density and blue negative one.

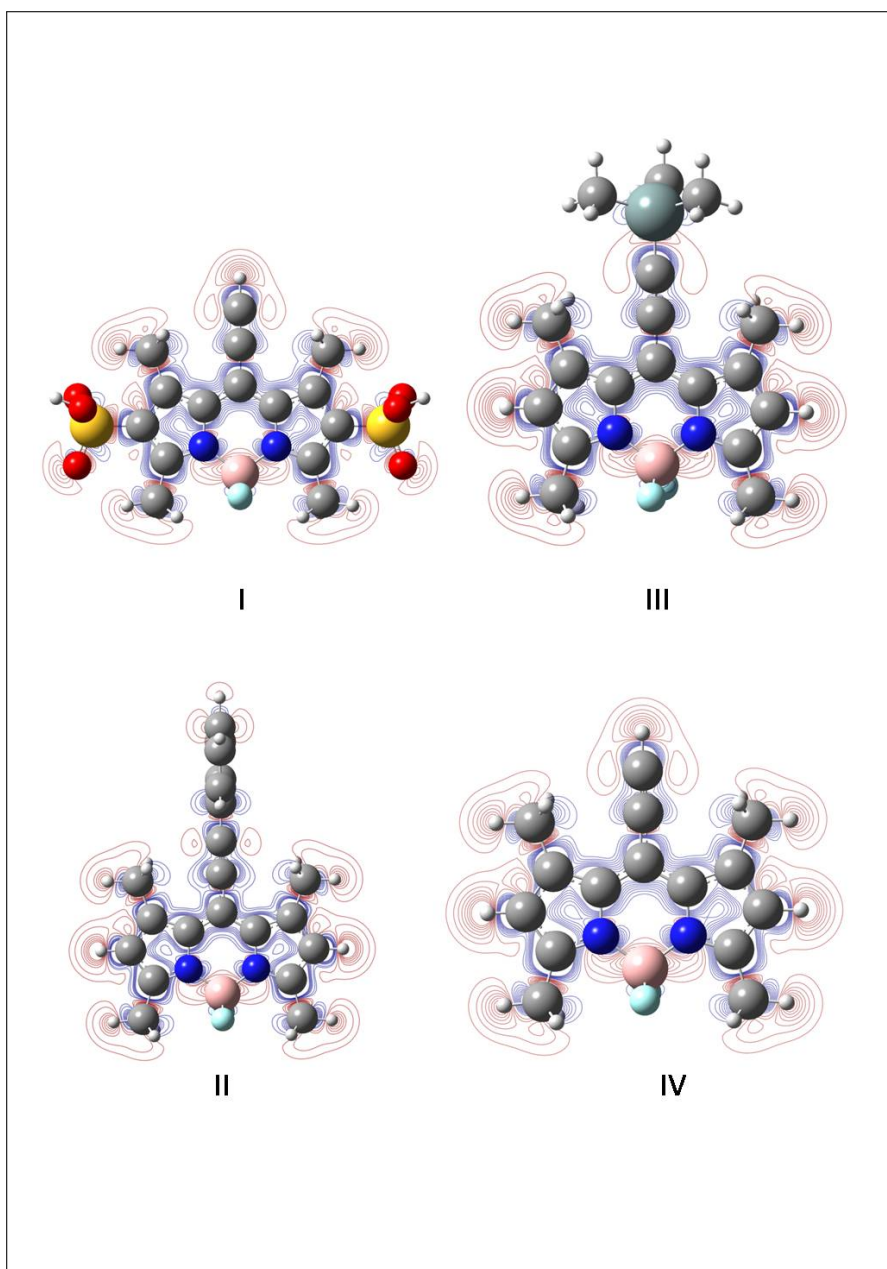


Figure 4.5: Fukui function for radical attack f^0 for the four compounds in aqueous phase Red color presents positive density and blue negative one

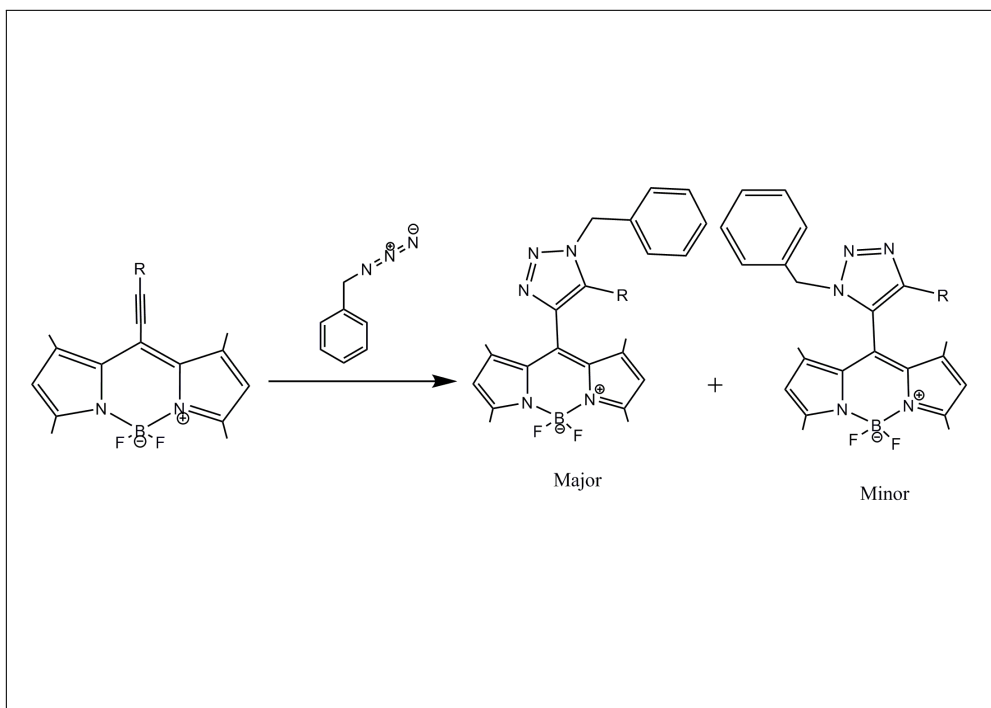


Figure 4.6: Representation of the click reaction (Reaction A) between BODIPY derivatives and benzyl azide).

4.2 Click reaction

The aim of this study was to understand the click reaction with the benzyl azide as presented in figure 4.6. In this part of the work, we will concentrate on compound II and III as the most reactive molecules which can react easily with benzyl azide.

Figure 4.7 and 4.8 present the two isomers energetically favored of the final products of compound II and III respectively. The formed triazole ring in that compound partially restricts the rotation of the molecule thanks to the interaction of the methyl groups of position 1 and 7 with the heterocyclic ring.

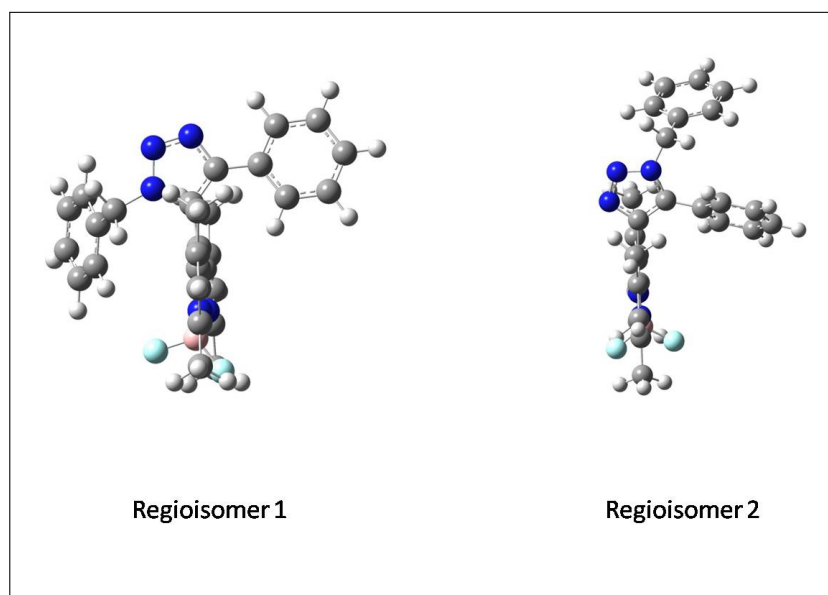


Figure 4.7: Regioisomers of compound II.

Even though both isomers are thermodynamically stable, one of them might be kinetically more favorable than the other, finally leading to dissimilar yield at the end of the reaction.

Two scenarios are possible depending on the particular conditions: first if the reaction is made in gas or solvent; and second, from which side of the triple bond exactly the benzyl azide is actually attacking. For the latter, the ring of the benzyl azide and the R of the BODIPY derivative are in the same plane but both are perpendicular to the core plane or the ring and the R are in two different planes and again perpendicular to the core's plane.

We determined the thermodynamic properties of the two reactions for each compound and compare them in gas and solvent. In here we will call Reaction A-1 the reaction leading the regioisomer 1 and Reaction A-2 the reaction leading to regioisomer 2.

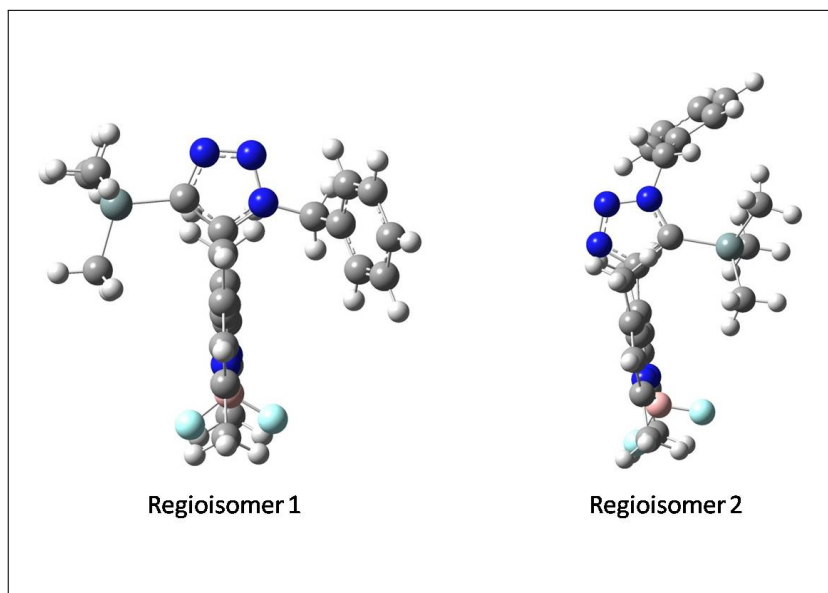


Figure 4.8: Regioisomers of compound III.

4.2.1 Reaction with compound II

The transition state (TS_1) of this reaction matches the formation of the triazole ring between the compound II and the Benzyl azide to form the regioisomer 1. The main structural changes of the molecule in transition state compared to regioisomer 1 are the increase of bond length N_1-C_1 from 1.37 Å (in the regioisomer 1) to 2.1 Å. Moreover, the angle $\angle N_1-N_2-N_3$ changes from 173.33° in the Benzyl azide to 138.76° in the transition state.

Simultaneously, angle $\angle C_2-C_1-C_R$ (C_R presents the carbon of the phenyl ring) and $\angle C_1-C_2-C_3$ change from 180° in compound II to 148.85° and 161.83° respectively.

However, the imaginary frequency of the transition state 2 (TS_2) characterizing the formation of the regioisomer 2 is -387.37 cm^{-1} . Structural changes

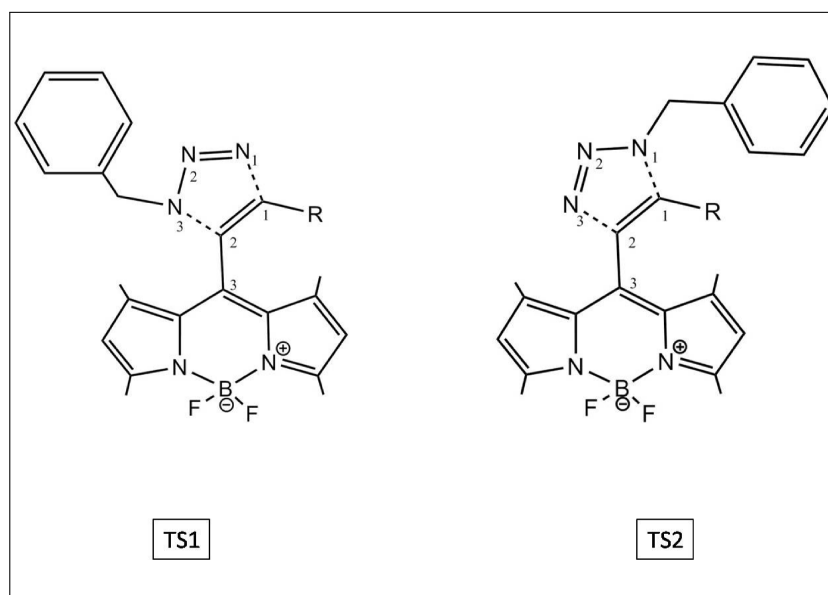


Figure 4.9: Representation of both structures of transition states of regioisomers 1 and 2. R changes for compound II and compound III (Numbering is only used as reference for this work).

are similar to what we found above: the bond length of the triazole ring N_1-C_1 increases from 1.37 Å to 2.12 Å while the bond N_3-C_3 increases from 1.37 Å to 2.25 Å. The angle $\angle N_1-N_2-N_2$ changes from 173.33° in the Benzylazide to 140.21° in the transition state, angle $\angle C_2-C_1-C_R$ and $\angle C_2-C_1-C_3$ change from 180° in compound II to 152.54° and 156.24° respectively. Table 4.4 presents the thermodynamic results for compound II. The activation energies are not significantly different. In gas phase regioisomer 1 has lower activation energy, however in solvent regioisomer 2 has lower barrier. Obviously, this low activation energy differences, 1.05 and 1.09 kcal/mol in gas and solvent respectively, decrease the regioselectivity choice.

Table 4.4: Energetic properties (kcal/mol) of the Reaction A-II at B3LYP/6-31G* level in aqueous and gas phase (blue color presents reaction A-1 and black color presents panel reaction A-2).

	<i>Gasphase</i>	<i>solvent</i>
Activation energy	21.10	23.06
	22.15	21.97
Reaction energy	-63.37	-61.44
	-60.24	-60.52

4.2.2 Reaction with compound III

As before, the transition state always leads to the formation of the triazole ring and several structural changes can follow.

Bond length N_1-C_1 changes from 1.38 Å in regioisomere 1(2) to 2.06 Å in TS₁ (2.23Å in TS₂), N_3-C_2 increases from 1.36 Å to 2.24Å in TS₁ (2.12) Å in TS₂). Besides, $\angle N_{27}-N_{24}-N_{25}$ decreases from 173.33° to 136.83° in TS₁ and 137.44° in TS₂ and finally the angles $\angle C_2-C_1-Si$ and $\angle C_1-C_2-C_3$ decrease from 180° to 152.91° in TS₁ (155.17° in TS₂) and 160.47° in TS₁ (151.73° in TS₂) respectively.

Schubert et al. [179] used DFT calculations at B3LYP/6-311++G (2d,2p) level to study the transition state in the click reaction of 1,5-substituted 4-(trimethylsilyl) triazole. In their TS₁, the angle $\angle N_{27}-N_{24}-N_{25}$ is found to be 134.7° while in TS₂ it is 135.4° starting from a relaxed azide angle of 173.3°. Same results are found by Houk's investigations [180], where the angle $\angle N_{27}-N_{24}-N_{25}$ angle about 135°.

Table 4.5: Energetic properties (kcal/mol) of the Reaction A-III at B3LYP/6-31G* level in aqueous and gas phase (blue color presents reaction A-1 and black color presents panel reaction A-2)

	<i>Gasphase</i>	<i>solvent</i>
Activation energy	19.39	21.41
	23.93	23.88
Reaction energy	-54.59	-54.10
	-55.22	-55.03

The activation and reaction energy values are collected in Table 4.5. The activation energy difference between the two isomers is 4.54 kcal/mol and 2.47 kcal/mol in gas and solvent phase, respectively. The regioisomer 1 has lower activation energy in both cases which make it the regioisomer likely to be formed. According to Schubert et al., the activation energy of the regioisomer 1 was found to be 24.02 kcal/mol and 29.92 kcal/mol for regioisomer 2 [179]. This difference in the activation energies is explained by the fact that in our calculations the SiMe₃ group is added to the BODIPY core but in here the SiMe₃ group is linked to smaller system.

4.3 Conclusion

As a conclusion for this working report we could confirm that using different substitutions attached to the reference molecule 8-Ethynyl-BODIPY can have remarkable consequences on the reactivity of derivatives studied in both gas and solvent phases.

We could identify by using global and local reactivity descriptors that two compounds (with TMS and Ph substitutions) are suitable to interact with benzyl azide to perform the click reaction as they can behave as electron donor and acceptor at the same time.

Moreover, our results of species' structures of the studied click reaction are in agreement with previous works. However, contrary to what the experimentalist found, we could determine a phase (gas or solution) dependency of the resulting products.

Bibliography

- [1] R.D. Hochkiss. The quantitative separation of purine, pyrimidine and nucleoside by paper chromatography. *J Biol Chem.*, pages 315–332, 1948.
- [2] M.W. Nirenberg and J.H. Matthaei. The dependence of cell-free protein synthesis in e. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci USA*, 47:1588–1602, 1969.
- [3] U. Kuhnlein, S. Linn, and W. Arber. Host specificity of dna produced by escherichia coli. xi. in vitro modification of phage fd replicative form. *Proc Natl Acad Sci USA*, 63:556–562, 2016.
- [4] P.H. Roy and A. Weissbach. A methylase from hela cell nuclei. *Nucleic Acids Res*, 2:1669–1684, 1975.
- [5] R. Stein, A. Razin, and H. Cedar. In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse l cells. *Proc Natl Acad Sci USA*, 79:3418–3422, 1982.
- [6] A. Bird. Dna methylation patterns and epigenetic memory. *Genes Dev*, 16:6–21, 2002.

- [7] M. Gardiner-Garden and M. Frommer. CpG islands in vertebrate genomes. *J Mol Biol*, 196:261–282, 1987.
- [8] C. Burge, A.M. Campbell, and S. Karlin. Over- and under-representation of short oligonucleotides in dna sequences. *Proc Natl Acad Sci USA*, 89:1358–1362, 1992.
- [9] K.J. Fryxell and W.J. Moon. CpG mutation rates in the human genome are highly dependent on local gc content. *Mol Biol Evol*, 22:650–658, 2005.
- [10] A. Gutierrez and R.J. Sommer. Evolution of dnmt-2 and mbd-2-like genes in the free-living nematodes *pristionchus pacificus*, *caenorhabditis elegans* and *caenorhabditis briggsae*. *Nucleic Acids Res*, 32:6388–6396, 2004.
- [11] J. Hubacek. Biological function of dna methylation. *Folia Microbiol (Praha)*, 37:323–329, 1992.
- [12] K.D. Robertson, S. Ait-Si-Ali, T. Yokochi, P.A. Wade, P.L. Jones, and A.P. Wolffe. Dnmt1 forms a complex with rb, e2f1 and hdac1 and represses transcription from e2f-responsive promoters. *Nat Genet*, 25:338–342, 2000.
- [13] J. Jerbi, M. Springborg, H. De Haan, and J.P. Ceron Carrasco. S-adenosyl-l-methionine analogs as enhanced methyl donor: Towards novel epigenetic regulators. *Chem. Phys. Lett*, 690:74–81, 2017.

-
- [14] T. Schmidt, T. Schwede, and M. Meuwly. Computational analysis of methyl transfer reactions in dengue virus methyltransferase. *J. Phy. Chem B*, 118:5882–5890, 2014.
- [15] R. Zangi, A. Arrieta, and F.P. Cossio. Mechanism of dna methylation: The double role of dna as a substrate and as a cofactor. *J. Mol. Biol.*, 400:632–644, 2014.
- [16] J.M. Mato, J. F. Corrales, C.S. Lu, and A.M Avila. S-adenosylmethionine: a control switch that regulates liver function. *FASEB Journal*, 16:15–26, 2002.
- [17] P.L. Jones, G.J. Veenstra, P.A. Wade, D. Vermaak, S.U. Kass, N. Landsberger, J. Strouboulis, and A.P Wolffe. Methylated dna and mecp2 recruit histone deacetylase to repress transcription. *FASEB Journal*, 19:187–191, 1998.
- [18] Z.M. Svedruzic and N.O Reich. The mechanism of target base attack in dna cytosine carbon 5 methylation. *Biochem*, 43:11460–11473, 2004.
- [19] J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, and J. Walter. Active demethylation of the paternal genome in the mouse zygote. *Curr Biol*, 10:475–478, 2000.
- [20] W. Mayer, A. Niveleau, J. Walter, R. Fundele, and T. Haff. Embryogenesis: Demethylation of the zygotic paternal genome. *Nature*, 403:501–502, 2000.

-
- [21] Z. Paroush, I. Keshet, and J. Yisraeli. Dynamics of demethylation and activation of the alpha-actin gene in myoblasts. *Cell*, 63:1229–1237, 1990.
- [22] F. Zhang, J.H. Pomerantz, G. Sen, A.T. Palermo, and H.M Blau. Active tissue-specific dna demethylation conferred by somatic cell nuclei in stable heterokaryons. *Proceedings of the National Academy of Sciences USA*, 104:4395–4400, 2007.
- [23] J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, AA. Margolin, S. Kim, CJ. Wilson, J. Lehar, GV. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, MF. Berger, JE. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jane-Valbuena, FA. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, IH. Engels, J. Cheng, GK. Yu, J. Yu, P Jr. Aspesi, M. de Silva, K. Jagtap, MD. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, RC. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, JP. Mesirov, SB. Gabriel, G. Getz, K. Ardlie, V. Chan, VE. Myer, BL. Weber, J. Porter, M. Warmuth, P. Finan, JL. Harris, M. Meyerson, TR. Golub, MP. Morrissey, WR. Sellers, R. Schlegel, and LA. Garraway. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Proceedings of the National Academy of Sciences USA*, 48:603–607, 2012.
- [24] K. Rai, I.J. Huggins, S.R. James, A.R. Karpf, D.A. Jones, and B.R. Cairns. Dna demethylation in zebrafish involves the coupling of a deam-

- inase, a glycosylase, and gadd 45. *Proceedings of the National Academy of Sciences USA*, 135:1201–1212, 2008.
- [25] N. Bhutani, J.J. Brady, M. Damian, A. Sacco, S.Y. Corbel, and H.M. Blau. Reprogramming towards pluripotency requires aid-dependent dna demethylation,. *Nature*, 463:1042–1047, 2010.
- [26] M. Tahiliani, K.P. Koh, Y. Shen, W.A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L.M. Iyer, D.R. Liu, L. Aravind, and A. Rao. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian dna by mll partner tet1. *Science*, 324:930–935, 2009.
- [27] G.P. Pfeifer, S. Kadam, and S.G. Jin. 5-hydroxymethylcytosine and its potential roles in development and cancer. *epigenetics & Chromatin*, 6:10–19, 2013.
- [28] J. Guo, Y. Su, C. Zhong, G.L. Ming, and H. Song. Hydroxylation of 5-methylcytosine by tet1 promotes active dna demethylation in the adult brain. *Cell*, 145:423–434, 2011.
- [29] S. Ito, L. Shen, Q. Dai, S.C. Wu, L.B. Collins, J.A. Swenberg, C. He, and Y. Zhang. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, 333:1300–1303, 2011.
- [30] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, K. Wessels, D. Soprano, L.K. Abramowitz, M.S. Bartolomei, F. Rainbow, M.R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A.J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Laure, and A. Bellacosa. Thymine dna glycosylase is

- essential for active dna demethylation by linked deamination-base excision repair. *Cell*, 146:67–79, 2011.
- [31] Y.F. He, B.Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, and L. Li. Tet-mediated formation of 5-carboxylcytosine and its excision by tdg in mammalian dna. *Science*, 333:1303–1307, 2011.
- [32] C.S. Nabel and R.M. Manning, S.A. and Kohli. the curious chemical biology of cytosine: Deamination, methylation, and oxidation as modulators of genomic potential. *Chem. Biol.*, 7:20–30, 2011.
- [33] P.M. Das and R. Singal. Dna methylation and cancer. *J. Clin. Oncol.*, 22:4632–4642, 2004.
- [34] Y.H. Jiang, J. Bressler, and A.L. Beaudet. Epigenetics and human disease. *Annu Rev Genomics Hum Genet.*, 5:479–510, 2004.
- [35] S. Houshdaran, S. Hawley, C. Palmer, M. Campan, M. Olsen, A.P. Ventura, B.S. Knudsen, C.W. Drescher, N.D. Urban, P.O. Brown, and P.W. Laird. Dna methylation profiles of ovarian epithelial carcinoma tumors and cell lines. *PLoS One*, 5:e9359, 2010.
- [36] P.A. Jones and S.B. Baylin. The fundamental role of epigenetic events in cancer. *Nat Rev Genet*, 3:415–428, 2002.
- [37] D. Rodenhiser and M. Mann. Epigenetics and human disease: translating basic biology into clinical applications. *CMAJ*, 174:341–348, 2006.

- [38] S. Xie, Z. Wang, M. Okano, M. Nogami, Y. Li, W.W. He, K. Okumura, and E. Li. cloning, expression and chromosome locations of the human dnmt3 gene family. *Gene*, 236:87–95, 1999.
- [39] J.P. Issa, P.M. Vertino, J. Wu, S. Sazawal, P. Celano, B.D. Nelkin, S.R. Hamilton, and S.B. Baylin. Increased cytosine dnamethyltransferase activity during colon cancer progression. *J.Natl. Cancer Inst.*, 85:1235–1240, 1993.
- [40] T. Nagao, K. Higashitsuji, H. ana Nonoguchi, T. Sakurai, S. Dawson, R.J. Mayer, K. Itoh, and J. Fujita. Mage-a4 interacts with the liver oncoprotein gankyrin and suppresses its tumorigenic activity. *J. Biol. Chem.*, 278:10668–10674, 2003.
- [41] K.D. Robertson, E. Uzvolgyi, G. Liang, C. Talmadge, J. Sumegi, F.A. Gonzales, and P.A. Jones. The human dna methyltransferases (dnmts) 1,3a and 3b: coordinate mrna expression in normal tissues and overexpression in tumors. *Nucleic Acids Res*, 27:2291–2298, 1999.
- [42] I. Girault, S. Tozlu, R. Lidereau, and I. Bieche. Expression analysis of dna methyltransferases 1, 3a, and 3b in sporadic breast carcinomas. *Clin Cancer Res.*, 9:4415–4422, 2003.
- [43] A.V. Bakin and T. Curran. Role of dna 5-methylcytosine transferase in cell transformation by fos. *SCIENCE*, 283:387–390, 1999.
- [44] N. Beaulieu, S. Morin, I.C. Chute, M.F. Robert, H. Nguyen, and A.R. MacLeod. An essential role for dna methyltransferase dnmt3b in cancer cell survival. *J. Biol. Chem.*, 277:28176–28181, 2002.

- [45] K. Soejima, W. Fang, and B.J. Rollins. Dna methyltransferase 3b contributes to oncogenic transformation induced by sv40t antigen and activated ras. *Ras. Oncogene*, 22:4723–47233, 2003.
- [46] A.P. Feinberg and B. Tycko. The history of cancer epigenetics. *Nat. Rev. Cancer*, 4:143–153, 2004.
- [47] S.E. Goelz, B. Vogelstein, S.R. Hamilton, and A.P. Feinberg. Hypomethylation of dna from benign and malignant human colon neoplasms. *Science*, 228:187–190, 1985.
- [48] A.P. Feinberg and B. Vogelstein. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, 301:89–92, 1983.
- [49] F. Gaudet, J.G. Hodgson, A. Eden, L. Jackson-Grusby, J.W. Dausman, J. an Gray, H. Leonhardt, and R. Jaenisch. Induction of tumors in mice by genomic hypomethylation. *Science*, 300:489–492, 2003.
- [50] A. Eden, F. Gaudet, A. Waghmare, and R. Jaenisch. Chromosomal instability and tumors promoted by dna hypomethylation. *Science*, 300:455, 2003.
- [51] H. Tsuda, T. Takarabe, Y. Kanai, T. Fukutomi, and S. Hirohashi. Correlation of dna hypomethylation at pericentromeric heterochromatin regions of chromosomes 16 and 1 with histological features and chromosomal abnormalities of human breast carcinomas. *Am J Pathol.*, 161:859–866, 2002.

- [52] W.A. Schulz, J.P. Elo, A.R. Florl, S. Pennanen, S. Santourlidis, R. Engers, M. Buchardt, H.H. Seifert, and T. Visakorpi. Genomewide dna hypomethylation is associated with alterations on chromosome 8 in prostate carcinoma. *genes chromosomes cancer. Am J Pathol.*, 35:58–65, 2002.
- [53] T. Ono, Y. Uehara, A. Kurishita, R. Tawa, and H. Sakurai. Biological significance of dna methylation in the ageing process. *Age Ageing*, 22:S34–43, 1993.
- [54] J. Aranda, K. Zinovjev, M. Swiderek, K. Roca, and I. Tunon. Unraveling the reaction mechanism of enzymatic c5-cytosine methylation of dna. a combined molecular dynamics and qm/mm study of wild type and gln119 variant. *Age Ageing*, 6:3262–3276, 2016.
- [55] C.B. Yoo and P.A. Jones. Dna methyltransferase inhibitors in cancer therapy. *Am. Assoc. Cancer Res. Educ. Book*, 2:333–337, 2004.
- [56] H. Gowher and A. Jeltsch. Mechanism of inhibition of dna methyltransferases by cytidine analogs in cancer therapy. *Cancer Biol. Ther.*, 3:1062–1068, 2004.
- [57] P.L. Lomen, L.H. Baker, G.L. Neil, and M.K. Samson. Phase i study of 5-azacytidine (nsc-102816) using 24-hour continuous infusion for 5 days. *Cancer Chemother Rep*, 59:1123–1126, 1975.
- [58] B.I. Shnider, M. Baig, and J. Colsky. A phase i study of 5-azacytidine (nsc-102816). *J. Clin. Pharmacol*, 16:205–212, 1976.

-
- [59] G. Egger, G. Liang, A. Aparicio, and P.A. Jones. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, 429:457–463, 2004.
- [60] A. Aparicio and J.S. Weber. Review of the clinical experience with 5-azacytidine and 5-aza-2'- deoxycytidine in solid tumors. *Curr Opin Investig Drugs*, 3:627–633, 2002.
- [61] C.A. Cooney, A.A. Dave, and G.L. Wolff. Maternal methyl supplements in mice affect epigenetic variation and dna methylation of offspring. *J. Nutr*, 132:2393S–2400S, 2002.
- [62] R.B. Klisovic, W. Stock, S. Cataland, M.I. Klisovic, L. Liu, W. Blum, M. Green, O. Odenike, L. Godley, J. Burgt, E.V. Laar, M. Cullen, A. Macleod, J. Besterman, G. Reid, J. Byrd, and G. Marcucci. A phase i biological study of mg98, an oligodeoxynucleotide antisense to dnammethyltransferase 1, in patients with high-riskmyelodysplasia and acutemyeloid leukemia. *Clinical Cancer Research*, 14:2444–2448, 2008.
- [63] R. Brown and G. Strathdee. Epigenomics and epigenetic therapy of cancer. *Trends Mol. Med.*, 8:S43–48, 2002.
- [64] P.W. Laird. Cancer epigenetics. *Human Mol. Genet*, 1:65–76, 2005.
- [65] J. R. Lakowicz. *Principles of fluorescence spectroscopy*. Springer, third edition, 2006.
- [66] W.E. Moerner and L. Kador. Optical detection and spectroscopy of single molecules in a solid. *Physical Review Letters*, 62:2535–2538, 1989.

- [67] phosphorescence. <http://www.futura-sciences.com/sciences/definitions/physique-phosphorescence-862/>. Accessed: 04-012-2017.
- [68] X. Zhang, M. Zhang, W. Li, D. and He, J. Peng, E. Betzig, and Xu P. Highly photostable, reversibly photoswitchable fluorescent protein with high contrast ratio for live-cell superresolution microscopy. *Proceedings of the National Academy of Sciences*, 113:10364–10369, 2016.
- [69] Photostability. <https://light-measurement.com/photostability/>. Accessed: 04-012-2017.
- [70] A. Reisch and A.S. Klymchenko. Fluorescent polymer nanoparticles based on dyes: Seeking brighter tools for bioimaging. *Small*, 12:1968–1992, 2016.
- [71] I.I. Shemiakina, G.V. Ermakova, P.J. Cranfill, R.A. Baird, E.A. Evans, R.A. and Souslova, D.B. Staroverov, A.Y. Gorokhovatsky, E.V. Putintseva, T.V. Gorodnicheva, T.V. Chepurnykh, S. Strukova, L. and Lukyanov, A.G. Zaraisky, M.W. Davidson, D.M. Chudakov, and D. Shcherbo. A monomeric red fluorescent protein with low cytotoxicity. *Nature Commun*, 3, 2012.
- [72] A. Jablonski. Über den mechanisms des photolumineszenz von farbstoffphosphoren. *Z Phys*, 94:38–46, 1935.
- [73] photophysique des proteines. <http://www.ibs.fr/recherche/groupe-dynamique-et-cinetique-des-processus-moleculaires-m-weik/>

- [pixel/photophysique-des-proteines/article/photophysique-des-proteines](#). Accessed: 04-012-2017.
- [74] Fluorescence and phosphorescence. https://chem.libretexts.org/LibreTexts/University_of_California_Davis/UCD_Chem_107B%3A_Physical_Chemistry_for_Life_Scientists/Chapters/7%3A_Spectroscopy/7.7%3A_Fluorescence_and_Phosphorescence. Accessed: 04-012-2017.
- [75] R.P. Haugland, J. Gregory, M.T.Z. Spence, and I.D. Johnson. *Handbook of fluorescent probes and research products; Molecular Probes*. 2002.
- [76] A. Treibs and F.H. Kreuzer. Difluorboryl-komplexe von di- und tripyrrylmethenen. *Liebigs Ann.Chem.*, 718:208–223, 1968.
- [77] C.T. Arranja, A. Aguiar, T. Encarnacao, S.M. Fonseca, L.G.L. Justino, A.E.R. Castro, A. Benniston, A. Harriman, D.H. Burrows, and Sobral J.F.N.A. Double-tailed long chain bodipys - synthesis, characterization and preliminary studies on their use as lipidfluorescence probes. *J. Molec. Struc*, 1146:62–69, 2017.
- [78] L. Zou, S. Guan, Li L., and L. Zhao. Dipyrin-based complexes for solution-processed organic solar cells. *Chem. Res. Chin. Univ.*, 31:801–808, 2015.
- [79] S. Fedeli, A. Brandi, L. Venturini, P. Chiarugi, E. Giannoni, P. Paoli, D. Corti, G. Giambastiani, G. Tuci, and S. Cicchi. The “click-on-tube” approach for the production of efficient drug carriers based on

- oxidized multi-walled carbon nanotubes. *J. Mater. Chem. B.*, 4:3823–3831, 2016.
- [80] Y. Ni, L. Zeng, N.Y. Kang, K.W. Huang, L. Wang, Z. Zeng, Y.T. Chang, and J. Wu. Meso-ester and carboxylic acid substituted bodipys with far-red and near infrared emission for bio-imaging applications. *Chem. Europ. J.*, 20:2301–2310, 2014.
- [81] Z.J. Zhao, B. Chen, J.L. Geng, Z.F. Chang, L. Aparicio-Ixta, H. Nie, C.C. Goh, L.G. Ng, A.J. Qin, Ramos-Ortiz. G., B. Liu, and B.Z Tang. Red emissive biocompatible nanoparticles from tetraphenylethene-decorated bodipy luminogens for two-photon excited fluorescence cellular imaging and mouse brain blood vascular visualization. *Partcl & Partcl. Syst. Charact.*, 31:481–491, 2014.
- [82] A. Loudet and K. Burgess. Bodipy dyes and their derivatives: syntheses and spectroscopic properties. *Chem. Rev.*, 107:4891–4932, 2007.
- [83] N. Boens, V. Leen, and W. Dehaen. Fluorescent indicators based on bodipy. *Chem. Soc.Rev.*, 41:1130–1172, 2012.
- [84] M. Li, H. Wang, X. Zhang, and H.S. Zhang. Development of a new fluorescent probe: 1,3,5,7-tetramethyl-8-(4'-aminophenyl)-4,4-difluoro-4-bora3a,4a-diaza-s-indacence for the determination of trace nitrite. *Chem. Soc.Rev.*, 60:987–993, 2004.
- [85] M. Baruah, W.W. Qin, N. Basaric, W.M. De Borggraeve, and N. Boens. Bodipy-based hydroxyaryl derivatives asfluorescent ph probes. *J. Org. Chem.*, 70:4152– 4157, 2005.

- [86] W.W. Qin, W. Dou, V. Leen, W. Dehaen, M. Van der Auweraer, and N. Boens. A ratiometric, fluorescent bodipy-based probe for transition and heavy metal ions. *Rsc Adv*, 6:7806–7816, 2016.
- [87] L. Qiu, Y.F. Ji, C.C. Zhu, Y.C. Chen, W.J. He, and Z.J. Guo. A bodipy-derived zn^{2+} fluorescent sensor: the enhanced ict effect. *Chin. J. Inorg. Chem.*, 30:169–178, 2014.
- [88] S.G. Awuah and Y. You. Boron dipyrromethene (bodipy)-based photosensitizers for photodynamic therapy. *Rsc Adv.*, 2:11169–11183, 2012.
- [89] A. Kamkaew, S.H. Lim, H.B. Lee, L.V. Kiew, L.Y. Chung, and K. Burgess. Bodipy dyes in photodynamic therapy. *Chem. Soc. Rev.*, 42:77–88, 2013.
- [90] D. Lide and G.W.A. Milne. *Handbook of Data on Organic Compounds: Compounds 10001-15600*. Cha-Hex, CRC Press, 1994.
- [91] M.S.T. Goncalves. Fluorescent labeling of biomolecules with organic probes. *Chem. Rev.*, 109:190–212, 2009.
- [92] X. Peng, J. Du, J. Fan, J. Wang, Y. Wu, S. Zhao, J. Sun, and T. Xu. A selective fluorescent sensor for imaging cd^{2+} in living cells. *J. Am. Chem. Soc.*, 129:1500–1501, 2007.
- [93] L. Zeng, E.W. Miller, A. Pralle, E.Y. Isacoff, and C.J. Chang. A selective fluorescent sensor for imaging cd^{2+} in living cells. *J. Am. Chem. Soc.*, 128:10–11, 2005.

- [94] O.A. Bozdemir, R. Guliyev, O. Buyukcakir, S. Selcuk, S. Kolemen, T. Gulseren, G. and Nalbantoglu, H. Boyaci, and E.U. Akkaya. Selective manipulation of ict and pet processes in styryl-bodipy derivatives: Applications in molecular logic and fluorescence sensing of metal ions. *J. Am. Chem. Soc.*, 132:8029–8036, 2010.
- [95] Q. Zheng, G. Xu, and P.N. Prasad. Conformationally restricted dipyrromethene boron difluoride (bodipy) dyes: Highly fluorescent, multicolored probes for cellular imaging. *Chem. Europ. J.*, 14:5812–5819, 2008.
- [96] T. Rousseau, A. Cravino, T. Bura, G. Ulrich, R. Ziessel, and J. Roncali. Multi donor molecular bulk heterojunction solar cells: improving conversion efficiency by synergistic dye combinations. *J. Mater. Chem.*, 19:2298–2300, 2009.
- [97] T. Rousseau, A. Cravino, T. Bura, G. Ulrich, R. Ziessel, and J. Roncali. Bodipy derivatives as donor materials for bulk heterojunction solar cells. *Chem. Comm.*, 0:1673–1675, 2009.
- [98] S. Erten Ela, M.D. Yilmaz, B. Icli, Y. Dede, S. Icli, and E.U. Akkaya. A panchromatic boradiazaindacene (bodipy) sensitizer for dye-sensitized solar cells. *Chem. Comm.*, 10:3299–3302, 2008.
- [99] S. Kolemen, Y. Cakmak, S. Erten-Ela, Y. Altay, J. Brendel, M. Thelakkat, and E.U. Akkaya. Solid-state dye-sensitized solar cells using red and near-ir absorbing bodipy sensitizers. *Org. Lett.*, 12:3812–3815, 2010.

-
- [100] Q. Zheng, G.S. He, and P.N. Prasad. A novel near ir two-photon absorbing chromophore: Optical limiting and stabilization performances at an optical communication wavelength. *Chem. Phys. Lett*, 475:250–255, 2009.
- [101] P. Didier, G. Ulrich, Y. Mely, and R. Ziessel. Improved push-pull-push e-bodipy fluorophores for two-photon cell-imaging. *Org. Biomol. Chem.*, 7:3639–3642, 2009.
- [102] G. Ulrich, R. Ziessel, and A. Harriman. The chemistry of fluorescent bodipy dyes versatility unsurpassed. *Angew. Chemie-Int. Ed.*, 47:1184–1201, 2008.
- [103] R. Ziessel, G. Ulrich, and A. Harriman. The chemistry of bodipy: a new el dorado for fluorescence tools. *New J. Chem.*, 31:496–501, 2007.
- [104] M.P. Rosconi, G. Zhao, and E. London. Analyzing topography of membraneinserted diphtheria toxin t domain using bodipy-streptavidin: at low ph, helices 8 and 9 form a transmembrane hairpin but helices 5 7 form stable nonclassical inserted segments on the cis side of the bilayer. *Biochemistry*, 43:9127–9139, 2004.
- [105] J. Rumin, H. Bonnefond, B. Saint-Jean, C. Rouxel, A. Sciandra, O. Bernard, J.P. Cadoret, and G. Bougaran. The use of fluorescent Nile red and bodipy for lipid measurement in microalgae. *Biochem. Biophys. Res. Commun.*, 8:8–42, 2015.
- [106] J. Kasurinen. A novel fluorescent fatty-acid, 5-methyl-bdy-3-dodecanoic acid, as a potential probe in lipid transport studies in lipid trans-

- port studies by incorporating selectively to lipid classes of bhk cells. *Biochem. Biophys. Res. Commun.*, 187:1594–1601, 1992.
- [107] J.S. Hansen, T. Petersen J.F., Hoeg-Jensen, and J.B. Christensen. Buffer and sugar concentration dependent fluorescence response of a bodipy-based aryl monoboronic acid sensor. *Tetrahedron Lett.*, 53:5852–5855, 2012.
- [108] M. Vendrell, D. Zhai, J.C. Er, and Y.T. Chang. Combinatorial strategies in fluorescent probe development. *Chem. Rev.*, 112:4391– 4420, 2012.
- [109] T. Kobayashi, T. Komatsu, M. Kamiya, C. Campos, M. Gonzalez Gaitan, T. Terai, K. Hanaoka, T. Nagano, and Y. Urano. Highly activatable and environment sensitive optical highlighters for selective spatiotemporal imaging of target proteins. *J. Am. Chem. Soc.*, 134:11153– 11160, 2012.
- [110] J.A. Levitt, M.K. Kuimova, G. Yahioglu, P.H. Chung, K. Suhling, and D. Phillips. Membrane-bound molecular rotors measure viscosity in live cells via fluorescence lifetime imaging. *J. Phys. Chem. C.*, 113:11634– 11642, 2009.
- [111] G. Hungerford, A. Allison, D. McLoskey, M.K. Kuimova, G. Yahioglu, and K. Suhling. Monitoring sol-to-gel transitions via fluorescence lifetime determination using viscosity sensitive fluorescent probes. *J. Phys. Chem. B.*, 113:12067– 12074, 2009.

- [112] W. Qin, M. Baruah, M. Van der Auweraer, F.C. De Schryver, and N. Boens. Photophysical properties of borondipyrromethene analogues in solution. *J. Phys. Chem. A.*, 109:7371–7384, 2005.
- [113] G. Ulrich, C. Goze, M. Guardigli, A. Roda, and R. Ziessel. Pyrromethene dialkynyl borane complexes for “cascadelle” energy transfer and protein labeling. *Angew. Chem.*, 44:3760–3764, 2005.
- [114] R. Ziessel, C. Goze, G. Ulrich, M. Cesario, P. Retailleau, A. Harriman, and J.P. Rostron. Intramolecular energy transfer in pyrene–bodipy molecular dyads and triads. *Chem. Europ.J.*, 11:7366–7378, 2005.
- [115] R. Ziessel, G. Ulrich, K. Elliott, and A. Harriman. Electronic energy transfer in molecular dyads built around boron–ethynesubstituted subphthalocyanines. *Chem. Europ. J.*, 15:4980–4984, 2009.
- [116] T. Kowada, H. Maeda, and Kikuchi. K. Bodipy-based probes for the fluorescence imaging of biomolecules in living cells. *Chem. Soc. Rev.*, 44:4953–4972, 2015.
- [117] B. Valeur and M.N. Berberan-Santos. Molecular fluorescence: Principles and applications. 2012.
- [118] M. Sauer, J. Hofkens, and J. Enderlein. Handbook of fluorescence spectroscopy and imaging. pages –, 2011.
- [119] X. Li, X. Gao, W. Shi, and H. Ma. Design strategies for water-soluble small molecular chromogenic and fluorogenic probes. *Chem. Rev.*, 114:590–659, 2014.

- [120] K. Lang and J.W. Chin. Bio-orthogonal reactions for labeling proteins. *Chem. Biol.*, 9:16–20, 2014.
- [121] D. M. Patterson, L.A. Nazarova, and J.A. Prescher. Finding the right (bio-orthogonal) chemistry. *Chem. Biol.*, 9:592–605, 2014.
- [122] T. Gronemeyer, G. Godin, and K. Johnsson. Adding value to fusion proteins through covalent labelling. *Curr. Opin. Biotechnol.*, 16:453–458, 2005.
- [123] S. Mizukami, Y. Hori, and K. Kikuchi. Small-molecule-based protein-labeling technology in live cell studies: Probe-design concepts and applications. *Acc. Chem. Res.*, 47:247–256, 2014.
- [124] L.V. Johnson, M.L. Walsh, and L.B. Chen. Localization of mitochondria in living cells with rhodamine 123. *Proc. Natl. Acad. Sci. USA*, 77:990–994, 1980.
- [125] H. Lu, J. Mack, Y. Yang, and Z. Shen. Structural modification strategies for the rational design of red/nir region bodipys. *Chem. Soc. Rev.*, 43:4778–4823, 2014.
- [126] Y. Ni and J. Wu. Far-red and near infrared bodipy dyes: synthesis and applications for fluorescent probes and bio-imaging. *Org. Biomol. Chem.*, 12:3774–3791, 2014.
- [127] K.E. Sapsford, L. Berti, and I.L. Medintz. Materials for fluorescence resonance energy transfer analysis: Beyond traditional donor–acceptor combinations. *Angew. Chem.*, 45:4562–4588, 2006.

- [128] H. Sunahara, H. Urano, Y. and Kojima, and T. Nagano. Design and synthesis of a library of bodipy-based environmental polarity sensors utilizing photoinduced electron-transfer-controlled fluorescence on/off switching. *J. Am. Chem. Soc.*, 129:5597–5604, 2007.
- [129] D. Dziuba, R. Pohl, and M. Hocek. Bodipy-labeled nucleoside triphosphates for polymerase synthesis of fluorescent dna. *Bioconjugate Chem.*, 25:1984–1995, 2014.
- [130] X. Li, F. Traganos, M. R. Melamed, and Z. Darzynkiewicz. Single-step procedure for labeling dna strand breaks with fluorescein- or bodipy-conjugated deoxynucleotides: Detection of apoptosis and bromodeoxyuridine incorporation. *Cytometry.*, 20:172–180, 1995.
- [131] H.C. Kolb, M.G. Finn, and K.B. Sharpless. Click chemistry: Diverse chemical function from a few good reactions. *Angew. Chem.*, 40:2004–2021, 2001.
- [132] Click chemistry and nucleic acids. <https://www.atdbio.com/content/23/Click-chemistry-and-nucleic-acids>. Accessed: 04-012-2017.
- [133] Click chemistry labeling and detection. <https://www.thermofisher.com/de/de/home/life-science/cell-analysis/labeling-chemistry/click-chemistry-labeling-and-detection.html>. Accessed: 04-012-2017.

- [134] V.V. Rostovtsev, L.G. Green, V.V. Fokin, and K.B. Sharpless. A stepwise huisgen cycloaddition process: Copper(i)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew. Chem.*, 41:2596–2599, 2002.
- [135] L. Landau and E. Lipchitz. *Mécanique quantique – Théorie non relativiste*. Editions Mir., second edition, 1967.
- [136] Bureau International des Poids et Mesures et Organisation Intergouvernementale de la Convention du Mètre. *Le système international d’unités (SI)*. 8th edition, 2006.
- [137] L.J. Butler. Chemical reaction dynamics beyond the born-oppenheimer approximation. *Annu. Rev. Phys. Chem.*, 49:125–171, 1998.
- [138] D.R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part i.theory and methods. *Proc. Cambridge Phil. Soc.*, 24:89–110, 1928.
- [139] D.R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part ii. some results and discussion. *Proc. Cambridge Phil. Soc.*, 24:111–132, 1928.
- [140] P.W. Atkins and R. Friedman. *Molecular quantum mechanics*. Oxford University Press, Oxford fifth edition, 2011.
- [141] R.G. Parr and W. Yang. *Density-functional theory of atoms and molecules*. Oxford University Press New York, first issue paperback., 1994.

-
- [142] I.N. Levine. *Quantum chemistry*. Brooklyn College, City University of New York. Pearson, sixth edition, 2009.
- [143] F. Jensen. *Introduction to computational chemistry*. Wiley, second edition, 2007.
- [144] C.J. Cramer. *Essentials of computational chemistry: theories and models*. Wiley, second edition, 2004.
- [145] R.G. Parr and W. Yang. Density-functional theory of the electronic structure of molecules. *Annu.Rev. Phys. Chem*, 46:701–728, 1995.
- [146] R.D. Luis, R.G. Mar, and P. Patricia. Applications of the conceptual density functional theory indices to organic chemistry reactivity. *Molecules*, 21:748–770, 2016.
- [147] R.G. Parr and R.G. Pearson. Absolute hardness: Companion parameter to absolute electronegativity. *J. Am. Chem. Soc.*, 105:7512–7516, 1983.
- [148] R.G. Parr and W. Yang. Density functional theory of atoms and molecules. *Oxford University Press: New York, NY, USA.*, 1989.
- [149] L. Pauling. The nature of the chemical bond. iv. the energy of single bonds and the relative electronegativity of atoms. *J. Am. Chem. Soc.*, 54:3570–3582, 1932.
- [150] L. Pauling and J. Sherman. A quantitative discussion of bond orbitals. *J. Am. Chem. Soc.*, 59:1450–1456, 1937.

- [151] R.G. Pearson. Hard and soft acids and bases. *J. Am. Chem. Soc.*, 85:3533–3539, 1963.
- [152] R.G. Pearson. Acids and bases. *Science*, 151:172–177, 1966.
- [153] R.G. Pearson and J. Songstad. Application of the principle of hard and soft acids and bases to organic chemistry. *J. Am. Chem. Soc.*, 89:1827–183, 1967.
- [154] R.G. Pearson. Hard and soft acids and bases. *Dowden, Hutchinson and Ross: Stroudsburg, PA, USA*, 1973.
- [155] R.G. Pearson. Chemical hardness: Applications from molecules to solids. *Wiley-VCH, Weinheim, Germany*, 1997.
- [156] Y. Zhao and D.G. Truhlar. Density functional for noncovalent interaction energies of biological importance. *J. Chem. Theory Comput*, 3:289–300, 2007.
- [157] R.G. Parr and W. Yang. Density functional approach to the frontier-electron theory of chemical reactivity. *J. Am. Chem. Soc.*, 106:4049–4050, 1984.
- [158] W.T. Yang and W.J. Mortier. The use of global and local molecular parameters for the analysis of the gas-phase basicity of amines. *J. Am. Chem. Soc.*, 108:5708–5711, 1986.
- [159] R.G. Parr, L. Von Szentpaly, and S. Liu. Electrophilicity index. *J. Am. Chem. Soc.*, 121:1922–1924, 1999.

-
- [160] P. Perez, L.R. Domingo, A. Aizman, and R. Contreras. The electrophilicity index in organic chemistry. *Theor. Comp Chem.*, 19:139–201, 2007.
- [161] R.K. Roy, S. Krishnamurti, P. Geerlings, and S. Pal. Local softness and hardness based reactivity descriptors for predicting intra- and intermolecular reactivity sequences: Carbonyl compounds. *J.Phys. Chem. A*, 102:3746–3775, 1998.
- [162] P.K. Chattaraj, B. Maiti, and U. Sarkar. Philicity: A unified treatment of chemical reactivity and selectivity. *J. Phys. Chem. A*, 107:4973–4975, 2003.
- [163] P.K. Chattaraj. Chemical reactivity and selectivity: Local h_{sab} principle versus frontier orbital theory. *J. Phys. Chem. A*, 105:511–513, 2001.
- [164] K. Umezawa, A. Matsui, D. Nakamura, Y. and Citterio, and K. Suzuki. Bright color tunable fluorescent dyes in the vis nir region: Establishment of new 'tailor made' multicolor fluorophores based on borondipyrromethene. *J. Chem. Eur.*, 15:1096–1106, 2009.
- [165] K. Umezawa, Y. Nakamura, H. Makino, D. Citterio, and K. Suzuki. Bright color tunable fluorescent dyes in the visible nearinfrared region. *J. Am. Chem. Soc.*, 130:1550–1551, 2008.
- [166] R. Misra, P. Chakraborty, S.C. Roy, D.K. Maity, and S.P. Bhattacharyya. Tailoring of spectral response and intramolecular charge

- transfer in β -enaminones through band gap tuning: Synthesis, spectroscopy and quantum chemical studies. *RSC Adv.*, 6:36811–36822, 2016.
- [167] K.S. Thanthiriwatte and K.M.N. De Silva. Non-linear optical properties of novel fluorenyl derivatives - *ab initio* quantum chemical calculations. *RSC Adv.*, 617:169–175, 2002.
- [168] L.K. Zhang, T. and Yan, C. Cong, W. Guan, and Z.M. Su. Prediction of second-order nonlinear optical properties of wells–dawson polyoxometalate derivatives $[x-c(ch_2o)_3 p_2 m_3 m_1 5 o_5 9]^{6-}$ ($x=no_2$, nh_2 , and ch_3 , $m=v$ and nb , $m=w$ and mo). *Inorg. Chem. Front.*, 1:65–70, 2014.
- [169] M. Albrecht, A. Lippach, M. P. Exner, J. Jerbi, M. Springborg, N. Budisa, and G.Wenz. Site-specific conjugation of 8-ethynyl-bodipy to a protein by [2 + 3] cycloaddition. *Organic Biomolecular Chemistry*, 15:6728–6736, 2015.
- [170] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo,

- R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision E.01, 2009. Gaussian Inc. Wallingford CT.
- [171] A.D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, 1988.
- [172] C. Lee, W. Yang, and R.G. Parr. Development of the colic-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 3:785–789, 1988.
- [173] W. Kohn, A.D. Becke, and R.G. Parr. Density functional theory of electronic structure. *Phys. Rev. B*, 100:12974–12980, 1996.
- [174] W. Kohn and L.J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, 1965.
- [175] R. Misra. Tuning of second-order nonlinear optical response properties of aryl- substituted boron-dipyrromethene dyes: Unidirectional charge transfer coupled with structural tailoring. *J. Phys. Chem. C*, 121:5731–5739, 2017.
- [176] R. Hu, E. Lager, A. Aguilar-Aguilar, J. Liu, J.W.Y. Lam, H.H.Y. Sung, I.D. Williams, Y. Zhong, K.S. Wong, E. Pena-Cabrera, and B.Z. Tang. Twisted intramolecular charge transfer and aggregation-induced emission of bodipy derivatives. *J. Phys. Chem. C*, 113:15845–15853, 2009.

-
- [177] J. Jerbi and M. Springborg. Reactivity descriptors for dna bases and the methylation of cytosine. *Int. J. Quant. Chem.*, -:e25538, 2017.
- [178] J. Jerbi and M. Springborg. Computational study of the reactivity of cytosine derivatives. *J. Comp. Chem.*, 38:1049–1056, 2017.
- [179] F. Kloss, U. Kohn, B.O. Jahn, M.D. Hager, H. Górls, and U.S. Schubert. Metal-free 1,5-regioselective azide–alkyne [3+2]-cycloaddition. *Chem. Asian J*, 6:2816 – 2824, 2011.
- [180] D.H. Ess and K.N. Houk. Theory of 1,3-dipolar cycloadditions: Distortion/interaction and frontier molecular orbital models. *J. AM. CHEM. SOC.*, 130:10187–10198, 2008.

FULL PAPER

Reactivity descriptors for DNA bases and the methylation of cytosine

Jihène Jerbi¹  | Michael Springborg^{1,2}

¹Physical and Theoretical Chemistry,
Saarland University, 66123 Saarbrücken,
Germany

²School of Materials Science and
Engineering, Tianjin University, Tianjin
300072, People's Republic of China

Correspondence

Jihène Jerbi, Physical and Theoretical
Chemistry, Saarland University, 66123
Saarbrücken, Germany. Email jihene.
jerbi@uni-saarland.de

Abstract

This work is devoted to the reactivity of DNA bases and base pairs. Various global molecular descriptors for reactivity were studied in addition to local descriptors for Cytosine and Cytosine-Guanine base pairs using three different ways of analysing the charge distribution. The process of methylation was selected since it is very important during the cell formation at the embryonic state and since much is already known about this system making an assessment of the quality of the theoretical approaches possible. The results reveal that Cytosine is the least reactive molecule among the bases although it can be attacked by the DNMT enzyme while paired with Guanine. A main finding of the present work is, however, that by analysing the Fukui function and a dual descriptor we can rationalise that carbon at position 5 of Cytosine is the most favorable site for the methyl attack.

KEYWORDS

cytosine, DFT calculation, DNA methylation, dual descriptor, Fukui function, reactivity

1 | INTRODUCTION

Alterations of deoxyribonucleic acid, DNA, such as oxidation, alkylation, hydrolysis, and mismatch lead to genetic mutations or to DNA replication blockage that ultimately may cause cancer. In fact, in most cases, these alterations can be repaired by specific enzymes. This repair must occur before the DNA replication takes place or the genome integrity will be lost and errors will take place during the transcription process.^[1]

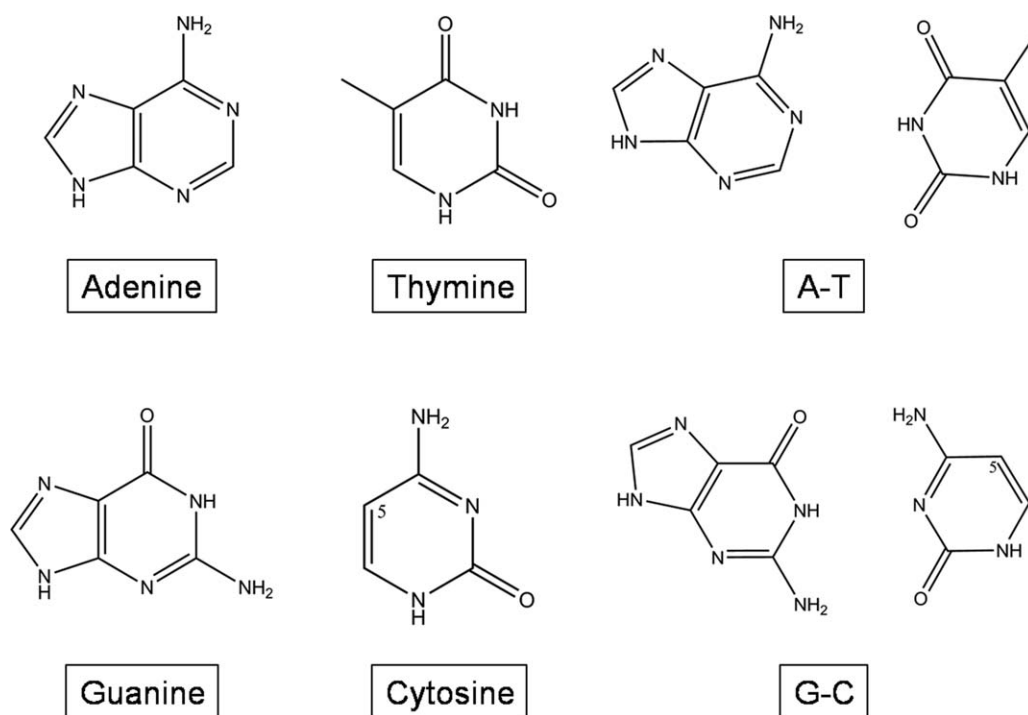
The best-known lesion and damages of DNA are deamination and methylation that concern the nucleobases shown in Scheme 1. In the present work, we will concentrate on the methylation.

DNA methylation can be defined as the addition of a methyl group from a DNA methyltransferase enzyme to a specific site of the nucleobase. It is known that this biological process is fundamental in bacteria, plants, and animal cells since it helps to encode the information in a stable and reversible manner.^[2]

S-adenosyl-L-methionine (SAM) is the DNA MTase catalysing the methyl transfer. However, the mechanism occurs differently in prokaryotes and mammals.^[3] Thus, in prokaryotes, the methyl group is transferred to an Adenine (at the N6 position) or to a Cytosine nucleobase (N4 or C5 positions), while in mammals and other vertebrates methylation takes place at only the C5 position of Cytosine.

This enzymatic methylation of Cytosine results in the fifth DNA base 5-methyl-cytosine that can be found only when Cytosine is paired to Guanine,^[4] which occurs with a frequency of about 5% in human cells.^[5,6] Moreover, 5-MeCyt is believed to cause about one-third of all transition mutations and is responsible for human genetic diseases like, in particular, cancer^[7,8] and damages to the immune system.^[9,10]

From computational point of view, the size of DNA makes it very difficult to carry through detailed and accurate electronic-structure calculations for the study of its chemical reactivity (see e.g., the QM/MM study by Tonon et al.^[11]). Very helpful would it be if it is possible to use simpler tools on smaller systems for quantifying the chemical reactivity, the question that we shall address here. We shall demonstrate that the isolated molecules can give the appropriate information needed. Specifically, purpose of the present study is to analyse the performance of such simpler tools by considering Cytosine as a target for the DNA MTase and thereby studying whether these tools can identify exactly the carbon atom at position 5 of the Cytosine as the location of the methyl addition. Various previous studies, both experimental and theoretical, have focused on DNA nucleobases as this subject has been of interest for more than fifty years.^[12-16]



SCHEME 1 Isolated and pairs of DNA bases. Leftmost is shown the purines (Adenine and Guanine), in the middle the pyrimidines (Thymine and Cytosine), and to the right the base pairs. The Carbon atom at position 5 in Cytosine is explicitly shown

At first, we shall present calculated values for global reactivity descriptors for the four bases Adenine, Guanine, Cytosine, and Thymine and the two base pairs Adenine-Thymine (A-T) and Cytosine-Guanine (C-G) to obtain information on the differences between the nucleobases. Subsequently, we focus on local reactivity descriptors, that is, the Fukui function and a dual reactivity descriptor to confirm the fact that the carbon at position 5 of the cytosine is the place where the addition of the methyl group occurs when Cytosine is methylated by the enzyme SAM.

2 | THEORETICAL BACKGROUND

Various global reactivity descriptors including the electronegativity χ ,^[17,18] the chemical potential μ ,^[19] the hardness,^[20,21] and the electrophilicity ω ^[22] were given a solid theoretical foundation by Parr and Yang^[23] within the concept of density functional theory.^[24] At first we, have (where IP is the ionization potential and EA represents the electron affinity):

$$\chi = -\mu = -\left(\frac{\delta E}{\delta N}\right)_{V(r)} \simeq \frac{IP+EA}{2} \quad (1)$$

Equivalently, one may define two different electronegativities,^[25]

$$\chi^- = \frac{1}{4}(3IP+EA) \quad (2)$$

$$\chi^+ = \frac{1}{4}(IP+3EA) \quad (3)$$

It is relevant to mention that a better electron donor has a lower value of χ^- whereas a larger value of χ^+ implies a larger capacity to accept electrons.

The hardness is given by

$$\eta = \left(\frac{\partial^2 E}{\partial N^2}\right)_{V(r)} = \left(\frac{\partial \mu}{\partial N}\right)_{V(r)} \simeq \frac{IP-EA}{2} \quad (4)$$

from which an electrophilicity can be introduced as

$$\omega = \frac{\mu^2}{2\eta} \quad (5)$$

Another quantity with relevance for chemical reactivity is the molecular polarizability. It is defined as the change in the dipole moment in response to an applied electric field.^[26] The orientationally averaged polarizability is calculated as^[27]

$$\alpha = \frac{1}{3} (\alpha_{xx} + \alpha_{yy} + \alpha_{zz}) \quad (6)$$

where α_{xx} , α_{yy} , and α_{zz} are the xx , yy , and zz components of the polarizability tensor, respectively. Also, the energy gap between the highest occupied molecular orbital and the lowest unoccupied molecular orbital can serve as a global reactivity descriptor.^[28]

For more detailed information it is useful to study local reactivity descriptors, that is, functions in position space or their decomposition into atomic contributions. The latter applies also to the electronic distribution. Here, we here used three methods to calculate atomic charges and local reactivity descriptors. These methods are the ones based on the Mulliken analysis,^[29] the Hirshfeld analysis,^[30,31] and the ESP method CHELPG. With the latter, partial charges are determined so that they reproduce the electrostatic potential generated by the molecular charge distribution of the nuclei and the electrons. The potential values are calculated at a set of test points located around the molecule according to a characteristic scheme, such as the CHELPG scheme by Breneman and Wiberg.^[32]

2.1 | Fukui function

More detailed is it to study spatially resolved response functions to which the Fukui function belongs that was proposed by Fukui.^[33,34] It is defined as the response of the electronic charge density of a system to a change in the number of electrons N under the constraint of a constant external potential,^[4]

$$f(r) = \left(\frac{\partial \rho(r)}{\partial N} \right)_{V(r)} \quad (7)$$

Since this function is nonanalytic as a function of N for integral N , one can define three different types of Fukui functions for a given system.^[4] With finite difference approximations these become:

1. f^+ for a nucleophilic attack measured by the electronic charge density change when accepting an electron:

$$f^+(r) = \rho(N+1, r) - \rho(N, r) \quad (8)$$

2. f^- for an electrophilic attack measured by the electronic charge density change after donating an electron:

$$f^-(r) = \rho(N, r) - \rho(N-1, r) \quad (9)$$

3. f° for a radical attack which can be approximated as the average of the other functions:

$$f^\circ(r) = \frac{\rho(N+1, r) - \rho(N-1, r)}{2} = \frac{f^+(r) + f^-(r)}{2} \quad (10)$$

In these expressions, $\rho(K, r)$ is the electronic charge density for the system with K electrons, and for the neutral system $K = N$

2.2 | Condensed Fukui function

Mortier and Yang proposed a simplified approach to analyse the Fukui function using the concept of a condensed Fukui function,^[35] which is related to the Mulliken atomic populations.^[36] They proposed an atomic condensed Fukui function for a given atom (no. n) as:

$$f_n^+ = q_n(N+1) - q_n(N) \quad (11)$$

$$f_n^- = q_n(N) - q_n(N-1) \quad (12)$$

Where f_n^+ and f_n^- are the condensed Fukui function for the n th atom and $q_n(M)$, $M = N-1$, N , and $N+1$, are the Mulliken charges for the n th atom in the molecule with $N-1$, N , and $N+1$ electrons, respectively^[36] and, again, the neutral system has $M = N$.

Sites at which f_n^+ is large and negative tend to accept electrons^[37] whereas sites with large and positive f_n^- are identified as donor sites since at those regions removing an electron destabilizes the molecule the least.^[37] Also condensed f_n^0 functions can be defined in a similar fashion.

When approximating $f^+(r)$ and $f^-(r)$ as being identical to the charge densities of the LUMO and HOMO, respectively, $f^0(r)$ becomes identical to their average. This makes it less obvious how to interpret large or small values of f_n^0 .

2.3 | Local softness

The local softness was defined by Yang and Parr^[38] as

$$s(r) = \frac{\partial \rho(r)}{\partial \mu} \quad (13)$$

which can be written in terms of the global softness, S , as

$$s(r) = f(r) S \quad (14)$$

The local softness and the Fukui function are accordingly closely related and they are considered as very important descriptors of chemical reactivity.^{24b} Although the Fukui function can be obtained from the local softness the reverse is not the case. Details are given in Ref. 39. Equivalent to the case of the Fukui function described above we can define three types of local softness,^[35] $s^-(r)$, $s^0(r)$, and $s^+(r)$. As for the Fukui function, also a condensed (atom-resolved) softness can be defined.

From the condensed softness, Roy et al.^[40] introduced a relative electrophilicity as s_n^+/s_n^- and a relative nucleophilicity as s_n^-/s_n^+ that provide the correct intramolecular reactivity trends which are not always the case when comparing only s_n^+ and s_n^- separately. According to this approach, the site having the highest s_n^+/s_n^- is the most probable site for an attack by a nucleophile, and the site having the highest s_n^-/s_n^+ ratio is the most probable site for an attack by an electrophile.^[40]

2.4 | Local electrophilicity

Also, a local electrophilicity can be defined^[41]:

$$\omega_n^\alpha = \omega f_n^\alpha \quad (15)$$

Where $\alpha = +, -, \text{ and } 0$ refer to nucleophilic, electrophilic, and radical attacks, respectively. Equation 15 highlights the role of the Fukui function and of the frontier orbital theory in general.^[35,42]

The local electrophilicity combines the global electrophilicity ω with the site selectivity contained in the Fukui function which makes an interpretation easier.^[41]

When the global electrophilicity of two different molecules is different, the best sites of the two molecules for a given reaction are at best identified through the local electrophilicity and not the Fukui function.^[41]

2.5 | Dual descriptor

Finally, in 2005, Morelle et al.^[43] proposed a new index of selectivity toward nucleophilic attack, that is, the dual index descriptor defined as

$$\Delta f(r) = [f^+(r) - f^-(r)] \approx [\rho^{\text{LUMO}}(r) - \rho^{\text{HOMO}}(r)] \quad (16)$$

Where $\rho^{\text{LUMO}}(r)$ and $\rho^{\text{HOMO}}(r)$ are the electronic charge densities of the LUMO and HOMO, respectively.

In regions where $\Delta f(r) > 0$ there is an increased tendency for a nucleophilic attack, whereas those regions where $\Delta f(r) < 0$ are hardly susceptible to a nucleophilic attack but rather to an electrophilic attack. As above, we can introduce a condensed version of this function f_n^Δ .

As indicated by Equation 16, $\Delta f(r)$ quantifies the changes in the electronic charge density upon an electronic excitation, a process that is relevant when using perturbation theory in studying weak interactions with other molecules, for instance. Thus, with this interpretation, sites with large negative values of f_n^Δ are sites that will receive electrons when the molecule is being perturbed. This is for instance relevant when a hydrogen atom is replaced by a methyl group, that is, the process we will study in the present work.

2.6 | Computational details

The three-parameter hybrid density functional B3LYP^[44] was used for all calculations within the framework of the density-functional theory^[45] and with the basis set 6-31++G* as implemented in the Gaussian 09 program suite.^[46] This choice was motivated by the fact that a flexible basis set is required for an accurate description of the dipole moment and its derivatives like polarizability.^[26] Moreover, as pointed out by Šponer and Hobza (see e.g., Ref. 47), B3LYP is expected to give accurate results on structures of base pairs.

In additional calculations, we included the dispersion-energy correction DFT-D3^[48] for all the molecules in order to study the importance of dispersive interactions.

At first, we optimized the molecular geometries of the neutral purines and pyrimidine bases and base pairs. That the structures correspond to total-energy minima was verified through vibrational frequency calculations. Subsequently, for the charged species we considered both the structures of the neutral species as well as the optimized ones whereby the effects of structural and electronic relaxations can be separated.

TABLE 1 Calculated polarizabilities (in Å³) of single and pairs of DNA nucleobases using DFT at the B3LYP/6.31++G* level

	α (Gas)	α (Solvent)	α/V_{mol} (Gas)	α/V_{mol} (Solvent)	α/N_{el} (Gas)	α/N_{el} (Solvent)
Adenine	92.59	126.79	0.426	0.584	1.852	2.536
Thymine	79.29	104.96	0.383	0.507	1.652	2.186
A-T	176.24	235.65	0.416	0.556	1.789	2.404
Cytosine	74.11	98.31	0.412	0.546	1.764	2.341
Guanine	98.80	133.72	0.426	0.577	1.764	2.387
C-G	176.91	237.15	0.429	0.576	1.805	2.419

Both the values themselves and those obtained by dividing by the molecular volume (V_{mol} , in Å³) and the number of valence electrons (N_{el}) are given.

The calculations were performed both for the gas phase and in aqueous solution, whereby we in the latter case treated water using the polarizable continuum model, PCM.^[49] Finally, for the sake of comparison, we shall also present results for the Adenine and Guanine bases as well as for the A-T and C-G base pairs.

3 | RESULTS AND DISCUSSION

Many excellent theoretical works have shed light on the DNA base pairs A-T and C-G structures. The reader is referred to the work of Poarter et al.^{[14]a} and Guerra et al.^{[14]b} for theoretical results on the A-T structures and the experimental work of Seeman et al.^{[15]c} However, we summarize our results and compare with previous works in the Supporting Information to establish the accuracy of our numerical approach and partly for comparison with other systems.

Moreover, Kurita et al.^{[50]a} gave a fair explanation about NH₂ group of the C-G base pair that gives rise to a nonplanarity of the latter which is caused mainly by the sp³ pyramidalization of the amino group of DNA as explained. The importance of this nonplanarity of the DNA amino group in the hydrogen bonds is discussed in detail by Sponer et al.^{[50]b} A detailed study of the effects of hydrogen bonds on the structure of C-G base pair and its energy, in gas and aqueous solution, was presented in our recent work^[51] where we used the same computational approach as here.

3.1 | Polarizability

It has been suggested that the polarizability is related to the reactivity.^[52] The polarizability is a measure for the changes in the charge distribution within a molecule induced by an electric field so that a large polarizability suggests that the system easily responds to the presence of other molecules. Thus, a system with a low polarizability is supposed to be more stable, that is, less reactive.^[51]

According to our calculations (cf. Table 1), in the gas phase Cytosine and Thymine are the molecules with the smallest values for the polarizability, that is, 74.11 and 79.29 Å³, respectively, although they are simultaneously also those with the smallest numbers of atoms. This result is relevant since they are active in many DNA mechanisms like methylation, demethylation of the Cytosine, amination and deamination of Thymine, and the formation of Uracil from RNA.^[53] It has been suggested that Thymine can contribute to the demethylation of DNA, too, using the Uracil pathway.^[54]

For the A-T and C-G base pairs, the polarizabilities are slightly larger than the sum of the values of the individual bases. Dispersion corrections have only small effects on the polarizabilities, whereas the solvent changes the polarizability by almost 35 Å³ for Guanine and Adenine, about 25 Å³ for Cytosine and Thymine, and around 60 Å³ for the CG and AT pairs.

When comparing the polarizabilities of different molecules with the purpose of identifying chemical reactivity, the comparison becomes obscured by the fact that the polarizability is an extensive quantity and, accordingly, in general increases with system size. To remove this dependency we follow two different approaches. In one, we divide the polarizability by an estimate of the molecular volume, V_{mol} , and in the other, we divide by the number of valence electrons.

In order to get a rough estimate of the molecular volume, we use the atomic radii, r_i as given by Bondi^[55] to obtain:

$$V_{\text{mol}} = \frac{4}{3} \pi \sum_i r_i^3 \quad (17)$$

For these scaled polarizabilities, the values in Table 1 show that the pyrimidine bases have the smallest values, whereas the purines and the base pairs have very similar values in the gas phase. In the solvent, Thymine still has the smallest value of 0.51 followed by the cytosine with the value 0.55. Adenine and Guanine have larger values. However, the C-G base pair is less stable (in the sense of having a larger value for the polarizability) than the A-T base pair. Finally, we also arrive at these conclusions when scaling the polarizability of each molecule with its number of valence electrons: again C-G is less stable or, equivalently, more reactive. These findings have more general consequences: when comparing polarizabilities

TABLE 2 The orbital energies of the HOMO and LUMO (E_{HOMO} and E_{LUMO}) in comparison with the total-energy differences of the neutral and charged species both for fixed geometries of the neutral system and for relaxed geometries

	Adenine	Thymine	A-T	Cytosine	Guanine	C-G
E_{HOMO}	-6.31 -6.44	-6.95 -6.83	-6.12 -6.36	-6.61 -6.75	-6.08 -6.15	-5.45 -5.93
$(E_{\text{tot}}(N) - E_{\text{tot}}(N-1))_{\text{unrelaxed}}$	-8.27 -6.37	-9.01 -6.79	-7.80 -6.29	-8.69 -6.69	-7.23 -5.85	-7.23 -5.85
$(E_{\text{tot}}(N) - E_{\text{tot}}(N-1))_{\text{relaxed}}$	-8.06 -6.15	-8.78 -6.57	-7.65 -6.09	-8.56 -6.56	-6.89 -5.61	-6.89 -5.61
E_{LUMO}	-0.94 -1.07	-1.54 -1.43	-1.42 -1.42	-1.32 -1.25	-0.87 -0.79	-1.64 -1.39
$(E_{\text{tot}}(N+1) - E_{\text{tot}}(N))_{\text{unrelaxed}}$	7.93 -1.22	8.38 -1.54	3.59 -1.54	3.74 -1.39	0.39 -1.52	0.39 -1.52
$(E_{\text{tot}}(N+1) - E_{\text{tot}}(N))_{\text{relaxed}}$	0.32 -1.48	-0.02 -1.90	-0.13 -1.92	0.12 -1.84	-0.36 -1.95	-0.36 -1.95

The results have been found using DFT at the B3LYP/6.31++G* level. The upper row presents the gas phase and the lower one the solution values and all values are given in eV. $E_{\text{tot}}(M)$ is the total energy for the system with M electrons and for the neutral system, $M = N$. relaxed/unrelaxed describes whether the structure of the charged systems has been optimized.

of different molecules it is important to compare properties that are scaled by the molecular size “somehow,” whereby the precise definition of this “somehow” is less relevant as long as it is realistic.

3.2 | Ionization potentials and electron affinities

In our previous work,^[51] we observed that the energies of the highest occupied and lowest unoccupied orbitals, HOMO and LUMO, as well as the energy gap separating those are useful descriptors for the reactivity of DNA bases and base pairs. Due to the delocalization of the mobile electrons, these energies provide information on the stabilization of the molecules. In fact, we found that the purines (here Guanine and Adenine) are better electron donors than the pyrimidines^[56] (here Cytosine and Thymine).

According to Koopmans' theorem, the HOMO and LUMO energies are related to the ionization potential and the electron affinity, respectively, when neglecting relaxation effects. Since we, however, have also studied the charged species both when keeping the structure as that of the neutral species as well as when relaxing the structure we can study the effects of the electronic and structural relaxations too.

In Table 2, we summarise the results for our systems in gas and solvent phases. The HOMO-LUMO energy gap lies between 3.8 eV for CG and 5.4 eV for Adenine in Thymine in the gas phase and between 4.5 eV for CG and 5.5 eV for Cytosine in the solvent. For the A-T and C-G base pairs, our HOMO and LUMO energies differ by 4.69 (4.94) and 3.81 (4.54) eV in the gas phase (solution) which agree well with the values of Najafi et al.,^[57] that is, 4.85 (5.03) and 3.76 (4.49) eV, respectively.

A discrepancy is seen when comparing our HOMO-LUMO energy differences with those of Labet et al.^[1] We can, however, reproduce their values by using Koopmans' theorem in calculating the hardness and omitting a factor of $1/2$, although we do not know for sure whether this was the approach of Labet et al.^[1]

The fact that the HOMO and LUMO energies of the base pairs differ from those of the individual bases is due partly to electronic interactions across the hydrogen bonds and partly to the structural changes due to the hydrogen bonding linking the two bases.

TABLE 3 Various global reactivity descriptors of DNA nucleobases and pairs in eV using DFT at the B3LYP/6.31++G* level in the gas phase and without the inclusion of dispersion corrections

	Adenine	Thymine	A-T	Guanine	Cytosine	CG
χ	3.96 (3.63)	4.32 (4.25)	3.82 (3.77)	3.61 (3.97)	4.26 (3.48)	3.61 (3.55)
χ^+	6.12	6.66	5.81	5.42	6.48	5.42
χ^-	1.81	1.98	1.83	1.79	2.05	1.79
ω	1.82	1.99	1.84	1.79	2.05	1.79
η	4.31	4.68	3.98	3.62	4.43	3.62

For the electronegativity χ , we also give (in parenthesis) the negative averages of the HOMO and LUMO orbital energies.

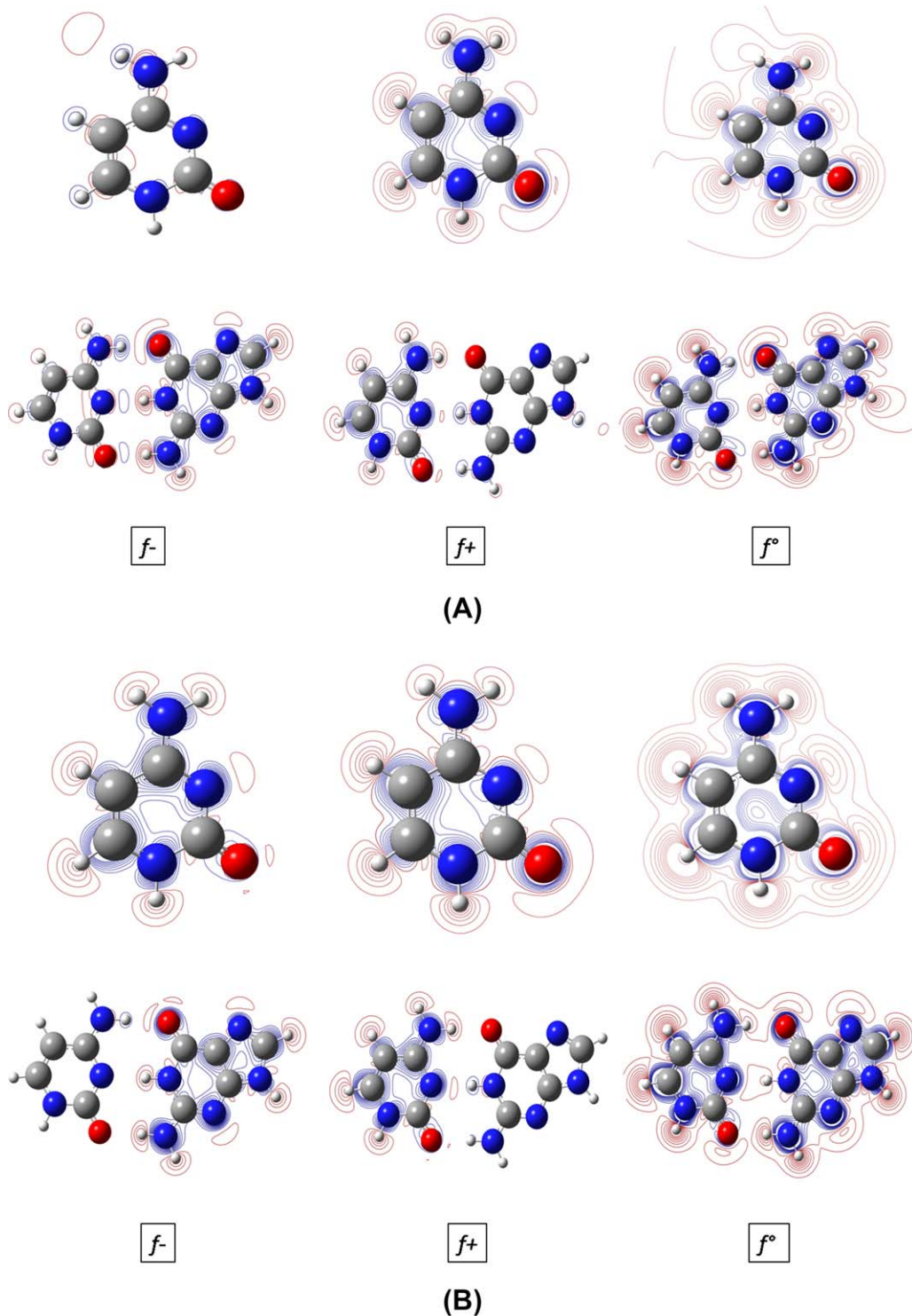


FIGURE 1 Fukui functions for Cytosine and CG in the gas phase (A) and an aqueous solution (B). The blue (red) color marks negative (positive) values. Isovalues used here are 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, and 0.008

The total-energy differences between the neutral systems and the cationic species with the same structure are seen to be close to the HOMO energies, in particular for the results for the solvated systems. Including structural relaxations does hardly change the total-energy differences. Thus, for the ionization potential, neither electronic nor structural relaxation effects are important for the present systems.

A different result is found when comparing the total-energy differences between the anionic and the neutral systems with the LUMO orbital energies. Here, quite large differences are seen, that is, the LUMO orbital energy is not a good approximation for the electron affinity. Again, this

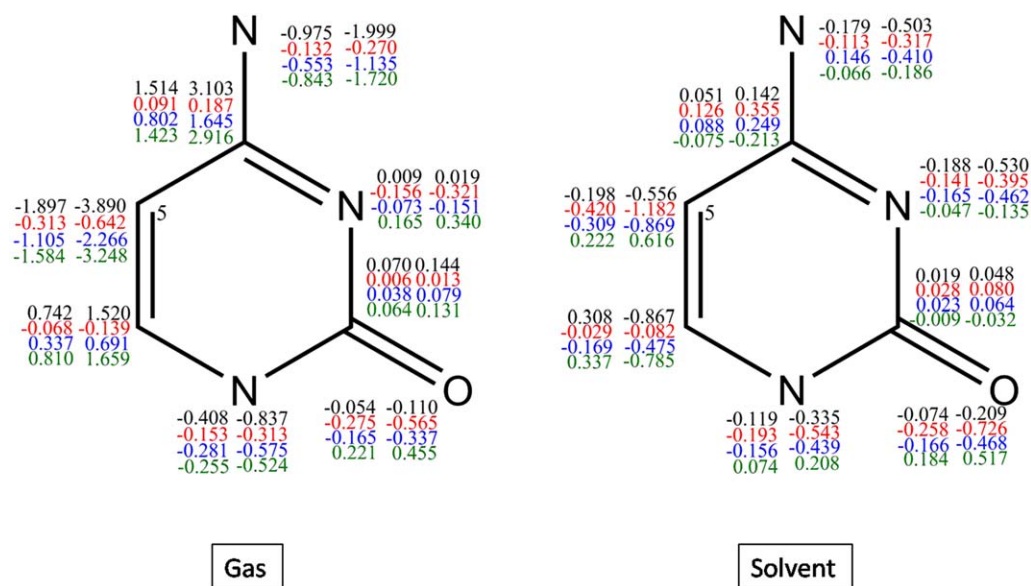


FIGURE 2 Condensed Fukui function (f_n^α , left numbers) and electrophilicity (ω_n^α , right numbers) in solvent and gas phase obtained by using Mulliken charges. In each column, from the top to the bottom the black, red, blue, and green numbers are for $\alpha = +, -, 0,$ and Δ , respectively. The contributions from the hydrogen atoms are included in those of the heavier atoms to which they are covalently bonded

founding is most pronounced for the gas-phase systems. Moreover, including structural relaxations changes the gas-phase total-energy differences markedly, although the structural changes are minor.

3.3 | Global reactivity descriptors

The HOMO-LUMO gap can give some information on the reactivity of a system but hardly on whether the system is a better electron donor or acceptor. Such information may be contained within the orbital energies of the HOMO and LUMO but, instead, it may be more useful to study the global reactivity descriptors. Our calculated values for those are given in Table 3. We found that the inclusion of dispersion corrections or solvent effects have only weak effects on the calculated properties so that we report only those found without dispersion corrections or solvent effects.

The calculated values of the electronegativity χ for the isolated bases range from 3.61 eV for Guanine to 4.32 eV for Thymine and for each base pair they are close to the lowest value of the individual bases. For the sake of comparison, we report also the average value of the negative of the HOMO and LUMO orbital energies which are seen to be quite close to the electronegativity with the Cytosine being an exception.

Our values are in good agreement with those of Najafi et al.^[57] who found a chemical potential (opposite of the electronegativity) of -3.77 eV for AT and -3.58 eV for CG.

The table shows also that Thymine has the largest value for χ^+ which indicates that this molecule is the best electron acceptor of the systems of our study followed by Cytosine. For the base paired systems, C-G is less electron accepting than A-T. Simultaneously, Cytosine has the largest value of χ^- , that is, is the best electron donor. Finally, in each case the value for a base pair is closest to the lowest values for the two isolated bases. A similar observation holds for the electrophilicity ω that possesses values very similar to those of χ^- . Since the descriptors are quantifying electronic responses it may not surprise that they for the base pairs take values that are close to the smaller values for the noninteracting bases, that is, that part of the pair that easier responds to electronic perturbations defines the response of the pair since the electronic and structural changes when forming the pair are fairly small.

Further information can be obtained from the absolute hardness (cf. Table 3). A molecule with a large value for the hardness is supposed to be more stable as proposed by Pearson.^[58] From the values in Table 3, we see that the pyrimidines, that is, Cytosine and Thymine, are the most stable ones with values 4.68 and 4.43 eV, respectively. Moreover, C-G is more reactive than A-T which confirms the idea that C-G is more active in chemical and biological reactions. Surprising is it that the hardness for the base pairs is smaller than those of the individual bases, suggesting that the base pairs are more reactive, a result that is in contrast with those found for the electronegativity-related descriptors.

When using the HOMO and LUMO orbital energies in calculating approximate values for the hardness we find values that are very close to those of Najafi et al.^[57] With their values in parenthesis, we find for A-T in gas phase 2.34 eV (2.42 eV) and in the aqueous solution 2.47 eV (2.51 eV), whereas for C-G we find 1.90 eV (1.88 eV) in the gas phase and 2.27 eV (2.24 eV) in aqueous solution.

As emphasized by Labet et al.^[1] and in agreement with our present results, pairing bases leads to significant changes in some of the global and local reactivity descriptors although the pairing involves only weaker hydrogen bonds. Conversely, the global descriptors are not able to give information on which parts of a base pair that is more reactive. To study this, we will need the local descriptors.

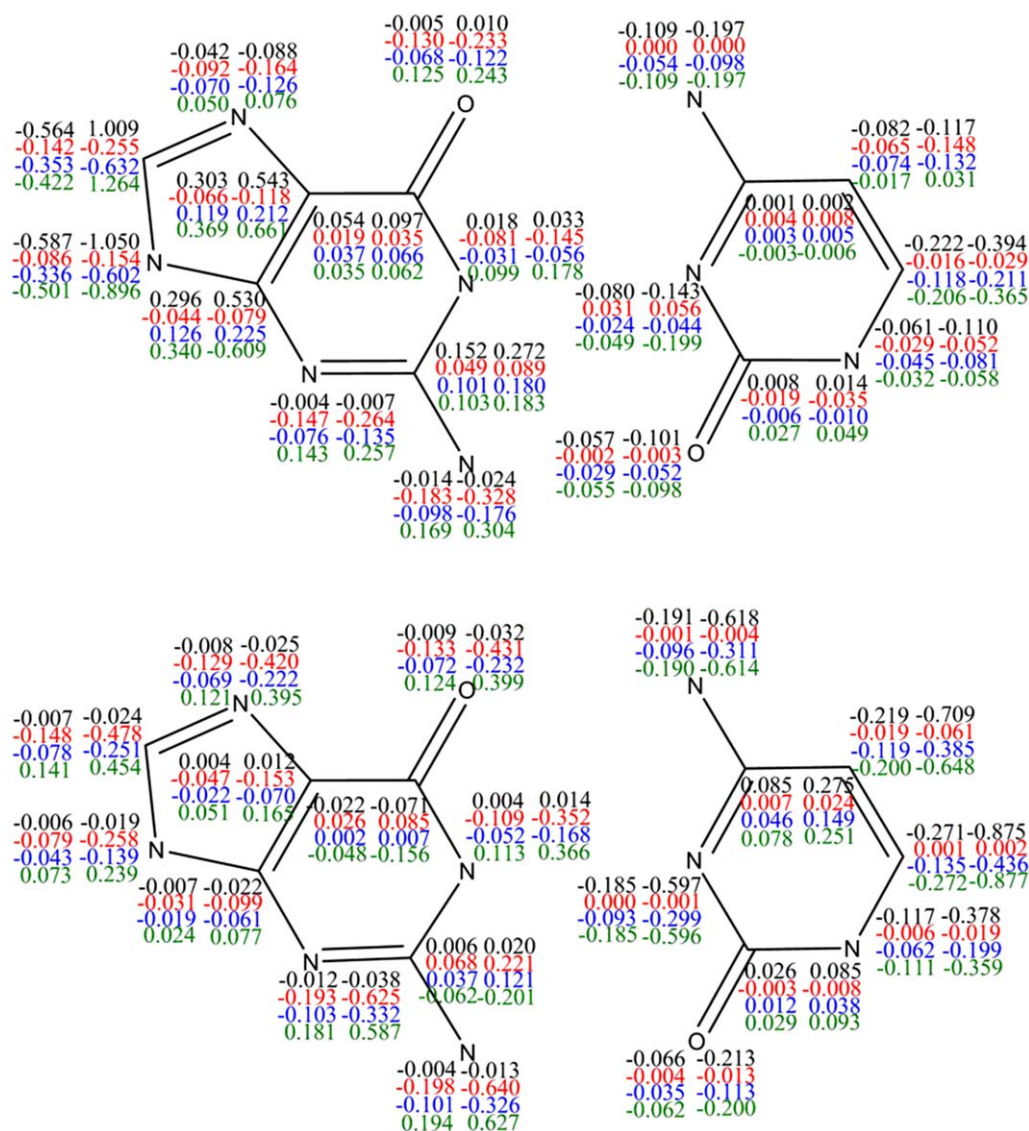


FIGURE 3 As Figure 2, but for the CG base pair in (upper part) gas phase and (lower part) aqueous solution

3.4 | Local reactivity descriptors

With the special aim of understanding the reason for the preference of adding a methyl group to the Carbon position 5 in Cytosine, we shall now study spatially resolved quantities as well as condensed functions.

At first, we show in Figure 1 the Fukui functions both for the isolated Cytosine and for the CG base pairs and for the systems both in the gas phase and in aqueous solution. For the base pairs, f^+ is localized to the Guanine base, whereas f^- is localized to Cytosine. Only for the aqueous solution, f^- resembles its analogue for the isolated Cytosine molecule.

Moreover, f^- for the isolated Cytosine differs between the gas and aqueous phases, whereas for the other cases there are only minor differences between the two phases. When focusing on the methylation, f° that quantifies a radical attack provides the most relevant information. In this case, it is seen in Figure 1 that f° for the isolated Cytosine has large contributions both for the Carbon atom at position 5 and for the neighboring C atom both for the gas and for the aqueous phase. Since, however, the neighboring C atom is not bonded to a hydrogen atom when being a part of DNA but rather to a co-enzyme during methylation process, our results show that the Fukui function is capable of identifying the Carbon atom at position 5 as the site for the methylation.

It turns out to be more useful to consider the condensed Fukui functions or electrophilicity. These are shown in Figure 2 for the isolated Cytosine base and in Figure 3 for the CG base pair. These results were obtained by using a Mulliken approach for separating the functions into atomic contributions.

In most cases, for the isolated Cytosine base, the condensed functions in the gas phase take significantly larger values than those in the aqueous solution. For the C-G base pair, the differences between the two phases are much smaller. Moreover, for the Cytosine part of the base pair, the

TABLE 4 Local reactivity descriptors for C atom at position 5 in Cytosine calculated using DFT at B3LYP/6.31++G* level for (upper part) Cytosine and (lower part) the CG base pair

	f^-	f^+	$\frac{s^+}{s^-}$	Δf
Mulliken	-0.31(0.262 ^a , 0.180 ^b) -0.42	-1.90(0.097 ^a) -0.20	6.06 0.47	-1.58(-0.164 ^a) 0.22
CHELPG	-0.31 -0.46	0.06 0.12	-0.20 0.26	0.37 0.32
Hirshfeld	-0.24 -0.19	-0.13 -0.20	1.05 0.54	0.11 -0.002
Mulliken	-0.06 -0.02	-0.08 -0.22	1.26 11.63	-0.02 -0.20
CHELPG	-0.04 -0.01	-0.06 -0.09	1.53 7.99	-0.02 -0.08
Hirshfeld	-0.03 -0.006	-0.90 -0.13	3.25 22.44	-0.06 -0.13

The second row gives the values obtained in aqueous solution.

^aReference [1].

^bReference [11]

condensed functions often have smaller values. Thus, both the aqueous solution as well as being a part of the base pair leads to a reduction of the chemical reactivity of Cytosine.

In a couple of cases, f^+ for the C atom at position 5 is larger (positive or negative) than that of most other atoms but in no case, except for the isolated Cytosine in the gas phase, it is substantially different from those of the other atoms.

Most interesting is it that f_n^+ for the Carbon at position 5 as well as for the neighboring C atom is particularly large for the C-G base pair in aqueous solution, that is, for the most realistic situation. As a consequence, also f_n^0 and f_n^Δ are particularly large for those atoms, implying that a radical attack, like that of methylene, will take place at one of those two positions. It then has to be remembered that the neighboring C atom in the DNA is not bonded to a hydrogen atom but to a co-enzyme so that an attack cannot take place at that position. This leaves us with only the C atom at position 5 as a possible position for the methylation, in agreement with experimental results.

Our results are in agreement with the results of Labet et al.^[1] who observed that the sign of the dual descriptor indicates that carbon C5 is more nucleophilic than electrophilic. Based on this finding they suggest that an electrophile will add more efficiently to carbon C5 than C6 which agrees with the fact that methylation occurs at carbon C5 of Cytosine.

Moreover, Zangi et al.^[59] emphasise that the carbon C5 of Cytosine is not a particularly strong nucleophile so that it can accept an electrophilic attack of the SAM. They suggested that some coenzymes can attack the cytosine at carbon C6 and then help the delocalization of electrons which can make carbon C5 more nucleophile and then more prone to become an electron acceptor.^[59]

Of course, using the condensed Fukui functions as calculated with a Mulliken-like method may be a strongly based approach. One may replace the Fukui function with, for example, the electrophilicity. However, as can be seen in the Figures 2 and 3, this will not change the conclusions: actually, for each system separately, the electrophilicity and the Fukui function differ by only a constant factor. Second, one may replace the Mulliken analysis through another approach of separating a spatial function into atomic contributions. In Table 4, we show the values of the condensed functions for the C atom at position 5 calculated using other approaches. We notice that the values possess a large scatter but that for the most relevant case, that is, the C-G base pair in aqueous solution, the Mulliken approach is capable of providing the largest numbers, that is, to emphasise the sensitivity of this position toward a radical attack.

4 | CONCLUSIONS

In this work, we have investigated different approaches for studying the reactivity of DNA bases and base pairs. Thereby, we focused on Methylation of Cytosine both in the gas phase and in aqueous solution and both when being isolated and when being a part of the C-G base pair. One goal of the study was to identify descriptors that could be used in identifying the chemically reactive sites, whereby we studied a well-known reaction.

A comparison of global reactivity descriptors confirmed that Cytosine is the most stable molecule, although, when linked to Guanine, it becomes more reactive and prone to chemical reactions.

Most important is, however, that we could identify the condensed Fukui function as a quantity that is capable of capturing the C atom at position 5 as the site for the methylation. Interesting is it that this result is found only when considering the most realistic model system, that is, the C-G base pair in aqueous solution. Moreover, in order to separate the Fukui function into atomic contributions, a Mulliken-based analysis was found to

be optimal. This is an important finding since it suggests that already the smaller, isolated base pairs contain the relevant information for their chemical reactivity without having to consider the true, large systems of the base pairs being embedded into DNA or interacting with solvent molecules.

ACKNOWLEDGMENT

The calculations were carried through at the C³MSaar at the University of Saarland.

ORCID

Jihène Jerbi  <http://orcid.org/0000-0002-5394-8570>

REFERENCES

- [1] V. Labet, C. Morell, V. Tognetti, O. A. Syzgantseva, L. Joubert, N. Jorge, A. Grand, J. Cadet, in *Structure, Bonding and Reactivity of Heterocyclic Compounds, Topics in Heterocyclic Chemistry*, Vol. 38 (Eds: P. DeProft, P. Greeling), Springer, Switzerland **2014**, Ch. 2, pp. 35–70.
- [2] X. Cheng, H. Hashimoto, J. R. Horton, X. Zhang, in *Handbook of Epigenetics* (Ed: T. O. Tollefsbol), Academic Press, Oxford, UK **2010**, pp. 9–24.
- [3] A. Jeltsch, *ChemBiochem* **2002**, *3*, 274.
- [4] C. R. Calladine, H. R. Drew, B. F. Luisi, A. A. Travers, in *Understanding DNA the Molecule and How it Works*, 3rd ed., Elsevier, UK **2004**, Ch. 11, pp. 270–294.
- [5] D. P. Barlow, *Science* **1993**, *260*, 309.
- [6] G. L. Verdine, *Cell* **1994**, *76*, 197.
- [7] A. P. Feinberg, B. Vogelstein, *Semin. Surg. Oncol.* **1987**, *3*, 149.
- [8] a) M. Esteller, J. G. Foncillas, E. Andion, S. N. Goodman, O. F. Hidalgo, V. Vanaclocha, S. B. Baylin, J. G. Herman, *N. Engl. J. Med.* **2000**, *343*, 1350; b) K. D. Robertson, *Oncogene* **2001**, *20*, 3139.
- [9] M. Groudine, R. Eisenman, H. Weintraub, *Nature* **1981**, *292*, 311.
- [10] W. Doerfler, *Biol. Chem. Hoppe Seyler* **1991**, *372*, 557.
- [11] J. Aranda, K. Zinovjev, K. Swiderek, M. Roca, I. Tonon, *ACS Catal.* **2016**, *6*, 3262.
- [12] D. Mishra, S. Pal, *J. Mol. Struct. THEOCHEM* **2009**, *906*, 96.
- [13] Y. Paukku, G. Hill, *J. Phys. Chem. A* **2011**, *115*, 4804.
- [14] a) J. Poater, M. Swart, C. F. Guerra, F. M. Bickelhaupt, *Comput. Theor. Chem.* **2012**, *998*, 57; b) C. F. Guerra, F. M. Bickelhaupt, J. G. Snijders, E. J. Baerends, *J. Am. Chem. Soc.* **2000**, *122*, 4117; c) I. R. Gould, P. A. Kollman, *J. Am. Chem. Soc.* **1994**, *116*, 2493; d) M. Zarei, A. Seif, K. Azizi, M. Zarei, J. Bahrami, *Int. J. Mod. Phys. C* **2016**, *27*, 1650119.
- [15] a) A. Kumar, M. K. Mohammady, P. C. Mishra, S. Suhai, *J. Comput. Chem.* **2003**, *25*, 1047; b) J. Bertran, A. Olivia, L. R. Santiago, M. Sodupe, *J. Am. Chem. Soc.* **1998**, *120*, 8159; c) N. C. Seeman, J. M. Rosenberg, F. L. Suddath, J. J. P. Kim, A. Rich, *J. Mol. Biol.* **1976**, *104*, 109.
- [16] M. Sassi, D. J. Carter, B. P. Uberuaga, C. R. Stanek, R. L. Mancera, N. A. Marks, *J. Phys. Chem. B* **2014**, *118*, 10430.
- [17] L. Pauling, *The Nature of the Chemical Bond*, 3rd ed., Cornell University Press, Ithaca, New York **1960**.
- [18] K. D. Sen, C. Jorgenson, Eds., *Structure and Bonding: Electronegativity*, Vol. 66, Springer-Verlag, Berlin **1987**.
- [19] R. G. Parr, R. A. Donnelly, M. Levy, W. E. Palke, *J. Chem. Phys.* **1978**, *68*, 3801.
- [20] K. D. Sen, D. M. P. Mingos, *Structure and Bonding: Chemical Hardness*, Vol. 80, Springer, Berlin **1993**.
- [21] R. G. Pearson, *Chemical Hardness: Application from Molecules to Solid*, Wiley –VHC, Weinheim, Germany **1997**.
- [22] R. G. Parr, L. V. Szentpál, S. Liu, *J. Am. Chem. Soc.* **1999**, *121*, 1922.
- [23] R. G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, New York **1989**.
- [24] a) P. Geerlings, F. DeProft, W. Langenaeker, *Chem. Rev.* **2003**, *103*, 1793; b) H. Chermette, *J. Comput. Chem.* **1999**, *20*, 129.
- [25] A. Martinez, *J. Phys. Chem. C* **2010**, *114*, 21240.
- [26] O. Lamarche, J. A. Platts, A. Hersey, *Phys. Chem. Chem. Phys.* **2001**, *3*, 2747.
- [27] P. K. Chattaraj, D. R. Roy, M. Elango, V. Subramanian, *J. Phys. Chem. A* **2005**, *109*, 9590.
- [28] L. H. M. Huizarand, C. H. R. Reyes, *J. Mex. Chem. Soc.* **2011**, *55*, 142.
- [29] R. S. Mulliken, *J. Chem. Phys.* **1955**, *23*, 1833.
- [30] J. P. Ritchie, S. M. Bachrach, *J. Comput. Chem.* **1987**, *8*, 499.
- [31] A. V. Marenich, S. V. Jerome, C. J. Cramer, D. G. Truhlar, *J. Chem. Theory Comput.* **2012**, *8*, 527.
- [32] C. M. Breneman, K. B. Wiberg, *J. Comput. Chem.* **1990**, *11*, 361.
- [33] K. Fukui, *Theory of Orientation and Stereoselection*, Springer, Berlin **1973**.
- [34] K. Fukui, *Science* **1982**, *217*, 747.
- [35] a) W. Yang, W. J. Mortier, *J. Am. Chem. Soc.* **1986**, *108*, 5708; b) R. G. Parr, W. Yang, *J. Am. Chem. Soc.* **1984**, *106*, 4049.

- [36] P. Fuentealba, C. Cardenas, *Density Functional Theory of Chemical Reactivity*. In *Chemical Modeling: Volume 11*. Roy. Soc. Chem. **2005**, 11, 151.
- [37] P. W. Ayers, W. Yang, L. J. Bartolotti, in *Chemical Reactivity Theory: A Density Functional* (Ed: P. K. Chattaraj), CRC Press, Boca Raton, Florida **2009**, Ch. 18, pp. 255–267.
- [38] W. Yang, R. G. Parr, *Proc. Natl. Acad. Sci. USA* **1985**, 82, 6723.
- [39] M. Galvan, A. DalPino, J. Wang, J. D. Joannopoulos, *J. Phys. Chem.* **1993**, 97, 783.
- [40] R. K. Roy, S. Krishnamurti, P. Geerlings, S. Pal, *J. Phys. Chem. A* **1998**, 102, 3746.
- [41] P. K. Chattaraj, B. Maiti, U. Sakar, *J. Phys. Chem. A* **2003**, 107, 4973.
- [42] P. K. Chattaraj, *J. Phys. Chem. A* **2001**, 105, 511.
- [43] a) C. Morell, A. Grand, A. Toro-Labbé, *J. Phys. Chem. A* **2005**, 109, 205; b) C. Morell, A. Grand, A. Toro-Labbé, *Chem. Phys. Lett.* **2006**, 425, 342.
- [44] a) A. D. Becke, *Phys. Rev. A* **1988**, 38, 3098; b) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, 37, 785.
- [45] a) W. Kohn, A. D. Becke, R. G. Parr, *J. Phys. Chem.* **1996**, 100, 12974; b) W. Kohn, L. J. Sham, *Phys. Rev. A* **1965**, 140, 1133.
- [46] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, *Gaussian 09, Revision A.02*, Gaussian Inc., Wallingford, CT **2009**.
- [47] P. Hobza, J. Šponer, *Chem. Rev.* **1999**, 99, 3247.
- [48] a) S. Grimme, *WIREs Comput. Mol. Sci.* **2011**, 1, 211; b) S. J. Grimme, *Comput. Chem.* **2004**, 25, 1463.
- [49] G. Scalmani, M. J. Frisch, *J. Chem. Phys.* **2010**, 132, 114110.
- [50] a) N. Kurita, V. I. Danilov, V. M. Anisimov, *Chem. Phys. Lett.* **2005**, 404, 164; b) P. Hobza, J. Sponer, *J. Am. Chem. Soc.* **2002**, 124, 11802; c) J. Sponer, J. Leszczynski, P. Hopza, *Biopolymers (Nucleic Acid Sciences)* **2002**, 61, 3.
- [51] J. Jerbi, M. Springborg, *J. Comput. Chem.* **2017**, 38, 1049.
- [52] P. Senet, in *Chemical Reactivity Theory: A Density Functional* (Ed: P. K. Chattaraj), CRC Press, Boca Raton, Florida **2009**, Ch. 24, pp. 331–361.
- [53] S. Seisenberger, J. R. Peat, T. A. Hore, F. Santos, W. Dean, W. Reik, *Phil. Trans. R. Soc. B* **2012**, 368, 20110330.
- [54] J. A. Hackett, J. J. Zylicz, M. A. Surani, *Trends Genet* **2012**, 28, 164.
- [55] A. Bondi, *J. Phys. Chem.* **1964**, 68, 441.
- [56] B. Pullman, and A. Pullman, *Proc. Natl. Acad. Sci. USA* **1958**, 44, 1197.
- [57] D. Farmanzadeh, M. Najafi, *Struct. Chem.* **2015**, 26, 831.
- [58] R. G. Pearson, *Acc. Chem. Res.* **1993**, 26, 250.
- [59] R. Zangi, A. Arrieta, F. P. Cossío, *J. Mol. Biol.* **2010**, 400, 632.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Jerbi J, Springborg M. Reactivity descriptors for DNA bases and the methylation of cytosine. *Int J Quantum Chem.* 2017;e25538. <https://doi.org/10.1002/qua.25538>

Supporting information for
Reactivity Descriptors for DNA bases and the Methylation of Cytosine

Jihène Jerbi* and Michael Springborg*⁺

**Physical and Theoretical Chemistry, Saarland University, 66123 Saarbrücken, Germany*

⁺Tianjin University, Tianjin 300072, P. R. China

jihene.jerbi@uni-saarland.de

m.springborg@mx.uni-saarland.de

Table of contents:

Computation information and results of A-T base pair	1-2
Table 1	3
Scheme 1	4
References	5

In the supporting information we present the structure of the DNA base pair A-T, at the B3LYP/6-31++G* level of theory in gas phase and in solvent.

The structures of the isolated molecules and the A-T base pair were found to be planar.

The hydrogen bonds connecting the A-T base pair can be identified in Scheme 1 of the main paper.

For these, the total energy changes the least when varying their lengths and therefore their lengths are connected with the largest uncertainties. In Table 1, our results for the hydrogen bonding are collected. One can see that dispersion corrections do hardly affect the results in gas phase with few exceptions including, for instance, the N—N distances. All our results are in line with published works, including that the N—O inter-atomic distance between the proton acceptor and donor is longer than the equivalent N—N distance, i.e., 2.95 and 2.91 Å, respectively. These values compare well with the experimental values of 2.93 and 2.85 Å. However, in aqueous solution the N—N inter-atomic distance becomes larger than the N—O distance, i.e., 2.94 and 2.96 Å, respectively, whereas the inclusion of dispersion corrections gives the same relative order as found in the gas phase. In all cases, the inter-atomic distances

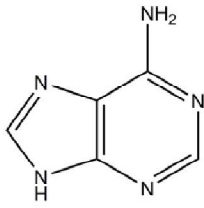
for the proton donor-acceptor pairs are found to be larger in the solution than in the gas phase which is understandable since it is known that the solvent stabilises the lone pairs that subsequently leads to weaker hydrogen bonds.

In Table1, results for the relative energy for the A-T base pair compared with the sum of the energies of the isolated, non-interacting bases are summarised. That the base pairs are more stable in gas phase than in the solvent is in agreement with the structural properties that suggest that the hydrogen bonds are weaker in the solvent. The difference between the gas phase energy and solvent energy is 9.2 kcal/mol, which agrees with the results of Poater et al. who found -17.9 kcal/mol in gas [1a] and -9.1 kcal/mol in solvent [1d]. When adding the correction due to basis-set superposition error (BSSE) these energies change by approximately 1 kcal/mol. Finally, we observe that the inclusion of dispersion corrections reduces those values by 4.2 kcal/mol both in gas and in solvent phases both without and with the inclusion of the BSSE correction.

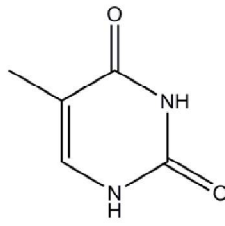
Table1: Structural and energetic properties of the proton donor-acceptor A-T pair in gas and solvent phase as found at the B3LYP/6-31++G* level. N-O and N-N give the interatomic distances in Å, whereas ΔE_{BSSSE} and ΔE give the total energy in kcal/mol of the base pair compared with the sum of those of the non-interacting bases without and with correcting for the basis-set superposition error

	N---O	N----N	ΔE_{BSSSE}	ΔE
<u>This work:</u>				
B3LYP/6-31++G*	2.95	2.91	-16.8	-15.8
B3LYP+D3/6-31++G*	2.93	2.85	-12.6	-11.7
B3LYP/6-31++G* (water)	2.94	2.96	-7.6	-6.6
B3LYP+D3/6-31++G*(water)	2.93	2.89	-11.6	-10.7
<u>Other theoretical work:</u>				
B3LYP/DZP++^a	2.91	2.85		
B3LYP/6-31G**^b	2.94	2.84		-12.3
BP86/TZ2P^c	2.85	2.81	-17.9	
BLYP/TZ2P^c	2.84	2.78	-14.5	-13.7
B3LYP/6-31++G**^d	2.95	2.88		
B3LYP/6-31++G** (water)^d	2.96	2.92	-9.1	
BP86/DZP^c			-15.8	-12.7
MP2/DZP//HF/6-31G**^e				-14.7
Experiment:^e	2.93/ 2.95	2.85/ 2.82		

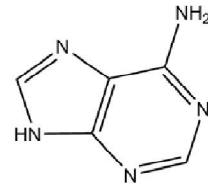
From references: ^a 2a, ^b 2b, ^c 1b, ^d 1d, ^e 1c, ^e 2c



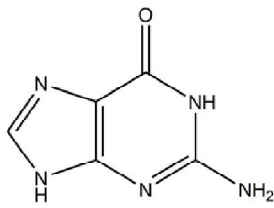
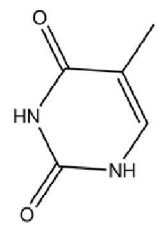
Adenine



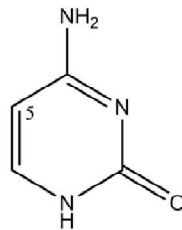
Thymine



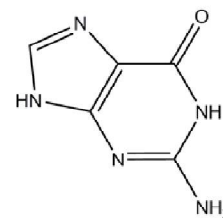
A-T



Guanine



Cytosine



G-C

Scheme 1: Isolated and pairs of DNA bases. Left most is shown the purines (Adenine and Guanine), in the middle the pyrimidines (Thymine and Cytosine), and to the right the base pairs. The Carbon atom at position 5 in Cytosine is explicitly shown.

References:

- [1] (a) Poater, J., Swart, M., Guerra, C.F. and Bickelhaupt, F.M. *Comp. Theo. Chem.* 2012, 998, 57-63.
- (b) Guerra, C.F., Bickelhaupt, F.M., Snijders, J.G. and Baerends, E.J. *J. Am. Chem. Soc.* 2000, 122, 4117-4128.
- (c) Gould, I.R. and Kollman, P.A. *J. Am. Chem. Soc.* 1994, 116, 2493-2499.
- (d) Zarei, M., Seif, A., Azizi, K., Zarei, M. and Bahrami, J. *Int. J. Mod. Phys. C* 2016, 27, 1650119-1650131.
- [2] (a) Kumar, A., Mohammady, M.K., Mishra, P.C. and Suhai, S. *J. Comp. Chem.* 2003, 25, 1047-1059.
- (b) Bertran, J., Olivia, A., Santiago, L.R. and Sodupe, M. *J. Am. Chem. Soc.* 1998, 120, 8159-8167.
- (c) Seeman, N.C., Rosenberg, J.M., Suddath, F.L., Kim, J.J.P. and Rich, A. *J. Molec. Biol.* 1976, 104, 109-145.



Research paper

S-adenosyl-L-methionine analogs as enhanced methyl donors: Towards novel epigenetic regulators

Jihène Jerbi^{a,*}, Michael Springborg^{a,b}, Helena den-Haan^c, José P. Cerón-Carrasco^{c,*}^a Physical and Theoretical Chemistry, Saarland University, Saarbrücken 66123, Germany^b School of Materials Science and Engineering, Tianjin University, Tianjin 300072, China^c Bioinformatics and High Performance Computing Research Group (BIO-HPC), Computer Engineering Department, Universidad Católica San Antonio de Murcia (UCAM), Campus de los Jerónimos, 30107 Murcia, Spain

ARTICLE INFO

Article history:

Received 1 September 2017

In final form 22 October 2017

Available online 24 October 2017

Keywords:

DNA

Mutations

Methylation

Molecular dynamics

Ab initio

ABSTRACT

Many efforts have been devoted to discover molecules able to halt methylation processes in DNA. However, less is known about the application of methyl promoters in the framework of hypomethylation diseases. Herein, we used molecular dynamics and *ab initio* calculations to assess the methylation ability of the parent S-adenosyl-L-methionine cofactor (SAM) and a series of analogues. Two molecules deposited in the PubChem database are shown to be promising candidates for increasing the methyl transfer rate of the original SAM. The reported data might be consequently used to guide further steps into the search of more efficient methyl donor-based drugs.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Most of the biochemical reactions that govern the cellular machinery are driven by DNA, which holds the genetic information encoded as a unique sequence of adenine (A), thymine (T), guanine (G) and cytosine (C) bases. According to the model originally proposed by Watson and Crick more than sixty years ago, the normal cellular behavior is thus guaranteed as long as the nucleobases are present as the canonical AT and GC base pairs [1]. However, the functionality of DNA might be drastically altered if other forms are introduced into the double helix architecture.

One of the possible sources of non-canonical structures arises from the methylation of DNA, a process in which a methyl group is added at the C5 position of the cytosine [2,3]. This reaction is regulated by DNA methyltransferases (DNMTs), the family of enzymes responsible of controlling most of the methylation equilibria in cells [4]. As shown in Fig. 1, DNMTs use S-adenosyl-L-methionine (SAM) as methyl donor factor in a four steps mechanism: (1) the process is initiated by the deprotonation of a cysteine located in the binding site (Cys81); (2) that residue subsequently undergoes the nucleophilic addition at the C6 position of the flipped-out cytosine [5]; (3) the resulting activated-cytosine yields

to a partial-negatively charged C5 that next reacts with SAM by transferring a methyl group towards DNA [6] (4); finally, the proton at C5 position is released in the β -elimination step to a surrounding base, e.g., hydroxide anion. Accordingly, the reaction finally leads to the C5-methylcytosine product (5mC) and the byproduct S-adenosyl-L-homocysteine (SAH) [7]. Although the complete picture of DNMTs activity is more complex than the simplified Fig. 1 [8], it is now well established that methyl transfer of step (3) is the rate-limiting reaction [9].

The methylation mechanism described above is naturally present in the cellular machinery at a low rate (less than 5% of the total bases in human DNA as 5mC forms) [10]. However, not only the number of 5mC forms but also the specific location of methylation events play a pivotal role in the regulation of gene expression, embryonic development, X-chromosome inactivation and genomic imprinting [11]. Indeed, 5mC is one of the mayor epigenetic modification of DNA in mammals [12] as that long-term genomic mark is maintained after DNA replication [13]. Unfortunately, the biological mechanisms that control the unmethylated/methylated rate in DNA are not flawless and the amount of 5mC might be out of the natural range. This aberrant regulation of 5mC rates (in both hypo- and hypermethylation senses) has been associated with severe human disorders, including cancer [14]. On the one side, it is known that cancer cell losses part of their methylated bases compared to its undamaged counterpart, which yields to an unstable hypomethylated genome [15]. On the other

* Corresponding authors.

E-mail addresses: jihene.jerbi@uni-saarland.de (J. Jerbi), jpceron@ucam.edu (J.P. Cerón-Carrasco).

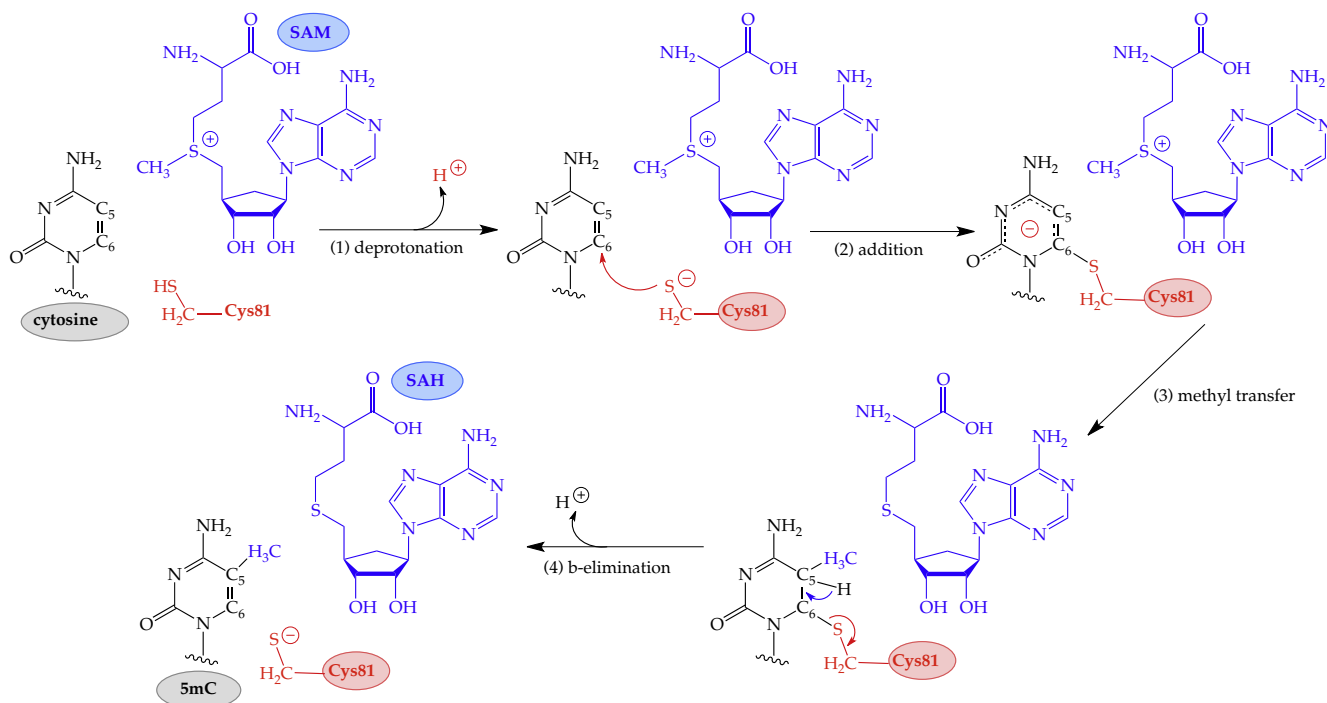


Fig. 1. Overview of the methylation mechanism of DNA catalyzed by DNMT. As sketched, the cytosine base and SAM are transformed into the methylated 5mC base and the SAH byproduct.

side, 5mC spontaneously mutates to thymine through a deamination reaction, which introduces a wrong sequence into DNA due to a complete C \rightarrow T exchange [16].

Recent experimental efforts to characterize the kinetic of the methylation reaction have been focused on determining the crystallographic structure of the DNMT enzymes and the enzyme-substrate interaction (see Fig. 2) [17]. Computational modeling, however, has also provided new meaningful insights into the chemical mechanisms behind the C + SAM \leftrightarrow 5mC + SAH methyl transfer equilibrium. An earlier study by Schramm and co-workers used a joint kinetic isotope effects (KIEs) and density functional theory (DFT) approach to assess the nature of the transition state (TS) in the human DNMT [18]. As confirmed by these authors, the catalyzed methyl transfer is associated to the highest energetic barrier along with all the process. The performed DFT calculations also allow to capture the changes in the bonding environment during the reaction. The located TS geometry shows a nearly complete covalent bond between the sulfur of close cysteine residue and the carbon-6, which fully supports the stepwise mechanism where the Cys81 attack (referred in that publication as Cys1226 residue) occurs prior to methyl transfer DNMT catalysis [18]. Aranda and co-workers have recently contributed to the field by applying molecular dynamics (MD) and hybrid quantum mechanics:molecular mechanics (QM:MM) calculations [19]. These simulations complete the picture of the reaction, and demonstrated that cysteine becomes unprotonated before attacking upon SAM, an equilibrium that might be mediated by DNA phosphate groups [19]. Other theoretical works have gone beyond ground state by exploring electronic transition upon UV–vis light radiation. In that framework, Improtá and co-workers have studied the stability of DNA under the exposure of highly energetic radiation by using time dependent-DFT (TD-DFT) approaches [20]. According to their calculations, the extra methyl at C5 position of cytosine results in a strong perturbation of the conformational and electronic structure of DNA, which eventually make 5mC more prone to undergo photo-damage [21–23].

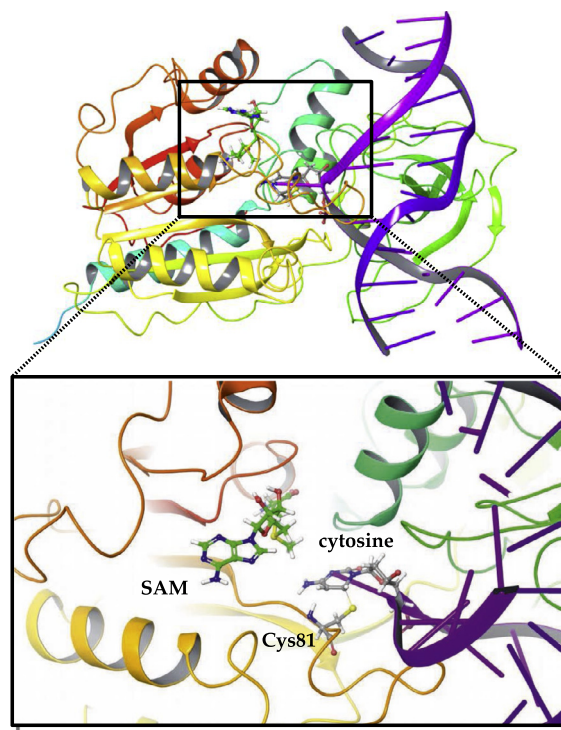


Fig. 2. Up panel: 2HR1 structure. SAM, Cys81 and residue and flipped-out cytosine are displayed as ball&sticks; DNA is plotted as purple ribbons while the DNMT enzyme is represented as colored ribbons. Low panel: zoom into the region of interest, which illustrate the relative position of the $-S(CH_3)$ respect the cytosine and Cys81 residue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In addition to all these mechanistic studies, theoretical tools have been also used to screen possible DNMT inhibitors that might

have antiproliferative activity against tumor cells [24]. More specifically, docking-based strategies have been extensively used to identify novel molecules able to block the Cys81-binding pocket of DNMT and consequently halt the methyl transference reaction [25–29]. On the contrary, recent *in vitro* studies have demonstrated that **SAM** can inhibit cancer cells' progression [30,31]. Although the debate is still opened, it has been hypothesized that the remethylation of proto-oncogene promoters might positively impact into the treatment of prostate cancer [32,33]. These studies suggest that DNA hypomethylation controls the activation of tumor and consequently **SAM** related methyl-donors molecules might provide valuable insights into designing enhanced therapeutic strategies against cancer [34]. To the best of our knowledge, the optimization of the original **SAM** structure to enhance the methyl transference has been never reported in the literature. Aiming to fill this gap in the literature, herein we perform MD and DFT calculations to screen the well-known PubChem chemical library [35] to search novel methylation promoters. Our calculations are a first step towards new molecules that, if selectively located in the damaged tissue, might be used to promote DNA-methylation with a higher degree than the parent **SAM**, and consequently can be used to shift down cancer-gene expression.

2. Computational details

We used a joint MD and DFT approach to guide the search of novel methyl transfer groups. More specifically, the selected MD protocol was adapted from the approach recently proposed by Aranda and co-workers [19]. Accordingly, the model system was built up from the experimental structure of the **SAH** embedded in the *Haemophilus hemolyticus* DNMT C5-MTase (*M.HhaI*). This structure shows a high homology to mammalian enzymes [36] and has been deposited in the Protein Data Bank at 1.96 Å of resolution (code 2HR1) [37]. Since there is not available a crystal structure for **SAM**-DNA, the coordinates of 2HR1 offer an alternative for designing an starting model. At an early stage, the original **SAH** ligand was converted into **SAM** by adding a methyl group at the sulfonium center. The modified 2HR1 structure is embedded in a solution model with a buffer distance of 10 Å in all directions, and that box was subsequently filled with water molecules SPC model and sodium cations in order to keep the system electronically balanced. Additional sodium and chloride ions were incorporated into the system to simulate the physiological salt concentration of 0.15 M NaCl. The resulting model contains ca. 45,000 atoms. The initial MD included a full minimization of 2000 steps using the steepest descent method, with a convergence threshold of 1.0 kcal mol⁻¹/Å, followed by several relaxation steps that include a solute-restrained minimization, free-restrain minimization, NVT simulation of 24 ps at T = 10 K, and NPT simulations at T = 10 K and P = 1 atm. For the MD production phase, the temperature was set to 300 K with the Nosé-Hoover algorithm, with a relaxation time of 1.0 ps [38,39]. Pressure was controlled to 1 bar with the Martyna-Tobias-Klein barostat using isotropic coupling and a relaxation time of 2.0 ps [40]. The RESPA integrator was used to integrate the motion equations with a 2.0 fs time step for bonded and near interactions and a 6.0 fs time step for far interactions [41]. A cut-off of 9 Å was applied to non-bonded interactions. Van der Waals interactions were evaluated using a cut-off radius 9 Å, and the electrostatic part was computed using the Particle Mesh Ewald (PME) method with tolerance value of 10⁻⁹ [42]. A first MD simulation is run for 75 ns using the recently developed OPLS3 force field as implemented in Desmond code [43,44], which is used to parametrize all entities including DNA, protein and ligands. The RMSD trend confirms that such trajectory is large enough to reach a dynamically stable structure, and consequently

the last snapshot is taken as initial structure for a second trajectory. The structure of the first trajectory is subsequently remodeled by binding the Cys81 to the DNA-cytosine moiety at C6 position, which corresponds to the addition step as shown in Fig. 1. We should stress here that our aim is the search of novel methyl transfer drugs and consequently we tackle the most critical step, namely the transference of the methyl from **SAM** to cytosine. That second trajectory is run over 75 ns, so that the total simulation time is 150 ns long. The **SAM** ligand, the cytosine entity and the Cys81 residue were eventually extracted from the MD trajectory to be further refined at a DFT level to determine the energetic barrier during the methyl transference. The reactant, TS and product are fully optimized at the M06-2X/6-311+G(d,p) level of theory [45], which has been previously shown to accurately mimic such reaction in DNA [19,18]. The vibrational analysis confirms the nature of all optimized structures as a true minimum in the potential energy surface of the stationary point (no imaginary frequencies) or a TS (one imaginary frequency associated to the normal mode of the methyl transfer) through. All DFT calculations were carried out with Gaussian09 Rev.E01 [46]. The computational protocol is eventually completed through docking calculations to ensure that novel methyl transfer drugs fit into the binding site. To this end, a grid size set of 25³ Å is centered at the cocrystallized SAH ligand resolved in the experimental PDB structure. The best ranked drugs are subsequently docked into that grid by using defaults parameters implemented in Leadfinder [47]. The two best ranked poses of each ligand were collected for visual inspection.

3. Results and discussion

Our first goal is to set up the configuration of the **SAM** within the binding site prior to the methylation step (see top right panel of Fig. 1). We assessed the reliability of the designed model to mimic the structure of the **SAM**-DNA-DNMT ensemble. As illustrated in Fig. 3, the evolution of the root-mean-square-deviation (RMSD) on all residues [black line] quickly reaches a stable value of ca. 2 Å during the early 0–75 ns of the MD trajectory. That stability suggests that the **SAH** → **SAM** replacement does not produce a significant impact in the overall structure of the system. However, a closer inspection of the flipped-out cytosine reveals that the Cys81–(C5) cytosine distance [red line] is increased by 1 Å after 20 ns of trajectory (see Fig. 1 for atomic numbering). The AMBER simulations performed by Aranda and co-workers showed a similar trend, with an average distance of 3.36 Å that is cyclically increased up to 5 Å [19]. The observed lengthening therefore seems not to depend on the particular force field but being an intrinsic feature of the system. The **SAM**(CH₃)–(C6) cytosine distance (a critical parameter for the methylation process) was also monitored during the MD simulation [blue line]. The methyl group is found to be orientated towards the C5(cytosine) atom at an average distance of ca. 4 Å along all the trajectory, which is compatible with the DNA methylation. The stability of the binding site is further confirmed by measuring the Arg165(NH₂)–(C=O) cytosine distance [green line], which remains stable during the whole trajectory with an average value of ca 2 Å. All these outcomes validate the choice of the X-ray structure deposited with PDB code 2HR1 as starting geometry for the DNMT-embedded **SAM** structure. The resulting system was next remodeled by deprotonation the Cys81 residue and induce a covalent bond with the C6(cytosine) position. As expected, that structural change is associated to an increase of the RMSD value. Interestingly, the RMSD reaches a value of 4 Å, that is, both the **SAH** → **SAM** replacement and the covalent Cys81–(C5) cytosine result in an accumulated RMSD of ca. 2 Å. In the last 75–150 ns window, the computed Cys81–(C5) cytosine, **SAM**(CH₃)–(C6) cytosine and Arg165(NH₂)–(C=O)

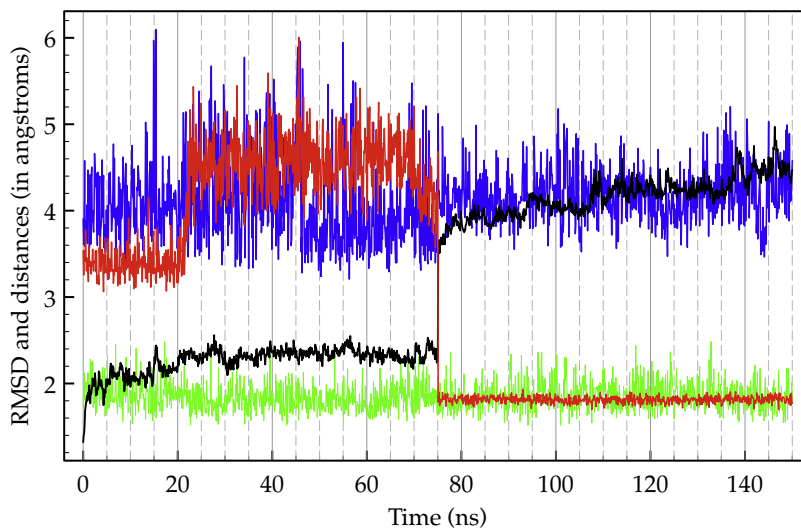


Fig. 3. Summary of the molecular dynamics trajectory after the two simulations of 75 ns (150 ns in total). The plots correspond to the evolution of the root-mean-square (RMSD) on all residues [black line] and the Cys81-(C5) cytosine [red line], SAM(CH₃)-(C6) cytosine [blue line] and Arg165(NH₂)-(C=O) cytosine [green line] distances in angstroms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cytosine distances are fully consistent with the simulation performed by [19].

Aiming to ascertain the impact of sulfur substituents in the methylation process, the chemical region of interest, that is, SAM drug, cytosine nucleobase and Cys81 residue, is extracted from the last snapshot of the MD trajectory. That truncated model is subsequently treated at the M06-2X/6-311+G(d,p) level of theory. It should be underlined that other residues might also impact into the absolute energies, e.g., Glu119, Arg163, and Arg165 [5,19]. The theoretical treatment of such larger model systems would require the use of more elaborated computational models as hybrid MD/QM:MM schemes, while herein the truncated model is used to compute relative rate constants. In our approach, reactives, products and TSs associated to the methyl transfer process (step 3 in Fig. 1) are fully optimized. Since the sulfonium center is a key moiety for transmethylation, we decided to screen the PubChem chemical library of compounds [35] by imposing the presence of a R-S⁺(CH₃)-R' fragment in the molecule. The resulting molecules from the screening are listed in Table 1 and Fig. 4.

All computed activation and reaction free energies (ΔG^\ddagger and ΔG in kcal mol⁻¹), forward and reverse rate constants (k_f and k_r in s⁻¹)

and equilibrium constants (K_{eq}) are given in Table 2. The distances of the transferred methyl group to its donor (SAM or an analogue) and to its acceptor (cytosine) are listed in Table 2. Let us consider first the results obtained with the original SAM molecule. The energy barrier associated to the methyl transfer from SAM to cytosine is predicted at $\Delta G^\ddagger = 13.54$ kcal mol⁻¹, a value that nicely matches the energy reported by Aranda et al. ($\Delta G^\ddagger = 16.8 \pm 0.4$ kcal mol⁻¹), who used the more refined and computational demanding MD/QM:MM sampling in the whole system. The obtained value also agrees with previous theoretical works. For instance, [18] used a DFT approach to scan the distances along with the methyl transference. The computed potential energy curve presents a barrier of $\Delta G^\ddagger = 10$ kcal mol⁻¹. The MD/QM:MM simulations conducted by [11] conducted MD/QM:MM simulation leads to $\Delta G^\ddagger = 15.8$ kcal mol⁻¹, and finally the calculations by [5] reported a barrier of $\Delta G^\ddagger = 18.6$ kcal mol⁻¹. An inspection of the optimized structures confirms that the full optimization of the truncated model does not yield to an over-distorted geometry. On the contrary, the orientation of reactants is preserved so that transference of the methyl group might take place in an efficient way. More specifically, the reactant has the methyl group of SAM

Table 1

Computed activation and reactions free energies (ΔG^\ddagger and ΔG , respectively, in kcal mol⁻¹), forward and reverse rate constants (k_f and k_r in s⁻¹), and equilibrium constants (K_{eq}) for methyl transference reaction. The unmethylated Cytosine is taken as reference to compare the free energies.

Molecule	Pubchem	ΔG^\ddagger	ΔG	k_f	k_r	K_{eq}
SAM		13.54	-46.14	7.27×10^{02}	1.04×10^{-31}	6.97×10^{33}
I	000161187	15.63	-47.55	2.13×10^{01}	2.83×10^{-34}	7.55×10^{34}
II	000446208	16.19	-48.19	8.28×10^{00}	3.72×10^{-35}	2.22×10^{35}
III	000447069	25.93	-38.84	5.94×10^{-07}	1.93×10^{-35}	3.08×10^{28}
IV	005327054	26.69	-42.21	1.65×10^{-07}	1.80×10^{-38}	9.14×10^{30}
V	025231313	20.64	-43.91	4.51×10^{-03}	2.79×10^{-35}	1.61×10^{32}
VI	118725204	12.36	-53.84	5.34×10^{03}	1.72×10^{-36}	3.10×10^{39}
VII	118725206	26.06	-39.57	4.77×10^{-07}	4.51×10^{-36}	1.06×10^{29}
VIII	118725209	12.15	-52.48	7.61×10^{03}	2.44×10^{-35}	3.12×10^{38}
IX	118725211	17.71	-46.77	6.35×10^{-01}	3.14×10^{-35}	2.02×10^{34}
X	118725215	14.31	-50.01	1.98×10^{02}	4.12×10^{-35}	4.81×10^{36}
XI	118725217	20.00	-43.88	1.33×10^{-02}	8.66×10^{-35}	1.53×10^{32}
XII	118725236	41.26	-27.27	3.38×10^{-18}	3.36×10^{-38}	1.01×10^{20}
XIII	118856003	18.24	-44.26	2.60×10^{-01}	8.91×10^{-34}	2.91×10^{32}

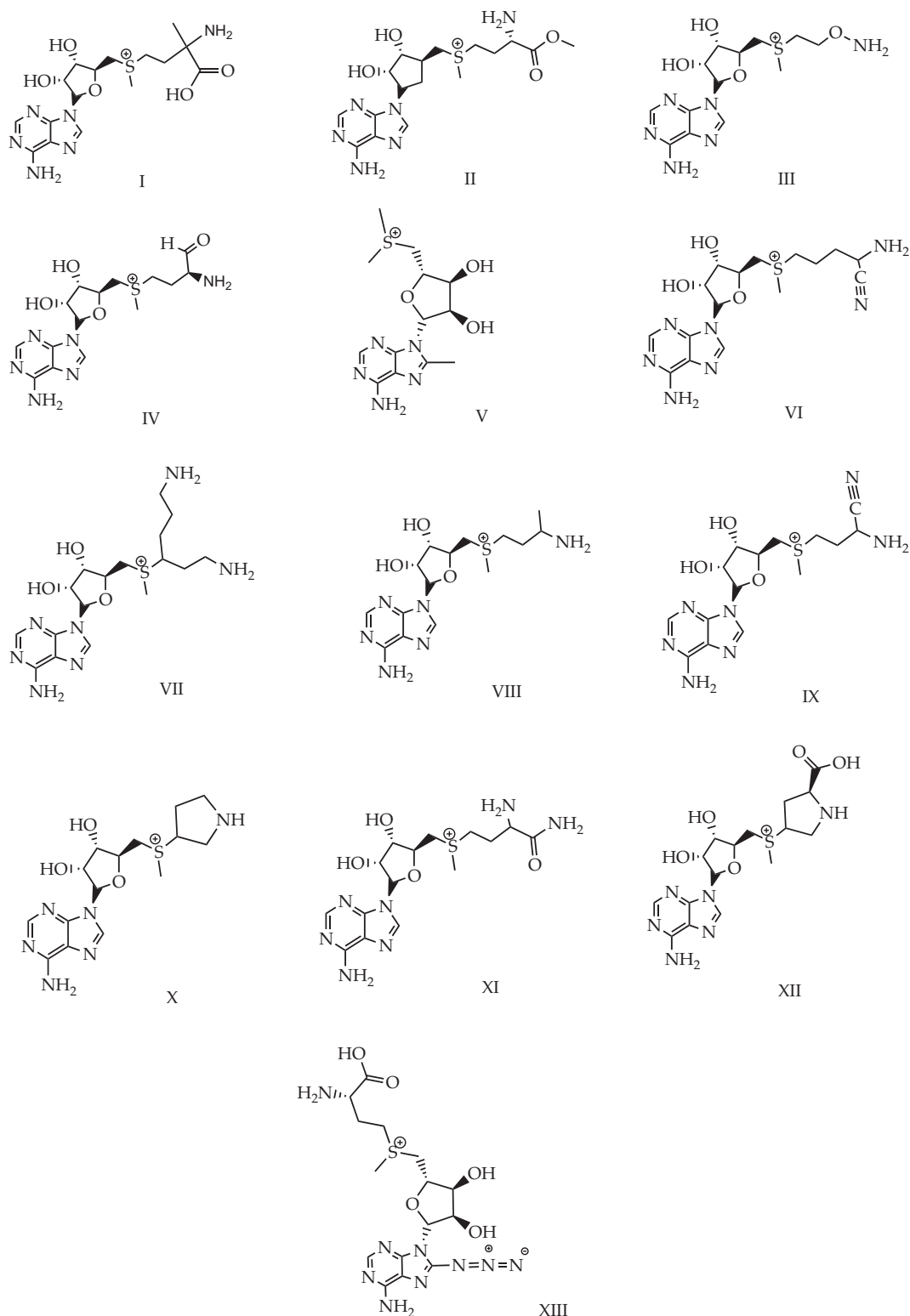


Fig. 4. Chemical structures of the molecules containing the $R-S^+(CH_3)-R'$ core in the PubChem library.

placed 1.81 Å from the C_5 carbon atom and at 3.45 Å of C_5 of cytosine, which are consistent to the average distances derived from the MD/QM:MM trajectory, 1.81 Å and 2.93 Å respectively [19]. The accuracy of the optimized TS geometries is also confirmed for the TS. As one can see in Table 2, the distances from the methyl to SAM and cytosine are 2.19 Å and 2.34 Å respectively, in agreement with distances reported by Du et al., 2.10 and 2.21 Å, respec-

tively. We should however note that a measurable discrepancy is observed in ΔG . The use of MD/QM:MM simulations yields to $\Delta G = -27.4 \text{ kcal mol}^{-1}$ while our DFT approach significantly overestimate the stability of the methylated product, with a computed $\Delta G = -46.14 \text{ kcal mol}^{-1}$. This is the consequence of the lack of surrounding residues in the truncated mode. The better rearrangement of product of the methylation allows for the final product

Table 2

Computed distances between the transferred methyl group and its donor (**SAM** or **SAM** analogue) and its acceptor (cytosine) in Å. React, TS and Prod stand for reactant, transition state and product geometry.

Molecule	React		TS		Prod	
	S-CH ₃	CH ₃ -C6	S-CH ₃	CH ₃ -C6	S-CH ₃	CH ₃ -C6
SAM	1.81	3.45	2.19	2.34	3.86	1.54
I	1.81	3.26	2.19	2.34	4.04	1.54
II	1.80	4.98	2.17	2.36	3.83	1.54
III	1.81	4.01	2.16	2.37	3.84	1.54
IV	1.81	5.51	2.15	2.36	3.71	1.54
V	1.80	4.41	2.16	2.31	4.41	1.54
VI	1.80	3.69	2.15	2.36	3.94	1.54
VII	1.81	6.54	2.18	2.34	3.86	1.54
VIII	1.81	4.01	2.17	2.35	3.72	1.54
IX	1.81	4.61	2.17	2.36	3.71	1.54
X	1.81	3.90	2.15	2.36	4.16	1.54
XI	1.80	3.83	2.17	2.36	3.64	1.54
XII	1.80	5.80	2.15	2.37	4.04	1.54
XIII	1.81	3.31	2.20	2.33	3.99	1.54

to reach a deeper minimum in the potential energy surface, which is not accessible in the protein-embedded environment due to the constraints imposed by the configuration of binding site. Consequently, ΔG and the associated k_f and K_{eq} values arising by a truncated model need to be cautiously analyzed, whilst more refined (larger) models will be required if absolute thermodynamic parameters are looked for. In spite of such limitation, all accumulated evidences confirm the selected protocol as a valid approach to provide reliable relative energetic barriers within the DNA methylation framework, which is the rate-limiting step of the process and the scope of the present contribution. Let us now consider the case of **SAM** analogues **I–XIII**, which were built up by directly remodeling the **SAM** moiety present in the truncated model. The proposed methyl-donors can be classified in 3 groups depending on the computed barrier listed in Table 1. The first set includes the pair **VI** and **VIII** with barriers of 12.36 and 12.15 kcal mol⁻¹, respectively, as they lie under the computed barrier for the parent **SAM**. The second group accounts for analogues that exceed the threshold barrier of **SAM** but still present a medium ΔG^\ddagger value (< 20 kcal mol⁻¹): **I**, **II**, **IX**, **X** and **XIII**, with barriers of 15.63, 16.19, 17.71, 14.31 and 18.24 kcal mol⁻¹, respectively. The final group is formed by **III**, **IV**, **V**, **VII**, **XI** and **XII** that present activation energies lying in the range of 20–41.26 kcal mol⁻¹. As shown in Table 2, both S-CH₃ and CH₃-C6 distances are similar in all TSs, with optimized values of 2.19 ± 0.04 Å for the distance to the sulfur atom and 2.34 ± 0.03 Å as the average distance to the cytosine base. In all cases, the donor and acceptor groups are almost linear during the methyl group displacement with a S-CH₃-C6 angle of ca. 174°. A contrasting behavior is observed in the CH₃-C6 distance of the reactant forms, in which the distance can be as large as 5.80 and 6.54 Å for **XII** and **VII**, respectively. Although it is not possible to draw a direct correlation between distance and energetic barrier, that variance hints that the analogues with the largest distance might be not optimal to fit into the binding pocket of the DNA-DNMT ensemble as they trend to be located farther than the parent **SAM**.

Let us finally analyze the obtained results in a biological framework with a focus on the extremes for the computed relative energies and rate constants. As discussed above, both **VI** and **VIII** induce a decrease of ΔG^\ddagger of 1.18 and 1.39 kcal mol⁻¹ respect to the parent **SAM**. This *a priori* minor change in the TS energy can however play a role in the global methylation process since the free energy barrier has a huge effect on the reaction rate. Indeed, the presence of this **SAM** analogues raises the forward rate constant by one order of magnitude ($k_f \sim 10^{02}$ for **SAM**, $k_f \sim 10^{03}$ for **VI** and **VIII**). In addition, **VI** and **VIII** have a shorter S-CH₃ distance compared to the initial **SAM** counterpart, so that these molecules are expected to

establish an effective contact with the flipped-out cytosine. As shown in Fig. 4, the structural change in **VI** implies the replacement of the carboxylic group in **SAM** by a nitrile group. This finding suggests that the presence of a strong electron withdrawing substituent as nitrile in the methionine moiety not only favor the interaction with the cytosine base but also significantly increase the rate constant of the methylation. Although as discussed above, thermodynamic data might be analyzed cautiously, it is remarkable that **VI** also presents the largest ΔG and K_{eq} values, being the latter six order of magnitude larger than for the one computed for **SAM**. The predicted back-barrier also demonstrated the large life-time of the methylated form ($k_r = 3.10 \times 10^{39}$). According to our data, **VIII** is the second-best ranked analogue in terms of methyl transfer ability. In this case, the carboxylic group is simply replaced by a methyl. As for **VI**, such structural change also induces an increase of both ΔG and K_{eq} , which present the second highest values. Accordingly, the optimization of the kinetic parameters is shown to be closely related to the thermodynamic of the reaction. Our theoretical protocol is finally completed through docking calculations to ensure that both compounds **VI** and **VIII** fit into the enzyme active site. As illustrated in Fig. 5, all three compounds are located in a similar way within the binding pocket: the adenosine moiety is perfectly superimposed while the -S(CH₃) group is orientated towards the flipped-out cytosine. From these results, it can be deduced that the decoration of the methionine moiety in the parent **SAM** structure might be an efficient synthetic route to tune its methyl donor ability. Among all assessed **SAM**

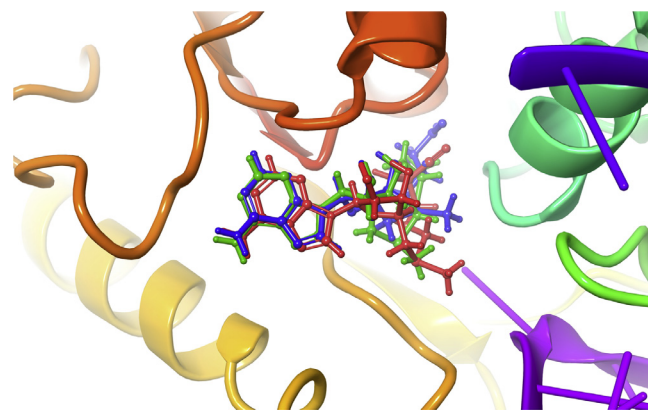


Fig. 5. Overlay of **SAM** (blue), **VI** (red) and **VIII** (green) compounds docked into the DNA-DNMT binding site. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

analogues, **VI** and **VIII** are the best ranked candidates to be used as novel methyl donor molecules with a potential use in regulation of epigenetic, a fascinating process that tunes the genomic data through gene regulation [48].

4. Conclusions and outlook

This letter touches upon the use of novel molecules as methyl promoters during the action of DNA methyltransferase enzymes. At an early stage, a joint molecular dynamics and *ab initio* protocol is used to describe the critical step in DNA methylation, namely, the transfer from the parent cofactor (S-adenosyl-L-methionine, **SAM**) towards the C5 position of the flipped-out cytosine nucleobase. The computed relative energies and rate constants match the evidences available in the experimental and theoretical literature, which backs up the proposed approach in the search of novel DNA-methyl analogs. Since the sulfonium center is a key moiety for transmethylation, the PubChem database is subsequently screened by imposing the presence of a R-S⁺(CH₃)-R' fragment in the molecule. Two molecules deposited in that database are shown to be promising candidates for enhancing the methyl transfer, labeled here as **VI** and **VIII** compounds. The former analog is built up by replacing the carboxylic group in **SAM** by a nitrile group and largely impact into both rate and equilibrium constants, which are one and six order of magnitude larger than for the ones computed for **SAM**, respectively. More striking is the latter compound, in which the carboxylic group is simply replaced by a methyl. As for **VI**, such structural change also induces a large increase of both rate and equilibrium constants by one a five orders of magnitude, respectively. Finally, optimized transition state structures confirm an effective contact between both **VI** and **VIII** with the flipped-out cytosine.

The reported results might be used to further rationalize the synthesis of improved drugs beyond classical **SAM** while simultaneously screening larger databases of bioactive compounds. We hope that the reported data help to guide next experimental and theoretical steps into the search of more efficient methyl donors in the framework of hypomethylation diseases.

Acknowledgement

J.P.C.-C. acknowledges the support provided by the Fundación Séneca-Agencia de Ciencia y Tecnología de la Región de Murcia under Project 19419/PI/14-1. This research used resources of (1) C³MSaar facilities installed at University of Saarland, (2) the Plataforma Andaluza de Bioinformática installed at the Universidad of Málaga, and (3) a local Galileo cluster installed at UCAM, Spain.

References

- [1] J.D. Watson, F.H.C. Crick, Molecular structure of nucleic acids, *Nature* 171 (1953) 737–738.
- [2] D.H.K. Lim, E.R. Maher, DNA methylation: A form of epigenetic control of gene expression, *Obstet. Gynaecol.* 12 (2010) 37–42.
- [3] J. Bohacek, I.M. Mansuy, Molecular insights into transgenerational non-genetic inheritance of acquired behaviours, *Nat. Gen.* 16 (2015) 641–652.
- [4] T.H. Bestor, The DNA methyltransferases of mammals, *Hum. Mol. Genet.* 9 (2016) 2395–2402.
- [5] R. Zangi, A. Arrieta, F.P. Cossio, Mechanism of DNA methylation: The double role of DNA as a substrate and as a cofactor, *J. Mol. Biol.* 400 (2010) 632–644.
- [6] T. Schmidt, T. Schwede, M. Meuwly, Computational analysis of methyltransfer reactions in dengue virus methyltransferase, *J. Phys. Chem. B* 118 (2014) 5882–5890.
- [7] J.M. Mato, F.J. Corrales, S.C. Lu, M.A. Avila, S-adenosylmethionine: A control switch that regulates liver function, *FASEB J.* 16 (2002) 15–26.
- [8] G.D. Jones, E.A. Lesnik, S.R. Owens, L.M. Risen, R.T. Walker, Investigation of some properties of oligodeoxynucleotides containing 4-thio-2-deoxynucleotides: Duplex hybridization and nuclease sensitivity, *Nucl. Acids Res.* 24 (1996) 4117–4122.
- [9] Ž.M. Svedružić, N.O. Reich, The mechanism of target base attack in DNA cytosine carbon 5 methylation, *Biochemistry* 43 (2004) 11460–11473.
- [10] W. Rideout, G. Coetzee, A. Olumi, P. Jones, 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes, *Science* 249 (1990) 1288–1290.
- [11] Y. Jin, L.-H. Lee, W. Shenglong, Z. Yingkai, B. Suse, DNA cytosine methylation: Structural and thermodynamic characterization of the epigenetic marking mechanism, *ACS Biochem.* 52 (2013) 2828–2838.
- [12] A. Bird, DNA methylation patterns and epigenetic memory, *Genes. Dev.* 16 (2002) 6–21.
- [13] R.M. Kohli, Y. Zhang, Tet enzymes, TDG and the dynamics of DNA demethylation, *Nature* 502 (2013) 472–479.
- [14] K. Marta, E. Manel, DNA methylation and cancer, *Adv. Genet.* 70 (2010) 27–56.
- [15] A.P. Feinberg, B. Tycko, The history of cancer epigenetics, *Nat. Rev. Cancer* 4 (2004) 143–153.
- [16] R. Holliday, G. Grigg, DNA methylation and mutation, *Mutat. Res.* 285 (1993) 61–67.
- [17] G.V. Avvakumov et al., Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1, *Nature* 455 (2008) 822–825.
- [18] Q. Du, Z. Wang, V.L. Schramm, Human DNMT1 transition state structure, *Proc. Nat. Acad. Sci. (PNAS) USA* 113 (2016) 2916–2921.
- [19] J. Aranda, K. Zinovjev, K. Świderek, M. Roca, I. Tuñón, Unraveling the reaction mechanism of enzymatic C5-cytosine methylation of DNA. A combined molecular dynamics and QM/MM study of wild type and Gln119 variant, *ACS Catal.* 6 (2016) 3262–3276.
- [20] L. Esposito, A. Banyasz, T. Douki, M. Perron, D. Markovitsi, R. Improta, Effect of C5-methylation of cytosine on the photoreactivity of DNA: A joint experimental and computational study of TCG trinucleotides, *J. Am. Chem. Soc.* 136 (2014) 10838–10841.
- [21] A. Banyasz, L. Esposito, T. Douki, M. Perron, C. Lepori, R. Improta, D. Markovitsi, Effect of C5-methylation of cytosine on the uv-induced reactivity of duplex DNA: Conformational and electronic factors, *J. Phys. Chem. B* 120 (2016) 4232–4242.
- [22] L. Martínez-Fernández, A.J. Pepino, J. Segarra-Martí, A. Banyasz, M. Garavelli, R. Improta, Computing the absorption and emission spectra of 5-methylcytidine in different solvents: A test-case for different solvation models, *J. Chem. Theory Comput.* 12 (2016) 4430–4439.
- [23] L. Martínez-Fernández, A. Banyasz, L. Esposito, D. Markovitsi, R. Improta, UV-induced damage to DNA: Effect of cytosine methylation on pyrimidine dimerization, *Signal Transduct Target Ther.* 2 (2017) 17021 EP.
- [24] G. Egger, G. Liang, A. Aparicio, Epigenetics in human disease and prospects for epigenetic therapy, *Nature* 429 (2004) 457–463.
- [25] A. Ceccaldi et al., C5-DNA methyltransferase inhibitors: from screening to effects on zebrafish embryo development, *Chem. Bio. Chem.* 12 (2011) 1337–1345.
- [26] J. Yoo, J.L. Medina-Franco, Homology modeling, Docking and structure-based pharmacophore of inhibitors of DNA methyltransferase, *J. Comput. Aided Mol. Des.* 25 (2011) 555–567.
- [27] S. Chen et al., Identifying novel selective non-nucleoside DNA methyltransferase 1 inhibitors through docking-based virtual screening, *J. Med. Chem.* 57 (2014) 9028–9041.
- [28] S. Krishna, S. Shukla, A.D. Lakra, S.M. Meeran, M.I. Siddiqi, Identification of potent inhibitors of DNA methyltransferase 1 (DNMT1) through a pharmacophore-based virtual screening approach, *J. Mol. Graph. Model.* 75 (2017) 174–188.
- [29] Z. Shao et al., Discovery of novel DNA methyltransferase 3a inhibitors via structure-based virtual screening and biological assays, *Bioorg. Med. Chem. Lett.* 27 (2017) 342–346.
- [30] N. Shukeir, P. Pakneshan, G. Chen, M. Szyf, S.A. Rabbani, Alteration of the methylation status of tumor-promoting genes decreases prostate cancer cell invasiveness and tumorigenesis in vitro and in vivo, *Cancer Res.* 66 (2006) 9202–9210.
- [31] S.M.K. Pulukuri, N. Estes, J. Patel, J.S. Rao, Demethylation-linked activation of urokinase plasminogen activator is involved in progression of prostate cancer, *Cancer Res.* 67 (2007) 930–939.
- [32] A. Fuso, R.A. Cavallaro, L. Orrù, F.R. Buttarelli, S. Scarpa, Gene silencing by S-adenosylmethionine in muscle differentiation, *FEBS Lett.* 508 (2001) 337–340.
- [33] N. Detich, S. Hamm, G. Just, J.D. Knox, M. Szyf, The methyl donor S-adenosylmethionine inhibits active demethylation of DNA: A candidate novel mechanism for the pharmacological effects of S-adenosylmethionine, *J. Biol. Chem.* 278 (2003) 20812–20820.
- [34] T. Schmidt, A. Leha, G. Salinas-Riester, Treatment of prostate cancer cells with S-adenosylmethionine leads to genome-wide alterations in transcription profiles, *Gene* 595 (2016) 161–167.
- [35] <<https://pubchem.ncbi.nlm.nih.gov>>, National Center for Biotechnology Information (accessed June 19, 2017).
- [36] X. Cheng, R.J. Roberts, AdoMet-dependent methylation, DNA methyltransferases and base flipping, *Nucl. Acids Res.* 29 (2001) 3784–3795.
- [37] F.-K. Shieh, B. Youngblood, N.O. Reich, The role of arg165 towards base flipping, base stabilization and catalysis in H.Hal, *J. Mol. Biol.* 362 (2006) 516–527.
- [38] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, *Mol. Phys.* 52 (1984) 255–268.
- [39] W.G. Hoover, Canonical dynamics: equilibrium phase-space distributions, *Phys. Rev. A* 31 (1985) 1695–1697.

- [40] G.J. Martyna, D.J. Tobias, M.L. Klein, Constant pressure molecular dynamics algorithms, *J. Chem. Phys.* 101 (1994) 4177–4189.
- [41] M. Tuckerman, B.J. Berne, G.J. Martyna, Reversible multiple time scale molecular dynamics, *J. Chem. Phys.* 97 (1992) 1990–2001.
- [42] T.A. Darden, D.M. York, L.G. Pedersen, Particle mesh Ewald: An n -log(n) method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092.
- [43] Desmond Molecular Dynamics System, Version 4.4, D.E. Shaw Research, New York, NY, 2015. Maestro-Desmond Interoperability Tools, Version 4.4, Schrödinger, New York, NY, 2015.
- [44] E. Harder et al., OPLS3: A force field providing broad coverage of drug-like small molecules and proteins, *J. Chem. Theory Comput.* 12 (2016) 281–296.
- [45] Y. Zhao, D.G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals, *Theor. Chem. Acc.* 120 (2008) 215–241.
- [46] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery, Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09 Revision E.01, gaussian Inc. Wallingford CT, 2009.
- [47] O.V. Stroganov, F.N. Novikov, V.S. Stroylov, V. Kulkov, G.G. Chilov, Lead Finder: An approach to improve accuracy of proteinligand docking, binding energy estimation, and virtual screening, *J. Chem. Inf. Model.* 48 (2008) 2371–2385.
- [48] M. Li, G.-H. Liu, J.C. Izpisua Belmonte, Navigating the epigenetic landscape of pluripotent stem cells, *Nat. Rev. Mol. Cell Biol.* 13 (2012) 524–535.

Computational Study of the Reactivity of Cytosine Derivatives

Jihène Jerbi ^[a] and Michael Springborg^[a,b]

The aim of the present study is to provide computational insight using dispersion-corrected density-functional calculations into the reactivity properties of modified cytosine in the gas phase and in aqueous solution, whereby special emphasis is put on systems that are obtained through demethylation and methylation. Since this field is relatively incipient, our goal is to identify relationships between reactivity and stability for the modified compounds to understand their biological

functionalities. Our results show that addition of a methyl, hydroxymethyl, formyl, or carboxyl group reduces the length of the nearest hydrogen bond between the cytosine–guanine (CG) base pair and increases the length of the longest hydrogen bond of the DNA base pair. © 2017 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.24781

Introduction

The effects of substitution on the chemical and biological properties of DNA may be profound. One of the most well-known cases is that of methylation of biochemical molecules that is a very important process for instance in epigenetics where 5-methylcytosine is considered as the fifth base of the genome. Moreover, during the embryonic stage, elimination of a larger number of methyl groups may occur.^[1,2] However, the demethylation may take place without leading to cell division and, hence, the creation of new DNA does not happen.

In the present work, we shall study the different modified forms of cytosine that are shown in Scheme 1. All of those are relevant for the 5-methylcytosine (5mC) and its oxidation products as schematically shown in Scheme 2. We shall now briefly outline the main mechanisms behind these reactions.

In case of demethylation, the methyl group has to be detached by enzymes as has been demonstrated recently.^[3,4] For instance, the so called Ten Eleven Translocation enzyme TET, discovered by Tahiliani et al.,^[4] is responsible for the demethylation (by oxidation) of 5-methyl-cytosine (5mC). Later it has been found that these proteins may be involved in the development of leukemia.^[5]

5-hydroxymethyl-cytosine (5hmC), whose existence was confirmed in 2009,^[4,6] is an oxidation product of 5methyl-cytosine. It has been suggested that this oxidization is the first step in the DNA demethylation pathway since the appearance of 5hmC reduces the level of 5mC at any nucleotide position.^[7]

DNA demethylation can be both passive and active. The passive process is related to a dilution of the 5hmC during cell divisions, whereas the active process involves successive TET-mediated conversions of 5-hmC to 5-formyl-cytosine (5fC) and 5-carboxyl-cytosine (5caC), both of which can be transformed back to the unmodified cytosine through the base excision repair (BER) mechanism.^[8,9]

Even though details of the demethylation process are still unclear, studies have demonstrated its importance for the

development of various types of tumors, especially in advanced stages of their development.^[10] The role of DNA demethylation in the development of cancer has been studied only little in the past, which has changed during the last years following the suggestion that for some cells the active demethylation can be self-fixed to redifferentiate.^[11]

The purpose of the present work is to use computational methods for studying the stability of these modified forms of cytosine. After having described our computational approach, we shall, at first, discuss the structural properties of each molecule both when being isolated and when being coupled to guanine. Thereby, we shall also discuss the role of hydrogen bonding. Finally, we shall use various descriptors like polarizability, reactivity descriptors, and the HOMO-LUMO gap to obtain further information on the properties of the different molecules, in particular with respect to their chemical reactivity. All molecules characteristics were compared to the characteristics of the cytosine which is the most stable molecule during this process (natural state). In general, we studied the stability of these derivatives during this cycle and not isolated.

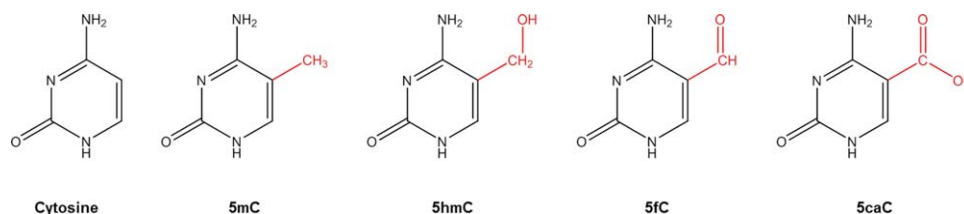
Reactivity Descriptors

Various global reactivity descriptors have been defined within the frame of conceptual density functional theory (DFT).^[12,13] In the present work, we shall use the electronegativity,^[14] the chemical potential,^[15] the absolute hardness,^[16] and the global electrophilicity.^[17] The electronegativity is the negative of the chemical potential, μ , and is defined^[15] as follows for an *N*-electron system with total energy *E* and an external potential $v(r)$,

[a] J. Jerbi, M. Springborg
Physical and Theoretical chemistry, Saarland University, Saarbrücken
66123, Germany
E-mail: jihene.jerbi@uni-saarland.de

[b] M. Springborg
School of Materials Science and Engineering, Tianjin University, Tianjin
300072, China

© 2017 Wiley Periodicals, Inc.



Scheme 1. The cytosine derivatives studied in the present work. [Color figure can be viewed at wileyonlinelibrary.com]

$$\chi = -\mu = \left(\frac{\delta E}{\delta N} \right)_{v(r)} \quad (1)$$

μ is the Lagrange multiplier associated with the particle-number-conservation constraint of DFT.^[18,19] The hardness is defined as the second partial derivative of the total energy with respect to the number of electrons,^[20]

$$\eta = \left(\frac{\partial^2 E}{\partial N^2} \right)_{v(r)} = \left(\frac{\partial \mu}{\partial N} \right)_{v(r)} \quad (2)$$

Using finite-difference approximations, these can be approximated through^[21]

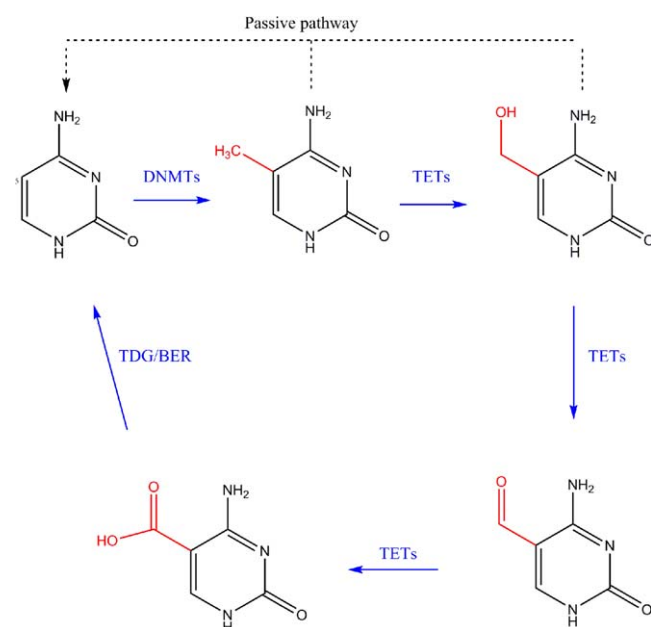
$$\chi = \frac{(I+A)}{2} \quad (3)$$

$$\eta = \frac{(I-A)}{2} \quad (4)$$

where I and A are the (vertical) ionization potential and electron affinity, respectively. Also the softness, defined as the inverse of the global hardness, can be useful,

$$S = \frac{1}{\eta} \quad (5)$$

Furthermore, Gazquez et al.,^[22,23] introduced two different chemical potentials that can be used in differentiating



Scheme 2. Schematic representation of methylcytosine and its oxidation products. For further details, see the text. [Color figure can be viewed at wileyonlinelibrary.com]

between the responses to charge donation and charge acceptance,

$$\mu^- = \frac{-1}{4}(3I+A) \quad (6)$$

$$\mu^+ = \frac{-1}{4}(I+3A) \quad (7)$$

Equivalent, one may define two different electronegativities,^[24]

$$\chi^- = \frac{1}{4}(3I+A) \quad (8)$$

$$\chi^+ = \frac{1}{4}(I+3A). \quad (9)$$

It is relevant to mention that a system with a larger electron-donating capacity has a lower value of χ^- whereas a larger value of χ^+ implies a larger capacity to accept electrons.^[25] We add that the chemical potentials and electronegativities are the negatives of each other and, therefore, we shall report only the latter.

With ϵ_{HOMO} and ϵ_{LUMO} being the energies of the highest occupied and lowest unoccupied molecular orbital, respectively, Koopman's theorem^[25] gives the approximate expression

$$\eta = \frac{(\epsilon_{\text{LUMO}} - \epsilon_{\text{HOMO}})}{2} \quad (10)$$

Another relevant descriptor is the electrophilicity, defined as^[17]

$$\omega = \frac{\mu^2}{2\eta} \quad (11)$$

that quantifies the capacity of the system of interest to accept electrons. In total, a good nucleophile is characterized by a low value of ω ^[26] and a good electrophile is characterized by a high value.

Computational Details

All our calculations were carried through using the three-parameters B3LYP^[27,28] density-functional within the frame of DFT^[29,30] together with a 6-31++G(d) basis set as implemented in the Gaussian 09 program suite.^[31] After optimizing the geometries of the different species the observation that all calculated vibrational frequencies were real demonstrated that each structure indeed corresponds to a total-energy minimum.

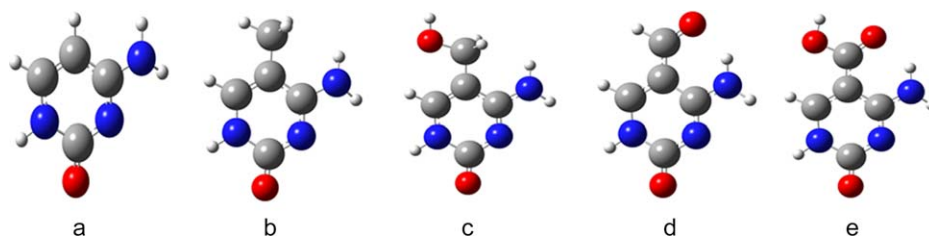


Figure 1. Structures of a) cytosine, b) 5mC, c) 5hmC, d) 5fC, and e) 5carC. Large red, blue, and gray spheres represent O, N, and C atoms, respectively, whereas small spheres represent H atoms. [Color figure can be viewed at wileyonlinelibrary.com]

We included a dispersion-energy correction DFT-D3^[32] for all the molecules to describe correctly the dispersive interactions. Moreover, the basis-set-superposition error (BSSE)^[33] was corrected only for the base-pair molecules. The reactivity descriptors described above were calculated using the eqs. (1)–(12). In addition, we also calculated the static polarizabilities of the molecules. Since the unmethylated cytosine (C) and un-methylated base pair cytosine-guanine (CG) correspond to the common units in the DNA we will use these as reference systems. We performed calculations both for the gas phase and in aqueous solution, whereby we in the latter case treated water using the polarizable continuum PCM model.^[34]

Result and Discussion

Structural properties

Geometry of the isolated molecules. At first, we optimized the structure of each of the five molecules of Scheme 1 both in gas phase and in aqueous phase. For all species (shown in Fig. 1) we studied the neutral molecule, the cation, and the anion. All the structures were found to have a planar backbone with the exception of the NH₂ group which can be explained by a partial sp³ pyramidalization of the amino group of DNA.^[35] Sponer et al.^[36] emphasized the importance of this nonplanarity for the hydrogen bonding in DNA.

Not surprising, in particular the geometry of the amino group depends strongly on the surroundings,^[35] that is, the structure changes from being more pyramidal to being less when we include the solvent. Through this flexibility, hydrogen bonds for the CG dimer become possible both for the neutral and for the charged molecules. The detailed explanation of

the relation between flexibility and nonplanarity of amino group is given by Sponer et al. in Ref. [36].

As expected, when subsequently including dispersion corrections, the molecular structures changed only very little.

Our calculations reveal also that the CH₃O group in anionic 5hmC and 5hmCG can rotate almost freely both in gas phase and in solution. Simultaneously, a graphical presentation of the frontier orbitals demonstrates that the electrons are delocalized in the cytosine ring and also an accumulation of negative charge in the CH₃O group as shown in Figure 2. (See Supporting Information for HOMO-LUMO figures of all the molecules).

Hydrogen bonding. Two different interactions affect the structure and dynamics of DNA: Hydrogen bonding and vertical stacking.^[36] Hydrogen bonds are known to play an important role in many biological systems and are crucial for the genetic code.^[37] In the latter case, electrostatic attractions in the form of dipole-dipole interactions form the dominating part.^[36] For the present systems, the interactions between the hydrogen-donating and -accepting groups are due to lone pair orbitals on O or N of one base and N–H σ^* orbitals of the other base.^[37]

The energy of hydrogen bonding for CG base pairs has been studied earlier both experimentally^[38] and computationally.^[39] Our calculated hydrogen-bond structures and energies for native and modified CG base pairs are listed in Tables (1–3). There, we also compare with previous results (experimental and theoretical).^[43]

Since all the modified molecules will be compared to the unsubstituted one, it is useful to start with analyzing the isolated unmodified CG base pair also because a comparison with results of other studies then allows for accessing the accuracy of our computations. The stabilization energy of the

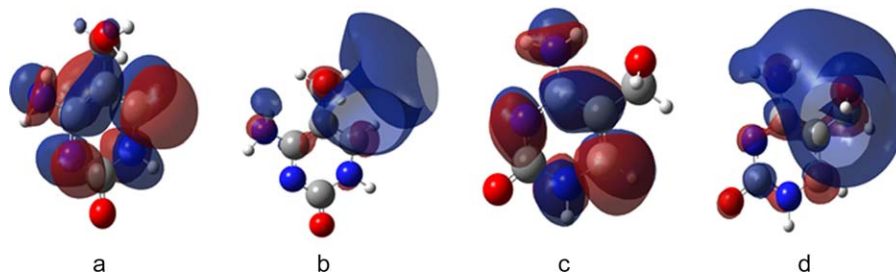


Figure 2. Frontier orbitals for 5formylcytosine: a) HOMO in gas phase, b) LUMO in gas phase, c) HOMO in solution, and d) LUMO in solution. Both orbitals are of pi (π) symmetry. [Color figure can be viewed at wileyonlinelibrary.com]

Table 1. Structural and energetic properties of CG and 5mCG in gas phase and aqueous solution as obtained in the present study and in other theoretical as well as experimental studies.

System	Method	O ₁₀ –N ₁₈ bond length (Å)	N ₁ –N ₁₄ bond length (Å)	N ₁₁ –O ₁₉ bond length (Å)	ΔE (kcal/mol)
CG	B3LYP/6-31++G*	2.82	2.96	2.94	–26.04
	B3LYP/6-31G** ^[a]	2.79	2.93	2.92	–25.5
	B3LYP-D3/6-31++G*	2.80	2.93	2.91	–30.99
	BP86-D/TZ2P ^[a]	2.70	2.84	2.83	–32.28
	B3LYP-D3/6-31++G* (PCM)	2.87	2.94	2.86	–18.13
	B3LYP-D3/6-31++G* (COSMO)	2.91	2.95	2.88	–13.10
	BP86-D/TZ2P (COSMO) ^[b]	2.80	2.85	2.79	–14.79
	B3LYP/6-31++G** ^[c]	2.82	2.95	2.93	–
	B3LYP/6-31++G** (PCM) ^[c]	2.91	2.96	2.88	–
	Experiment ^[d]	2.86	2.95	2.91	–
5mCG	B3LYP/6-31++G*	2.82	2.96	2.93	–26.48
	BP86/TZ2P ^[a]	2.73	2.84	2.83	–32.28

The calculations marked COSMO or PCM are for the aqueous phase and COSMO and PCM denote the theoretical method used for treating the solvents. ΔE is the energy of the base pair compared to the isolated bases. [a] Ref. 40. [b] Ref. 37. [c] Ref. 41. [d] Ref. 42.

hydrogen bonding is –26.0 and –13.3 kcal/mol in vacuum and water, respectively. Our results are in good agreement with the previous results by Poater et al.^[37] and Guerra et al.^[40] as demonstrated in Table 1.

The optimized structure of the unsubstituted CG base pairs is shown in Scheme 3. As can be seen in Table 1, there is a satisfying agreement between our calculated values for structural and energetic properties and those of early studies.^[37,40,41] Furthermore, also experimental values obtained through X-ray crystallography^[42] agree well with the theoretical results.

As discussed by Sassi et al.,^[44] any modification of the donor-acceptor order that perturbs the Watson–Crick base pair

(here, guanine cytosine) may have large impact on the genetic code, making it very important to study the modified cytosine during the demethylation process.

We studied the interaction energies between the different modified cytosine molecules and a single guanine molecule. The resulting hydrogen-bond strengths, both in gas phase and in solution and both for the neutral and for the charged species, are reported in Tables 2 and 3. For the unsubstituted cytosine we found that the backbone of the anionic CG pair is nonplanar both in the gas phase and in the solution, whereas it is planar in the neutral and cationic cases. This observation applies to 5mCG and to 5hmCG, too. Conversely, different structures can be observed for 5fCG and 5carCG that are planar in gas phase (with and without dispersion corrections) and also in aqueous solution (see Table 3). Overall, the inclusion of dispersion interactions leads to only small structural changes.

Substitution at position 5 leads to a re-distribution of the charge which also affects the structural properties (See Fig. 3). Thereby, the hydrogen bond between N₁₈ and O₁₀ becomes shorter, whereas that between N₁ and N₁₄ becomes longer.^[41] In particular the length of the O₁₀–H₂₆N₁₈ hydrogen bond is affected by the substitution being between 1.687 Å for 5mCG and 2.217 Å for 5fCG in the gas phase and from 1.928 Å for 5fCG to 2.045 Å for 5mCG in water. The hydrogen bonds N₁

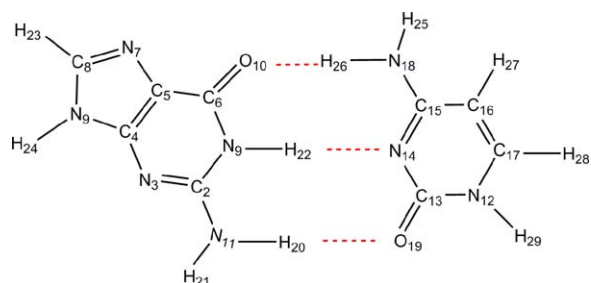
Table 2. Hydrogen bond lengths (in Å) of modified, anionic CG base pairs at the B3LYP/6-31++G* level.

	O ₁₀ –H ₂₆ N ₁₈	N ₁ H ₂₂ –N ₁₄	N ₁₁ H ₂₀ –O ₁₉
CG	2.026	1.814	1.703
	1.991	1.773	1.702
	2.041	1.825	1.780
	2.009	1.775	1.777
5mCG	1.687	1.835	2.014
	1.978	1.794	1.685
	2.045	1.831	1.763
	2.012	1.780	1.760
5hmCG	2.018	1.850	1.702
	1.979	1.810	1.698
	2.029	1.847	1.771
	1.999	1.793	1.768
5fCG	2.217	1.710	1.362
	1.926	1.853	1.704
	1.928	1.917	1.789
	1.887	1.883	1.769
5carCG	1.962	1.882	1.710
	1.920	1.848	1.701
	1.931	1.907	1.783
	1.897	1.864	1.771

In each case, the first, second, third, and fourth row shows the results for the gas phase without the inclusion of dispersion corrections, for the gas phase with the inclusion of dispersion corrections, for the aqueous phase without the inclusion of dispersion corrections, and for the aqueous phase with the inclusion of dispersion corrections, respectively.

Table 3. Hydrogen-bond energies (in kcal/mol) of CG base pairs in gas phase (upper row) and in aqueous solution (lower row) as well as the energies when including BSSE and dispersion corrections at the B3LYP/6-31++G* level.

System	ΔE	ΔE _{BSSE}	ΔE _{Disp}
CG	–26.0	–24.7	–31.0
	–13.3	–11.9	–18.1
5mCG	–26.5	–25.4	–31.4
	–13.4	–12.3	–18.3
5hmCG	–26.7	–24.9	–31.6
	–13.4	–11.6	–18.3
5fCG	–23.2	–21.7	–28.1
	–12.3	–10.8	–17.2
5carCG	–23.7	–21.7	–28.5
	–12.4	–10.5	–17.4



Scheme 3. Optimized structure of guanine–cytosine base pair. [Color figure can be viewed at wileyonlinelibrary.com]

$\text{H}_{22}\text{—N}_{14}$ and $\text{N}_{11}\text{H}_{20}\text{—O}_{19}$ become shorter for the functionalized systems, which may be explained through a charge-neutrality argument.^[37]

In general, both for the neutral and for the ionized species, solvation causes a smaller elongation of the hydrogen bonds compared to the gas phase with anionic 5fCG being the only exception in the present study. An explanation is that the addition of the solvent causes a stabilization of lone pairs which is converted into weaker hydrogen-bond interactions and, consequently, longer hydrogen bonds.^[37]

In Table 3, we list the calculated bonding energies ΔE for the systems of our interest. By comparing the bond energies in gas phase with those in aqueous solution, it is seen that the base pairs in solution always are less stable than in the gas which is in line with the longer hydrogen bonds found in the former case. The reason is that solvent stabilization of the individual bases dominates over solvent stabilization of the base pairs.^[37]

Ignoring the corrections due to basis-set superposition error (BSSE) and dispersion, for the unmodified CG base pair the bonding energy is reduced from $-26.0 \text{ kcal}\cdot\text{mol}^{-1}$ in the gas phase to $-13.3 \text{ kcal}\cdot\text{mol}^{-1}$ in aqueous solution whose

difference is very similar to what is found for the 5mCG and 5hmCG systems, and slightly larger than the differences for the 5fCG and 5carCG systems. When adding BSSE and dispersion corrections, these conclusions remain essentially unaltered.

These findings are in agreement with those of the recent experimental work of Tokmakoff et al.^[45] They proposed that the electron-withdrawing —CHO and —COOH groups in 5fC and 5carC decrease the electron density at N_3 in cytosine. This leads to weaker hydrogen bonding of the 5fCG and 5carCG and accordingly a destabilization of the base pair.

Reactivity descriptors of the molecules

The polarizability has been proposed as a parameter that quantifies chemical reactivity,^[46] that is, an atom with a low polarizability is supposed to be more “stable” and, accordingly, less “reactive.”

In the present work, we shall consider the spherically averaged polarizability,^[25]

$$\alpha = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz}) \quad (13)$$

In Table 4 this quantity, together with several others, is listed for the systems of the present study, both in gas phase and in solution. The first observation from the table is that the effects of the solvent are very weak.

Moreover, using α as a descriptor for chemical activity, 5fC/CG and 5carC/CG are more reactive than the other molecules, a result that is in accord with those found above from the analysis of the hydrogen bonding energies.

The energies and spatial distributions of the highest occupied and the lowest unoccupied orbitals of the molecules are important for the electron-donating and electron-accepting

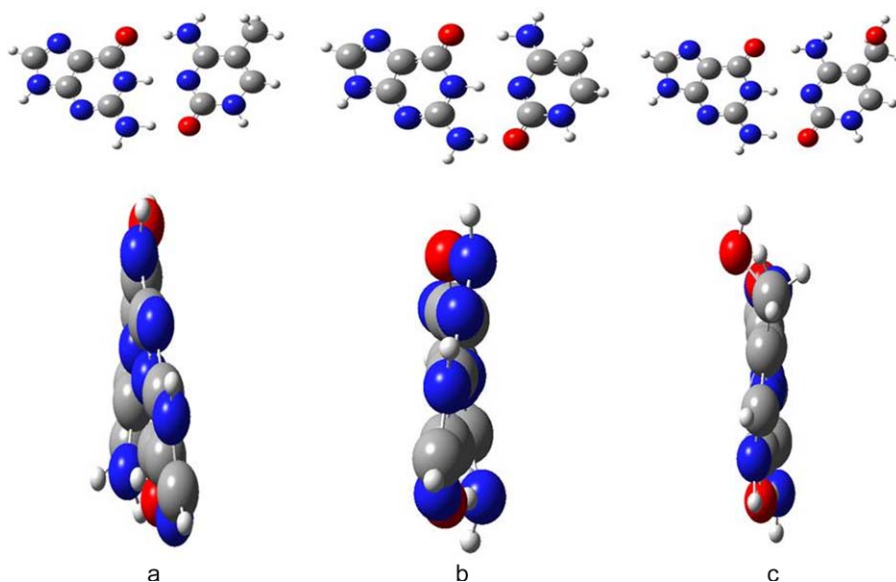


Figure 3. Optimized structures of a) 5mCG, b) 5CG, and c) 5hmCG as anion species in gas phase at B3LYP/6-31++G* level. In all cases, guanine is in the left part of the panel. The lower panel shows a side view of the molecules that illustrates the non-planarity due to the change in hydrogen bonds shape. [Color figure can be viewed at wileyonlinelibrary.com]

Table 4. Global reactivity descriptors of the cytosine and its four derivatives at the B3LYP/6-31++G* level.

	C	5mC	5hmC	5fC	5CarC
α (\AA^3)	74.11 (74.12) 176.91 (177.20)	86.50 (86.47) 189.72 (190.01)	90.50 (90.44) 193.99 (194.28)	92.57 (92.85) 194.96 (195.38)	95.36 (95.57) 198.19 (198.57)
ΔE_{gap} (eV)	5.3 (5.5) 3.8 (4.5)	5.1 (5.3) 3.9 (4.8)	5.1 (5.3) 3.9 (4.5)	4.8 (5.1) 3.0 (3.7)	4.9 (5.2) 3.1 (3.9)
ω (eV)	2.23 1.47	2.18 1.49	1.88 1.39	1.74 2.34	1.77 1.06
χ (eV)	4.41 3.26	4.19 3.26	4.00 3.18	4.11 3.84	4.11 2.94
χ^- (eV)	6.45 5.09	6.20 5.06	6.09 5.03	6.56 5.41	6.48 4.95
χ^+ (eV)	2.26 1.36	2.17 1.39	1.82 1.224	1.49 2.42	1.58 0.653
η (eV)	2.64 1.90	2.58 1.93	2.59 1.96	2.42 1.49	2.45 1.55

In each case, the first row gives the results for the isolated molecule and the second row gives those for the CG pair and its derivatives. Moreover, all results are for the gas phase except for those in parentheses is that are for the aqueous solution.

properties. In fact, it has been demonstrated that the purines (here guanine) are better electron donors than pyrimidine (here cytosine and its derivatives).^[47] The fact that the HOMO is localized to the cytosine and the LUMO to the guanine (see Fig. 4 as well as the Supporting Information) is a further argument for this observations.

When comparing our calculated value of the gap for the native Watson–Crick base pair CG in gas phase (3.8 eV) with that of previous work (3.8 eV)^[43] a good agreement is found. This is also the case for the values obtained in aqueous phase, that is, 4.5 eV (PCM) [4.6 eV (COSMO)] compared to 4.5 eV (COSMO).

According to the HSAB (hard and soft acids and bases) principle, a molecule with a smaller HOMO–LUMO energy gap is more reactive and can be considered as a soft molecule.^[48] In Table 4 we see that 5fC/CG and 5carC/CG have the smallest gap, that is, 4.8 (5.1) and 4.9 (5.2) eV, respectively, so, like the polarizability, also the values of the HOMO–LUMO energy gap suggest that these two systems are the most reactive ones.

The other measures for chemical reactivity of cytosine and its derivatives are collected in Table 4, too. It is worth to mention that in all our calculations the neutral species have

positive electron affinities. Here, it is seen that out of the three different electronegativities, only one, χ^+ , shows some, albeit weak, dependence on the substitution. For the individual molecules, the fact that this quantity is lowest for 5fC and 5carC implies that these two systems are those with the lowest electron-accepting capability, whereas the electron-donating capabilities are similar for all five systems. Conversely, for the guanine cytosine base pair, 5fCG becomes the one with the largest electron-accepting capability. For the unsubstituted base pair, we find a value of -4.41 eV for the chemical potential, 2.23 eV for the electrophilicity, and 1.90 eV for the hardness in gas phase, that are in fair agreement with the values obtained by Najafi et al.^[43]: -3.14 , 3.41, and 1.88 eV, respectively.

By analyzing the chemical potential (the negative of the electronegativity), we observe that the nucleophilicity decreases from 5hmC to unsubstituted cytosine indicating that 5hmcytosine is the most nucleophilic molecule of this series with the highest chemical potential followed by 5fC and 5carC and, finally, 5mC which makes cytosine the least nucleophilic molecule. This may be associated with the electron-withdrawing groups added to the cytosine at position 5.

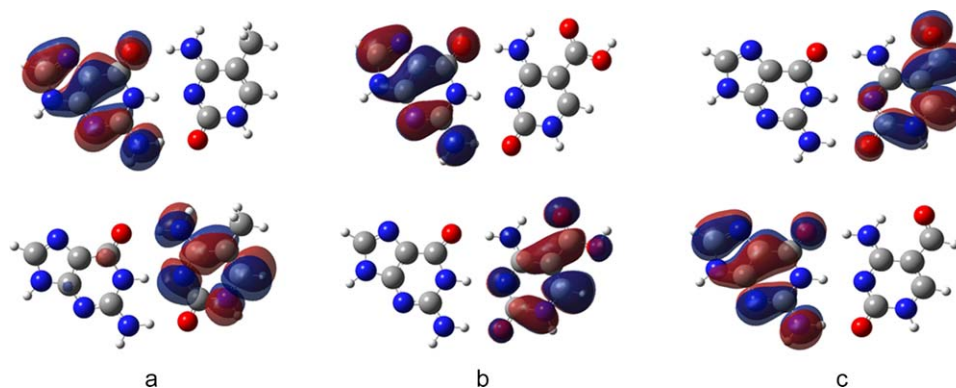


Figure 4. HOMO (top) and LUMO (bottom) of a) 5mC, b) 5carC, and c) 5fC in gas phase at the B3LYP/6-31++G* level. [Color figure can be viewed at wileyonlinelibrary.com]

The calculated ω values of the native and modified molecules in gas phase ranges from 1.74 to 2.23 eV. The values decrease in the order cytosine, 5mC, 5hmC, 5CarC, and 5fC. The obtained values show that cytosine, 5mC, and 5hmC have a high capacity to accept electrons while for the CG pairs, this is the case for 5fCG, CG, and 5mCG.

According to the results in Table 4, the best electron acceptor (large χ^+) is cytosine followed by 5mC and 5hmC. Conversely, 5carC is considered as the worst electron donor (large χ^-) and more reactive toward electrophilic attack followed by the unsubstituted cytosine, and 5hmC is the best electron donor with an χ^- of 6.09 eV. However, when paired with guanine the 5carCG becomes the most nucleophilic molecule.

When studying the Mulliken charges (See Supporting Information) for each molecule, isolated or paired, it is found that the nucleophilic character of position 5 in cytosine is induced by the presence of different functional groups. The addition of hydroxyl or carboxyl groups leads to a delocalization of the electrons in the ring and makes the carbon 5 of the cytosine more negative, that is, -0.37 and -0.36 , respectively, compared to charges of 0.21 and 0.17 for the formyl- and methyl-containing compounds. Mulliken populations may not provide absolute values for atomic charges but a comparison between related systems can give information on changes in the atomic charges.

Further information is obtained from the values of the absolute hardness, also listed in Table 4. As proposed by Pearson,^[49] a larger value of the hardness implies increased stability. We see that for the isolated molecules, cytosine, 5methylcytosine, and 5hydroxymethylcytosine are then the most stable ones although their hardness values actually are very close. For the pairs with guanine the same order is found: the first three molecules of the cycle are the most stable ones, whereas 5carCG and 5fCG are the softest and thus the most reactive ones. These interpretations remain unaltered also after adding a solvent and/or dispersion corrections to our calculations.

Conclusions

In the present work, we have studied the reactivities of five derivatives of cytosine isolated or paired with together with guanine, all systems that are relevant for methylated cytosine. We have considered global reactivity descriptors, the polarizability, and the hydrogen bonding energies. According to our results, all presented functional groups influence the structure and geometry of the Watson–Crick base pairs, whereby hydroxymethyl leads to the largest structural and energetic changes. 5hmC has the lowest hydrogen bonding energy of -26.7 kcal·mol⁻¹ compared to -26.5 kcal·mol⁻¹ for 5mC and -26.0 kcal·mol⁻¹ for the unsubstituted cytosine. Within the HSAB principle, the stability may be related to softness and hardness which in turn may be related to the energies of the frontier orbitals. With those assumptions, the difference in the stability for 5hmC and 5mC can be also related to the variation in the energy of the HOMO-LUMO gap.

The addition of different functional groups to the cytosine at its carbon 5 position can change the charge distribution. Furthermore, our results on the structural properties of each derivative show that ionization, solvent, and functional groups can have a considerable effect on the hydrogen bonds and the pyramidal behavior of the NH₂ group of isolated cytosine.


Finally, according to our results 5hmC has larger hydrogen-bond energies than the other derivatives of cytosine which is in agreement with results of experimental studies^[7,50,51] where it is observed that 5hmC plays a crucial role in gene expression and is found in larger amounts in the brain of developing embryo at early stages.

Acknowledgments

The calculations were carried through at the C³MSaar at University of Saarland. The authors would like to thank M.Sc. Mahdi Louati for his fruitful advices.

Keywords: dispersion-correction · density functional theory calculations · DNA modifications · reactivity

How to cite this article: J. Jerbi, M. Springborg. *J. Comput. Chem.* **2017**, *38*, 1049–1056. DOI: 10.1002/jcc.24781

 Additional Supporting Information may be found in the online version of this article.

- [1] A. Meissner, *Nat. Biotechnol.* **2010**, *28*, 1079.
- [2] S. Feng, S. E. Jacobsen, W. Reik, *Science* **2010**, *330*, 622.
- [3] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129.
- [4] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930.
- [5] R. Ono, T. Taki, T. Taketani, M. Taniwaki, H. Kobayashi, Y. Hayashi, *Cancer Res.* **2002**, *62*, 4075.
- [6] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929.
- [7] G. P. Pfeifer, S. Kadam, S.-G. Jin, *Epigenetics Chromatin* **2013**, *6*, 10.
- [8] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300.
- [9] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334.
- [10] C. De Smet, A. Loriot, *Epigenetics* **2010**, *5*, 206.
- [11] C. S. Nabel, S. A. Manning, R. M. Kohli, *ACS Chem. Biol.* **2011**, *7*, 20.
- [12] R. G. Parr, W. Yang, *Density Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, UK, **1989**.
- [13] (a) P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.* **2003**, *103*, 1793; (b) H. Chermette, *J. Comput. Chem.* **1999**, *20*, 129.
- [14] (a) L. Pauling, *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, New York, **1960**; (b) K. D. Sen, C. Jorgenson, *Structure and Bonding: Electronegativity*, Vol. 66; Springer: Berlin, **1987**.
- [15] (a) R. G. Parr, R. A. Donnelly, M. Levy, W. E. Palke, *J. Chem. Phys.* **1978**, *68*, 3801; (b) R. S. Mulliken, *J. Chem. Phys.* **1934**, *2*, 782.
- [16] K. D. Sen, D. M. P. Mingos, *Structure and Bonding: Chemical Hardness*, Vol. 80; Springer: Berlin, **1993**.
- [17] R. G. Parr, L. V. Szentpály, S. Liu, *J. Am. Chem. Soc.* **1999**, *121*, 1922.
- [18] P. Pérez, L. R. Domingo, M. Duque-Noreña, E. Chamorro, *J. Mol. Struct. THEOCHEM* **2009**, *895*, 86.
- [19] P. Linda, G. Marino, *J. Chem. Soc. B* **1968**, *1968*, 392.
- [20] H. Chermette, *J. Comput. Chem.* **1999**, *20*, 129.
- [21] R. G. Parr, R. G. Pearson, *J. Am. Chem. Soc.* **1983**, *105*, 7512.
- [22] J. L. Gazquez, A. Cedillo, A. Vela, *J. Phys. Chem. A* **2007**, *111*, 1966.
- [23] J. Z. Ramirez-Ramirez, R. Vargas, J. Garza, J. L. Gazquez, *J. Phys. Chem. A* **2010**, *114*, 7945.

- [24] A. Martinez, *J. Phys. Chem. C* **2010**, *114*, 21240.
- [25] P. K. Chattaraj, D. R. Roy, M. Elango, V. Subramanian, *J. Phys. Chem. A* **2005**, *109*, 9590.
- [26] H. Luis, H. Mendoza, H. R. R. Clara, *J. Mex. Chem. Soc.* **2011**, *55*, 142.
- [27] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
- [28] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
- [29] W. Kohn, A. D. Becke, R. G. Parr, *Phys. Chem.* **1996**, *100*, 12974.
- [30] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.
- [31] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09, revision A.02*; Gaussian, Inc.: Wallingford, CT, **2009**.
- [32] (a) S. Grimme, *WIREs Comput. Mol. Sci.* **2011**, *1*, 211; (b) S. Grimme, *J. Comput. Chem.* **2004**, *25*, 1463.
- [33] G. Chalasin, M. M. Szczesniak, *Chem. Rev.* **1994**, *94*, 1723.
- [34] G. Scalmani, M. J. Frisch, *J. Chem. Phys.* **2010**, *132*, 114.
- [35] N. Kurita, V. I. Danilov, V. M. Anisimov, *Chem. Phys. Lett.* **2005**, *404*, 164.
- [36] (a) P. Hobza, J. Sponer, *J. Am. Chem. Soc.* **2002**, *124*, 11802; (b) J. Sponer, J. Leszczynski, P. Hopza, *Biopolymers* **2002**, *61*, 3.
- [37] J. Poater, C. F. Guerra, F. M. Bickelhaupt, *Comput. Theor. Chem.* **2012**, *998*, 57.
- [38] (a) Y. Kyogoku, R. C. Lord, A. Rich, *Proc. Natl. Acad. Sci. USA* **1967**, *57*, 250; (b) S. B. Paterson, J. J. Led, *J. Am. Chem. Soc.* **1981**, *103*, 5308.
- [39] (a) B. Pullman, P. Claverie, J. Caillet, *Proc. Natl. Acad. Sci. USA* **1996**, *55*, 904; (b) V. I. Danilov, I. S. Tolokh, V. I. Potlov, G. G. Malenkov, *FEBS Lett.* **1984**, *167*, 245.
- [40] C. F. Guerra, F. M. Bickelhaupt, J. G. Snijders, E. J. Baerends, *J. Am. Chem. Soc.* **2000**, *122*, 4117.
- [41] M. Zarei, A. Seif, K. Azizi, M. Zarei, J. Bahrami, *Int. J. Mod. Phys. C* **2016**, *27*, 1650119.
- [42] (a) I. Yonson, A. Teplitsky, L. Sukhodub, *Biopolymers* **1979**, *18*, 1149; (b) J. M. Rosenberg, N. C. Seeman, R. O. Day, A. Rich, *J. Mol. Biol.* **1976**, *104*, 145.
- [43] D. Farmanzadeh, M. Najafi, *Struct. Chem.* **2015**, *26*, 831.
- [44] M. Sassi, D. J. Carter, B. P. Uberuaga, C. R. Stanek, R. L. Mancera, N. A. Marks, *J. Phys. Chem. B* **2014**, *118*, 10430.
- [45] Q. Dai, P. J. Sanstead, C. S. Peng, D. Han, C. He, A. Tokmakoff, *ACS Chem. Biol.* **2016**, *11*, 470.
- [46] P. Senet, In *Reactivity and polarizability response*; P. K. Chattaraj, Ed.; *Chemical Reactivity Theory: A Density Functional View*; CRC Press Taylor & Francis Group, **2009**; pp. 331–362.
- [47] B. Pullman, A. Pullman, *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 1197.
- [48] T. Rajamani, S. Muthu, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2013**, *115*, 654.
- [49] R. G. Pearson, *Acc. Chem. Res.* **1993**, *26*, 250.
- [50] W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Nature* **2011**, *473*, 394.
- [51] Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicheolson, T. Chen, G. Xu, Y. Shi, K. Zhang, Y. G. Shi, *Mol. Cell.* **2011**, *42*, 451.

Received: 29 November 2016
Revised: 15 February 2017
Accepted: 16 February 2017
Published online on 20 March 2017

Computational study of the stability of cytosine derivatives

Jihène Jerbi* and Michael Springborg*⁺

*Physical and Theoretical chemistry, Saarland University, 66123 Saarbrücken, Germany

⁺Tianjin University, Tianjin 300072, P. R. China

jihene.jerbi@uni-saarland.de

m.springborg@mx.uni-saarland.de

Here we present the HOMO/LUMO orbitals of the cytosine and its derivatives while isolated and paired to guanine in gas phase then in solvent. In all figures: the orbitals of the gas phase are shown to the right and those of the aqueous phase to the left. a, b, and c corresponds to the anion, the cation, and the neutral species, respectively.

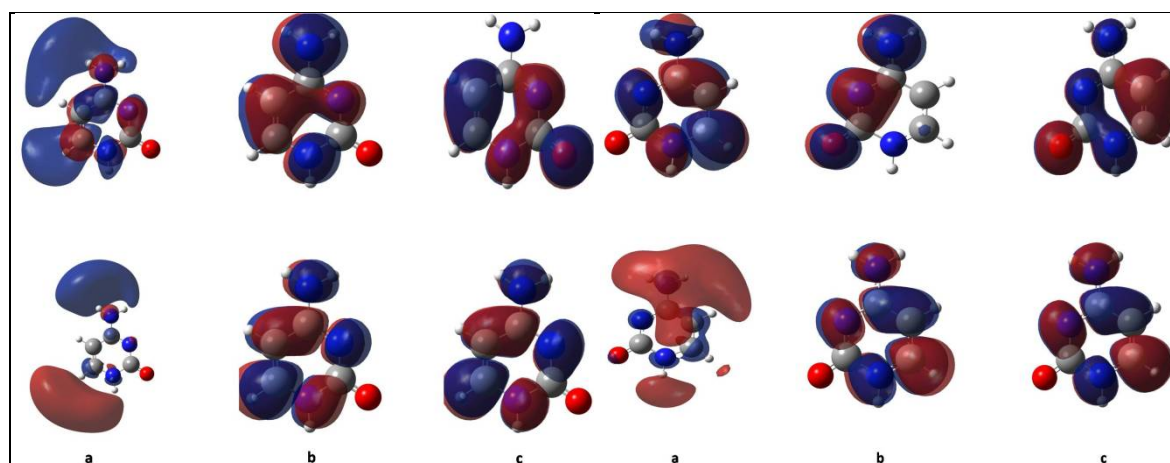


Figure1: HOMO/LUMO orbitals for Cytosine

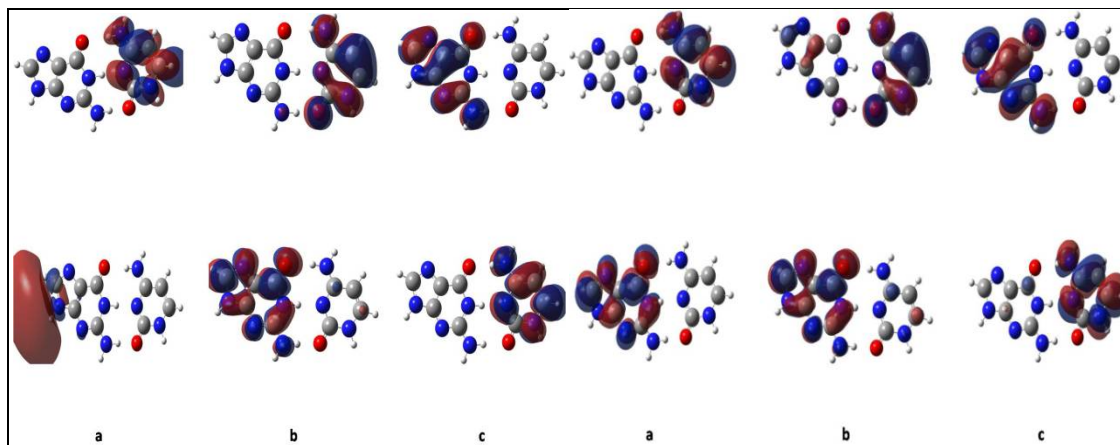


Figure2: HOMO/LUMO orbitals for CG

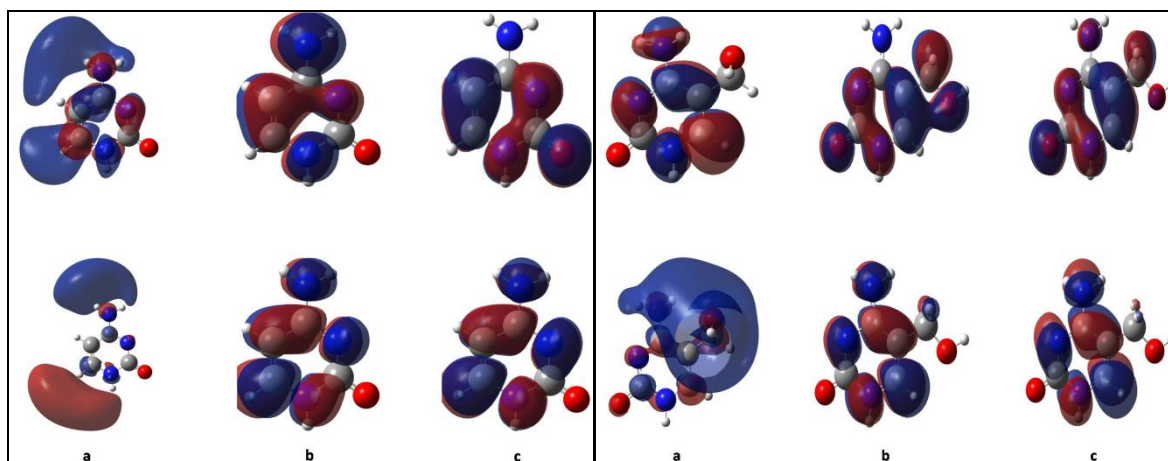


Figure3: HOMO/LUMO orbitals for 5mCytosine

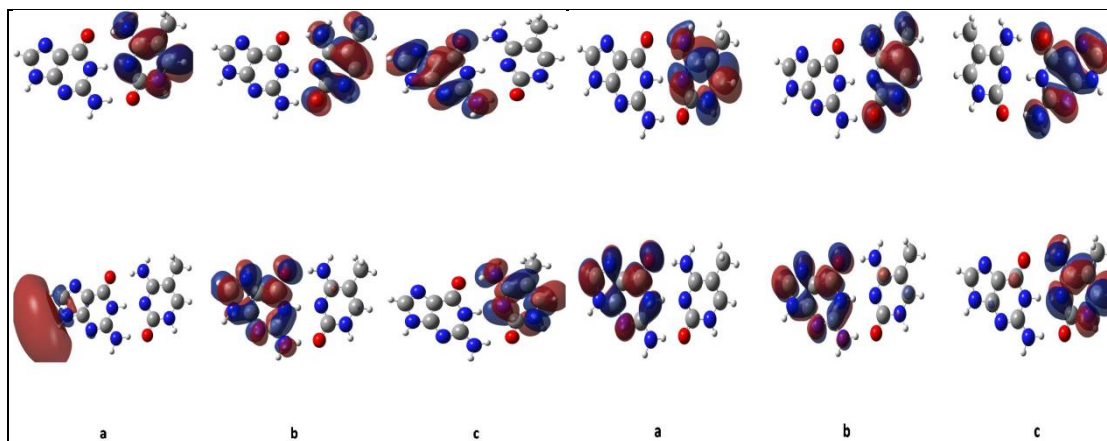


Figure4: HOMO/LUMO orbitals for 5mCG

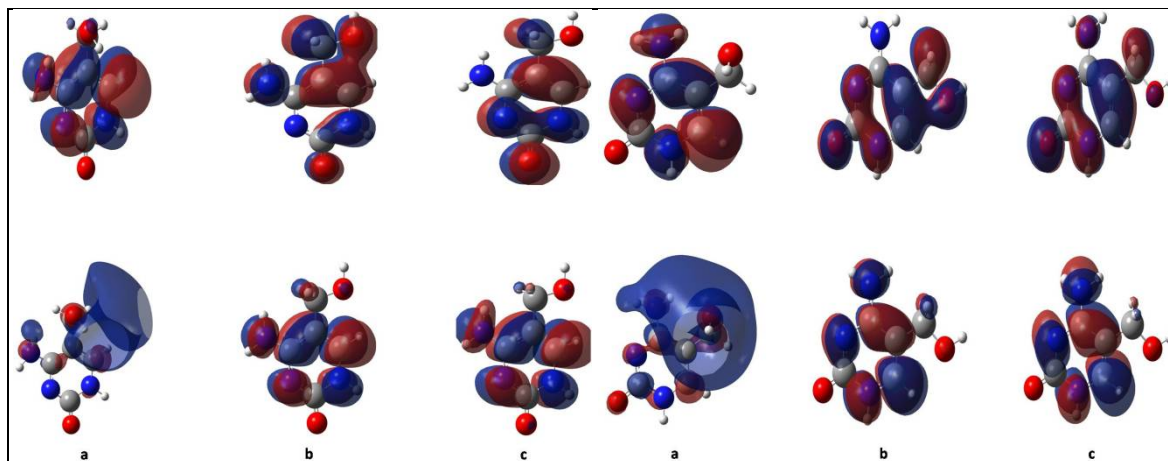


Figure5: HOMO/LUMO orbitals for 5hmC

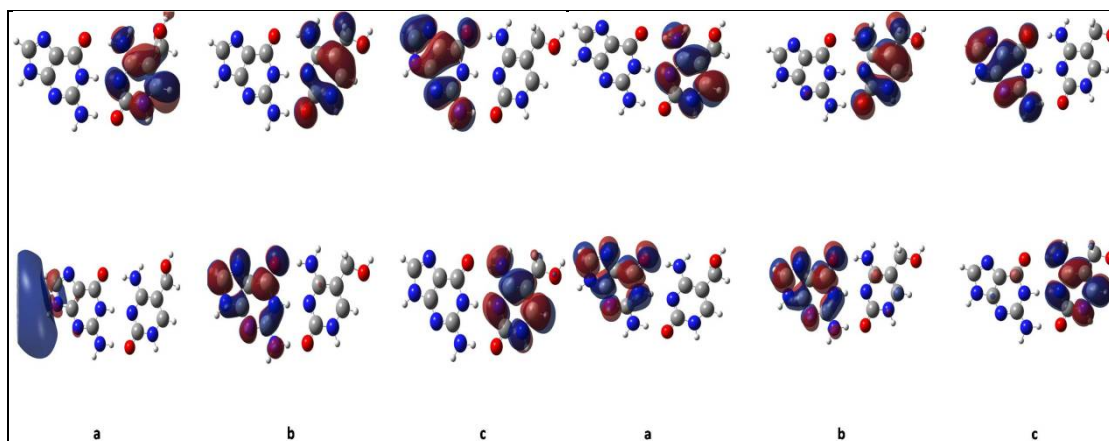


Figure6: HOMO/LUMO orbitals for 5hmCG

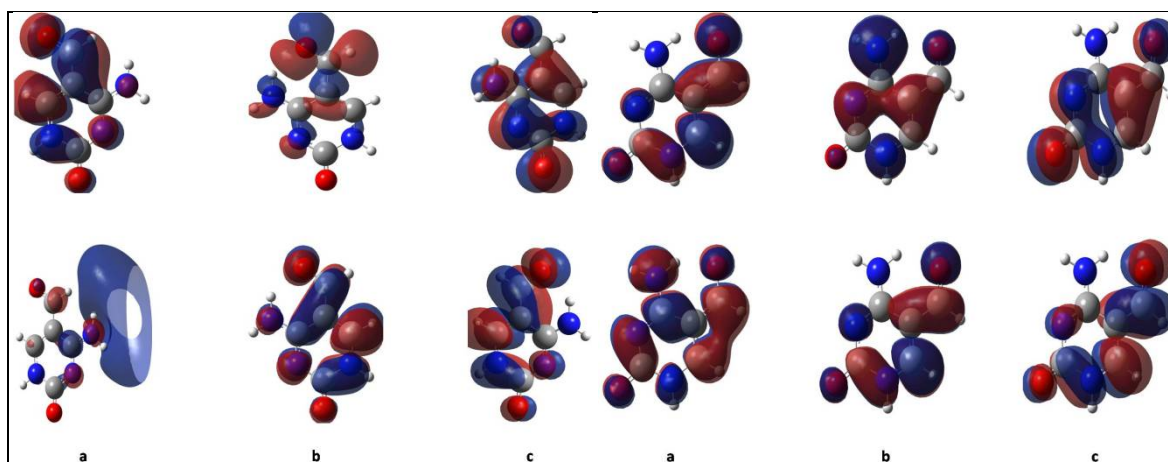


Figure7: HOMO/LUMO orbitals for 5fC

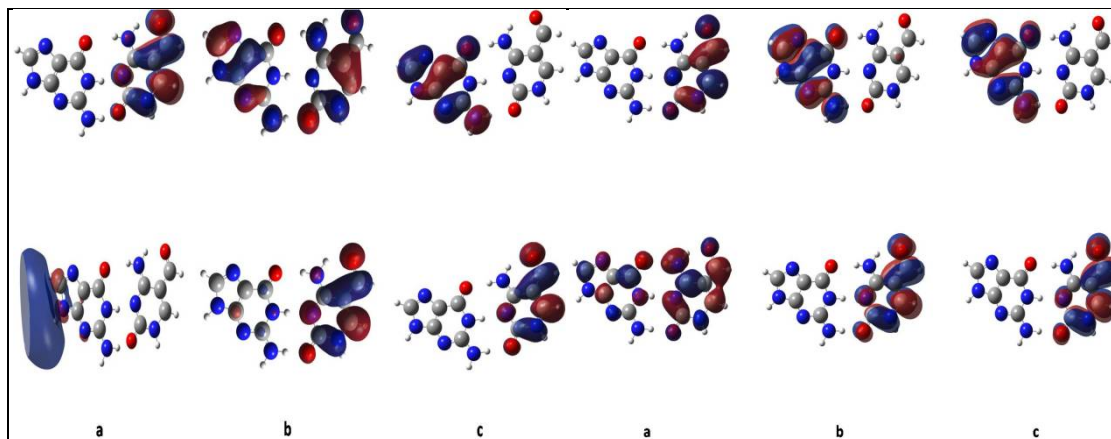


Figure8: HOMO/LUMO orbitals for 5fCG

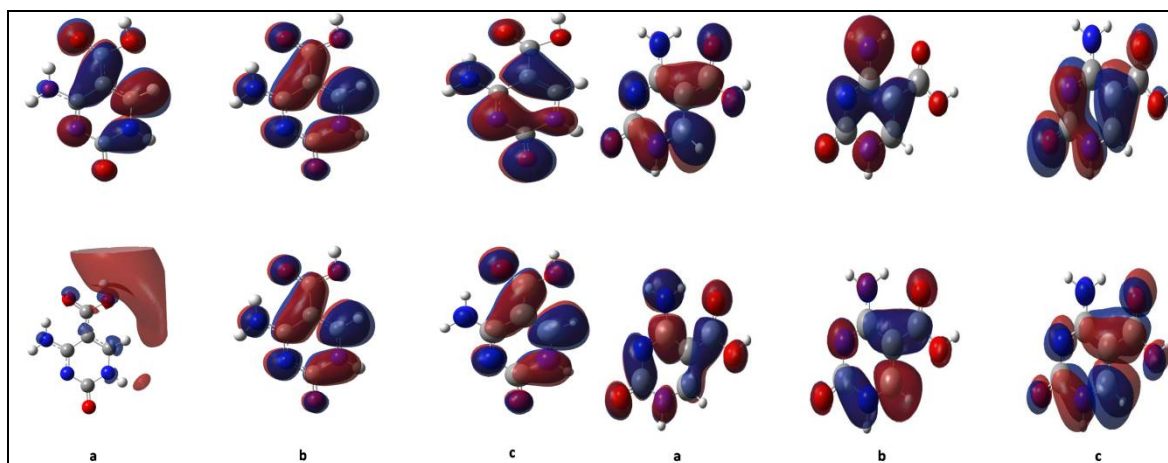


Figure9: HOMO/LUMO orbitals for 5caC

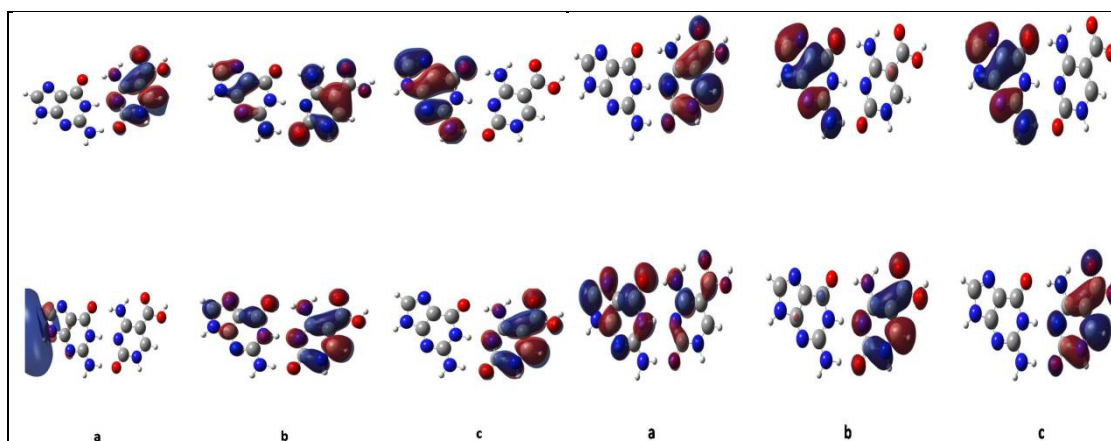
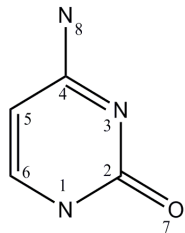
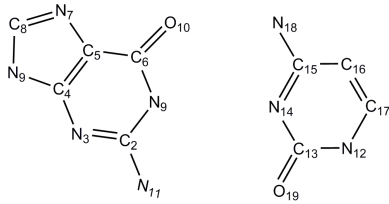


Figure10: HOMO/LUMO orbitals for 5caCG

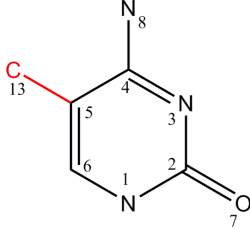
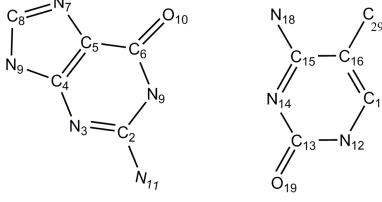
Next, we list the Mulliken charges for the all systems both in gas and in aqueous phase

Cytosine/ CG

	Gas	Aqua
Cytosine 	1 N -0.100327 2 C 0.508491 3 N -0.442786 4 C 0.113877 5 C 0.227940 6 C 0.044671 7 O -0.493123 8 N 0.141257	1 N -0.052207 2 C 0.594144 3 N -0.613403 4 C 0.192394 5 C 0.308208 6 C 0.055794 7 O -0.644735 8 N 0.159805
Cytosine-Guanine 	1 N 0.038575 2 C 0.508179 3 N -0.532267 4 C 0.002799 5 C -0.235908 6 C 0.409507 7 N -0.334806 8 C 0.634449 9 N -0.102076 10 O -0.611007 11 N 0.266238 12 N -0.039071 13 C 0.544321 14 N -0.538134 15 C 0.358422 16 C 0.106375	1 N 0.060468 2 C 0.543058 3 N -0.570407 4 C 0.102984 5 C -0.232277 6 C 0.340018 7 N -0.431726 8 C 0.684479 9 N -0.064349 10 O -0.688284 11 N 0.291156 12 N -0.008094 13 C 0.592214 14 N -0.609547 15 C 0.328156 16 C 0.191624

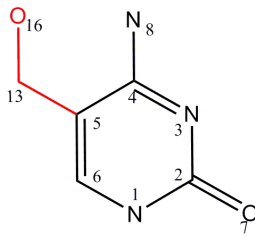
	17 C -0.112848	17 C -0.117415
	18 N 0.211152	18 N 0.220137
	19 O -0.573900	19 O -0.632194

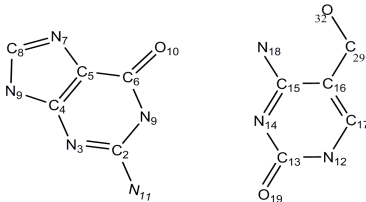
5mC/ 5mCG

	Gas	Aqua
5methyl-Cytosine	1 N -0.087271	1 N -0.039343
	2 C 0.433506	2 C 0.485365
	3 N -0.535692	3 N -0.639402
	4 C 0.556446	4 C 0.579262
	5 C 0.173665	5 C 0.229957
	6 C 0.220011	6 C 0.236472
	7 O -0.508408	7 O -0.643765
	8 N 0.170395	8 N 0.173610
	13 C -0.422653	13 C -0.382155
		
5mCG	1 N 0.028417	1 N 0.060711
	2 C 0.520526	2 C 0.515230
	3 N -0.509801	3 N -0.565909
	4 C 0.047623	4 C 0.146777
	5 C -0.269062	5 C -0.243377
	6 C 0.364854	6 C 0.345773
	7 N -0.311777	7 N -0.431885
	8 C 0.658332	8 C 0.679135
	9 N -0.125434	9 N -0.065041
	10 O -0.605090	10 O -0.682752
	11 N 0.268749	11 N 0.282430
		

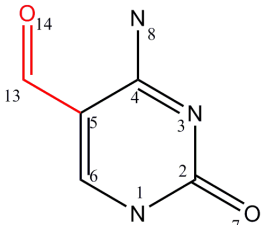
	12 N -0.086697	12 N -0.040324
	13 C 0.563099	13 C 0.588537
	14 N -0.621513	14 N -0.644414
	15 C 0.850734	15 C 0.871998
	16 C 0.088747	16 C 0.082185
	17 C -0.234739	17 C -0.230146
	18 N 0.232824	18 N 0.228929
	19 O -0.548114	19 O -0.629359
	29 C -0.311678	29 C -0.268497

5hmC/5mCG

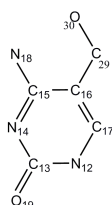
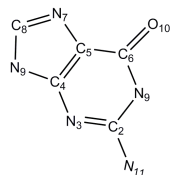
	Gas	Aqua
5Hydroxymethyl	1 N -0.088781	1 N -0.040933
	2 C 0.439424	2 C 0.484013
	3 N -0.531188	3 N -0.638655
	4 C 0.424714	4 C 0.483422
	5 C -0.369114	5 C -0.290555
	6 C 0.563408	6 C 0.509957
	7 O -0.508836	7 O -0.640948
	8 N 0.176694	8 N 0.183139
	13 C 0.043486	13 C 0.120255
	16 O -0.149807	16 O -0.169694
		
5hmCG	1 N 0.024862	1 N 0.054830
	2 C 0.508442	2 C 0.507057
	3 N -0.507281	3 N -0.563204
	4 C 0.065752	4 C 0.163647
	5 C -0.289869	5 C -0.261213

	6 C	0.380120	6 C	0.356446
	7 N	-0.311558	7 N	-0.431072
	8 C	0.660380	8 C	0.681253
	9 N	-0.125918	9 N	-0.065423
	10 O	-0.602867	10 O	-0.680277
	11 N	0.268270	11 N	0.282016
	12 N	-0.101225	12 N	-0.052389
	13 C	0.576115	13 C	0.603412
	14 N	-0.626552	14 N	-0.646388
	15 C	0.887897	15 C	0.938771
	16 C	-0.660834	16 C	-0.657986
	17 C	0.139307	17 C	0.042108
	18 N	0.236503	18 N	0.242820
	19 O	-0.545990	19 O	-0.624258
	29 C	0.165358	29 C	0.270092
	32 O	-0.140913	32 O	-0.160242

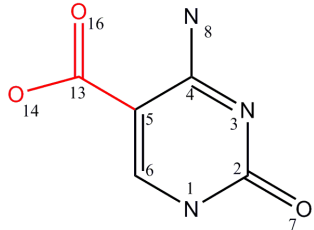
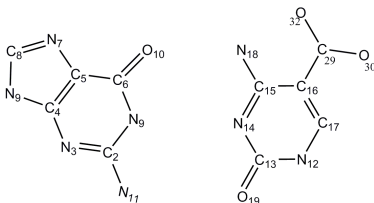
5fC/5fCG

	Gas	Aqua
5fomilcytosine	1 N -0.064740	1 N 0.001361
	2 C 0.435753	2 C 0.490393
	3 N -0.539800	3 N -0.616082
	4 C 0.766018	4 C 0.736461
	5 C 0.212789	5 C 0.317954
	6 C 0.059397	6 C 0.079975
	7 O -0.476710	7 O -0.592095

	8 N 0.200873 13 C -0.149989 15 O -0.443590	8 N 0.214635 13 C -0.129202 15 O -0.503400
5fCG	1 N 0.009487 2 C 0.533961 3 N -0.505705 4 C 0.044218 5 C -0.235316 6 C 0.331193 7 N -0.307515 8 C 0.662077 9 N -0.126211 10 O -0.596993 11 N 0.262927 12 N -0.065343 13 C 0.561591 14 N -0.595333 15 C 0.684111 16 C 0.450699 17 C -0.368970 18 N 0.289176 19 O -0.521740 29 C -0.100598 30 O -0.405714	1 N 0.050651 2 C 0.521373 3 N -0.561434 4 C 0.148525 5 C -0.212349 6 C 0.310854 7 N -0.429591 8 C 0.681203 9 N -0.065884 10 O -0.676820 11 N 0.282071 12 N 0.000235 13 C 0.591599 14 N -0.605343 15 C 0.769294 16 C 0.500182 17 C -0.397737 18 N 0.265035 19 O -0.583404 29 C -0.113212 30 O -0.475248



5caC/5caCG

	Gas	Aqua
5caboxylmethyl	1 N -0.063004	1 N -0.001055
	2 C 0.435818	2 C 0.485699
	3 N -0.542331	3 N -0.618166
	4 C 0.519413	4 C 0.467680
	5 C -0.356220	5 C -0.235565
	6 C 0.508674	6 C 0.509455
	7 O -0.481785	7 O -0.599592
	8 N 0.212185	8 N 0.226532
	13 C 0.396349	13 C 0.418561
	14 O -0.093415	14 O -0.077196
	16 O -0.535684	16 O -0.576353
	1 N 0.008505	1 N 0.046352
	2 C 0.515578	2 C 0.506075
	3 N -0.504063	3 N -0.559447
	4 C 0.065616	4 C 0.168204
	5 C -0.272772	5 C -0.245245
	6 C 0.364127	6 C 0.339656
	7 N -0.307911	7 N -0.429800
	8 C 0.660203	8 C 0.679992
	9 N -0.126979	9 N -0.066141
	10 O -0.591917	10 O -0.672215
	11 N 0.263616	11 N 0.281828
	12 N -0.075469	12 N -0.012842
	13 C 0.565554	13 C 0.593932
	14 N -0.591091	14 N -0.603480
		

	15 C 0.504020	15 C 0.520231
	16 C -0.414225	16 C -0.345970
	17 C 0.203280	17 C 0.160078
	18 N 0.301251	18 N 0.281197
	19 O -0.525583	19 O -0.588631
	29 C 0.528472	29 C 0.553180
	30 O -0.072745	30 O -0.057614
	32 O -0.497467	32 O -0.549341



Cite this: *Org. Biomol. Chem.*, 2015, **13**, 6728

Site-specific conjugation of 8-ethynyl-BODIPY to a protein by [2 + 3] cycloaddition†

Marcel Albrecht,^a Andreas Lippach,^a Matthias P. Exner,^b Jihene Jerbi,^c Michael Springborg,^{c,d} Nediljko Budisa^b and Gerhard Wenz^{*a}

We report a straightforward synthesis of 8-ethynyl-BODIPY derivatives and their potential as fluorescent labeling compounds using an alkyne–azide click chemistry approach. The ethynyl substituted BODIPY dyes at the *meso*-position were reacted under Cu⁺ catalysis and mild physiological conditions in organic and biological model systems using benzyl azide and a Barstar protein which was selectively modified by a single amino acid substituted methionine at the N-terminus (Met1) → azidohomoalanine (Aha). Conjugation with the protein and the model azide was indicated by a significant blue shift upon formation of the triazole moiety system, which allowed easy distinction between free and coupled dyes. This blue shift was rationalized by the perpendicular orientation of the triazole relative to the chromophore using time dependent density functional theory (TDDFT) calculations. A full spectroscopic and thermodynamic characterization of the protein revealed that a fluorophore was incorporated without the cross influence of protein stability and functional integrity. Furthermore, model reactions of 8-ethynyl-BODIPY derivatives with benzyl azide under copper-free conditions indicate second order kinetics with high rate constants comparable with those found for the strain-promoted azide–alkyne cycloaddition (SPAAC). In this way, we establish a unique and highly efficient method to introduce alkyne-BODIPY into a protein scaffold potentially useful for diverse applications in areas ranging from fundamental protein dynamics studies to biotechnology or cell biology.

Received 13th March 2015,
Accepted 6th May 2015

DOI: 10.1039/c5ob00505a

www.rsc.org/obc

Introduction

The visualization of enzyme and protein mediated biochemical processes is one of the most challenging tasks in the research field of life sciences. The introduction of fluorophores into proteins is the method of choice as it allows both *in vitro* and *in vivo* real time imaging by commonly used spectroscopic methods with relatively low costs and high selectivities. Fundamental strategies for the combination of fluorescent dyes with proteins include *self-labeling tags* in which a specific peptide sequence of the protein of interest (POI) is directly modified

by the labeling compound, self-labeling protein (SLP) in which the fluorophore contains both the label and the functional moiety that are able to react with the SLP, and enzyme-mediated labeling of tags.¹

A breakthrough in the detection of fluorophores in living cells was achieved through the discovery of fluorescent proteins (FPs) as highly selective and stable protein tags. In this context, the green fluorescent protein (GFP) was the first example which was introduced into the POI by using efficient, genetic approaches.² During the last few decades, based on their easy accessibility various fluorescent proteins with tunable photochemical properties have been employed in cell imaging techniques.³ However, both in *in vivo* and *in vitro* experiments the FP approach often suffers from dramatic drawbacks caused by the relatively large size of the used FPs resulting in a limitation on potential applications. To overcome size-induced side effects, various synthetic methods for the selective binding of chromophores to POIs such as modifications of cysteine containing peptides,⁴ metathesis reactions,⁵ palladium catalyzed cross-coupling reactions⁶ or azide/acetylene click-chemistry functionalizations⁷ have been established.

Based on these general considerations the site-directed and residue-specific co-translational decoration of proteins with

^aOrganic Macromolecular Chemistry, Campus Saarbrücken C4.2, Saarland University, D-66123 Saarbrücken, Germany. E-mail: g.wenz@mx.uni-saarland.de

^bDepartment of Chemistry-Biocatalysis, TU Berlin, Mueller-Breslau-Strasse 10/L1, D-10623 Berlin, Germany

^cPhysical and Theoretical Chemistry, Campus Saarbrücken B2.2, Saarland University, D-66123 Saarbrücken, Germany

^dSchool of Materials Science and Engineering, Tianjin University, Tianjin, 300072, People's Republic of China

†Electronic supplementary information (ESI) available: Procedure for barstar biosynthesis, ¹H and ¹³C NMR spectra and UV-vis spectra of 3. See DOI: 10.1039/c5ob00505a

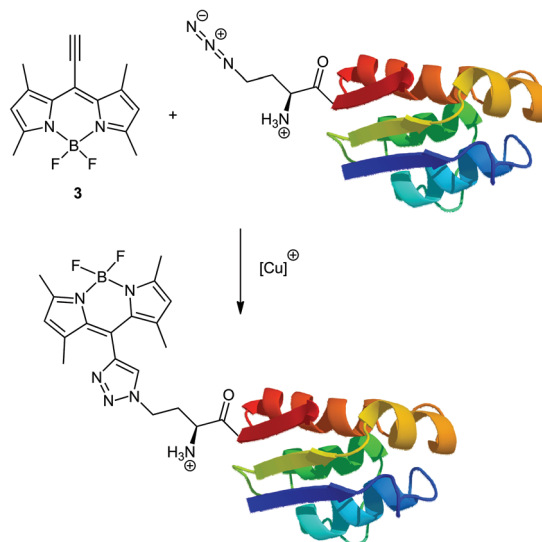


small, easily accessible fluorophores have recently become widely used conjugation approaches in molecular and cellular biology.⁸ The main motivation for these experiments is that such modifications caused by addition of only a few atoms to the amino acid side chain lower the risk of compromising protein structural and functional integrity.⁹ Furthermore, many organic dyes have better photophysical properties and are usually more than 20 times smaller than the widely used fusion-tagged auto-fluorescent proteins.¹⁰ For biological applications, such fluorescent dyes, however, need to be fully biocompatible and meet some specific requirements such as the ability to pass through the plasma membrane and to avoid nonspecific adsorption and cross-reactivity with cellular components and structures. Furthermore, protein conjugation *via* genetic code engineering is in general a two-step process whereby a non-canonical amino acid (ncAA) with a chemically distinct (in an ideal case bio-orthogonal) side chain should be efficiently inserted into a target protein followed by specific coupling of the fluorescent dye.¹¹ In addition to genetic encoding, the reactive side chain functionalities of the dye and ncAA need to be orthogonal to each other (*i.e.*, azides only reacting with alkynes and *vice versa*). In particular, the copper catalyzed [2 + 3] cycloaddition of azides and alkynes provides fundamental advantages compared to other synthetic techniques as it allows a regioselective and economic combination of two reaction partners with high conversion rates under mild reaction conditions.¹²

In particular, 4,4-difluoro-4-bora-3a,4a-diaza-s-indacenes (BODIPYs) are suitable fluorescent dyes for the incorporation into proteins as they are chemically stable compounds with remarkable photophysical properties, such as high quantum yields, sharp absorption and fluorescence peaks and relatively large Stokes shifts.¹³ Based on these results a wide variety of synthetic approaches of protein labelling with BODIPY derivatives have been developed.¹⁴ Furthermore, several azide and ethynyl substituted BODIPY derivatives have been applied in copper catalyzed [2 + 3]-cycloadditions (the so-called click-reaction) for the synthesis and modification of cell imaging probes,¹⁵ fluorescent surfaces,¹⁶ sensor systems,¹⁷ or nanoparticles.¹⁸

In the following we describe a new class of functional BODIPYs, 8-ethynyl-BODIPYs, which show high reactivity and a significant blue shift with regard to the click-reaction. The general proof of principle of the site-specific conjugation of the ethynyl substituted BODIPYs with biologically relevant systems was illustrated using a pseudo-wild type barstar, a small ribonuclease inhibitor of 90 amino acids that is widely used for folding studies, as a model protein (Scheme 1).

Specifically, an engineered cysteine-free 'pseudo-wild type' barstar (ψ -b*), Pro28Ala/Cys41Ala/ Cys83Ala with only one Met residue at the N-terminus (Met1) is used, making the incorporation of methionine analogs and subsequent coupling reaction site-specific. Its N-terminal modification can generally be expected to retain a functional protein structure while introducing novel functions.¹⁹

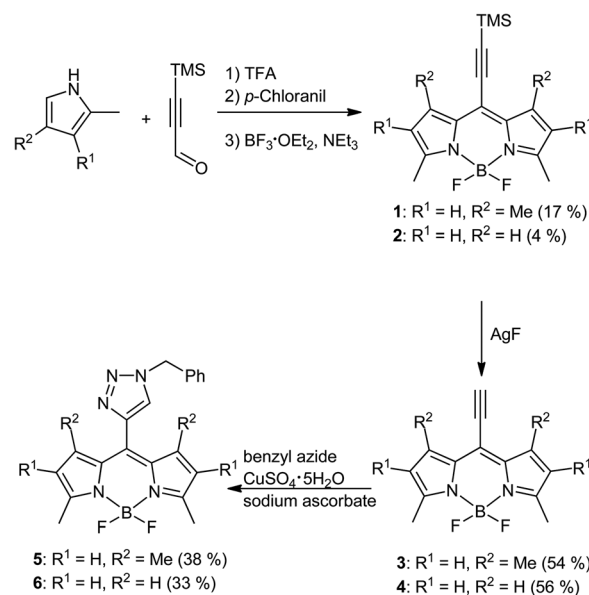


Scheme 1 Bioorthogonal conjugation of 8-ethynyl-BODIPY **3** to pseudo-wild type barstar with azidohomoalanine at position 1.

Results and discussion

Synthesis of 8-ethynyl BODIPY derivatives

TMS protected 8-ethynyl-BODIPYs **3/4** were synthesized starting from 3-trimethylsilylpropynal, 2,4-dimethylpyrrole and 3-ethyl-2,4-dimethylpyrrole by the condensation reaction using trifluoroacetic acid, subsequent oxidation with *p*-chloranil and complexation with boron trifluoride diethyl etherate (Scheme 2) as dark red solids in 17 (**1**) and 4% (**2**) yield, respectively. Due to their instability under basic conditions,



Scheme 2 Syntheses of 8-ethynyl-BODIPYs **3/4** and their conjugation to benzyl azide.



standard deprotection techniques using tetrabutylammonium fluoride completely failed because complete decomposition occurred. In contrast, the mild deprotection agent AgF furnished the 8-ethynyl-BODIPYs **3/4** in yields of up to 56%. As an alternative to the literature known synthesis of **3** through Stille coupling of the corresponding 8-chloro BODIPY,²⁰ we describe a more straightforward synthetic strategy for 8-ethynyl BODIPYs in only two steps from commercially available starting materials.

Copper catalyzed model reactions

To get some primary information about their reactivities in azide-alkyne cycloadditions, model reactions of BODIPYs **3/4** with benzyl azide were performed under copper catalyzed reaction conditions. The progress of each reaction was analyzed through both ¹H NMR measurements of the crude reaction mixture and through isolation of the reaction products **5/6**. As expected for copper catalyzed click-reactions only the formation of the 1,4-triazole regioisomer indicated by one singlet of the triazole moiety at 7.47 (**5**) and 7.45 ppm (**6**) was observed. In general, due to the high affinity of BODIPY derivatives towards silica gel during the purification process significantly lower yields of the isolated products were obtained as determined by NMR. For example, the conjugate **5** was isolated after a reaction time of 3 d at 50 °C in 38% yield while the yield based on the NMR measurements was 52%.

Photophysical properties of BODIPY derivatives

The progress of the click-reaction of the BODIPY derivatives was also observed through the changes within the absorption and fluorescence spectra since the reactive ethynyl group is conjugated with the BODIPY chromophore (Table 1 and Fig. 1). In comparison with the 8-ethynyl-BODIPY **3**, the 8-triazolyl-BODIPY **5** showed hypsochromic shifts both in their absorption and emission spectra which can be explained by prevention of the conjugation between the triazole and the BODIPY moiety in **5** due to steric hindrance of coplanar alignment of triazole and the BODIPY core. This is in contrast to the literature known substituted BODIPYs in which an introduction of a triazole moiety at the pyrrole positions induces a red-shift in absorption and emission.^{13c} Surprisingly, a similar but weaker shift was also found for triazole derivative **6** bearing no bulky methyl groups at the 1- and 7-positions of the BODIPY moiety. The quantum yields ϕ of compounds **3-6** are generally high which is typical for BODIPY derivatives.¹³

Table 1 Absorption maxima λ_{abs} , emission maxima λ_{em} and quantum yields ϕ of the 8-ethynyl-BODIPY (**3/4**) and triazole BODIPY (**5/6**) derivatives in ethanol solution, excitation at absorption maximum each

BODIPY derivatives	λ_{abs} [nm]	λ_{em} [nm]	Stokes shift [nm]	ϕ
3	540	550	10	0.61
4	540	551	11	0.64
5	506	522	16	0.18
6	520	540	20	0.75

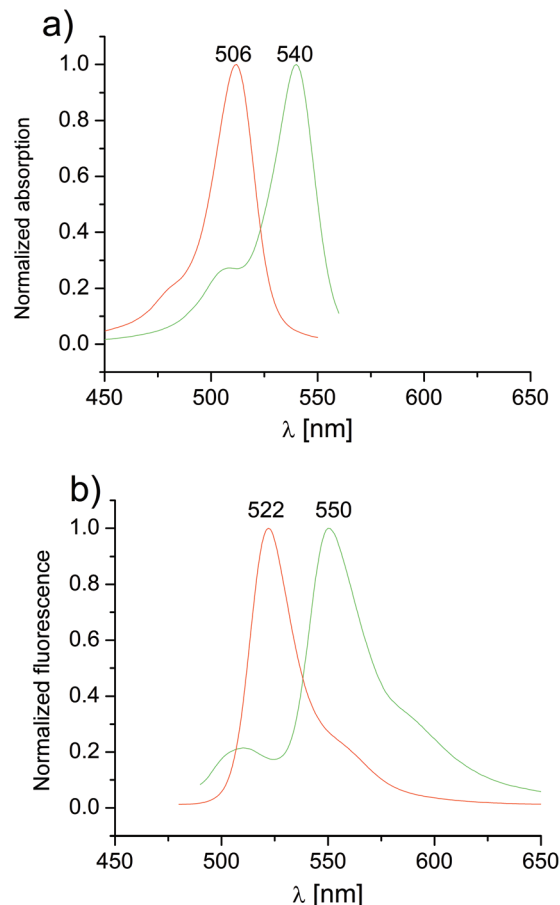


Fig. 1 Absorption (a) and emission (b) spectra of — 8-ethynyl-BODIPY **3** and the corresponding — triazole substituted BODIPY **5** in ethanol solution, excitations at 500 nm (**3**) and 481 nm (**5**), respectively.

Interestingly, quantum yield of the methyl substituted BODIPY **5** is significantly lower than that of **6**. This fact is in contrast to the literature known observation in which an increased restriction at the *meso*-position is connected with higher quantum yields.²¹ We consider that twisted internal charge transfer (TICT)²² between the BODIPY core and the triazole ring might be responsible for this abnormal spectral characteristic. Depending on the conformation of both parts of the molecule the process might be affected by the presence of the methyl groups at the BODIPY moiety at 1- and 7-positions.

Computational analysis

The absorption and emission energies of the ethynyl- and triazole-modified BODIPY derivatives in the ground state in the ethanol environment were calculated using time dependent DFT (TDDFT). The calculations were performed with the Gaussian09 suite of programs.²³ The molecular orbitals were visualized using GaussView05. All geometries were optimized by the B3LYP^{24,25} method with the standard 6-311G* basis set. We optimized the structures of the energetically lowest singlet and



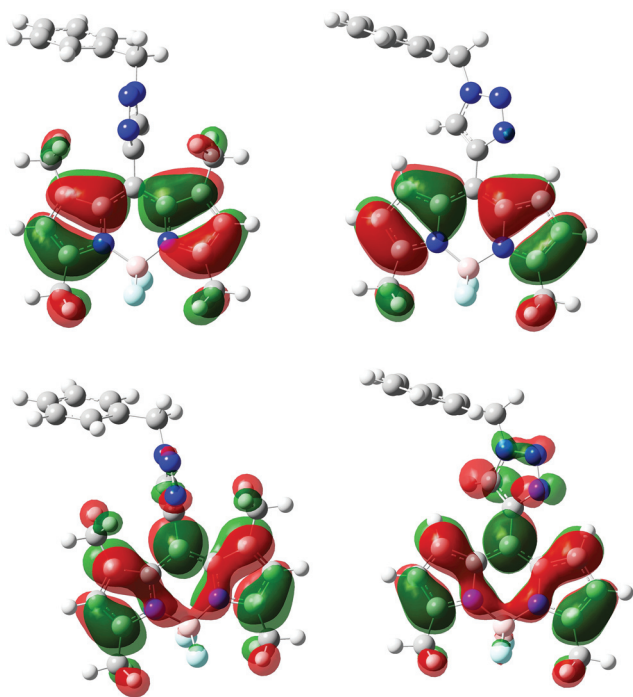


Fig. 2 Calculated HOMOs (top) and LUMOs (bottom) of BODIPY derivatives 5 (left) and 6 (right).

triplet states using the latter as an approximation to that of the second-lowest singlet state.

While the ethynyl substituted BODIPYs 3 and 4 revealed free rotation of the acetylene moiety, the triazole ring of BODIPY 5 shows restricted rotation due to the interaction of the heterocyclic ring with the methyl groups attached to the BODIPY moiety at 1- and 7-positions. In the most stable conformation of 5, the triazole ring is oriented perpendicular to the BODIPY unit, as shown in Fig. 2. As expected, in BODIPY derivative 6, with lacking methyl groups at positions 1 and 7, the triazole ring and the BODIPY unit adopt an energetically favored conformation in which both rings are located in a more planar structure with a relatively small torsion angle to each other.

Subsequently, using TDDFT calculations we determine the excitation spectra for BODIPYs 3/4 and 5/6 starting from the electronic ground state. Thereby the structure of the energetically lowest triplet state is used as an approximation to that of the first excited singlet state. Relevant data are specified in Table 2.

Table 2 Calculated and experimental absorption maxima of BODIPYs 3/4 and 5/6

Compound	Calc. λ_{abs} [nm]	Exp. λ_{abs} [nm]
3	482	540
4	476	540
5	458	506
6	455	520

In general, the wavelengths of the calculated absorption maxima were lower than the experimental data, but the calculated hypsochromic shifts from 482 to 458 nm ($\Delta\lambda = 24$ nm) and 476 to 455 nm ($\Delta\lambda = 21$ nm) for the transformations of BODIPYs 3 to 5 and 4 to 6, respectively, are in good agreement with the experimental hypsochromic shifts of 34 nm (3 \rightarrow 5) and 20 nm (4 \rightarrow 6). The high shift for the sterically hindered methyl derivatives 3/5 can be explained through the disruption between the frontier orbitals of the triazole and BODIPY moieties in compound 5. Due to the reduced rotation barrier between the triazole and BODIPY in the non-methylated compound 6 the interconnection of the frontier orbitals is stronger than that in 5 as shown in Fig. 2.

Copper-free cycloadditions

Surprisingly, the regioselective triazole formation was also observed under copper free reaction conditions. As for the copper catalyzed click-reaction of BODIPY 3 the sole formation of the 1,4-regioisomer was evidenced by a single signal for the triazole ring at 7.47 ppm. The reaction kinetics could be followed through UV/vis measurements. A mixture of BODIPY 3 (25 μM), and benzyl azide (15 mM) in ethanol/water (2 : 1 v/v) was incubated in a 10 mm quartz cuvette at 25 $^{\circ}\text{C}$. The reaction progress was visualized by a characteristic decrease in the absorption maximum of the starting material 3 at 540 nm and a simultaneous increase in the absorption band of product 5 at 506 nm (Fig. 3). Furthermore, the appearance of an isosbestic point at 521 nm indicates the formation of one defined species with high conversion rates without the formation of any side products. Similar results were obtained for the copper-free click reaction of BODIPY 4 to the corresponding triazole substituted BODIPY 6.

The apparent pseudo first order rate constants were determined from the decay of the absorption at 540 nm of a 250 μM solution of 3 with various concentrations of excess benzyl

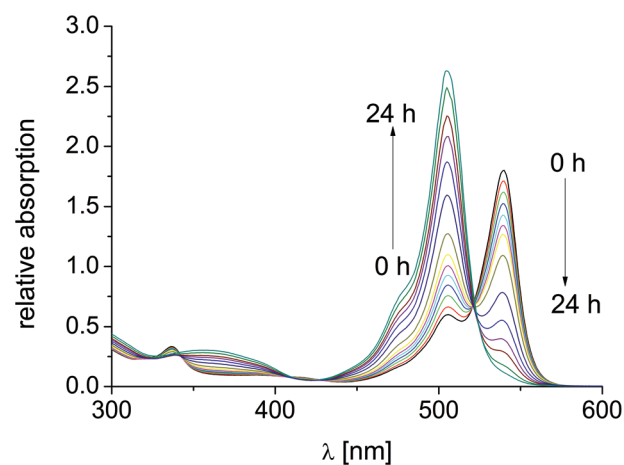


Fig. 3 UV/vis spectra during the reaction of BODIPY 3 (25 μM) and benzyl azide (15 mM) with BODIPY 5 under copper free reaction conditions in water/ethanol (1 : 2 v/v).



azide. Clear second order kinetics was found after plotting the apparent rate constants as a function of the benzyl azide concentration. The obtained second order rate constant k_2 $3.56 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ was even in the same order of magnitude like those found for the strain-promoted azide-alkyne cycloaddition (SPAAC), which usually exhibit values between 3.0×10^{-3} and $4 \text{ M}^{-1} \text{ s}^{-1}$.²⁶ The observed extraordinary reactivity of alkyne **3** may be rationalized by its conjugation with the electron deficient BODIPY core. In addition compared to the conversion of **3** to triazole **5** a ten times smaller rate constant of $3.0 \times 10^{-4} \text{ M}^{-1} \text{ s}^{-1}$ for the reaction of ethynyl BODIPY **4** with benzyl azide was observed under copper-free coupling conditions. The presence of an intensive peak at 509 nm in the normalized absorption spectrum of **4** indicates an increased aggregate formation of BODIPY **4** in aqueous ethanol solution which inhibits the triazole formation (ESI Fig. S13†).

Conjugation with the protein

An azide-labeled protein Aha- ψ -b* was expressed in yields of approximately 50% (5 mg L^{-1}) compared to ψ -b* (10 mg L^{-1}).²⁷ This is significantly better than yields of proteins containing unnatural amino acids incorporated *via* suppression-based methods, which typically do not exceed 20% of the wild-type protein.²⁸ Aha- ψ -b* was purified to >80% with single ion exchange chromatography, as determined by SDS-PAGE, which was sufficient for chemical modification. Conjugation of alkyne-BODIPY **3** and Aha- ψ -b* occurred with 50% total yield after desalting, during which the purity increased to >90% (see Fig. 4A and B). Electrospray mass-spectrometric analysis (ESI-MS) of Aha- ψ -b* clearly revealed a quantitative or near-quantitative level of replacement without detectable amounts

of the parent protein as the contaminant (see Fig. 4C). While the yield is sufficient for most experimental setups, it should be noted that such conjugations usually exhibit efficiencies of approx. 80% (see Li *et al.*, 2010²⁹ for an example). However, as there is no pronounced mass peak corresponding to the starting material Aha- ψ -b*, it can be concluded that the low yield is caused by denaturation of proteins and losses during desalting, and can potentially be increased by optimization of reaction conditions.

It can be seen in Fig. 4D that the conjugate exhibits a slightly decreased thermal stability (the melting point is lowered by 7 °C). However, this effect may be reduced by optimization of expression and reaction conditions.

The expression of the corresponding protein samples was induced with 0.5 mM of Aha and 1 mM of IPTG at $\text{OD}_{600} = 0.7$. After 4 h at 30 °C, cells were harvested. A sample of uninduced and induced culture (corresponding to $\text{OD}_{600} = 1$ in 1 mL) was applied to a 17% SDS-PAGE gel to determine ψ -b* expression. The induced culture shows a distinct band at the expected migration length ($M(\text{Aha-}\psi\text{-b}^*) = 10\,247.9 \text{ Da}$, indicated by the arrow). Protein conjugation was analyzed through standard expression techniques. The samples were run on a 17% tris-glycine SDS-gel. Bands were visualized at 365 nm (Fig. 4B, right) with Coomassie staining (Fig. 4B, left). Only the band corresponding to BODIPY-Aha- ψ -b* exhibits fluorescence without staining (Fig. 4B, right). Furthermore, success in protein conjugation was controlled using ESI-MS measurements of Aha- ψ -b* and BODIPY-Aha- ψ -b*. The mass spectrum of Aha- ψ -b* shows a single peak at 10 246.8 Da (theoretical mass 10 247.9 Da), the spectrum of BODIPY-Aha- ψ -b* exhibits a pronounced peak at 10 518 Da (theoretical mass 10 520 Da). To obtain further the same information about their thermal stability samples of modified and unmodified proteins were thermally denatured from 25 °C (Aha- ψ -b*) or 40 °C (BODIPY-Aha- ψ -b*) to 95 °C. Circular dichroism was monitored at 222 nm (characteristic minimum for α -helices). The melting point only slightly dropped from 62.2 °C for Aha- ψ -b* to 55.2 °C for BODIPY-Aha- ψ -b*.

From the UV/vis and fluorescence spectroscopy data shown in Fig. 5 it can be concluded that conjugation with alkyne-

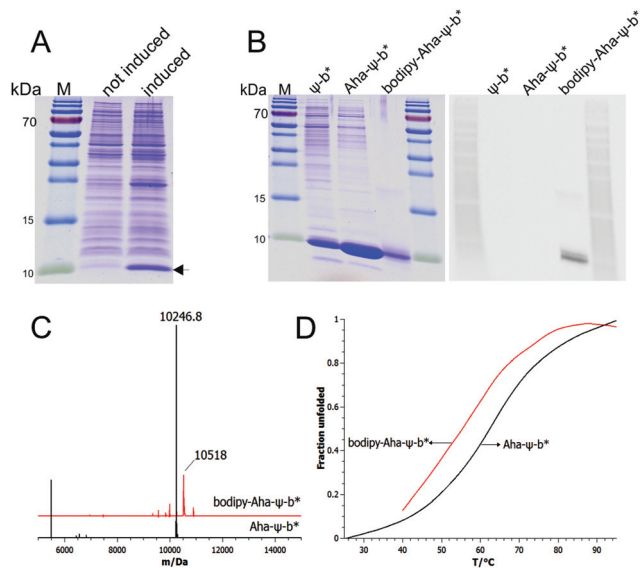


Fig. 4 SDS-PAGE gels of expression of Aha- ψ -b* (A), dito of ψ -b*, Aha- ψ -b* and BODIPY-Aha- ψ -b* with and without Coomassie staining (B), ESI-MS spectra (C) and thermal stability profiles (D) of Aha- ψ -b* and BODIPY-Aha- ψ -b*, respectively.

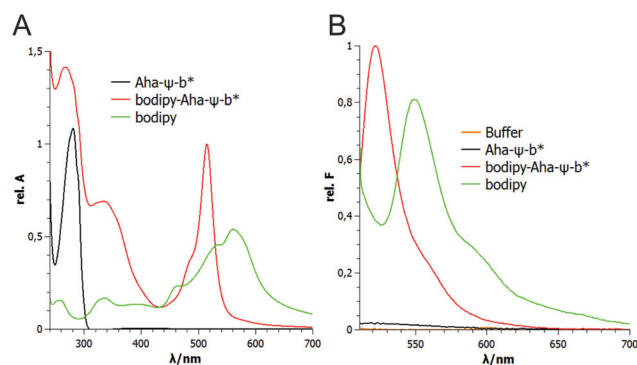


Fig. 5 Absorption (A) and emission (B) spectra of coupling partners and conjugates.



BODIPY 3 endows Aha- ψ -b* with a distinct fluorescence that is significantly blue-shifted compared to the unconjugated BODIPY. These results are comparable with those found for the model reactions (see Fig. 1).

As fluorescence tagging of proteins is very desirable, several attempts have been made to modify amino acids that have been installed *via* suppression techniques,^{30,31} or to directly incorporate fluorescent amino acids.³² However, a full spectroscopic characterization of proteins with installed fluorophores is rarely pursued, probably due to the low amounts of substances that can be typically obtained by stop-codon suppression.

In the absorption spectra both BODIPY-Aha- ψ -b* and Aha- ψ -b* show a peak at 280 nm, as expected for proteins (see Fig. 5A). BODIPY-Aha- ψ -b*, however, shows an additional peak at 508 nm, which is in correspondence with conjugated BODIPY and clearly differs from the spectrum of unconjugated BODIPY 3, which shows a broad maximum between 540 and 570 nm. Similar results were observed in the corresponding emission spectra of all coupling components. The fluorescence was determined with an excitation wavelength of 500 nm. A background spectrum of Tris buffer at pH 8 is also shown. Aha- ψ -b* shows a minimal emission signal due to light scattering. BODIPY-Aha- ψ -b* exhibits a maximum at the expected wavelength (522 nm) that is blue-shifted from the emission of unconjugated BODIPY 3 at 550 nm.

Self-aggregation of BODIPY 3 was demonstrated by the appearance of a UV/vis absorption band at 509 nm. With an increasing content of an organic solvent this band decreased and an increasing absorption of the free dye at 540 nm appeared (ESI Fig. S12[†]). Significant amounts of the free dye are necessary for the copper-free coupling reaction. The model reactions already revealed that a content of 66 vol% of ethanol in water is required for a sufficient coupling rate. On the other hand this solvent is highly non-polar for the Barstar protein leading to its denaturation. The copper-catalyzed click-reaction was successful, because it is much faster and requires much lower concentrations of the free dye than the copper-free coupling. Currently we are looking for ethynyl-BODIPY derivatives with improved solubility in water which would allow conjugation of proteins under copper-free physiological conditions.

Conclusion

8-Ethynyl-BODIPY 3 is a very promising new fluorescence label which allows site-specific conjugation with and without copper catalysis. The induced blue shift provides a simple test for the success of the coupling which can be monitored using optical spectroscopy techniques. As an example for its potential as a fluorescent probe the protein Aha- ψ -b* was successfully conjugated with alkyne-BODIPY 3 *via* a co-translationally introduced Aha. While only specific in the absence of methionine codons other than the start codon, this constitutes a general protein modification technique, as internal methionines are rarely structurally or functionally relevant. Based on

these facts the designed BODIPYs 3/4 are promising labeling components in various biological systems under copper free conditions with direct control over the formation of the resulting bioconjugates.

Experimental

Synthesis of ethynyl-BODIPYs

A mixture of 3-trimethylsilylpropynal (1.76 mL, 12 mmol) and 2,4-dimethylpyrrole (2.47 mL, 24 mmol) dissolved in 50 mL dry dichloromethane was treated with a catalytic amount (3 drops) of trifluoroacetic acid. After stirring at RT for 18 h *p*-chloranil (2.95 g, 12 mmol) was added and the resulting mixture was stirred at RT for 4 h. Then triethylamine (6.1 mL, 44 mmol) was added and stirring was continued for an additional 45 min. After addition of boron trifluoride diethyl etherate (6.0 mL, 48 mmol) the mixture was stirred for 1 h. The solvent was evaporated under reduced pressure and the crude reaction product was dissolved in a mixture of water and diethyl ether. The aqueous phase was separated and extracted with diethyl ether (2 \times). The combined organic fractions were dried over magnesium sulfate and purified by column chromatography (SiO₂, *n*-hexane/dichloromethane 1 : 1), trimethylsilyl-ethynyl-BODIPY 1: a dark red solid (702 mg, 17%), ¹H NMR (CDCl₃): δ = 6.07 (s, 2H, pyrrole-H), 2.53 (s, 6H, methyl-H), 2.45 (s, 6H, methyl-H), 0.30 (s, 9H, TMS-H) ppm. HRMS: calc. for C₁₆H₁₉N₂BF₂Si: 316.1379; found: 316.1385.

2 was synthesized according to the procedure for the preparation of 1 using 2-methylpyrrole as the starting material: trimethylsilyl-ethynyl-BODIPY 2: a dark red solid (151 mg, 4%), ¹H NMR (CDCl₃): δ = 7.13 (d, *J* = 4.1 Hz, 2H, pyrrole-H), 6.25 (d, *J* = 4.1 Hz, 2H, pyrrole-H), 2.60 (s, 6H, methyl-H), 0.31 (s, 9H, TMS-H) ppm. ¹³C NMR (CDCl₃): δ = 157.9, 150.6, 136.5, 128.6, 119.3, 109.1, 97.8, 14.9, -0.5 ppm. HRMS: calc. for C₁₈H₂₃N₂BF₂Si: 344.1692; found: 344.1689.

3: A mixture of trimethylsilyl-ethynyl-BODIPY 1 (172 mg, 0.5 mmol) and silver(I) fluoride (96 mg, 0.75 mmol) dissolved in 5 mL dry methanol was stirred under exclusion of light at RT for 24 h. After the addition of hydrochloric acid (1 M, 2.25 mL, 2.25 mmol) and stirring for an additional 10 min the mixture was filtered. Water was added and the mixture was extracted with ethyl acetate (3 \times). The organic phase was washed with a saturated sodium chloride solution and dried over magnesium sulfate. The solvent was evaporated under reduced pressure and the crude product was purified by column chromatography (SiO₂, *n*-hexane/ethyl acetate 10 : 1), ethynyl-BODIPY 3: a red solid (93 mg, 54%), ¹H NMR (CDCl₃): δ = 6.08 (s, 2H, pyrrole-H), 3.92 (s, 1H, C \equiv CH), 2.54 (s, 6H, methyl-H), 2.46 (s, 6H, methyl-H) ppm. ¹³C NMR (CDCl₃): δ = 155.1, 142.6, 133.3, 121.0, 110.0, 94.4, 79.2, 15.4, 14.6 ppm. IR (ATR): 3265, 2922, 2108, 1540, 1466, 1404, 1306, 1192, 1048, 976, 801, 704 cm⁻¹. HRMS: calc. for C₁₅H₁₅N₂BF₂: 272.1296; found: 272.1286.

4: A mixture of trimethylsilyl-ethynyl-BODIPY 2 (385 mg, 1.22 mmol) and silver(I) fluoride (232 mg, 1.83 mmol)



dissolved in 10 mL dry methanol was stirred under exclusion of light at RT for 24 h. After the addition of hydrochloric acid (1 M, 5.50 mL, 5.50 mmol) and stirring for an additional 10 min the mixture was filtered. Water was added and the mixture was extracted with ethyl acetate (3×). The organic phase was washed with a saturated sodium chloride solution and dried over magnesium sulfate. The solvent was evaporated under reduced pressure and the crude product was purified by column chromatography (SiO₂, dichloromethane), ethynyl-BODIPY 4: a red solid (167 mg, 56%), ¹H NMR (CDCl₃): δ = 7.15 (d, *J* = 4.1 Hz, 2H, pyrrole-H), 6.27 (d, *J* = 4.1 Hz, 2H, pyrrole-H), 3.64 (s, 1H, C≡CH), 2.61 (s, 6H, methyl-H) ppm. ¹³C NMR (CDCl₃): δ = 158.5, 136.7, 128.8, 120.5, 119.7, 88.5, 77.0, 15.0 ppm. IR (ATR): 3253, 2106, 1560, 1429, 1342, 1265, 1223, 1198, 1073, 998, 934, 779, 716 cm⁻¹. HRMS: calc. for C₁₃H₁₁N₂BF₂: 244.0983; found: 244.0982.

Copper(I) catalyzed click-reaction

5/6: A mixture of the corresponding ethynyl-BODIPY (0.145 mmol), benzyl azide (19 mg, 0.145 mmol), copper(II) sulfate pentahydrate (4 mg, 0.015 mmol) and sodium ascorbate (9 mg, 0.045 mmol) in 10 mL dimethyl sulfoxide and 0.1 mL water was stirred at 50 °C for 3 d. Dichloromethane was added and the organic phase was successively washed with a saturated sodium chloride solution and water. The mixture was dried over magnesium sulfate and the crude reaction product was purified by column chromatography (SiO₂, *n*-hexane/ethyl acetate 4 : 1 → 2 : 1).

Triazole substituted BODIPY 5, a dark red solid (22 mg, 38%), ¹H NMR (CDCl₃): δ = 7.47 (s, 1H, triazole-H), 7.38 (m, 3H, phenyl-H), 7.32 (m, 2H, phenyl-H), 5.96 (s, 2H, pyrrole-H), 5.64 (s, 2H, benzyl-H), 2.53 (s, 6H, methyl-H), 1.39 (s, 6H, methyl-H) ppm. ¹³C NMR (CDCl₃): δ = 156.6, 142.7, 140.8, 134.5, 132.3, 129.2, 129.1, 128.0, 123.2, 121.4, 110.0, 54.6, 14.6, 14.2 ppm. IR (ATR): 2923, 1729, 1545, 1507, 1467, 1405, 1307, 1180, 1154, 1044, 971, 816, 711, 697 cm⁻¹. HRMS: calc. for C₂₂H₂₂N₅BF₂: 405.1936; found: 405.1942.

Triazole substituted BODIPY 6, a dark red solid (18 mg, 33%), ¹H NMR (CDCl₃): δ = 7.75 (s, 1H, triazole-H), 7.42 (m, 3H, phenyl-H), 7.35 (m, 2H, phenyl-H), 7.10 (d, *J* = 4.2 Hz, 2H, pyrrole-H), 6.27 (d, *J* = 4.2 Hz, 2H, pyrrole-H), 5.64 (s, 2H, benzyl-H), 2.63 (s, 6H, methyl-H) ppm. ¹³C NMR (CDCl₃): δ = 158.2, 141.8, 133.9, 133.5, 129.9, 129.4, 129.2, 128.2, 125.7, 119.6, 110.0, 54.5, 15.0 ppm. IR (ATR): 2923, 1735, 1551, 1455, 1261, 1181, 1002, 860, 763, 709 cm⁻¹. HRMS: calc. for C₂₀H₁₈N₅BF₂: 377.1623; found: 377.1617.

Copper free click-reaction

5: A mixture of the corresponding ethynyl-BODIPY (0.145 mmol) and benzyl azide (19 mg, 0.145 mmol) in 10 mL dimethyl sulfoxide and 0.1 mL water was stirred at 50 °C for 3 d. Dichloromethane was added and the organic phase was successively washed with a saturated sodium chloride solution and water. The mixture was dried over magnesium sulfate and the crude reaction product was purified by column chromatography (SiO₂, *n*-hexane/ethyl acetate 4 : 1 → 2 : 1).

Triazole substituted BODIPY 5, a dark red solid (14 mg, 24%). The spectroscopic data are identical to those obtained using the copper(I) catalyzed approach.

Triazole substituted BODIPY 6, a dark red solid (21 mg, 35%). The spectroscopic data are identical to those obtained using the copper(I) catalyzed approach.

Expression and purification of ψ-b* and Aha-ψ-b*

The proteins ψ-b* and Aha-ψ-b* were modified as described in the previously described protocol.³³ Details are provided in the ESI†.

Cycloaddition on a model protein

Copper-induced cycloaddition was carried out as follows: 0.2 μg μL⁻¹ Aha-ψ-b* (or ψ-b* as a negative control), 0.02 μg μL⁻¹ alkyne-BODIPY 3, 1 mM CuSO₄, 2 mM TCEP and 2 mM BCA were mixed in PBS at pH 8 (137 mM NaCl, 2.68 mM KCl, 4.3 mM Na₂HPO₄, 1.47 mM KH₂PO₄) in a total volume of 30 μL. A spatula tip of powdered copper was added. The reaction mixture was incubated at 37 °C for 4 h with agitation. Several reaction supernatants were combined and purified on a Zeba Desalt Spin Column (Pierce Biotechnology, Rockford, USA) equilibrated with 50 mM Tris-HCl at pH 8.

Analysis of protein-dye conjugates

MS analysis of full length proteins was performed on an LTQ-FT Ultra mass spectrometer (Thermo Scientific) coupled online to an Ultimate 3000 HPLC Instrument (Thermo Scientific). Desalting was carried out with MassPREP Online Desalting Cartridges (Waters). Briefly, proteins were loaded in 1% formic acid and eluted in a 5 min gradient from 6 to 95% acetonitrile, 1% formic acid. Spectra were acquired in full scan mode with a resolution of 200 000 at *m/z* 400 and afterwards deconvoluted with the software Promass (Thermo Scientific) using basic deconvolution default settings.

The melting curves of Aha-ψ-b* before and after conjugation were measured at a concentration of 20 μM by monitoring the changes in residual ellipticity (*i.e.* unfolding) at 220 nm. The protein solutions were heated from 25 °C or 40 °C to 95 °C with a rate of 30 °C h⁻¹ in 110-QS Hellma quartz cells with an optical path length of 0.1 cm. These experiments were performed on a Jasco J-715 dichrograph equipped with a Peltier-type FDCD and water bath attachment (model PTC-423S/15 and F250).

Fluorescence spectra of samples were recorded on a LS 55 (PerkinElmer Life Sciences, Boston, USA) with an excitation/emission slit of 5 nm. The concentrations of the samples were 0.2 μM for BODIPY-Aha-ψ-b* and Aha-ψ-b*, and 1 μM for BODIPY. The proteins were excited at 500 nm and the fluorescence was measured in the range of 510–700 nm at 20 °C. At least three spectra were accumulated for each sample.

Absorption spectra of samples were recorded on a Lambda 35 (PerkinElmer Life Sciences, Boston, USA) UV/vis spectrophotometer in the range of 250–700 nm at 20 °C.



Analysis of the model reactions

All NMR spectra including ^1H , ^{13}C , H/H-COSY and C/H-COSY were recorded at room temperature by using a Bruker 400 MHz Ultrashield Plus BioSpin Spectrometer Magnet System (^1H : 400 MHz, ^{13}C : 100.6 MHz). The chemical shifts are given in parts per million (ppm) in relation to the corresponding solvent signal.

IR spectra of solid samples were recorded on a Bruker Tensor 27 FT-IR-Spectrometer equipped with a Golden Gate Diamond ATR accessory using the software OPUS by Bruker.

UV kinetics were performed using an Evolution 220 UV/vis Spectrophotometer (Thermo Scientific), 1 mM quartz glass cuvettes (Hellma 110-QS) and ultrapure solvents.

The fluorescence measurements and the determination of the quantum yields were performed according to the literature known procedures established by Jung *et al.*, 2014.^{13c}

Acknowledgements

The authors wish to thank the following persons: Dr Patrick Durkin for synthesis of Aha; Nina Bach and Katja Bäuml from the group of Professor Sieber for mass spectroscopy; Anett Pfeiffer for initial protein expression and conjugation experiments. Blandine Boßmann is gratefully acknowledged for performing the fluorescence measurements. Furthermore, we would like to thank Dennis Meier and Ronny Heisel for their synthetic support for preparing various BODIPY derivatives. Many thanks also to Peter Wiater for his support concerning the DFT calculations. The calculations were carried out at C³M-Saar at Saarland University.

Notes and references

- For recent reviews in the field of tag-mediated protein labeling see: (a) M. J. Hinner and K. Johnsson, *Curr. Opin. Biotechnol.*, 2010, **21**, 766–776; (b) J. M. Chalker, G. J. L. Bernardes and B. G. Davies, *Acc. Chem. Res.*, 2011, **44**, 730–741; (c) C. Jing and V. W. Cornish, *Acc. Chem. Res.*, 2011, **44**, 784–792; (d) D. W. Romanini and V. W. Cornish, *Nat. Chem.*, 2012, **4**, 248–250; (e) Y. Takaoka, A. Ojida and I. Hamachi, *Angew. Chem., Int. Ed.*, 2013, **52**, 4088–4106; (f) D. Jung, K. Min, J. Jung, W. Jang and Y. Kwon, *Mol. Biosyst.*, 2013, **9**, 862–872.
- R. Heim, D. C. Prasher and R. Y. Tsien, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 12501–12504.
- (a) I. V. Yampolsky, A. A. Kislyukhin, T. T. Amatov, D. Shcherbo, V. K. Potapov, S. Lukyanov and K. A. Lukyanov, *Bioorg. Chem.*, 2008, **36**, 96–104; (b) R. M. Wachter, J. L. Watkins and H. Kim, *Biochemistry*, 2010, **49**, 7417–7427; (c) J. Conyard, M. Kondo, I. A. Heisler, G. Jones, A. Badrigde, L. M. Tolbert, K. M. Solntsev and S. R. Meech, *J. Phys. Chem. B*, 2011, **115**, 1571–1577; (d) B. L. Grigorenko, A. V. Nemukhin, I. V. Polykov and A. I. Krylov, *J. Phys. Chem. Lett.*, 2013, **4**, 1743–1747.
- (a) H. Nonaka, S. Fujishima, S. Uchinomiya, A. Ojida and I. Hamachi, *J. Am. Chem. Soc.*, 2010, **132**, 9301–9309; (b) A. Tirat, F. Freuler, T. Stettler, L. M. Mayr and L. Leder, *Int. J. Biol. Macromol.*, 2006, **39**, 66–76.
- (a) Y. A. Lin and B. G. Davis, *Beilstein J. Org. Chem.*, 2010, **6**, 1219–1228; (b) Y. A. Lin, J. M. Chalker and B. G. Davis, *J. Am. Chem. Soc.*, 2010, **132**, 16805–16811.
- (a) V. Böhrsch and C. P. R. Hackenberger, *ChemCatChem*, 2010, **2**, 243–245; (b) J. M. Chalker, C. S. C. Wood and B. G. Davis, *J. Am. Chem. Soc.*, 2009, **131**, 16346–16347; (c) N. Li, R. K. V. Lim, S. Edwardraja and Q. Lin, *J. Am. Chem. Soc.*, 2011, **133**, 15316–15319.
- (a) E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa and T. Carrel, *Angew. Chem., Int. Ed.*, 2012, **51**, 4466–4469; (b) Y. Kim, S. H. Kim, D. Ferracane, J. A. Katzenellenbogen and C. M. Schroeder, *Bioconjugate Chem.*, 2012, **23**, 1891–1901; (c) N. W. Nairn, K. D. Shanebeck, A. Wang, T. J. Graddis, M. P. VanBrunt, K. C. Thornton and K. Grabstein, *Bioconjugate Chem.*, 2012, **23**, 2087–2097; (d) J. Szychowski, A. Mahdavi, J. J. L. Hodas, J. D. Bagert, J. T. Ngo, P. Landgraf, D. C. Dieterich, E. M. Schuman and D. A. Tirrell, *J. Am. Chem. Soc.*, 2010, **132**, 18351–18360; (e) J. T. Ngo and D. A. Tirrell, *Acc. Chem. Res.*, 2011, **44**, 677–685; (f) K. L. Kiick, E. Saxon, D. A. Tirrell and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 19–24.
- J. H. Bae, M. Rubini, G. Jung, G. Wiegand, M. H. J. Seifert, M. K. Azim, J.-S. Kim, A. Zumbusch, T. A. Holak, L. Moroder, R. Huber and N. Budisa, *J. Mol. Biol.*, 2008, **328**, 1071–1081.
- L. Merkel, M. Schauer, G. Antranikian and N. Budisa, *ChemBioChem*, 2010, **11**, 1505–1507.
- S. van de Linde, M. Heilemann and M. Sauer, *Annu. Rev. Phys. Chem.*, 2012, **63**, 519–540.
- S. Lepthien, L. Merkel and N. Budisa, *Angew. Chem., Int. Ed.*, 2010, **49**, 5446–5450.
- (a) V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2002, **41**, 2596–2599; (b) C. W. Tornøe, C. Christensen and M. Medal, *J. Org. Chem.*, 2002, **67**, 3057–3064; (c) J. C. Jewett and C. R. Bertozzi, *Chem. Soc. Rev.*, 2010, **39**, 1272–1279; (d) C. R. Becer, R. Hoogenboom and U. S. Schubert, *Angew. Chem., Int. Ed.*, 2009, **48**, 4900–4908; (e) J. E. Hein and V. V. Fokin, *Chem. Soc. Rev.*, 2010, **39**, 1302–1315.
- (a) A. Loudet and K. Burgess, *Chem. Rev.*, 2007, **107**, 4891–4932; (b) G. Ulrich, R. Ziessel and A. Harriman, *Angew. Chem., Int. Ed.*, 2008, **47**, 1184–1201; (c) M. Wirtz, A. Grüter, P. Rebmann, T. Dier, D. A. Volmer, V. Huch and G. Jung, *Chem. Commun.*, 2014, **50**, 12694–12697.
- (a) S. L. Niu, C. Massif, G. Ulrich, R. Ziessel, P.-Y. Renard and A. Romieu, *Org. Biomol. Chem.*, 2011, **9**, 66–69; (b) T. Kobayashi, T. Komatsu, M. Kamiya, C. Campos, M. González-Gaitán, T. Terai, K. Hanaoka and T. Nagano, *J. Am. Chem. Soc.*, 2012, **134**, 11153–11160; (c) K. E. Wilke, S. Francis and E. E. Carlson, *J. Am. Chem. Soc.*, 2012, **134**, 9150–9153; (d) J.-J. Lee, S.-C. Lee, D. Zhai, Y.-H. Ahn, H. Y. Yeo, Y. L. Tan and Y.-T. Chang, *Chem. Commun.*, 2011,



- 47, 4508–4510; (e) R. Weinstein, J. Kanter, B. Friedman, L. G. Ellies, M. E. Baker and R. Y. Tsien, *Bioconjugate Chem.*, 2013, **24**, 766–771.
- 15 (a) X. Zhang, Y. Xiao, J. Qi, J. Qu, B. Kim, X. Yue and K. B. Belfield, *J. Org. Chem.*, 2013, **78**, 9153–9160; (b) T. Uppal, N. V. S. D. K. Bhupathiraju and M. G. H. Vicente, *Tetrahedron*, 2013, **69**, 4687–4693; (c) C. W. Cunningham, A. Mukhopadhyay, G. H. Lushington, B. S. J. Blagg, T. E. Prisinzano and J. P. Krise, *Mol. Pharm.*, 2010, **7**, 1301–1310.
- 16 A. M. Hansen, A. L. Sewell, R. H. Pedersen, D.-L. Long, N. Gadegaard and R. Marquez, *Tetrahedron*, 2013, **69**, 8527–8533.
- 17 S. Ast, T. Fischer, H. Müller, W. Mickler, M. Schwichtenberg, K. Rurack and H.-J. Holdt, *Chem. – Eur. J.*, 2013, **19**, 2990–3005.
- 18 (a) Q. M. Kainz, A. Schätz, A. Zöpfl, W. J. Stark and O. Reiser, *Chem. Mater.*, 2011, **23**, 3606–3613; (b) S. Zhang, Z. Li, S. Samarajeewa, G. Sun, C. Yang and K. L. Wooley, *J. Am. Chem. Soc.*, 2011, **133**, 11046–11049.
- 19 R. Golbik, G. Fischer and A. R. Fersht, *Protein Sci.*, 1999, **8**, 1505–1514.
- 20 H. Wang, M. G. H. Vicente, F. R. Fronczek and K. M. Smith, *Chem. – Eur. J.*, 2014, **20**, 5064–5074.
- 21 S. Zhui, J. Zhang, G. Vegesna, F.-T. Luo, S. A. Green and H. Lui, *Org. Lett.*, 2011, **13**, 438–441.
- 22 Z. Lou, S. Yang, P. Li, P. Zhou and K. Han, *Phys. Chem. Chem. Phys.*, 2014, **16**, 3749–3756.
- 23 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian, Inc*, Wallingford CT, 2009.
- 24 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 25 C. T. Lee, W. T. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter*, 1988, **37**, 785–789.
- 26 (a) J. A. Codelli, J. M. Baskin, N. J. Agard and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2008, **130**, 11486–11493; (b) M. Chigrinova, C. S. McKay, L.-P. B. Beaulieu, K. A. Udachin, A. M. Beauchemin and J. P. Pezacki, *Org. Biomol. Chem.*, 2013, **11**, 3436–3441; (c) S. V. Orski, G. R. Sheppard, S. Arumugan, R. M. Arnold, V. V. Popik and J. Locklin, *Langmuir*, 2012, **28**, 14693–14702; (d) M. F. Debets, J. S. Prins, D. Merckx, S. S. van Berkel, F. L. van Delft, J. C. M. van Hest and F. P. J. T. Rutjes, *Org. Biomol. Chem.*, 2014, **12**, 5031–5037; (e) C. G. Gordon, J. L. Mackey, J. C. Jewett, E. M. Sletten, K. N. Houk and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2012, **134**, 9199–9208.
- 27 L. Merkel, H. S. G. Beckmann, V. Wittmann and N. Budisa, *ChemBioChem*, 2008, **9**, 1220–1224.
- 28 J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. W. Zhang and P. G. Schultz, *Science*, 2003, **301**, 964–967.
- 29 X. Li, T. Fekner and M. K. Chan, *Chem. – Asian J.*, 2010, **5**, 1765–1769.
- 30 Y. J. Lee, B. Wu, J. E. Raymond, Y. Zeng, X. Q. Fang, K. L. Wooley and W. R. S. Liu, *ACS Chem. Biol.*, 2013, **8**, 1664–1670.
- 31 Y. M. Li, M. Pan, Y. T. Li, Y. C. Huang and Q. X. Guo, *Org. Biomol. Chem.*, 2013, **11**, 2624–2629.
- 32 M. J. Schmidt and D. Summerer, *Angew. Chem., Int. Ed.*, 2013, **52**, 4690–4693.
- 33 Y. Ma, H. Biava, R. Contestabile, N. Budisa and M. L. di Salvo, *Molecules*, 2014, **2**, 1004–1022.

