



universität
wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Simulation of light-control DNA binding and damage“

verfasst von / submitted by

Dott. Dott. mag. Davide Avagliano

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2021/ Vienna 2021

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on the student
record sheet:

A 796605419

Dissertationsgebiet lt. Studienblatt /
field of study as it appears on the student record sheet:

Chemie

Betreut von / Supervisor:

Prof. Leticia González

ABSTRACT

The deoxyribonucleic acid, known as DNA, is one of the most important molecules in our life, as it stores the genetic code and thus the information needed for the functionality of our organism. When DNA interacts with external agents, like small molecules, or stimuli, like light irradiation, it can be damaged, leading to severe problems. At the same time, these very same interactions can be exploited to stop the reproduction of damaged fractions of genetic code and kill those fragments. Consequently, binding the DNA with external molecules and studying the response of the macromolecule to light irradiation represents an active and attractive field of research. The purpose of these studies is not only therapeutical. It is of equal importance to gain deeper understanding about the properties of DNA, as well as to identify and characterise new possible structures, like the so-called G-quadruplex. In the present thesis, different state-of-the-art computational techniques were applied to study DNA binding and damage. Classical and hybrid quantum/classical simulations were used to study the interaction mechanism of a family of compounds, called spiropyran, known to bind the DNA after a light-controlled isomerisation. The binding modes between different derivatives and different DNA structures were characterised. An ad-hoc computational protocol was applied to simulate and characterise the absorption spectra of several ligand-DNA complexes, to understand the role of the ligands in possible light-induced damaging processes. Light-induced damage was also the focus of a joint experimental-computational study, where the calculations provided an important contribution in the understanding of the photo-oxidation in gas phase of DNA derivatives. In summary, this thesis offers a collective insight in different mechanisms of both DNA binding and damage induced by small binders, light irradiation and a combination of the two.

ZUSAMMENFASSUNG

Die Desoxyribonukleinsäure, bekannt als DNA, ist eines der wichtigsten Moleküle für unser Leben, da sie den genetischen Code und damit die notwendigen Informationen für die Funktionalität unseres Organismus speichert. Wenn DNA mit äußeren Stoffen wie kleinen Molekülen oder Reizen wie Lichtbestrahlung wechselwirkt, kann sie beschädigt werden, was zu schwerwiegenden Problemen führen kann. Gleichzeitig kann dieselbe Wechselwirkung dafür genutzt werden, die Reproduktion bereits geschädigter Teile des genetischen Codes zu unterbinden und diese zu eliminieren. Daher stellen sowohl die Bindung von DNA an fremde Moleküle als auch die Untersuchung der Reaktion der Makromoleküle auf äußere Reize wie Lichtbestrahlung ein aktives und attraktives Forschungsfeld dar. Die Ziele dieser Studien sind nicht nur therapeutischer Natur. Von ebenso großer Bedeutung ist es, immer mehr Informationen über die Eigenschaften von der DNA zu gewinnen und zusätzlich neue mögliche Strukturen – wie beispielsweise die sogenannten G-Quadruplexe – zu identifizieren und zu charakterisieren. In der vorliegenden Arbeit wurden verschiedene modernste computerbasierte Methoden angewandt, um DNA-Bindungen und DNA-Schäden zu untersuchen. Klassische und hybrid quantenmechanisch/klassische Simulationen wurden verwendet, um den Wechselwirkungsmechanismus einer Familie von Stoffen, genannt Spiropyran, zu untersuchen, von denen bekannt ist, dass sie nach einer lichtregulierten Isomerisation an die DNA binden. Die Arten der Bindungen zwischen verschiedenen Derivaten und verschiedenen DNA Strukturen wurden charakterisiert. Ein rechnerisches Ad-hoc-Protokoll wurde angewandt, um die Absorptionsspektren von verschiedenen Ligand-DNA Komplexen zu simulieren und zu charakterisieren, um die Rolle der Liganden in möglichen lichtinduzierten Beschädigungsprozessen zu verstehen. Lichtinduzierte Schäden waren auch im Fokus einer gemeinsamen experimentell-rechnerischen Studie, in der die Rechnungen einen wichtigen Beitrag zum Verständnis der Fotooxidation in der Gasphase von DNA Derivaten lieferten. Zusammenfassend bietet diese Arbeit einen kollektiven Einblick in DNA-Bindungen und DNA-Schäden, sowohl durch kleine Binder oder Lichtbestrahlung als auch durch die Kombination der beiden.

Contents

1	Introduction	3
1.1	DNA: The Molecule of Life	3
1.2	DNA Binding	6
1.3	Photoinduced DNA Damage	8
1.4	Spiropyran Derivatives as DNA Binders	10
2	Theory and Methods	13
2.1	Basic Concepts	13
2.1.1	The Schrödinger Equation and the Wavefunction	13
2.1.2	The Born-Oppenheimer Approximation	15
2.2	<i>Ab initio</i> Electronic Structure Methods	16
2.2.1	Hartree-Fock Theory and the Correlation Problem	16
2.2.2	Recovering Electronic Correlation	18
2.2.3	Electronic Structure Methods to Calculate Electronically Excited States	20
2.3	Density Functional Theory	22
2.3.1	Density Functional Theory for Ground State Calculation	22
2.3.2	Time-Dependent DFT for the calculation of excited states	27
2.3.3	Tight-Binding approximation applied to DFT	31
2.4	Force Field Approach to Describe Potential Energy	33
2.5	Hybrid Quantum Mechanics/Molecular Mechanics approach	36
2.6	Molecular dynamics techniques	39
2.6.1	Classical Molecular dynamics	39
2.6.2	Binding Energy Calculation	42
2.6.3	Enhanced Sampling Techniques	43
2.7	Computational strategy to study excited states of DNA-ligand complexes	45
3	Simulation of DNA Binding	49
3.1	Merocyanine binding mechanism to (poly-dAT) ₂	49
3.2	Quinolizidine-substituted Spiropyran Thermal Isomerisation and G-Quadruplex binding	53

4	Simulation of DNA Damage	59
4.1	Simulation of photooxidation of ATP dianion	59
4.2	Absorption spectrum of c-Myc Guanine quadruplex and ligand effect	62
4.3	Excited state characterisation of merocyanine and (poly-dAT) ₂ complexes	65
5	Conclusions and Outlook	69
6	Appendix: Reprinted Publications	71
6.1	DNA-binding Mechanism of Spiropyran Photoswitches: The Role of Electrostatics	72
6.2	Spiropyran Meets Guanine Quadruplexes: Isomerization Mecha- nism and DNA Binding Modes of Quinolizidine-Substituted Spiry- opyran Probes	90
6.3	Site-Specific Photo-oxidation of the Isolated Adenosine-5'-triphosphate Dianion Determined by Photoelectron Imaging	112
6.4	Enhanced Rigidity Changes Ultraviolet Absorption: Effect of a Merocyanine Binder on G-Quadruplex Photophysics	130
6.5	Directional and regioselective hole injection of spiropyran photo- switches intercalated into A/T-duplex DNA	144
	Bibliography	157

List of Acronyms

(poly-dAT)₂	Double stranded poly-deossiadenosine-deossithymidine
1TDM	One electron transition density matrix
A	Adenine
ADC	Algebraic diagrammatic construction
ATP	Adenosine triphosphate
C	Cytosine
CAM-B3LYP	Coulomb-attenuating method - Becke, 3-parameter, Lee-Yang-Parr
CASPT2	Second order multiconfigurational perturbation theory
CASSCF	Complete active space self consisted field
CI	Configuration interaction
CSF	Configuration states functions
CT	Charge transfer
DFT	Density functional theory
DFTB	Density functional tight binding
DNA	Deoxyribonucleic acid
FF	Force field
FWHM	Full width half maximum
G	Guanine
G4	Guanine-quadruplex
GAFF	General Amber force field
GGA	Generalised gradient approximation
GPU	Graphical processor units
HF	Hartree-Fock
LDA	Local density approximation
LR-TD-DFT	Linear response time-dependent density functional theory
MC	Merocyanine
MCH	Protonated Merocyanine
MD	Molecular Dynamics
MM-ISMSA	Molecular Mechanics Implicit Solvent Model Surface Area
MM-PBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
MM	Molecular Mechanics

MP	Møller-Plesset
ONIOM	Our-N-layered Integrated molecular Orbital and Molecular mechanics schemes
QM/MM	Quantum mechanics/ molecular mechanics
QM	Quantum Mechanics
QMC	Quinolizidine-merocyanine
QMCH	Protonated quinolizidine-merocyanine
QSP	Quinolizidine-spiropyran
RC	Reaction coordinate
RCB	Repulsive Coulomb barrier
SCF	Self consisted field
SP	Spiropyran
T	Thymine
TB	Tight binding
TD-DFT	Time-dependent density functional theory
TD	Time-dependent
TDSE	Time-dependent Schrödinger equation
TIP3P	Transferable intermolecular potential with 3 points
TISE	Time-independent Schrödinger equation
US	Umbrella sampling
UV	Ultra-violet
VdW	Van der Waals
vFEP	Variational free energy profile method
WHAM	Weighted histogram analysis method

Introduction

Theoretical and computational chemistry has nowadays become an essential discipline to solve chemical problems. The use of atomistic-scale simulations based on applying mathematical models developed and encoded in computer software, can find many applications, ranging from astro to biochemistry. In this thesis, different computational approaches are applied to contribute to the understanding of the central molecule of our life: the deoxyribonucleic acid, commonly known as DNA. In particular, this thesis focuses on the relationship between DNA, light and DNA binders. First of all, light is used to activate a family of organic molecules, called spiropyran, to bind the macromolecule. Once the binding is established, the effect of light on the complex formed by the DNA and the ligand is investigated. Finally, the direct photo-induced damage is studied with a combination of computational methods and photoelectrons spectroscopy experiments. In its totality, this thesis tries to investigate light-controlled DNA binding and damage with different computational strategies and to give an overview on how to simulate different possible current problem in DNA-light interaction. This work is part of an European Training Network named LightDyNAmics, funded from the European Union's Horizon 2020 Research and Innovation Programme under a Marie Skłodowska-Curie grant agreement. The goal of that project is connecting thirteen experimental and theoretical research groups and ten external partners in eight different European countries, to study and understand unsolved problems in the interaction between light and DNA. This thesis is structured as follows: first the research topic is introduced, then the theories and methods used in the simulation are reviewed, finally the results obtained are reported.

1.1 DNA: The Molecule of Life

In May 1952 Raymond Gosling and Rosalind Franklin obtained a X-ray diffraction pattern, called "Photo 51"^{1,2}, which represents the first image and experimental proof of the DNA structure. This image was part of a series of experiments that showed several conformational and structural parameters^{3,4} of the DNA and opened the door to the great current knowledge of DNA, revolutionising modern

science. Nowadays, scientific knowledge about DNA is well advanced, but its implication for medicine and biology or the possible applications for biomedical⁵ or optoelectronic⁶ purposes are only some of the reason why DNA remains a hot field of research. Indeed, DNA is one of the systems scientists of different fields are most interested in, because it is the storage of the genetic code, the information it contains are at the base of life of any organism, and because its particular complex structure and its properties can be exploited in multiple applications⁷. This intriguing structure and its properties come from how DNA is formed, a sequence of building blocks, called nucleotides (Figure 1.1). Nucleotides are composed of

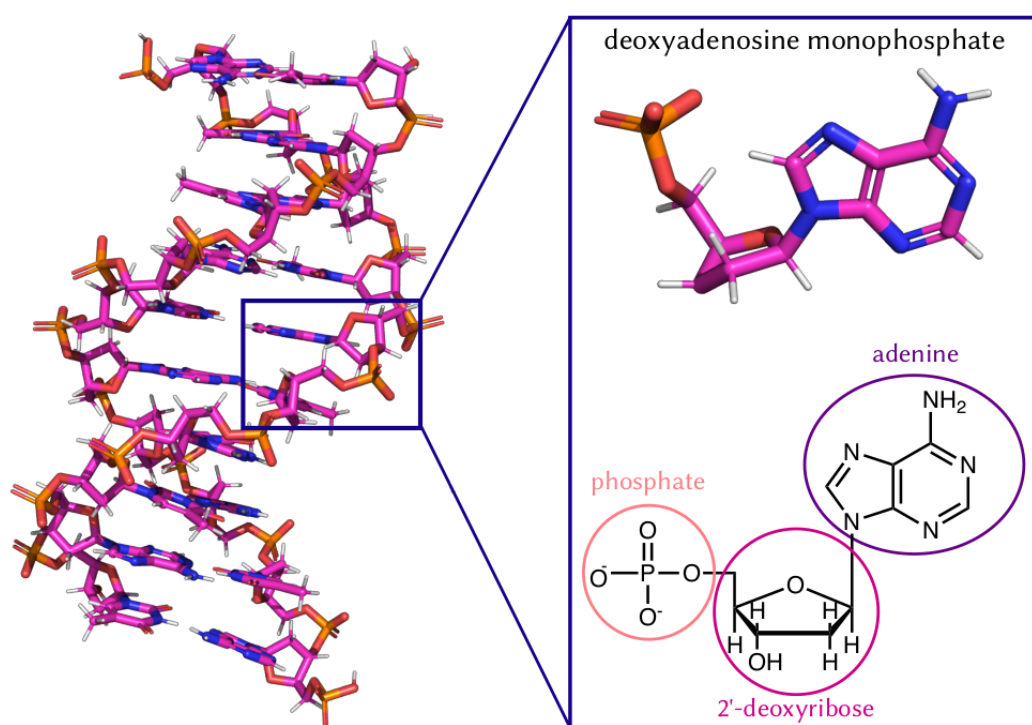


Fig. 1.1: Cartoon representation of double strand DNA. Zoom on one of the four nucleotides forming DNA, a deoxyadenosine monophosphate, represented in its 3D structure and 2D were the base (adenine), the sugare (2-deoxyribose) and the phosphate group are highlighted

a nucleobase, a 2'-deoxyribose sugar and a phosphate group that each base to another one (Figure 1.1). Four possible nucleobases are found in DNA, two purine bases, adenine (A) and guanine (G), and two pyrimidine, thymine (T), and cytosine (C), ones. These bases alternate forming strands, whose sequences represents the primary structure of DNA. Every single strand terminates in one side with a 5'-end, a nucleotide bearing a phosphate group attached to the carbon 5' of the sugar, and with a 3'-end on the other termination, where the chain ends with a hydroxyl group at the third carbon of the sugar. Nucleobases of different strands

interact selectively with each via hydrogen-bonding. In particular, in a canonical DNA structure, an A can pair a T and a G can pair a C (Figure 1.2.a), but other possible bases pairing are possible. As example, Gs can interact with each other with a so-called Hoogsteen base pairing (Figure 1.2.b). Once the nucleobases of

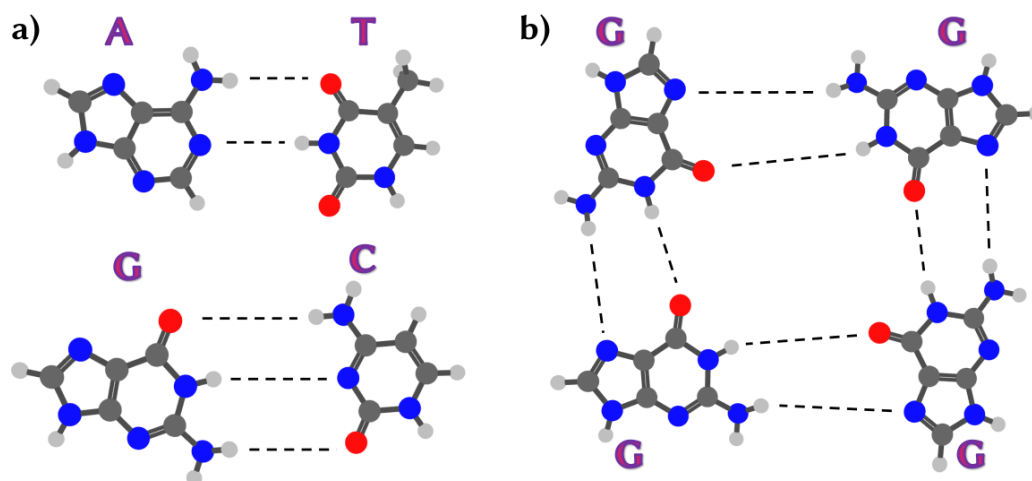


Fig. 1.2: Nucleobase pairing. Colour code: carbon atoms gray, nitrogens blue, oxygens red, H-bond dashed black lines. a) canonical Adenine (A) - Thymine (T) and Guanine (G) - Cytosine (C) pairing ; b) Hoogsteen pairing among four G interacting each others

two strands interact with each other, one of the possible DNA secondary structures is formed. The most common is the well-known double helix B-DNA. The main characteristic of these structures are i) the complementary of the two strands, given by the selective A-T and G-C base pairings; ii) the stacking of the nucleobases of the same strand, due to the interactions of their π electrons clouds, iii) the antiparallel directionality of the two strands with the 3'-end and a 5'-end alternated in the two strands and iv) the formation of two grooves, called major and minor one according to their relative size. Other helix structures can be formed, like the Z-DNA and A-DNA. In G-rich sequences, four Gs can associate in planar tetrads via Hoogsteen-pairing. In the presence of potassium or sodium cations that coordinate them, these tetrads can stack to each other, forming a three dimensional structure called G-quadruplex (G4)⁸. A huge variety of G4 structures have been characterised, according to the sequence and the stacking interactions among the tetrads⁹. These structures have been found to be present in the human telomeric region¹⁰ as well as in protooncogenes¹¹ and genomes of viruses¹². Due to the emerging awareness of the biological involvement and role of these structures, their scientific interest has lately been growing immensely.

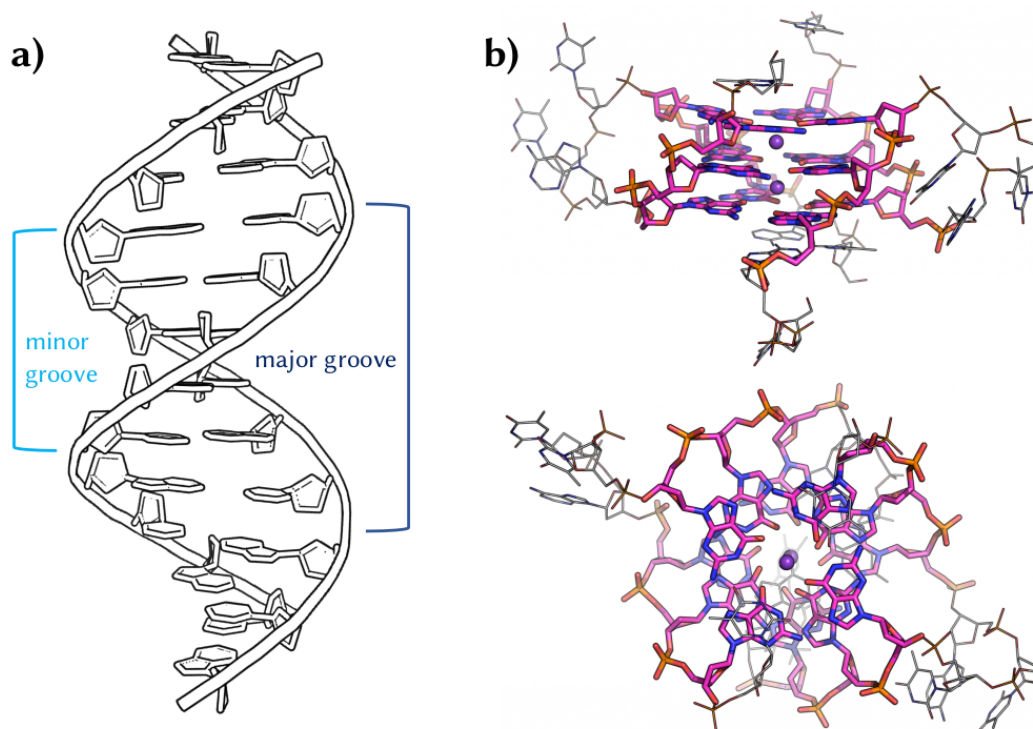


Fig. 1.3: Cartoon representation of B-DNA and G-quadruplex structures. a) B-DNA and minor and major groove highlighted; b) two views of c-Myc G4 structure, PDB I.D. 2L7V. Gs are coloured and other nucleobases are in grey lines; two coordinating K^+ ions are shown in pink spheres

The above-summarised informations are important in the context of the result of this thesis, but clearly do not represent the full knowledge about DNA. In the next sections the importance of probing DNA with small organic molecules and the damage induced by light irradiation on the genetic code will be quickly introduced.

1.2 DNA Binding

The study of the interaction between an external agent and the DNA represents a very broad field of research. Broad in terms of research fields and expertises involved in it, and broad related to number of potential implications and applications. For example, a molecule, with high affinity that the DNA might be exposed to, can poisonously interact with the macromolecule inducing a damage in the genetic code. Despite the bad consequences the interaction with an external molecule can bring to the DNA, its binding can be exploited in a favourable way for im-

portant applications¹³. Binder-DNA interactions can be of different nature. The poly-anionic nature of the DNA backbone offers potential electrostatic interactions with cationic binders as well as the electron-rich π system of the nucleobases can interact with aromatic moieties of external molecules. The double helix structure offers different sites for interaction. The two grooves can be both be attacked by a molecule, but the two different sizes of the grooves lead to different possible interactions (Figure 1.3). The wider major groove represents a good target region for proteins that bind the DNA¹⁴, but it cannot offer strong enough interaction with small molecules. On the opposite, several small drugs have been proposed and applied binding the DNA in the minor groove¹⁵. Another site for interaction between small molecules with the double helix is the space between two base pairs. Indeed, planar aromatic molecule/moieties can approach the DNA from either the major or minor groove, create room between two pairs and stack with the four surrounding nucleobases via π - π stacking interactions. The energy required to stretch the double strand structure is well rewarded by a combination of van der Waals, electrostatic and hydrophobic interactions that make intercalation of molecules into the DNA one of the most stable and exploited binding mode¹⁶. Example of DNA intercalators used for therapeutical application are danuomycin¹⁷ or doxorubicin¹⁸ or several intercalators based on transition metal complexes¹⁹. Indeed, as a source of lesion, DNA can be targeted for pharmaceutical applications with small molecules: whether an external agent or the inner damaged cells' growth are causing a disease, intervening by inhibiting and stopping the reproduction of damaged DNA by ligand interaction has been shown to be an efficient therapeutical action²⁰. Clearly, from synthesising and proposing a potential DNA binder until the clinical applications, many research and testing steps are needed as well as different expertises. Computational chemistry is definitely one of these, since computer simulations can offer atomistic models of DNA-ligand complexes, can estimate the energies of interactions with several possible molecules faster and cheaper than experimentally, and can look in detail at specific physical and chemical interactions that may or may not be favourable for the binding. Besides pharmaceutical applications, DNA binding can be used for other applications. A big scientific debate evolves around the electrochemical definition of DNA. According to condition, structure etc., DNA can behave as an insulator as well as a superconductor²¹. Without entering in the details of this controversial topic, several works focused on studying electronic properties of the macromolecule using intercalators, which can be both electron donor and acceptor. For example, oxidative damage of the genetic code was observed after a long-distance charge migration induced by an oxidant intercalated into the double helix²². DNA-ligand

binding can also be a powerful recognition tool. A ligand can be only selective for a certain DNA primary, due to higher affinity to one pair of nucleobases over the other²³, or secondary structure, due to specific interactions with differently folded strands²⁴. These specific interactions are highly used for diagnostic applications, where the successful binding of a specific ligand indicates the presence of a certain structure. These are only some examples of the importance of developing and optimising DNA binders and motivate the main part of this work, focused on studying DNA-ligand interaction and on the characterisation of DNA-ligand complexes.

1.3 Photoinduced DNA Damage

DNA components, the nucleobases can absorb UV light via their π systems. As our bodies are highly exposed to UV radiation coming from the Sun, one of the reasons behind the mystery of the natural choice of those four nucleobases, among possible others, to compose our genetic code might be the high efficiency with which the absorbed radiation does not produce any negative effect in the DNA²⁵. Indeed, after absorption of UV light, the nucleobases have efficient deactivation mechanisms that prevent detrimental photoreactions^{26,27}. Nonetheless, stable products of photochemical reactions can happen with very small quantum yields ($< 1\%$), but still inducing lesions in the genetic code and leading to mutations and/or carcinomas²⁸. Examples of lesions are the formation of radical after guanine oxidation²⁹ or the formation of cyclobutane pyrimidine dimers³⁰. The seriousness of the UV-induced damage to the DNA strongly necessitates an in-depth knowledge of the optical properties of the macromolecule and its components. The photochemistry of such a complex multi-chromophoric system as DNA is highly complicated and changes if a single base, a single strand, a double strand or a non-canonical secondary structure is considered. Gas phase studies, as reported in this work as well, are fundamental to understand the basis of the mechanism of damaging of DNA components³¹. Additionally, it is important to focus on the effect of the base pairing and/or the hydrogen bonding to the photophysics of DNA³². In particular, once the nucleobases stack to each other and their electronic systems interact, the electronic nature of the excited states changes drastically. In a stranded DNA in addition to local excitation involving only one chromophore, it is possible to find collective excitations involving more than one nucleobases (Figure 1.4). When more bases absorb light in a concerted way, the resultant excitation can be seen as a combination of local excitations as well as highly delocalised excitation, and due

to their electronic nature and features, in analogy with solid-state physics³³, are called exciton state (Figure 1.4.b). Another possible scenario after light irradiation is the inter-base charge separation. In particular, an electron hole can be formed on one nucleobases while an excited electron can be promoted on a different one. These states are called charge transfer states (CT, Figure 1.4.c). CT states are usually dark, i.e. they cannot be populated directly by light irradiation due to the vanishing transition dipole moment, but they have been shown to play a crucial role in the excited state decays of multimeric DNA and their population represents a key step to understand the photochemistry of DNA structures³⁴. In

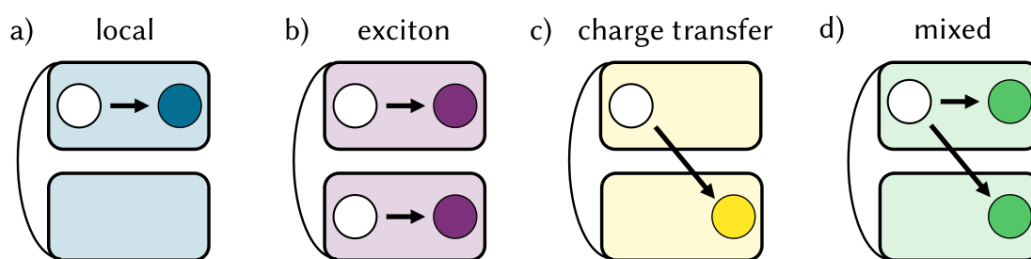


Fig. 1.4: Possible electronic excitations induced by UV light in stranded DNA. a) local excitation where both the electron hole (white circle) and the excited electron are localised on the same nucleobases; b) exciton state, linear combination of local excitations; c) charge transfer state, where the electron hole is formed on one nucleobases but the excited electron is promoted on a different one; d) mixed state, combination of excitonic and charge transfer states

addition to the above described states, it needs to be mentioned that combination of excitonic and charge transfer states have been proposed³⁵ (Figure 1.4.d), further complicating the characterisation and identification of the different states formed in DNA structures following UV irradiation. Both exciton and CT states have been hypothesised and seen experimentally to be involved in the excited state decays²⁶ and possibly responsible in detrimental photochemical pathways²⁸. The complicated nature of the electronic excitations, the large number of chromophores involved, the high conformational flexibility of the DNA structure³⁶ and the effect of the environment³⁷ represent, both if considered individually and, even more, altogether, current challenges in the computational description of the electronically excited states of a full DNA model. First approaches, applied to an adenine-rich single strand, have been successfully employed to look in detail to the nature of electronic excitations after absorption of UV light³⁸ and to unveil the first steps of the excited states relaxation³⁹. In this thesis, an ad-hoc computational strategy (section 2.7) is applied to the study of DNA in the presence of a ligand. This is of particular importance if the ligand is an oxidative chromophore and can absorb light and induce charge separation and migration into the DNA strand.

These interaction can be very useful for electrochemical studies, to measure the electron/hole migration along the DNA or to study the mechanism of induced electron/hole charge separation⁴⁰. Generally, as the ligand-DNA interaction alters physical and chemical properties of the molecule, this can be extended to optical properties that can be better understood by studying the photochemical properties of DNA interaction with chromophores. For all these reasons, in this thesis great attention will be given to the photophysics of DNA-ligand complexes.

1.4 Spiropyran Derivatives as DNA Binders

Spiropyran (SP) is a chemical compound composed by an indoline and and chromene moiety bound together by a spiro carbon (Figure 1.5)⁴¹. It is widely used in chemistry, material science and technology due to its versatility and its chemical and physical properties, being part of the family of molecular switches⁴². Molecular switches are molecules that can be easily interconverted reversibly between two forms that show drastically different properties to be used in chemical contexts⁴³. SP can be converted in its isomeric ring-opened form, merocyanine (MC, Figure 1.5). The equilibrium can be both regulated thermally and controlled by light. Indeed, UV light absorption induces the ring opening, which can be reversibly closed with visible light. Additionally pH, solvent and temperature can induce

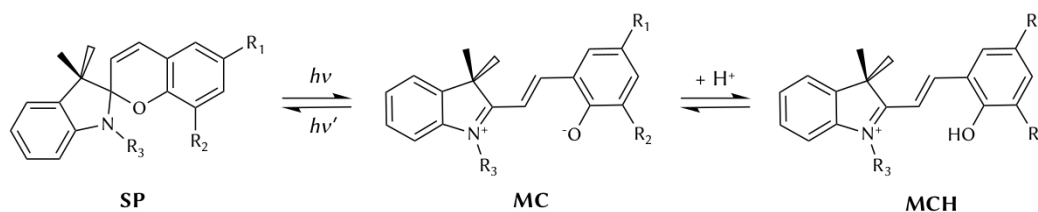


Fig. 1.5: Spiropyran (SP) isomerisation to merocyanine (MC) form and consequential protonation (MCH)

the isomerisation from SP to MC, which is accompanied by colouration⁴⁴. Indeed, SP shows absorption in the UV range and it is consequently colourless, while MC absorbs in the visible showing high colouration. This colour change is the most exploited property of SP, but other important properties are redox behaviour and fluorescence properties⁴², with well separated wavelength of emission that make the two compound easily distinguishable in, for example, biological contexts⁴⁵. All these properties can be enhanced and modulated by interchanging functional groups of the indoline and the especially the chromene moieties are

substituted with⁴². In 2008, Andersson et al. showed DNA-binding properties of a nitro-substituted spiropyran⁴⁶. Interestingly, only the open MC form showed DNA affinity and an intercalative binding mode was proposed according to linear dichroism experiments. That opened the door to more studies about the possibility of binding the DNA with spiropyran derivatives, in a binding interaction that might be controlled externally by light. More than one problem arose from the usage of nitro-substituted spiropyran. First of all, the intersystem crossing, which leads to the population of SP triplet excited state and consequently to the C-O bond breaking⁴⁷, is quenched by the presence of DNA⁴⁸, reducing the isomerisation yields. Additionally, MC undergoes hydrolysis and its degradation decreases the efficiency as DNA binder⁴⁸. Consequently, other derivatives have been proposed, with an amidinium-substituted SP that show stronger resistance to hydrolysis and stronger binding with DNA as well⁴⁹. In particular, it was shown that once the MC is protonated at the phenolic oxygen (MCH form, Figure 1.5) the binding to the macromolecule is enhanced. Additionally, a quinolizidine-substituted spiropyran was synthesised to selectively bind and recognize G4 *in vivo* after thermal isomerisation in the cell⁵⁰. Despite the promising application, the DNA-MCH complexes were never crystallised. Furthermore, computational studies were never adopted to give structural insights on the binding mode, or on the mechanism of action, or the characterisation of the main interactions between the macromolecule and the ligand. In this thesis, the DNA binding properties of several SP derivatives were investigated computationally. The mechanism of action was studied as well as the most favourable binding modes characterised. Additionally, the photochemistry of DNA:MCH complexes was studied in order to understand the implication on the optical properties of the DNA due to the interaction with the ligand.

Theory and Methods

In this chapter, the theory and methodologies used in this thesis are reported. Obviously, the proper understanding of the basic and advanced principles of quantum mechanics requires mathematical concepts that for synthetic reasons cannot be reviewed in this single work. An exhaustive description of the needed linear algebra can be found elsewhere⁵¹. A recommended read about required mathematical concepts applied to quantum chemistry, is the textbook "Modern Quantum Chemistry"⁵². This chapter is structured as the follows: First of all, basic concepts of quantum mechanics will be reviewed; successively, the approximations behind the methods to calculate electronic energies will be introduced, with particular attention to methodologies able to calculate energies and properties of electronically excited state. Furthermore, the force field approach and quantum-classical hybrid methodology to calculate electronic energies in dynamics simulations will be introduced. Finally, different molecular dynamics methods employed in this work will be presented. Hoping that this chapter will give a full overview to understand strengths and limitations of the methods used in this thesis, for a full treatment of the topic are however recommended as additional reading⁵²⁻⁵⁴.

2.1 Basic Concepts

2.1.1 The Schrödinger Equation and the Wavefunction

Whether a single hydrogen atom or a multimeric DNA double strand is considered, a quantum system under the action of a potential is described by the same fundamental equation, the time dependent Schrödinger equation (TDSE)⁵⁵:

$$i\hbar \frac{\partial \Psi(\vec{R}, \vec{r}, t)}{\partial t} = \hat{\mathcal{H}}(\vec{R}, \vec{r}, t) \Psi(\vec{R}, \vec{r}, t) \quad (2.1)$$

This equation introduces another fundamental concept, the position-space wavefunction $\Psi(\vec{R}, \vec{r}, t)$, which represents the quantum physical state of a system, projected in the nuclear (\vec{R}) and electronic (\vec{r}) coordinate space. $\Psi(\vec{R}, \vec{r}, t)$ is a complex

function and thus only its squared modulus $|\Psi(\vec{R}, \vec{r}, t)|^2$ is a real number that represents the probability to find the system in a specific state. Wavefunctions are defined in a specific Hilbert space, where the vectors spanning the space each represent a pure state of the system. Applying the Hamilton operator $\hat{\mathcal{H}}(\vec{R}, \vec{r}, t)$ to $\Psi(\vec{R}, \vec{r}, t)$ gives a set of eigenstates of the system and the correspondent spectrum of eigenvalues, which describe the energy of the system in the given state. The spatial and the temporal contributions of the TDSE can be split if a not explicitly time-dependent potential is considered affecting the system. The spatial part corresponds to the time-independent Schrödinger equation (TISE)

$$\hat{\mathcal{H}}(\vec{R}, \vec{r}) |\Psi(\vec{R}, \vec{r})\rangle = E_{tot} |\Psi(\vec{R}, \vec{r})\rangle \quad (2.2)$$

whose solutions are called stationary states, eigenstates of $\hat{\mathcal{H}}(\vec{R}, \vec{r})$. Solutions of the TISE are eigenvectors of $\Psi(\vec{R}, \vec{r})$ and the energies, as well as all the other properties, are independent of time. The energy is then given by the eigenvalue of the time-independent Hamiltonian $\hat{\mathcal{H}}(\vec{R}, \vec{r})$. In a molecular system, formed by N nuclei and n electrons, $\hat{\mathcal{H}}(\vec{R}, \vec{r})$ contains two kinetic energy terms, the contribution of the nuclei

$$\hat{\mathcal{T}}_{nucl}(\vec{R}) = - \sum_{\alpha=1}^N \frac{\nabla_{\alpha}^2}{2M_{\alpha}} \quad (2.3)$$

and of the electrons

$$\hat{\mathcal{T}}_{el}(\vec{r}) = - \sum_{i=1}^n \frac{\nabla_i^2}{2} \quad (2.4)$$

and three potential energy terms, the nuclear repulsion

$$\hat{\mathcal{V}}_{nucl,nucl}(\vec{R}) = \sum_{\alpha=1}^N \sum_{\beta>\alpha}^N \frac{M_{\alpha}M_{\beta}}{|R_{\alpha} - R_{\beta}|} \quad (2.5)$$

the electronic repulsion

$$\hat{\mathcal{V}}_{el,el}(\vec{r}) = \sum_{i=1}^n \sum_{j>i}^n \frac{1}{|r_i - r_j|} \quad (2.6)$$

and the interaction given by the nuclei-electrons attraction:

$$\hat{\mathcal{V}}_{nucl,el}(\vec{R}, \vec{r}) = - \sum_{\alpha=1}^N \sum_{i>1}^n \frac{M_{\alpha}}{|R_{\alpha} - r_i|} \quad (2.7)$$

which summed up give the definition of $\hat{\mathcal{H}}$:

$$\hat{\mathcal{H}}(\vec{R}, \vec{r}) = \hat{\mathcal{T}}_{nucl}(\vec{R}) + \hat{\mathcal{T}}_{el}(\vec{r}) + \hat{\mathcal{V}}_{nucl,nucl}(\vec{R}) + \hat{\mathcal{V}}_{el,el}(\vec{r}) + \hat{\mathcal{V}}_{nucl,el}(\vec{R}, \vec{r}) \quad (2.8)$$

In the equations listed before, atomic units are used, α and β are indices for nuclei, i and j for electrons, M is the nuclear mass. Nonetheless, analytical solutions of TISE cannot be given for multi-electronic systems neither can an exact numerical solution be found. The work of a theoretical and computational chemist is then focused on finding appropriate and meaningful approximation to solve the TISE for the chemical system of interest.

2.1.2 The Born-Oppenheimer Approximation

Describing the motion of coupled N nuclei and n electron is a highly complex problem from both a mathematical and numerical point of view. However, since the mass of the electron is substantially smaller than the nuclei one (more than three orders of magnitude), but the forces acting on them are comparable, it is reasonable to approximate the nuclei as fixed objects with respect to the electrons. That means the electrons can be considered moving in the field of fixed nuclei, while the nuclei move independently on the potential energy surfaces expressed by the electron energy for a certain nuclear position. This is called Born-Oppenheimer approximation⁵⁶. The uncoupling of the motions allows to split the $\Psi(\vec{R}, \vec{r})$ in two terms:

$$|\Psi(\vec{R}, \vec{r})\rangle = |\Psi^{el}(\vec{r}; R)\rangle |\chi^{nucl}(\vec{R})\rangle \quad (2.9)$$

and to find the solution for the electronic Schrödinger equation:

$$\hat{\mathcal{H}}_{el}(\vec{r}; R) |\Psi^{el}(\vec{r}; R)\rangle = E_{el}(R) |\Psi^{el}(\vec{r}; R)\rangle \quad (2.10)$$

and for the nuclear equation:

$$(T_{\alpha} + E_{el}(\vec{R})) |\chi^{nucl}(\vec{R})\rangle = E_{tot} |\chi^{nucl}(\vec{R})\rangle \quad (2.11)$$

2.2 *Ab initio* Electronic Structure Methods

In this section the basics behind the most common methods for the resolution of the electronic TISE are introduced.

2.2.1 Hartree-Fock Theory and the Correlation Problem

The Hartree Fock (HF) method is central in theoretical chemistry, because it represents the basics of more modern methods. Its cornerstones are⁵²:

- The variational principle is used to find an approximate solution for the TISE. According to the variational principle, a given $|\Psi_0\rangle$ is used to approximate the ground state of a molecule. The expectation value $\langle\Psi_0|H|\Psi_0\rangle$ will be always higher in energy than the real ground state energy. Based on this principle in HF the calculator will try to minimize the energy of the given approximate function in order to get as close as possible to the real energy.
- The trial $|\Psi_0\rangle$ is described by a Slater determinant⁵⁷, whose elements are represented by molecular spin-orbitals. HF is based on finding the best set of spin-orbitals, according to the variational principle. spin-orbitals are the product of a spatial orbital and a spin function. They are built as a linear combination of monoelectronic functions. The collection of these basis functions is called basis set. A basis set can be composed by, for example, Slater-type orbitals or more common, due to their easy computational applicability, by Gaussian-type orbitals. Different families of basis sets have been developed, but all the basis sets are formed by a series of functions to describe core electrons and one or more series of functions to describe valence electrons. The higher the number of valence functions, the closer the energy will be with respect of the real one of the ground state. Polarised and diffuse functions further increase the quality of the basis set. A minimum of a double shell of valence basis and a set of polarised function is needed to obtain reliable energies in any calculation, but the choice of a proper basis set still represents one of the most fundamental tasks for any computation chemist.
- HF is a method based on independent particles. The interaction between one electron and all others is local and it is represented by the instantaneous interaction with the electronic cloud formed by the other electrons. The energy variation in the HF procedure is associated with a monoelectronic

Hamiltonian called Fock operator. It is composed of a sum of monoelectronic and bielectronic operators. The bielectronic operators are the Coulomb \hat{J} and exchange operator \hat{K} . \hat{J} has a classical interpretation and gives the energy of the interaction between an electron and the charged electron cloud. \hat{K} does not have a classical interpretation, but it represents quantum interactions between identical particles that can change their symmetry upon interaction.

- Practically, the energy calculation is based on the self-consistent field procedure (SCF). It is an iterative scheme that, starting from an initial guess minimizes the wavefunctions, calculates the eigenvalues and uses the new wavefunction to start a new iteration cycle, until the difference in energy between the current and previous step satisfies a convergence criteria.
- In the case of two electrons sharing the same orbital, the same spatial function can be used to describe both of them. It is possible to describe any electron with an independent spatial orbital and this leads to the Unrestricted Hartree-Fock method, which uses analogous equation and SCF procedure for the calculation of electronic energies. Unrestricted HF is useful to describe semi-occupied orbitals and open-shell systems.

Being a model based on independent particles, in HF the interaction between electrons is only included through the effect of a field given by the electronic cloud at a certain position. That implies HF intrinsically neglects the energy contribution given by the effective instantaneous interaction between the electrons. This energetic contribution is called dynamic electronic correlation. Although it represents in total a minimal part of the electronic energy (ca. 1-5%), it is a crucial contribution to obtain realistic energies and to study chemical properties since is in the order of magnitude of energy at which chemical reactions happen. Additionally, HF uses a single Slater determinant $|\Psi_0\rangle$ to describe the electronic configuration, restricting each electron to be described by a single spatial orbital and excluding the role of other possible orbitals occupation. The contribution of more configurations includes the so-called static electronic correlation. This is crucial to describe processes where more than one possible orbital occupation has an important weight, such as a bond breaking.

2.2.2 Recovering Electronic Correlation

Configuration Interaction based methods

Several approaches have been developed to include the effect of correlation⁵⁸. The conceptually easiest and most efficient way is to consider $|\Psi_0\rangle$ as a linear combination of Slater determinants, each of them representing a possible configuration. These determinants are obtained from extensions to the virtual orbitals result of the HF calculation on a reference determinant. Those are called configurations state functions (CSF) and can represent single, double, triple etc. excited configurations, according to the degree of occupation of virtual orbitals. This method is called Configuration Interaction⁵⁹. The inclusion of several determinants allows to describe any electronic configuration and, in the limit of using an infinite basis set, the energy calculated is exact. Unfortunately, the computational cost of such a procedure makes it unfeasible for most chemical applications. The more electronic configurations, more mobility and more occupation of anti-bonding orbitals, allow to recover more of the correlation energy missing in HF, but the computational cost becomes at the same time higher. To limit the computational cost, it is possible to only include few selected CSF to $|\Psi_0\rangle$ and simultaneously optimize the molecular orbitals included. Commonly used implementations of this approach are applied to calculate the energy of electronically excited states (section 2.2.3). The inclusion of not a single, but a set of reference Slater determinants to obtain the CSF, leads to the so-called mult-reference methods⁶⁰, and increases even more the quality of the calculation, allowing to fully recover the correlation, but simultaneously increases the computational cost notably. An important deficiency of any multi-configurational method based on truncation of the full CI series is the lack of size-consistency: when two molecular units are at an infinite distance, e.g. in case of a dissociation, the energy of the full molecule does not correspond to the sum of the energies of the two units. Including for example only double excited determinants for the calculation of the energy of two molecular parts, would require an inclusion of quadruple excitations in the description of the full molecule, not considered in a truncation at the second order. Although based on a truncation of CI, Coupled Cluster method solves the problem of the size inconsistency⁶¹. This method uses an exponential cluster operator, which is the sum of the operators relative to single, double etc. excitations.

Perturbation theory based method

Perturbation theory is at the basis of the widely used Møller-Plesset (MPn) methods^{54,62}, applied in this thesis for the calculation of merocyanine isomers energies (section 3.2) and for the study of photo-oxidation of adenosine derivatives (section 4.1). To find a solution of the TISE the Hamiltonian is split in an unperturbed term $\hat{\mathcal{H}}_{el}^0$ and a small perturbation $\hat{\mathcal{P}}$

$$\hat{\mathcal{P}} = \sum_{i < j} r_{ij}^{-1} - \sum_i \sum_j \hat{\mathcal{J}}_j(i) - \hat{\mathcal{K}}_j(i) \quad (2.12)$$

which gives the correlation energy correction. The Hamiltonian is then given by the sum of the two operators

$$\hat{\mathcal{H}}_0 - \hat{\mathcal{H}}_{el} = \lambda \hat{\mathcal{P}} \quad (2.13)$$

where the parameter λ measures the intensity of the perturbation. The Hamiltonian, the wavefunction and the energies can be expanded in the power series of λ . The truncation at the second order gives for example the MP2 method, which is often enough to recover the dynamic correlation. The TISE for a truncation at the second order is given by

$$(\hat{\mathcal{H}}_0 + \lambda \hat{\mathcal{P}})(\Phi^0 + \lambda \Phi^1 + \lambda^2 \Phi^2) = (E^0 + \lambda E^1 + \lambda^2 E^2)(\Phi^0 + \lambda \Phi^1 + \lambda^2 \Phi^2) \quad (2.14)$$

since λ must be the same, for the same power, the correction on the first order to the energy is

$$E^1 = \langle \Phi^0 | \hat{\mathcal{P}} | \Phi^0 \rangle \quad (2.15)$$

which does not apply any correction to the HF energy, and the correction at the second order is

$$E^2 = \langle \Phi^0 | \hat{\mathcal{P}} - E^1 | \Phi^1 \rangle = \langle \Phi^0 | \hat{\mathcal{P}} - E^1 | \sum_i \frac{\langle \Phi^0 | \hat{\mathcal{P}} | \Phi^0 \rangle}{E^0 - E_i^0} \Phi_i^0 \rangle = \sum_i \frac{|\langle \Phi_i^0 | \hat{\mathcal{P}} | \Phi^0 \rangle|^2}{E^0 - E_i^0} \quad (2.16)$$

and the wave function, written as an expansion of eigenstate of \mathcal{H} becomes

$$\Phi^1 = \sum_{i \neq k} \frac{\langle \Phi^0 | \hat{\mathcal{P}} | \Phi^0 \rangle}{E^0 - E_i^0} \Phi_i^0 \quad (2.17)$$

at the first order and so on. For many systems, MP2 represents a very suitable choice of method to calculate the electronic energy. Indeed, it shows a good compromise

between accuracy and computational cost. It allows to recover efficiently a good portion of the dynamic correlation, but due to the mono-determinant nature of the method, it is still lacking in the recovering of static correlation. It represents an excellent choice for geometry optimisation, energy evaluations and calculation of many other molecular properties as long as the system does not need a multi-configurational description.

2.2.3 Electronic Structure Methods to Calculate Electronically Excited States

Multi configurational methods

Studying computationally the photochemical and photophysical properties of a system requires the calculation of electronically excited states. Excited states are intrinsically represented by multiple electronic configurations and, consequently, the solution of the TISE to reliably describe the excited states requires either a multi-configurational approach or an alternative ad-hoc method. Introduced in the previous section, multi-configurational self-consistent space (MCSCF) methods are based on the inclusion of more than one electronic configuration⁶³, where the mixing coefficients of these orbitals as well as the orbitals themselves are optimised in each iteration of the calculation. Due to the absolute computational unfeasibility of applying this approach to the full system, it is possible to restrict the procedure to a determined active space, composed by a fix number of selected electrons and orbitals, and keeping the rest of the molecular orbitals frozen in their configuration. Complete Active Space SCF (CASSCF)⁶⁴ and its Restricted Active Space extension⁶⁵ approaches are based on this procedure. The excited state energy can be calculated optimising the orbital for each of the states required, called state-specific approach, or using a single set of orbitals optimised to minimize the averaged energy of each of the states, the state-average approach. One big problem of CASSCF is the lack of dynamical correlation. Nonetheless, CASSCF wavefunctions can be used as a reference for a more accurate method where the correlation is recovered applying to second order perturbation theory to the reference (CASPT2)⁶⁶. Nowadays, CASPT2 represents the gold standard of accuracy for an excited states calculation, but its cost represents a big limitation to its application. Generally, the active space size, the number of electron and orbitals included in the procedure, represents the bottleneck of this family of methods.

Algebraic Diagrammatic Construction of the Polarization Propagator

An alternative approach to the excited states calculation is represent by studying the time evolution of a multi electronic system, knowing the ground state electron density. This time evolution can be represented by the polarization propagator⁶⁷, a function that expresses the molecular polarizability over time. In the frame of the TDSE a one-electron propagator

$$(\hat{\mathcal{H}} - i\hbar \frac{\partial \Psi(\vec{R}, \vec{r}, t)}{\partial t})K(x, t; x' t') = i\hbar \delta(x - x')\delta(t - t') \quad (2.18)$$

gives the probability of finding a single electron for a certain time in a specific set of coordinates. In a many-body problem, a two-particle propagator needs to be defined, which gives the same probability, but for a pair of interacting electrons. The polarization is given by the reciprocal interaction between the two single-particle transitions that are described in the two-particle propagator function. Introducing the creation \hat{c}^+ and annihilation \hat{c} operators, an intuitive mathematical representation of the polarization propagator is given by the Lehmann representation

$$\Pi_{ab,st}(\epsilon) = \sum_{n \neq 0} \frac{\langle \psi_0 | \hat{c}^+ + \hat{c}_a | \psi_n \rangle \langle \psi_n | \hat{c}_s^+ + \hat{c}_t | \psi_0 \rangle}{\epsilon + E_0 - E_n} + \sum_{n \neq 0} \frac{\langle \psi_0 | \hat{c}_s^+ + \hat{c}_t | \psi_n \rangle \langle \psi_n | \hat{c}^+ + \hat{c}_a | \psi_0 \rangle}{-\epsilon + E_0 - E_n} \quad (2.19)$$

where ψ are the wavefunctions and E the energies of the ground state 0 and excited states n , and the excitation energies ϵ are given by the poles of the first sum ($\Pi_+(\epsilon)$). The second sum has poles at the deexcitations energies and it is neglected in the practical algorithms implemented. The Algebraic Diagrammatic Construction (ADC) scheme calculates the vertical excitation using the non-diagonal form of the PP that contains a matrix corresponding to $\hat{\mathcal{H}}$ and one transition moment matrix^{68,69}. Those matrices can be expanded with diagrammatic perturbation theory, which is based on the Feynman's diagrams used to describe many-body systems. The derivation of the theory can be found in⁷⁰. The application of the diagrammatic perturbation theory allows the construction of algebraic expression for the two matrices and reduces finding the vertical excitations to a linear eigenvalue problem. The expansion can be truncated at the second order (ADC2) or, less common, at the third (ADC3)⁷¹. For the goal of this work, it is important to underline some practical aspect of the ADC procedure. The method requires the knowledge of an initial excited states wavefunction, which can be built as a linear combination of excited states bases. Those are calculated starting from a

ground state wavefunction, usually a MP2 wavefunction. Knowing the ground state wavefunction, an operator containing the n-ple excitation is defined, which acts on the ground state wavefunction and generates the excited states basis. In conclusion is important to underline the strengths and limitations of the method. The method allows the calculation of excited states energies and properties for medium-size systems with a moderate computational cost. The single configurational nature of the methods is also its main deficiency, not being able to recover static correlation. Nonetheless ADC(2) represents a good method to benchmark cheaper approaches^{72,73} and in general to obtain reliable energies, if affordable.

2.3 Density Functional Theory

An alternative approach to calculate electronic energies and molecular properties is based on a very simple concept: the energy of a system, as well as all the other quantities related to it, are functionals of the electron density of the system. This is called Density Functional Theory (DFT)^{54,74–76}. Despite the apparent simplification of abandoning the concept of a wavefunction to rely on an experimental observable, the DFT based methods carry unavoidable limitations which are important to always be considered. Nonetheless, DFT allows the calculation of medium to big size system, in the ground and excited states, and even offers the possibility of calculating hundreds atoms systems with more drastic approximations.

2.3.1 Density Functional Theory for Ground State Calculation

Basics

For a system of n electrons, the probability of finding any of those, with arbitrary spin, in a region of space is given by

$$\rho(\vec{r}) = n \int d\vec{r}_1 \dots d\vec{r}_n |\Psi(\vec{r}_1 \dots \vec{r}_n)|^2 \quad (2.20)$$

which is called electron density. In order to take into account the interaction of quantum electron, the pair density is defined as

$$\rho_2(\vec{r}_1, \vec{r}_2) = n(n-1) \int d\vec{r}_3 \dots d\vec{r}_n |\Psi(\vec{r}_1 \dots \vec{r}_n)|^2 \quad (2.21)$$

and takes into account the probability of finding two electrons with different spin at the same time in a certain region of space. All the effects resulting from the electronic correlation are included in $\rho_2(\vec{r}_1, \vec{r}_2)$. Indeed, the Coulombic repulsion affects the motion of the pairs of electrons. This correlation effect is missing in both HF and DFT theories. Additionally, $\rho_2(\vec{r}_1, \vec{r}_2)$ includes another type of correlation, called exchange correlation, which is given by the antisymmetric nature of the wavefunction and the Pauli exclusion principle. This correlation is included in HF due to the antisymmetric nature of Slater determinants, but missing in DFT. The exchange correlation consists in the null probability of find two indistinguishable particle, two electrons with the same spin, at the same point in space. The conditional probability

$$\Omega(\vec{r}_2; \vec{r}_1) = \frac{\rho_2(\vec{r}_2, \vec{r}_1)}{\rho(\vec{r}_1)} \quad (2.22)$$

returns the probability of find an electron in a point of the space if another electron is in a second point. This can be summed up to the self interaction, to the exchange and the Coulomb correlation, giving a function

$$h_{xc}(\vec{r}_2; \vec{r}_1) = \frac{\rho_2(\vec{r}_2, \vec{r}_1)}{\rho(\vec{r}_1)} - \rho(\vec{r}_2) \quad (2.23)$$

called exchange-correlation hole. This can be seen as a region of space around an electron where the probability of finding another electron is minimised and it will be a central concept in the following derivation of DFT. Indeed, h_{xc} appears in the electronic repulsion term in $\hat{\mathcal{H}}$ as

$$\mathcal{V}_{el-el}(\vec{r}) = \frac{1}{2} \int \int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{\rho_{12}} d\vec{r}_1 d\vec{r}_2 + \frac{1}{2} \int \int \frac{\rho(\vec{r}_1)h_{xc}(\vec{r}_1; \vec{r}_2)}{\rho_{12}} d\vec{r}_1 d\vec{r}_2 \quad (2.24)$$

h_{xc} can be split in its two exchange and Coulombic components, and represents the bottleneck of the whole method, as it does not allow to obtain an exact solution for the electronic energies within the DFT formalism.

The Hohenberg-Kohn theorems

DFT is based on two basic theorems, the Hohenberg-Kohn theorems. The first theorem says that the total energy of the system is functional of the density and that the knowledge of the ground state density is sufficient to know all the properties of the system⁷⁷. The second theorem is equivalent to the variational principle and

states that the lowest energy of the system is given by the functional of the real ground state density. The minimum in energy of the system is then reached when the density is the correct ground state one. The full demonstration of the theorem can be found in the original paper⁷⁷.

The Kohn-Sham approach

Practically, the energy of the system is composed by a kinetic energy, a Coulombic and a correlation term. Out of these terms, only the electrostatic energy can be evaluated directly classically, while the kinetic energy can be estimated, wrongly, with analytical functions like in the Thomas-Fermi method⁷⁸. Kohn and Sham proposed a practical approach to calculate the kinetic energy as a functional of density and in general to calculate the electronic energy of the ground state⁷⁹. The method consists in defining a non-interacting system, formed by independent electrons, where all the interactions among electrons are neglected. The density of this system is set equal to the density of the real system. Setting an equation equivalent to TISE, the Hamiltonian $\hat{\mathcal{H}}_{non}$ of the non-interacting system contains a local potential and a kinetic energy term $T_{non}[\rho(\vec{r})]$, which obviously excluded part of the energy due to the missing inter-particle interaction. Similarly to HF, the wavefunction is expressed as a Slater determinant, whose elements are called Kohn-Sham orbitals and represent the orbital for the non-interacting system. The energy functional becomes

$$E[\rho(\vec{r})] = T_{non}[\rho(\vec{r})] + E_{nucl,el}[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{xc}[\rho(\vec{r})] \quad (2.25)$$

with

$$E_{nucl,el}[\rho(\vec{r})] = - \sum_{\alpha}^N \int \frac{M_{\alpha}(\vec{R})\rho(\vec{r})}{|\vec{R}_{\alpha} - \vec{r}|} d\vec{r} \quad (2.26)$$

and

$$J[\rho(\vec{r})] = \frac{1}{2} = \int \int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2 \quad (2.27)$$

where the leftover of the kinetic energy not considered in $T_{non}[\rho(\vec{r})]$ is simply added to the already unknown term $E_{xc}[\rho(\vec{r})]$. This term is called exchange-correlation functional and contains all the non-classical interactions that makes its form inaccessible and its exact evaluation impossible. In a procedure analogue to the HF method, an iterative scheme based on variational principle is used. Key step is the definition of the local potential V_{non} that should be generated in order to equal the density of the reference system with the true one. V_{non} corresponds to

the potential given by the nuclei interaction, the Coulombic potential and the unknown exchange-correlation potential generated by $E_{xc}[\rho(\vec{r})]$

$$V_{non} = \int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2 - \sum_{\alpha}^N \frac{M_{\alpha}}{r_{1\alpha}} \quad (2.28)$$

The Kohn-Sham approach is in principle exact, knowing the exact form of $E_{xc}[\rho(\vec{r})]$. The impossibility of solving this problem and the need of finding an approximate analytical solution for it, represent the main drawbacks of the full DFT. Several solutions have been proposed and different families of functionals have been developed. The total dependency of the results from the nature of the functional, the ability of some functional to recover some properties excellently, but failing miserably in the description of others, makes it absolutely crucial to know the features of each functional (or better family of functionals) and make a conscious choice of the functional to be used.

Choosing the right functional

$E_{xc}[\rho(\vec{r})]$ and the exchange-correlation hole function are obviously connected. Although $E_{xc}[\rho(\vec{r})]$ contains the correction for the kinetic energy, knowing h_{xc} would mean giving a correct form to $E_{xc}[\rho(\vec{r})]$, since can be demonstrated that

$$E_{xc}[\rho(\vec{r})] = \frac{1}{2} \int \int \frac{\rho(\vec{r}_1) \bar{h}_{xc}(\vec{r}_1; \vec{r}_2)}{r_{12}} \quad (2.29)$$

with

$$\bar{h}_{xc}(\vec{r}_1; \vec{r}_2) \equiv \int_0^1 h_{xc}^k(\vec{r}_1; \vec{r}_2) dk \quad (2.30)$$

with k that indicates the coupling strength between two electrons. Modelling a functional for DFT calculation means modelling $\bar{h}_{xc}(\vec{r}_1; \vec{r}_2)$. The easiest way to describe E_{xc} is to consider the density as an uniform electron gas, as done in the local density approximation (LDA). The exchange and correlation energies depend, in this approximation, only on the density. Even in its most accurate derivation, this leads to poor energies, directly comparable with HF ones. Including the dependency on the derivatives of the density improves remarkably the performance of a functional, as done in the Generalised Gradient Approximation (GGA) functionals. These functionals are generally composed by a formulation of the exchange part and one for the correlation one, and they can contain parametrised functions on semi-empirical values or empirical parameters. Additionally, adding a

Laplacian of the density or the orbital kinetic energy term, leads to the meta-GGA functionals. A very popular family of functionals is represented by the hybrid functionals, which add HF exchange to the series of parameters of exchange and correlation parts of already existing functionals. That improves the performance for many applications, but also increases the computational cost. HF is in theory able to calculate the exchange energy, which represents the main component of the exchange-correlation energy, but it is important to remember that Kohn-Sham orbitals are used and that exchange and correlation holes are not fully independent. For this reason HF exchange can be either included fully or only a fraction of it can be used, like in the range-separated hybrid functionals. In these functionals, the HF exchange is weighted differently according to the distance of the interaction. Once DFT is chosen to solve a chemical problem a few considerations about the functionals need to be assessed: i) the functional should have been developed, or showed reliable results, for the system and the properties of interest; ii) a proper benchmark of possible functionals is needed, or need to be present in literature; iii) chemical and physical conclusion should never be drawn by the simple numerical comparison of results obtained with different functionals, since different families of functionals can perform drastically differently on similar systems.

Main strengths and limitations of the method

The mono-determinant nature of the Kohn-Sham procedure implies all the limitations already listed for other single-references method. The lack of the direct interaction between electrons leads to a wrong estimation of non-covalent interactions, but with the usage of dispersion corrections scheme this issue is recovered. The major drawback is definitely the impossibility of finding an exact functional, which leads to the development and application of multiple analytical expressions to choose from. Indeed, while for pure *ab initio* methods, within the limit of the basis set, the error is systematic and coherent for any system, more attention needs to be given to the choice of the functional and interpretation of the results⁸⁰. DFT is widely used and represents often the first, or the only, choice to solve computationally chemical problems for medium-large systems, due to its low cost and thus its applicability to medium-large system, as can be seen in the large number of studies, benchmarking and applications that are possible to find in literature.

2.3.2 Time-Dependent DFT for the calculation of excited states

The original DFT formalism can be applied only to obtain the eigenvalue corresponding to the ground state density. In order to calculate the excited states energies and properties, it is possible to consider the time evolution of the ground state density under the action of an external potential, like an electromagnetic radiation. Indeed, according to the Time-Dependent DFT (TD-DFT) theory^{81,82}, there is an unique correspondence between a time-dependent (TD) potential where the system evolves and the density of this system $\rho(\vec{r}, t)$. Accordingly, the wavefunction depends on the TD density and the initial Ψ_0

$$\Psi(\vec{r}, t) = \Psi[\rho(\vec{r}, t), \Psi_0](t) + w \quad (2.31)$$

with w an exponential TD phase factor. This is stated and formulated in the first Runge-Gross theorem⁸³, which basically states that the expectation value of any operator can be evaluated by the TD density

$$O(t) = \langle \Psi[\rho(\vec{r}, t), \Psi_0](t) | \hat{O}(t) | \Psi[\rho(\vec{r}, t), \Psi_0](t) \rangle = O[\rho(\vec{r}, t), \Psi_0](t) \quad (2.32)$$

including the Hamiltonian and the transition dipole moment (TDM). This is followed up by a second theorem that states that the variational principle can be applied to obtain the exact TD density. Nonetheless, if the potential suddenly changes at a certain time, this does not effect the density at the previous time along the real-time propagation. This aspect needs to be considered in finding the solution with the variational principle. A stable procedure that keeps trace of the causality between the TD potential and response was proposed by Vignale⁸⁴. On the other way around, the density at any time depends on the time evolution of the density, from the initial state until that moment. This is called memory effect⁸⁵ and it affects the form of the TD $E_{xc}[\rho(\vec{r}, t)]$ functional, which can be expressed as the combination of a time-independent functional and the correction due to the memory effect. Neglecting this last term leads to the adiabatic approximation. Although some TD functionals have been developed^{85,86}, employing the adiabatic approximation and using a ground state functional in the TD-DFT framework still represents the most affordable and widely used approach to calculate excitation energy within TD-DFT theory.

Kohn-Sham theory for TD-DFT

In analogy to the standard DFT formulation, the TD-DFT equations are solved using a TD formulation of the Kohn-Sham approach: a non-interacting system is defined that evolves in time as the real system and the perturbation of its density corresponds to the perturbation of the real density. Analogously to ground state DFT, an iterative procedure is developed, with the inclusion of the external TD potential, which leads to the TD Kohn-Sham equations⁸⁷

$$i\hbar \frac{\partial \phi_i(\vec{r}, t)}{\partial t} = \hat{h}^{KS}[\rho](t) \phi_i(\vec{r}, t) \quad (2.33)$$

The main difference with standard DFT is the TD dependency of the density of \hat{h}^{KS}

$$\hat{h}^{KS}(t) = \hat{t}_{KS} + \hat{v}_{ext}(t) + \hat{v}_H[\rho](t) + \hat{v}_{xc}[\rho](t) \quad (2.34)$$

with \hat{t}_{KS} being the non-interacting kinetic energy operator, $\hat{v}_H[\rho](t)$ the TD Hartree potential

$$\hat{v}_H[\rho](t) = \int \frac{\rho(\vec{r}', t)}{|\vec{r} - \vec{r}'|} \quad (2.35)$$

and $\hat{v}_{ext}[\rho](t)$ and $\hat{v}_{xc}[\rho](t)$ the external and exchange-correlation potentials, respectively.

Linear-Response TD-DFT

Two approaches have been proposed to solve the KS equations for TD-DFT. The first one is to integrate the KS equations in the real-time using a time-evolution operator, called Real-Time TD-DFT⁸⁸. However, the most common way to solve the TD Kohn-Sham equations is using linear-response theory⁸⁹, which says that when an external perturbation is acting on the system, the poles of the response occur when in resonance with the eigenvalue of the unperturbed system. That means that measuring the linear response of the density to an electric field will give the energy of the excited states. This is valid until the perturbation stays weak and does not destroys the ground state density. The LR of the density is given by

$$\delta\rho(\vec{r}, t) = \int_{-\infty}^t \int \sigma[\rho_0](\vec{r}, \vec{r}', t - t') \delta V_{ext}(\vec{r}', t') d\vec{r}' dt' \quad (2.36)$$

and applying the Fourier transform to it, the response function $\sigma(t - t')$ becomes in the frequency (ω) domain

$$\hat{\sigma}(\vec{r}, \vec{r}', \omega) = \sum_n \frac{\langle \Psi_0 | \hat{\rho}(\vec{r}) | \Psi_n \rangle \langle \Psi_n | \hat{\rho}(\vec{r}') | \Psi_0 \rangle}{\omega - \Omega_n + 0^+} - \frac{\langle \Psi_0 | \hat{\rho}(\vec{r}') | \Psi_n \rangle \langle \Psi_n | \hat{\rho}(\vec{r}) | \Psi_0 \rangle}{\omega + \Omega_n + 0^+} \quad (2.37)$$

where Ω_n corresponds to the excitation energy ($E_n - E_0$). A response function for the non-interacting system $\hat{\sigma}_{KS}(\omega)$ can be expressed in the same way in the frequency domain. According to the KS theory, we can obtain all the information about the perturbation of the real system based on the perturbation of the non-interacting system. The central equation of LR-TD-DFT becomes

$$\hat{\sigma}(\omega) = \hat{\sigma}_{KS}(\omega) + \hat{\sigma}_{KS}(\omega) \star f_{Hxc}(\omega) \star \hat{\sigma}(\omega) \quad (2.38)$$

where \star denotes a spatial convolution integral and $f_{Hxc}(\omega)$ is the Hartree-exchange-correlation kernel, which is given by the sum of

$$f_H(\omega) = \frac{\partial v_H(\omega)}{\partial \rho(\omega)} = \frac{1}{|\vec{r} - \vec{r}'|} \quad (2.39)$$

and

$$f_{xc}(\omega) = \frac{\partial v_{xc}(\omega)}{\partial \rho(\omega)} \quad (2.40)$$

and it represents a key quantity to obtain exact energies and oscillator strengths. This is obtained shifting the poles of the non-interacting system response to the true excitations energy, since this is true only when $\hat{\sigma}_{KS}(\omega) \star f_{Hxc}(\omega) \star \hat{\sigma}(\omega)$ is equal to zero. Different approaches have been proposed to solve equation 2.38, but the most employed one remains the one proposed by Casida⁸⁹. This approach is based on finding the eigenvalues of the matrix

$$\mathbf{M}_{ab,cd}(\omega) = \omega_{ab}^2 \delta_{ab,cd} + 4\sqrt{\omega_{ab}\omega_{cd}} f_{Hxc}^{ab,cd}(\omega) \quad (2.41)$$

since the squares of the excitations energy correspond to the eigenvalues of the matrix, while the oscillator strengths are obtained from the eigenvectors. In equation 2.41 ab, cd correspond to pairs of occupied and unoccupied orbitals, respectively, and $f_{Hxc}^{ab,cd}(\omega)$ is

$$f_{Hxc}^{ab,cd}(\omega) = \int \phi_a(\vec{r}) \phi_b(\vec{r}) f_{Hxc}(\vec{r}, \vec{r}', \omega) \phi_a(\vec{r}') \phi_b(\vec{r}') d^3r d^3r' \quad (2.42)$$

that introduces the correlation effects to the energy, which would be otherwise just the difference in energy between the occupied and unoccupied orbitals as in

an independent-particle system. This equation can be solved for both excitation and de-excitation energies, but since the de-excitation terms are relatively small compared to the excitations one, the Tamm-Dancoff approximation⁹⁰ neglects those and the equation is simplified.

Main strengths and limitations

The knowledge of the initial KS density and of the correct $\hat{v}_{xc}(\omega)$ would be enough to determine the excitation energy of single excitations from the first-order response function of the density. The main limitations of the methods stem from the adiabatic representation. The most intuitive limitation is the lack of double or multiple excitation, since a multiple excitation requires multiple photons and cannot be described by a linear-response function⁹¹. In the approximation of using a time-independent exchange-correlation functional, the problem of missing multiple excitations does not have a solution. Additionally, LR-TD-DFT encounters big problems when it needs to calculate charge transfer excitations and Rydberg states. Charge transfer states are characterised by electronic density localised on different part of the system, which can be highly distant in space. That implies that the overlap between the orbitals involved can be very limited and the correction to the simple orbital energy difference, given by equation 2.42 results to be null. The exact energy of a charge transfer state would tend to the difference between the ionisation potential of the donor unit, the electronic affinity of the acceptor and an electrostatic corrective term based on the distance

$$E_{CT} \rightarrow IP - EA - \frac{1}{R} \quad (2.43)$$

LDA, GGA and many hybrid functionals lead to an overestimation of ϵ_{CT} due to the wrong description of the $1/R$ term given by the local or semi-local dependence on the density^{92,93}. A similar error is found in the description of Rydberg states because the term $1/R$ tends to zero at the large distance at which Rydberg orbitals are usually delocalised⁹⁴. Both problems find a solution in the application of range-separated hybrid functionals. The Coulombic repulsion is separated in two components

$$\frac{1}{R} = \frac{1 - \text{erf}(\gamma r)}{r} + \frac{\text{erf}(\gamma r)}{r} \quad (2.44)$$

The first component represents the short range interaction and is usually treated with a semi-local functional. The second component represents the long-range interaction and is usually treated with exact or a fraction of Hartree exchange.

The parameter γ modulates the two components and can also be the resultant of a multiple parametrisation, like in the CAM-B3LYP functional⁹⁵ used in this work to properly describe charge transfer states, or can be optimally tuned ad hoc for the system⁹⁶. Another limitation needs to be considered when nonadiabatic dynamics is propagated on TD-DFT potentials and consists in the poor description of near-degeneracy point, conical intersection, between excited states and the ground state⁹⁷. Beyond the mono-determinantal nature of the method, the main limit is given by the null coupling element between the ground and the first excited state, since TD-DFT only calculates excitation energies, summed up to the energy of the ground state.

It is fundamental to always keep in mind the limitations of the method before running a TD-DFT calculation. However, TD-DFT represents often the best possibility to describe optical properties of medium-large system, especially to calculate excitation energies for several conformers, a big number of electronic excitations, big systems or propagate excited states dynamics for thousands of time steps. This makes TD-DFT an excellent choice for the description of excited states of multichromophoric system, as example like in this work, where was employed to study the absorption spectra of ligand-DNA complexes.

2.3.3 Tight-Binding approximation applied to DFT

Although not as computationally demanding as more accurate *ab initio* methods, DFT encounters a limit cost for system composed by tens of atoms, in particular if it is the chosen method for dynamics simulations where thousands of time steps need to be calculated. Semiempirical methods speed up the calculations, allowing the treatment of bigger systems, thanks to meaningful approximations aimed to give as much as possible accurate energies. A possible way to obtain energies based on DFT with a substantially lower price is to apply a semiempirical method, called Tight Binding (TB) within the DFT framework^{98,99}. The TB approximation consists in expressing the system wavefunction as a linear combination of atomic functions very delocalised within the confines of a minimal set of basis functions. This is applied only to the valence electrons that are then tightly bound to a frozen core. The elements of the Hamiltonian matrix are simply approximated and become just numerical elements, according to the different implementations of the method. In density functional tight binding approach (DFTB) a reference density ρ^0 is built

based on the superimposition of parametrised atomic densities. The density of the system is then evaluated as a sum of the reference density and its fluctuations

$$\rho = \rho^0 + \Delta\rho \quad (2.45)$$

where the fluctuations are the deviation of the real density from the reference density. The DFTB energy is given by three terms

$$E_{DFTB} = E^{H^0} + E^{rep} + E^{SCC} \quad (2.46)$$

In particular, the first term contains the matrix elements of the Hamiltonian depending on ρ^0

$$E^{H^0} = \sum_i \langle \psi_i | \hat{\mathcal{H}} | \psi_i \rangle = \sum_i c_\mu^i c_\nu^i H_{\mu\nu}^0 \quad (2.47)$$

where the Hamiltonian matrix elements $H_{\mu\nu}^0$ are equal either to the free atom energy (for same atomic orbitals $\phi_\mu = \phi_\nu$), or to $\langle \phi_\mu | \hat{\mathcal{H}}(\rho_\alpha^0 + \rho_\beta^0) | \phi_\nu \rangle$ (for atomic orbitals ϕ_μ and ϕ_ν belonging to different atoms), or are null otherwise. The second term is called repulsive energy contribution and it represents an approximation of all the double-body potentials, including the exchange-correlation potential and functional and it is *a priori* parametrised using a fitted potential. The first two terms give the standard DFTB energy, while the third one consists in the self-consistent charge contribution (SCC), which leads to SCC-DFTB¹⁰⁰. This term allows to improve the energy beyond the description of a sum of atomic densities, but including the effect of Coulombic interaction, fundamental to describe a proper charge fluctuating system, and including self-interaction as well. This term is given by

$$E^{SCC} = \frac{1}{2} = \int' \int \left(\frac{1}{|r - r'|} + \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'} \right) \Delta\rho\Delta\rho' = \frac{1}{2} \sum_{ab} \Delta q_a \Delta q_b \eta_{ab} \quad (2.48)$$

where Δq_a and Δq_b correspond to the net charge and η_{ab} is a function that includes the electron pairs interaction. This is parametrised fitting between the two extreme cases of electrons sharing the same site and being at infinite distance, considering the product of two normalised spherical atomic densities. The addition of this term improves the energy calculation, but the exchange-correlation energy and the SCC energy contribution can be expanded to the third order, leading to the so-called DFTB3 method¹⁰¹. The inclusion of the third order improves generically the energy description without loosing in efficiency and additionally removes shape restrictions of the atoms based on the reference density, allowing a better

description of anions. A η_{ab} function is similarly derived and parametrised as for the second order case. A dispersion correction term is also added to recover the correct non-covalent interactions. DFTB shows similar strengths and limitations of DFT, with the additional advantage of being three order of magnitude faster than standard DFT with a medium-size basis set, due to the complete removal of integral calculation, but at the cost of poorer energies and geometries, being based on parametrised densities. As used in this work, this method is still a good choice for conformational sampling or free energy calculations where long dynamical simulation are needed with less necessity of highly accurate energies and geometries.

2.4 Force Field Approach to Describe Potential Energy

Even using drastic approximations, the computation of the electronic energy with quantum mechanical methods is not feasible for systems with hundreds of degrees of freedom and a quantum chemical description of macromolecules and/or environment cannot be obtained. However, potential energies can be approximated as a sum of analytical functions that describe the different contributions to the potential energy a specific atom is subjected to. The collection of functions that describe the energy of an atom is called Force Field (FF)¹⁰². The extreme speed up of the calculation using a FF is achieved through the classical approximation of its terms. Indeed, molecular degrees of freedom like bond stretching and bending are described using harmonic oscillators. This neglects the possibility of describing chemical reactions as well as the lack of zero point energy. FF are widely used in Molecular Mechanics (MM) approaches, where the properties of a complex systems are investigated using long time scale simulation (section 2.6.1).

General approach

A FF is composed of a selection of terms including the main interaction in a chemical system. These terms can be collected in two categories. The first one contains the bonding terms, the second one the non-bonding ones. The total energy is then given by

$$U^{TOT} = U^{bonding} + U^{non-bonding} \quad (2.49)$$

The bonding term includes bonds stretching, bond angles bending and dihedral angles rotation

$$\begin{aligned}
 U_{stretching} &= \sum_{bonds} k_b (r - r_{eq})^2 \\
 U_{bending} &= \sum_{angles} k_\theta (\theta - \theta_{eq})^2 \\
 U_{dihedrals} &= \sum_{dihedrals} \frac{k_d}{2} [\cos(n\tau)]
 \end{aligned}
 \tag{2.50}$$

where $k_{b,\theta,d}$ are parametrised constants and r_{eq} and θ_{eq} are parametrised equilibrium values. All these values can be obtained from empirical data or quantum mechanical calculations and they are specifically obtained for any of the so-called atom type¹⁰³. An atom type represents the chemical nature of the single atom, i.e. a sp^1 , sp^2 or sp^3 carbon will each be represented by different atom types. Non-bonding energy contribution are given by van der Waals (VdW) and electrostatic interactions. The VdW term is usually described by a Lennard-Jones potential¹⁰⁴

$$U_{LJ} = 4u \left[\left(\frac{\zeta}{R_{1,2}} \right)^{12} - \left(\frac{\zeta}{R_{1,2}} \right)^6 \right] \tag{2.51}$$

where $R_{1,2}$ is the distance between two atoms, u represents the parametrised potential well and ζ the distance between the two atoms at which the potential is zero. The function contains a repulsive term (term to the power of 6) and an attractive term (to the power of 12). The electrostatic interactions are described by a Coulombic term¹⁰⁵

$$U_{electrostatic} = \frac{q_1 q_2}{4\pi e_0 R_1 R_2} \tag{2.52}$$

with q_1 and q_2 indicating partial charges of pairs of atoms, R_1 and R_2 their positions and e_0 the vacuum permittivity. The different level of parametrisation of these functions and the definitions of atom types differentiates between FFs¹⁰⁶. In the following, the basic aspects of the family of FF used in this work will be given.

General Amber Force Field

A widely used set of parameters is the so-called General Amber Force Field (GAFF)¹⁰⁷, parametrised by the developing team of AMBER. This FF does not add any particular potential with respect of the standard defined ones, but it contains specific parameters for a wide collection of different atom types. Indeed, almost the full chemical space of common organic molecules is covered and extra

parametrisation can be easily included by the user. The parameters were obtained at MP2/MP4 level and the charge distribution evaluated at HF level.

FF for the description of DNA

In order to model biochemical systems like enzymes, RNA and DNA, specific FFs are necessary in order to recover structural, physical and biochemical properties of the different macromolecules^{108,109}. Developing accurate FF for biomolecules is a wide research topic due to its fundamental importance. Indeed an apparently irrelevant inaccuracy in the parametrisation would still lead to convergence of the simulation, but also to wrong description, difficult to be noticed. Regarding FF for nucleic acids and DNA, several specific atom types need to be defined, since an heteroatom in the sugar or in the phosphate or in the termination should be described by a specific set of parameters¹¹⁰. At the same time, a correct description of sugar-backbone torsion¹¹¹, of backbone conformation and of charge distribution is required. One of the most used and reliable families of FF for biomolecular system is the AMBER FF¹¹². In particular, recent modifications, based on quantum mechanical calculations including the effect of the solvent with implicit models, show reliable results and a good reproduction of conformational DNA structures compared, for example, to NMR data^{111–114}. Additionally, they contain specific modification and re-parametrisation to include a proper description of non canonical secondary structures, e.g. G-quadruplexes¹¹³. Based on the AMBER FF is for example the FF called ParmBSC1 used in this work¹¹⁵. This FF has been parametrised on different high level QM calculation, up to coupled cluster method, in both gas phase and solvated DNA and tested on a hundred of different systems.

FF for the description of water molecules

One of the main advantages of approximating the energy with FF is the possibility of including environment and solvent effects explicitly. Indeed, thousands of molecules can be used to solvate the system of interest and their effect on the problem evaluated directly. This leads to the need of developing ad hoc FFs for solvent molecules which are able to capture the main physical and thermodynamic features of the solvent. In particular, several models to describe water molecules have been parametrised. One of the simplest but most used one is called Transferable Intermolecular Potential with 3 Points (TIP3P)¹¹⁶ which is composed by three sites, corresponding to the three atoms, where a point charge is collocated.

Stretching and bending terms are included as well as a Lennard-Jones potential on the oxygen. This simple model results in a poor geometrical description but high computational performances and it still represents one of the most popular choices to describe an environment formed by thousands of water molecules. The model can be improved by adding additional sites on dummy atoms, improving the description of the electrostatic distribution of H_2O ¹¹⁷. An additional improvement is given by polarizable FFs that account for changes in the charge distribution of the molecule. Their computational cost often poses a limiting factor, but several polarizable FF for water molecules have been for example developed, which improved the description of the solvent effects¹¹⁸.

2.5 Hybrid Quantum Mechanics/Molecular Mechanics approach

While on one side an accurate quantum mechanical (QM) description, which gives accurate energy calculation, describes bond breaking and forming and can calculate energy of excited states, is limited to small-medium system with limited number of degrees of freedom, on the other side FF can give extremely fast evaluation of large system including environmental effects, at the cost of reduced accuracy. The two advantages of the two methods can be combined and their limitations overcome thanks to hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) approaches^{119–121}. In this methodology, a system is partitioned in two regions, treated at different levels of theory. In particular, the region where a higher accuracy is required, e.g. a chromophore or a dissociating molecule, can be treated at high level quantum mechanically, while its surrounding, e.g. the solvent or inactive residues of a molecule, are treated classically with the goal of including their effect on the QM region of interest. The implementation of this methodology revolutionised the computational chemistry world, opening the door to, for example, the study of biochemical systems, restricting the accuracy necessary to few residues, but without losing the effect of the environment. In this work an extensive use of this methodology is done in order to include the environment effects on the simulation of DNA binding and photo-induced damage.

Energy definition

The QM/MM energy can be expressed according to two different schemes. In the additive scheme the energy

$$E_{QM/MM}^{add} = E_{QM} + E_{MM} + E_{QM-MM} \quad (2.53)$$

is given by the sum of the energy of the QM part (E_{QM}), the energy of the MM part (E_{MM}) and an additional term that includes the coupling between the two regions (E_{QM-MM}). In the subtractive scheme first the energy of the full system is calculated at the MM level ($E_{MM(tot)}$). Then the energy of the QM at the QM level (E_{QM}) is added and finally the energy of the QM part, but now at the MM level is subtracted ($E_{MM(QM)}$). The total energy is then given by

$$E_{QM/MM}^{sub} = E_{QM} + E_{MM(tot)} - E_{MM(QM)} \quad (2.54)$$

This scheme allows the generalisation to n-possible layers, as implemented in the Our-N-layerd Inegrated molecular Orbital and Molecular mechanics (ONIOM) schemes¹²². The general energy expression in the subtractive scheme does not explicitly need a term for the coupling between the two region, as it is already included at the MM level. This usually leads to a poor description of the electrostatic coupling and an additional calculation of this term can be included following one of the possible scheme proposed for the electrostatic QM-MM interaction¹²³.

Electrostatic interaction definition

Since the main goal of the whole QM/MM approach is to include the effect of the environment on the QM region, an accurate description of the coupling between the electron density of the QM part and the electrostatic model used in the MM calculation is needed. Three main approaches are used to this aim. The most rudimental approach is to extend the charge model for the MM region also to the QM part. This is called mechanical embedding and allows a very quick evaluation of the contribution. This approach is not really used due to its severe limitations. The QM charge distribution can change very drastically during a simulation, but the MM model would not account for it, and the density of the QM region remains stable and is not polarised by the surrounding. This is overcome in the electrostatic embedding scheme, where the MM point charges are included in the Hamiltonian of the QM region, whose electron density is explicitly affected by the environment

via a series of one-electron terms. This scheme can be used both in dynamical simulations as well as in static QM calculations involving only the QM part, where the explicit inclusion of the environment is still crucial to describe environmental-dependent properties, like the formation of excitations of different character in DNA³⁷. The following natural step to improve the description of electrostatic interaction is to allow the polarizability of the MM charges as well, which is done using polarizable models¹²⁴, e.g. polarizable FFs¹²⁵, for the MM part in the so-called polarizable embedding schemes.

Boundary region

Treating the boundary region between the QM and the MM regions is a delicate task, since it represents an unphysical situation and any possible artificial effect needs to be limited. First of all, the boundary region is susceptible to over-polarisation problems¹²⁶, especially when highly polarised and diffuse basis sets are used for the QM calculation. This problem arises due to the close proximity of MM point charges to the QM region. The addition of local Lennard-Jones terms, opportunely parametrised to increase the repulsive part, showed to tackle efficiently the problem¹²⁷, as well as the use of screening functions to screen the electrostatic interaction according to the distance between the point charges and the QM density¹²⁸. Another delicate aspect is the system partition. This is rather straightforward in case of an isolate molecule surrounded by solvent molecules. This becomes more complicated when the partition involves the breaking of a covalent bond ($\text{QM}_A\text{-MM}_B$) between a QM atom (QM_A) and a MM atom (MM_B)¹²⁹. A practical example, used in this work for DNA simulations, is the inclusion of only the nucleobases in the QM region, leaving the sugar and the phosphates in the MM part. Different schemes have been proposed to deal with this issue, including placing frozen hybrid orbitals in the QM part or boundary atom schemes that replace the (MM_B) with a new atom present in both the QM and MM region¹²⁰. The most common approach remains the introduction of a link-atom¹³⁰. A link-atom, usually a H atom, is linked to QM_A forming a new covalent bond $\text{QM}_A\text{-H}$. Clearly, the simple addition of an extra atom would change physical and chemical properties of the QM part, adding degrees of freedom and altering the charge distribution. However the charge can be redistributed on the surrounding atoms to reproduce the correct original distribution and, in case of dynamical simulations, the position of H can be restrained and its gradient projected over QM_A and MM_B .

The conceptual and practical ease of this approach makes it the most popular one to treat the cutting of covalent bonds in the partitions.

2.6 Molecular dynamics techniques

Molecular dynamics (MD) is referred to a series of techniques where a chemical system is investigated as a function of time. The range of methods and applications is wide and can range from real time propagation with quantum dynamics techniques to the evaluation of thermodynamic properties on a series of conformers obtained. In this chapter MD methodologies used in this work will be illustrated, starting from classical dynamics up to MD in the excited states.

2.6.1 Classical Molecular dynamics

Equation of motion and its integration

The Born-Oppenheimer approximation allows to uncouple the electronic and nuclear motion and solve the two problems independently. Obviously, the quantum nature of the nuclei would require a full quantum treatment, which induces a huge increase of the computational cost and limits full quantum dynamics to systems with only few degrees of freedom. A widely used and accepted approximation is to treat the nuclei classically, propagating them following the classical equation of motion

$$\vec{F} = \frac{d}{dt}(m\vec{v}) \quad (2.55)$$

This equation can be integrated numerically with different propagation algorithm, famous are the leapfrog algorithm¹³¹ and the velocity-Verlet algorithm¹³². In both these integration schemes, velocities and positions are updated independently. The two algorithms are similar, but the main difference is the integration step. In the leapfrog algorithm, positions are calculated at time t while the velocities at half step after the considered time step $t - \frac{1}{2}\Delta t$. In the velocity-Verlet algorithm are both calculated for the time step. The expressions for r and v in the leapfrog algorithm are

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}(t + \frac{1}{2}\Delta t)\Delta t \quad (2.56)$$

$$\vec{v}(t + \frac{1}{2}\Delta t) = \vec{v}(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m}\vec{F}(t) \quad (2.57)$$

while for the velocity-Verlet are

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}(t)\Delta t + \frac{\Delta t^2}{2m}\vec{F}(t) \quad (2.58)$$

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \frac{\Delta t}{m}[\vec{F}(t) + \vec{F}(t + \Delta t)] \quad (2.59)$$

The force corresponds to the negative of the gradient of the potential energy which can be calculated at any level of theory. In case a potential is calculated at any time step with any QM theory, the method is called *ab initio* MD. In case of fully classical potential like FF is called Molecular Mechanics (MM).

Molecular Mechanics

Hundreds of millions of time steps can be calculated with MM and the trajectory can be propagated in the microsecond time scale, according to the size of the system, the computer codes and resources. However, according to the ergodic hypothesis¹³³, the time spent by a system in a region of its phase space is proportional to the volume of the space. In other words, if the system is sampled for a long enough simulation time (ns, ms..), a good approximation of the full conformational space is sampled and thermodynamic and physical conclusions can be drawn from its analysis. Although conceptually very simple, running a classical MD simulation requires attention to several aspects to obtain meaningful simulations and to limit the drawbacks of the approximations used. A brief list of the most important ones is:

- Even if a substantial number of molecules can be included, this number will always be far from simulating an infinite-like system. In order to approximate that, the system can be included in a so-called unit cell that can be replicated infinite times in order to reproduce an infinite system. However, the calculation of the properties and the propagation is done for only the original system and only reflected to all the replica of the system, saving computational time and disk space. This is done by implementing a series of periodic boundary conditions in the three dimensional space. The practical implementation consists in storing the position, rotation and translation of the atoms of the unit cells, applying them to the replica and finally reconstructing the unit cell, in case part of the system overcame the borders of the cell.

- In case of periodic boundary conditions the total charge of the system needs to be zero in order to avoid artificial effects due to summation of infinite charges. If the system is not neutral, it can be neutralised adding counter ions.
- The electrostatic interaction is the most computational demanding part to calculate. In case of periodic boundary conditions, with repeating images to simulate an infinite system, the number of interactions to evaluate grows substantially. Nevertheless, the calculation of the electrostatic interactions can be sped up dividing them in short and long-range interactions, according to a user-defined cut off value. Only the short-range interactions are calculated directly, in the real space, while the long-range ones are evaluated in the Fourier space. To this end, it is only necessary to evaluate the potential and charge density on a discrete lattice in space. This approach is the most commonly implemented method to calculate electrostatic interaction on periodic systems. It is called Particle Mesh Ewald¹³⁴ and is the implemented version of the Ewald summation method to describe long-range interaction¹³⁵.
- In MM, restraints can be adopted i) to save computational time, neglecting the calculation of certain degrees of freedom, ii) to prolong the simulation time, increasing the time-step, iii) to enhance the sampling, imposing external forces to constrain the system in a certain state, iv) to prevent wrong description due to models that poorly describe certain degrees of freedom. Popular restraints in MM, used in this work too, are the SHAKE algorithm¹³⁶, to restrain only the positions in the integration, and its extension RATTLE to include restriction to velocities¹³⁷. Both algorithms are applied to restrain O-H bonds in water, which with TIP3P model would lead to unrealistic bond descriptions.
- Before running a MM simulation, once the system is built it needs to be minimised. In order to do that, several possible algorithms can be used, with the conjugate gradient¹³⁸, and the steepest descent methods¹³⁹, being among the most used ones.
- MM can be run in the microcanonical (NVE), isothermal-isobaric (NPT) or canonical (NVT) ensemble. The different ensembles can be used in the preparation runs of the simulation, to equilibrate the volume, the pressure and the temperature of the unit cell, or for the actual propagation. Barostat and thermostat are used to regulate pressure and temperature respectively, as

well as to preserve the correct thermodynamical behaviour along the simulation. Choosing the right thermostat or barostat is a delicate task and care is always recommended to choose the right algorithm, congruent with the system and problem of interest. Thermostats can be based on a simple rescaling of velocities every interval of times steps, or according to a friction coefficient of collision with other particles (e.g. Langevin hermostat¹⁴⁰), or to the coupling with a thermal bath (e.g. Berendsen thermostat¹⁴¹), or to random rescaling of some velocities according to a Boltzmann distribution (e.g. Andersen thermostat¹⁴²) or implemented in other alternative integrator based on Lagrangian approach to classical mechanics (e.g. Nosé-Hoover thermostat¹⁴³). Pressure can be controlled with volume rescaling at periodic intervals or according to the coupling with a weak pressure bath (e.g. Berendsen barostat¹⁴¹)

2.6.2 Binding Energy Calculation

In the wide range of MM applications, the statistical evaluation of the thermodynamic properties on an ensemble of conformations is a powerful tool for computational chemists. This is can be done only if a statistically relevant number of different conformers is collected in a section of volume of the phase space. One exemplary, and used in this work, application is the evaluation of the gain in Gibbs free energy after a ligand-receptor binding. Different methods can be used to this goal, the two used in this work are called Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA)¹⁴⁴ and Molecular Mechanics Implicit Solvent Model Surface Area (MM-ISMSA)¹⁴⁵ analysis. Both methods are based on the evaluation of the $\Delta G_{binding}$ according to a thermodynamic cycle. First, the energy of ligand and receptor are calculated at the MM level in gas phase as well as the binding energy in vacuum (ΔE_{MM}), as a gain in energy forming the complex from the ligand and the receptor. Then the solvation free energy is calculated for the receptor (ΔG_{solv}^{rec}), for the ligand (ΔG_{solv}^{lig}) and for the complex (ΔG_{solv}^{tot}). From these quantities, $\Delta G_{binding}$ is calculated as

$$\Delta G_{binding} = \Delta E_{MM} + (\Delta G_{solv}^{tot} - \Delta G_{solv}^{rec} - \Delta G_{solv}^{lig}) \quad (2.60)$$

The solvation energies consider two contributions: a non-polar and a polar one. The non-polar one is given by the difference between the energy gain from the attractive ligand-receptor interactions and the loss in energy given by the formation

of the space for the binding. The polar component values the gain/loss in energy due to the change of phase of the charged molecules. The different ways of calculating this last term gives the difference between the two different methods. In MM-PBSA the polar term is calculated with the Poisson-Boltzmann equation, while MM-ISMSA uses an implicit solvation model, where the calculation is notably sped by not considering the solvent molecules explicitly. The entropy term can be evaluated separately from a normal mode analysis at the MM level¹⁴⁶ and later added to the other terms.

2.6.3 Enhanced Sampling Techniques

One big limitation of classical MD is the sampling of only a circumscribed portion of the full conformational space. The exploration of the full conformational space is indeed often limited by the impossibility of overcoming potential energy barrier higher than the thermal energy, which hinders the population of regions other than the initial one. This limitation can make several minima of the potential inaccessible that might be actually populated in reality, compromising the validity of the ergodicity of the simulation. This limitation can be overcome with enhanced sampling techniques that allow the population of high probability states separated by high energy barriers. Connecting these regions means being able to calculate the free energy profile between these states. As shown in this work, that can be employed, for example, to study the binding process of a ligand to DNA or a chemical reaction. Different methods have been developed, whether the final state to sample is known or not. If the end state is unknown, techniques such as replica exchange¹⁴⁷, metadynamics¹⁴⁸ or hyperdynamics¹⁴⁹ can be used. If the initial and ending states are known, the free energy profile can be explored imposing artificial restraints that force the system along an ad-hoc defined reaction coordinate (RC). This is done for example in the Umbrella Sampling (US) method¹⁵⁰. In a US simulation RC is defined by the user, according to the problem of interest, a bond distance, a torsional, a collective variable and so on. The RC is then partitioned in several windows and a biased potential is defined for each of these windows. For each of these windows, an independent simulation is run using the corresponding biased potential. The potential V^{biased} is given by the sum of the unbiased potential $V^{unbiased}$ and an artificial potential centred at the middle of the window (RC_{min})

where the dynamic is run on. Commonly, a harmonic potentials is used as the artificial one, and the biased potential for each of the windows becomes:

$$V^{biased} = V^{unbiased} + K(RC - RC_{min})^2 \quad (2.61)$$

A constant force K is defined to restrain the system to move within the artificial potential. Once the simulation is run for a long enough time to sample each of the windows, the probability of the system being in one point of the RC is calculated for every window and later summed up for the whole RC. A biased probability distribution ($\Xi^{biased}(RC)$) can be then defined as a function of the RC. From this function the unbiased free energy profile can be obtained, often called Potential of Mean Force¹⁵¹, as

$$PMF = k_B T \ln \Xi^{biased}(RC) - K(RC - RC_{min})^2 - k_B T \ln \left\langle e^{(-k_B T K(RC - RC_{min})^2)} \right\rangle \quad (2.62)$$

where k_B is the Boltzmann's constant and the third term is the ensemble average of the artificial potentials.

$$\left\langle e^{(-k_B T K(RC - RC_{min})^2)} \right\rangle = \int \Xi^{unbiased}(rc) e^{(-k_B T K(RC - RC_{min})^2)} dRC \quad (2.63)$$

The full derivation, based on probability theory, of how the biased probability function and the Potential of Mean Force are linked can be found in Ref.¹⁵⁰. The first two terms of equation 2.62 are easily computable after the simulation, while more attention is required to compute the last term. Different methods can be used to solve this term. One of the most popular is the Weighted Histogram Analysis Method (WHAM)¹⁵² where $\Xi^{unbiased}(RC)$ is calculated from a weighted average of the distribution of each individual window. A computationally more efficient approach is given by the variational Free Energy Profile method (vFEP)¹⁵³, used in this work as well. This method is based on maximum likelihood method¹⁵⁴, which consists on finding the optimal set of parameters that minimize the likelihood of the probability distribution function that represents a given set of collected data. Starting from a trial Ξ , the method aims to find the unbiased probability distribution with variational principle. In contrast to WHAM, where a re-weighting is necessary in a self-consistent procedure, vFEP assumes first and second derivatives of the Potential of Mean Force between the windows as continuous via cubic spline functions to approximate the likelihood of the system, expressed as a combination of the likelihood of each individual windows. The whole US protocol can be extended in two dimensions, as well as vFEP method, which show excellent performance in reconstructing Potential of Mean Force in 2D. In both 1D and 2D

cases, a crucial step is the definition and partition of the RC, along which initial conformations for each of the window need to be given. These can be obtained with non-equilibrium technique like Steered Molecular Dynamics¹⁵⁵, where the system is driven from the initial to the final state under the action of an external force. After a certain simulation time, the coordinate of the system matches with the required value along the reaction coordinate and the output of the simulation can be used as input for US simulation.

2.7 Computational strategy to study excited states of DNA-ligand complexes

Increasing the size and the complexity of the system, from an isolated molecule in the gas phase to a full model including DNA and solvent, automatically increases the difficulties to study the photophysical process leading to the damage of the genetic code from a computational perspective. The first important step is to properly understand the response these systems have immediately after light irradiation, in particular which states are created and populated, how is their character and how they are delocalised. A proper description of the UV absorption of a multi-chromophoric system, like DNA or a DNA-ligand complex, it is not a straightforward but a challenging task and requires careful considerations on multiple levels.

First of all, a proper sampling of the phase space of the system is required to consider conformational, thermal and vibrational effects on the absorption. For a system with an extremely large number of degrees of freedom, the most reliable quantum-based sampling methods are not feasible and alternative strategies need to be developed. According to the ergodic hypothesis, classical MD allows a sampling of the conformational space of the system if this is propagated for a long enough period of time (section 2.6.1). However, using a FF the potential energy of the system is described by harmonic functions that do not represent the real potential of the bonding interactions. Moreover, the application of restraining algorithm, which speeds up the simulation time keeping frozen the fastest modes, like vibrations including H atoms, overlooks the sampling of potentially important modes involving hydrogen atoms. In our approach to study the excited states of DNA-ligand complexes, we run a first long (ns timescale) fully classical MM to obtain a first ensemble of representative geometries. These geometries are

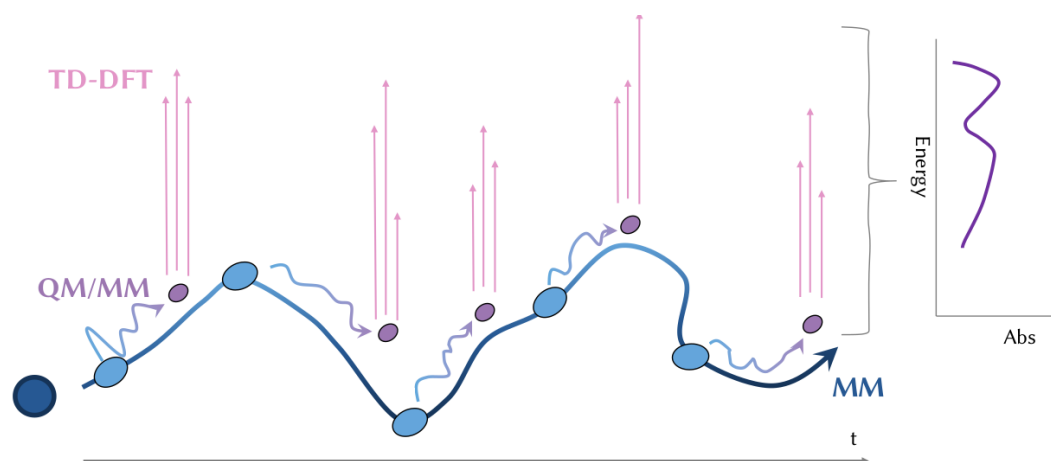


Fig. 2.1: Schematic representation of the workflow used in this work to calculate the excited states. A MM trajectory is initially propagated over nanosecond time scale; a certain number of conformations are extracted and propagated in the picosecond time scale with a QM/MM approach; the final trajectory are used for the calculation of the electronically excited state at the TD-DFT level of theory and the transitions obtained used to convolute the absorption spectrum.

selected along the MM trajectory and further propagated with a hybrid QM/MM approach (section 2.5), where the QM region includes at least the chromophores of interest. The usage of semi-empirical approaches, like DFTB3 (section 2.3.3), allows to include the full DNA in the QM region, taking into account the effect of the sugars and backbone on the nucleobases at a QM level. Additionally, using DFTB3 we can propagate the QM/MM trajectories for in the picosecond time scale and thereby ensure a full equilibration of the vibrations on the QM potentials. Subsequently, the vertical excitations of the final geometries of these trajectories are calculated within within an electrostatic embedding QM/MM scheme, where the effect of the environment on the QM region is considered by including it as an ensemble of point charges in the Hamiltonian. The level of theory for the excited states has to be chosen with great care. It needs to be a good balance between accuracy and cost, since treating multi-chromophoric system leads to on the one hand to a large amount of electrons to be included in the calculation, which requires large computational resources, and on the other hand it means the presence of complex excitation, like high delocalised or CT states. LR-TD-DFT (section 2.3.2) often represents the best choice in terms of accuracy/cost for these problems, where more than one unit are included in the QM region, with a large number of electrons. It is important to chose the proper functional to capture properly the effects that arise in the absorption of multi-cromophoric system, and in the present work we used CAM-B3LYP functional (section 2.3.2), proved to

be a suitable choice⁹⁵ for these problems. Running the calculation on graphical processor units (GPU) based codes notably speeds up the calculation and allows to include a bigger number of chromophores. The absorption spectrum is convoluted from the vertical transitions as a sum of Gaussian functions:

$$ABS(E) = \sum_i^{geom} \sum_j^{state} f_{ij} \exp(-4 \ln(2))(E_{ij} - E_{g.s.})^2 (FWHM)^{-2} \quad (2.64)$$

where i and j indicates a geometry and a state respectively, f_{ij} is the oscillator strength and FWHM is the full width at half maximum of the Gaussian. Due to the high number of coupled excitations, it is necessary to calculate a large amount of vertical excitation in order to include all the states contributing in the absorption in the UV/vis range. In the case of hundreds of geometries considered, this leads to an ensemble of thousands of vertical excitations to be analysed. A one-electron transition density matrix (1TDM) can be employed to analyse a collection of excitations in a fast and orbital-free way. The 1TDM between the ground and an excited state I is given by

$$D^{0I}(r_h, r_e) = n \int \dots \int \Psi_0(r_h, r_2, \dots, r_n) \Psi_I(r_e, r_2, \dots, r_n) dr_2 \dots dr_n \quad (2.65)$$

where r_h and r_e represent the position of the electron hole and the excited electron, respectively. Exciton and CT states (section 1.3) can be analysed by dividing the system in meaningful fragment, in our case each chromophore represents a single fragment, and decomposing the excitation in terms of total, hole and electron delocalisation lengths, and in terms of CT number, given by

$$CTN_{a,b} = \int_a \int_b |D^{0I}(r_h, r_e)|^2 dr_e dr_h \quad (2.66)$$

for a fragment a bearing the hole and b the excited electron. This TDM analyse is automatised in the program TheoDOR¹⁵⁸. Thanks to this protocol we were able to calculate the UV absorption of all the twelve guanines forming c-Myc G4 in the absence and presence of the ligand and the absorption of four nucleobases in (poly-dAT)₂ and the ligand intercalated in between as reported in the following chapters 4.

Simulation of DNA Binding

In this chapter the mechanism of MCH binding to DNA will be discussed as well as the most favourable binding modes will be described. Two examples will be considered. The first one is the binding to a dodecamer double strand, formed by alternating adenosines and thymidines (poly-dAT)₂ of different SP derivatives that showed affinity for this sequence. The second case is the binding of a quinolizidine-substituted derivative to a G4. The binding mechanisms were studied with US simulations (section 2.6.3), while the possible binding interactions were investigated with unrestrained MM and the Gibb's free energy of the binding modes were estimated with both MM-PBSA and MM-ISMSA methods (section 2.6.2).

3.1 Merocyanine binding mechanism to (poly-dAT)₂

Different SP derivatives were proposed as intercalators and showed good affinity for A-T rich sequences⁴⁶. However, neither the mechanism of action nor the binding modes, have been characterised experimentally or computationally yet. Our first goal was to study the mechanism of action of the actual DNA binders, the MCH forms. We studied two proposed intercalators, a nitro (MCH1, Figure 3.1) and amidinium-substituted (MCH2, Figure 3.1) SP, both with an aminopropyl tail attached to the indoline nitrogen. The two compounds differ in their total net charge, +2 in case of NO₂ and +3 in case of CN₂H₄⁺ substitution.

We started considering two possible intercalation pathways for compound MCH1, one passing through the major (path M) and one passing through the minor (path m) groove. Since no reference was provided describing the intercalative mode, except for the measured angle between the intercalative ring and the helical axis of ca. 80 degrees, we manually superimposed MCH with known intercalators of similar chemical structure and properties for which the crystal structure was given. That was the starting point of US simulations used to calculate the free energy profile of the two pathways, while the ending point was the unbound, solvated ligand. The pathways were interpolated between these two points increasing

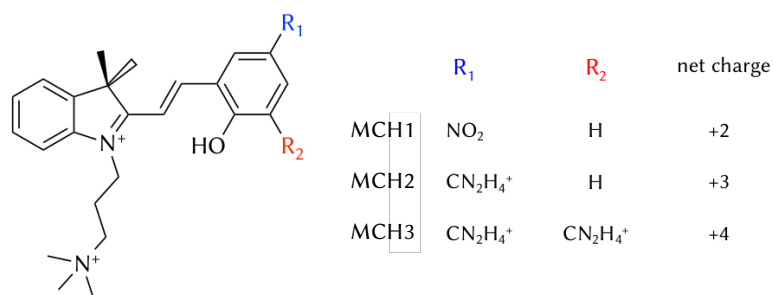


Fig. 3.1: Merocyanine derivatives investigated in this work as DNA binder. All the compounds have an aminopropyl functional group bound to the indoline nitrogen and they differ for the substituents on the phenolic ring, bearing one nitro group (MCH1) or one/two (MCH2/MCH3) amidinium groups. The three compounds have net charge +2, +3 and +4, respectively

the distance between the center of mass of the ligand and the four nucleobases forming the intercalative pocket. In both cases we can identify three minima in the paths (Figure 3.2), one corresponding to the ligand fully solvated (M1 and m1), one to the groove binding mode (M2 and m2) and one to intercalative binding (M3 and m3). Our simulations show a preferential pathway through the minor groove, forming a stable intermediate only in the minor groove (m2), before establishing the most stable interaction in the intercalative pocket (m3). We

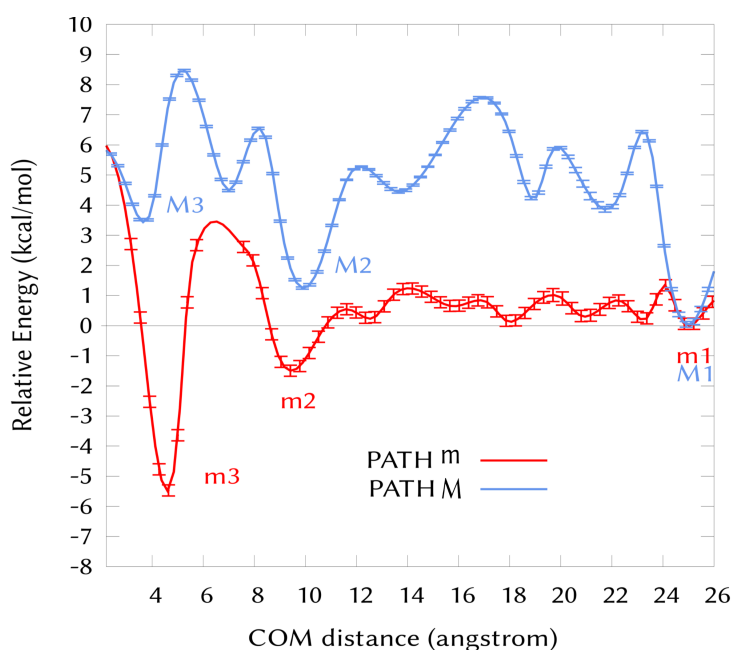


Fig. 3.2: Free energy profile of NO₂-substituted MCH1 intercalation pathway through the major (path M) and the minor groove (path m). Adapted from Phys. Chem. Chem. Phys., 2019,21, 8614-8618 with permission from the Royal Society of Chemistry.

removed the restrains from the systems in these two conformations and carried out classical MD for 100 ns to better characterize the two binding modes (Figure 3.3). The phenol ring is intercalated symmetrically between the base pairs and establishes a combination of repulsive and attractive electrostatic interactions as well as favourable Van der Waals interactions with the surrounding nucleobases (Figure 3.3a). The positively charged aminopropyl tail helps the stabilisation by interacting with the negatively charged DNA backbone (Figure 3.3b). This interaction is important in the minor groove as well, where the VdW connections are reduced with respects of the intercalation and the electrostatic term becomes more relevant to obtain a stable mode (Figure 3.3c-d). The favourable pathway passing from the minor groove was confirmed for MCH2 so we study the two binding modes for this compound as well and calculated the binding energy. Once

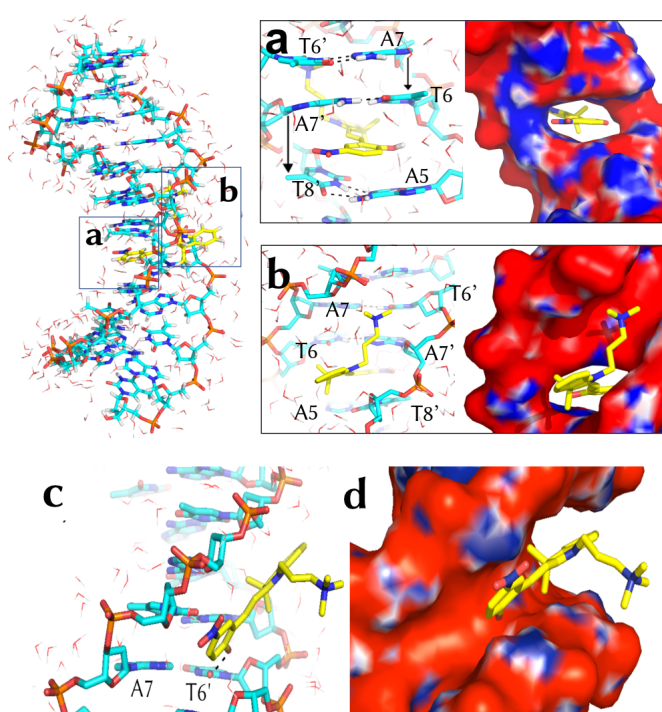


Fig. 3.3: Top: graphical representation of the intercalative binding mode and, with zoom on the interaction between the intercalative phenol ring and the surrounding nucleobases (a) and on the interaction between the positive tail of the ligand and the polyanionic DNA backbone. The two insets show the electrostatic potential maps with attractive (red) and repulsive (blue) regions; bottom: graphical representation of the minor groove binding mode with focus on the interaction stabilising the mode (c) and on the electrostatic potential map (d). Adapted from Phys. Chem. Chem. Phys., 2019,21, 8614-8618 with permission from the Royal Society of Chemistry.

MCH2 is intercalated, the VdW contribution is found to be similar to MCH1, but due to a slightly smaller electrostatic repulsion, the total gain in free energy turns

out to be bigger in the m3 mode (Table 3.1). In m2 the VdW interactions are instead slightly reduced, but the increase of the total net charge of the ligand makes the electrostatic term in the minor groove an overall favourable contribution to the total binding energy, which results to be higher than for MCH1. Based on this finding, we inserted as a proof of concept a third compound bearing two amidinium groups on the phenol ring (MCH3, Figure 3.1) in both m2 and m3 positions, to see the effect of an additional positively charged functional group. In m3, electrostatic, VdW and non polar contributions are similar to MCH1, but the entropic penalty, due to the different size of the systems, makes this mode less favourable. The same increase in entropy is present in m2, where the VdW interactions are also reduced, but the additional charge of the ligand leads to a notably stronger electrostatic attraction, making the minor groove binding more favourable than the intercalation for this compound.

Tab. 3.1: Binding energy (kcal mol⁻¹) at 300K of compounds MCH1-MCH3 in binding mode m2 and m3. Energy is computed with MM-PBSA method along a window of 20 ns of MD. The errors represent the standard deviations of the mean values

ligand	$\Delta E^{electr.}$	ΔE^{VdW}	$\Delta E^{nonpolar}$	$T\Delta S$	ΔG^{bind}
mode m3					
MCH1	15.81 ± 0.14	-55.17 ± 0.25	-4.32 ± 0.01	-21.73 ± 0.09	-21.96 ± 0.24
MCH2	10.10 ± 0.40	-55.78 ± 0.23	-4.62 ± 0.40	-20.96 ± 0.17	-29.35 ± 0.33
MCH3	11.09 ± 0.59	-50.71 ± 0.42	-4.57 ± 0.03	-26.42 ± 0.37	-17.77 ± 0.74
mode m2					
MCH1	10.82 ± 0.22	-37.72 ± 0.41	-3.23 ± 0.01	-20.97 ± 0.02	-9.29 ± 0.47
MCH2	-1.02 ± 0.60	-34.16 ± 0.26	-3.22 ± 0.01	-23.88 ± 0.13	-14.52 ± 0.36
MCH3	-12.07 ± 0.39	-29.91 ± 0.49	-3.19 ± 0.02	-24.71 ± 0.26	-20.46 ± 0.61

In conclusion, we first indicated how MCH intercalates into poly(d-AT)₂ passing through a stable intermediate into the minor groove and excluded the pathway passing through the major groove. We further confirmed and characterised a favourable intercalation mode, where VdW interactions play a major role in the stabilisation of the binding. However, the possibility of modulating the electrostatic potential on the ligand means modulating the interaction with the DNA as well. Indeed, an increase of the positive charge of the ligand can induce a stronger intercalation mode or even an inversion of relative energy between intercalation and minor groove binding, due to an extra stabilisation of this last one due to favourable electrostatic interactions. Full discussion, methods and additional details can be found in the Appendix (section 6.1), under the reprinted publication titled "DNA-binding Mechanism of Spiropyran Photoswitches: The Role of Electrostatics".

3.2 Quinolizidine-substituted Spiropyran Thermal Isomerisation and G-Quadruplex binding

The colour change in SP derivatives can be observed in case of thermal isomerisation, depending on specific environmental conditions. Very recently, a quinolizidine-substituted spiropyran (QSP) was proposed as a diagnostic tool for the identification of G4 in cells, showing isomerisation in live cells, with a particularly strong response and high selectivity for the protooncogene c-Myc G4⁵⁰. Thanks to a change in wavelength of emission, it was demonstrated that, as for



Fig. 3.4: Quinolizidine-substituted spiropyran (QSP) isomerisation to the open and protonated merocyanine form (QMCH) in the presence of G4. α , β and γ represents the bonds for the definition of the QMCH isomers. Adapted from Chem. Eur. J. 2020, 26, 13039-13945 with permission of the authors and Wiley-VCH GmbH.

other known SP-based DNA binders, the actual form that binds the DNA is the open merocyanine form (QMC) and, in particular, the protonated form (QMCH). The role of the pKa of the cells might play a key roles, as the thermal isomerisation can follow a proton attack, that consequently favours the binding to the macro-molecule. We wanted to provide more insight into the isomerisation mechanism, into the role of the environment and into the possible QMC isomers that can be produced following the ring opening, according to the rotation around the β bond (Figure 3.4).

To these aims, we investigated the isomerisation mechanism using 2D-US simulations (section 2.6.3) with a QM/MM approach (section 2.5) both in water and a G4 environment. The QM part consisted of SP and the closest water molecule to the $C_{\text{spiro}}\text{-O}$ bond, treated at DFTB3 (section 2.3.3) level, while the solvating water molecules were treated at TIP3P level (section 2.4) and G4 at ParmBSC1 force field level (section 2.4). Two possible pathways, proton assisted and unassisted, are considered and two reaction coordinates are defined to describe these pathways. (Figure 3.5.a). The first one is the ring opening, defined through the distance between the C and the O of the breaking bond; the second describes the proton attack, defined as a linear combination of distances between the SP oxygen, the oxygen of the QM water molecule and the H transferred between those. Our 2D free

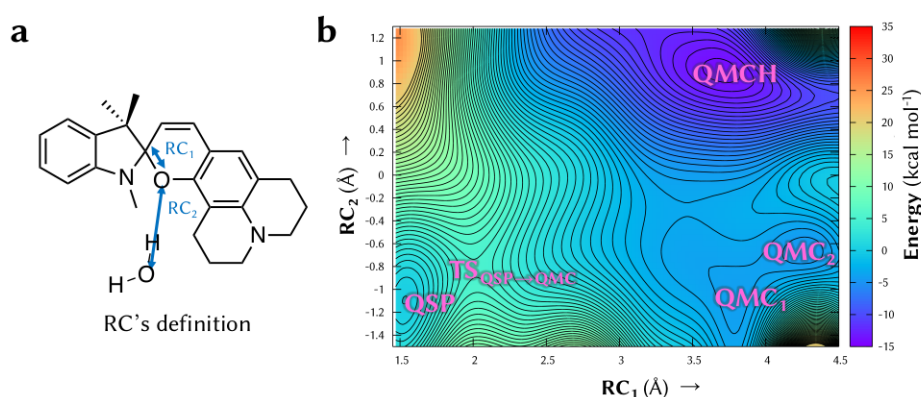


Fig. 3.5: 2D-US exploration of QSP isomerisation in water. a) reaction coordinates definition, b) free energy profile obtained with assignment of the stationary points; c) geometrical redefinition at MP2/def2-TZVP level of theory of the most stable isomers found. Adapted from Chem. Eur. J. 2020, 26, 13039-13945 with permission of the authors and Wiley-VCH GmbH.

energy profile shows a thermal isomerisation not assisted by proton transfer. Once the energy barrier is overcome, QSP can form two QMC isomers of comparable energy (Figure 3.5.b). These isomers are identified as *trans-trans-cis* (TTC, Figure 3.6) and *trans-trans-trans* (TTT, Figure 3.6) with respect of the exocyclic double bond. Static MP2/def2-TZVP calculations on possible QMC isomers confirm the almost degeneracy of these two isomers in water and higher energies for other possible isomers (Figure 3.6). Once QMC is formed, the proton transfer from the

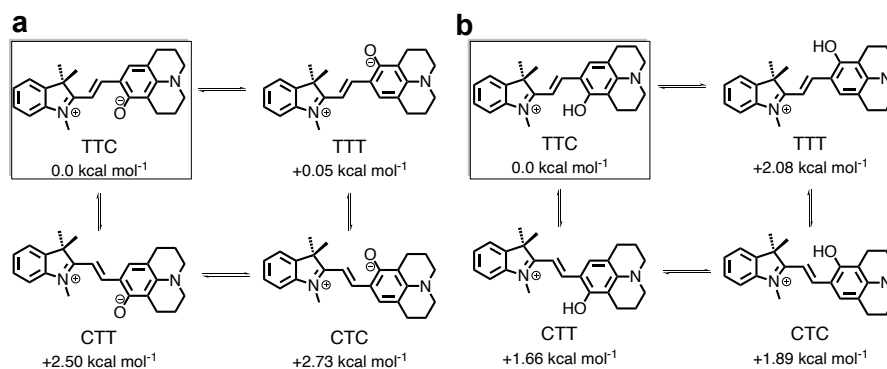


Fig. 3.6: QMC (a) and QMCH (b) possible *E*-rotomers around the α/β bond and their energy calculated at MP2/def2-TZVP level of theory, where the zero of the energy is set to the Gibbs energy of the most stable compound at 300K. Adapted from Chem. Eur. J. 2020, 26, 13039-13945 with permission of the authors and Wiley-VCH GmbH.

molecule forms the most stable QMCH form. For QMCH, only one rotomer can be found, the TTC form, which is confirmed by MP2 calculations showing that TTC is lower in energy relative to other rotomers, including TTT (Figure 3.6). Having unveiled the mechanism in water, we wanted to extend our protocol to

study the possible effect of G4 on the QSP opening process. We repeated the 2D-US simulation with the same parameters, but allocating SP in the presence of c-Myc G4. Interestingly, the presence of G4 does not change the overall isomerisation

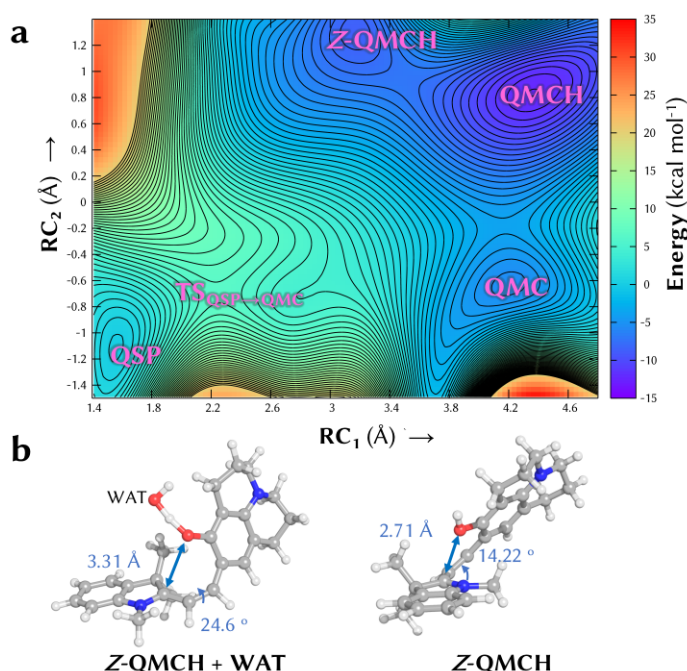


Fig. 3.7: 2D-US exploration of QSP isomerisation in the presence of c-Myc G4. a) free energy profile obtained with assignation of the stationary points; b) geometrical redefinition at MP2/def2-TZVP level of theory of the Z-isomer found. Adapted from Chem. Eur. J. 2020, 26, 13039-13945 with permission of the authors and Wiley-VCH GmbH.

mechanism of QSP, which still does not need a proton to be activated (Figure 3.7.a). However, the effect of the macromolecule can be seen in the presence and stabilisation of only the TTC form of the QMC and QMCH isomers (which can be found in water too), as well as of an additional QMCH isomer, which is a Z-isomer with respect to the β bond (Figure 3.7.b). This Z-isomer can be accessed only by UV light for other SP derivatives¹⁵⁶, but the presence of the poly-anionic G4 manages to stabilize this form. Again, the presence of all these species were confirmed by MP2 calculations.

To characterise the binding modes of both species we performed microsecond time scale classical molecular dynamics and estimated the binding energy for different ensembles of conformations for both QMC and QMCH. In particular, we carried out three independent simulations of 300ns each for both, the deprotonated and protonated probe. MD simulations with QSP in G4 environment was ran as well but, as expected, no stable binding mode was found. Both QMC and QMCH can form stable binding modes with G4 stacking on the upper tetrad of the structure

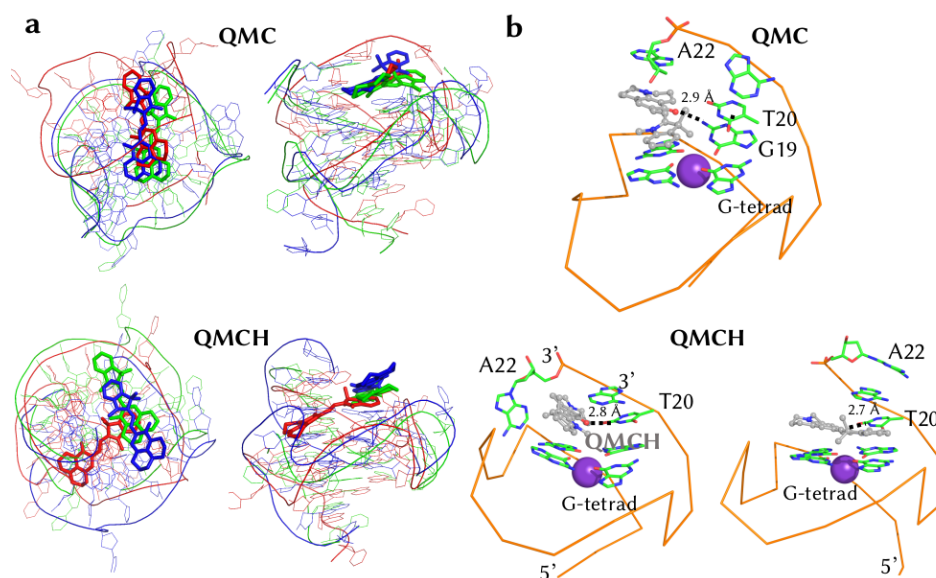


Fig. 3.8: a) Superimposition of three independent simulation of binding modes of QMC and QMCH and G4. b) Details into the interaction between the probe and the upper tetrad and the side A and T nucleobases. Adapted from Chem. Eur. J. 2020, 26, 13039-13945 with permission of the authors and Wiley-VCH GmbH.

and. In case of QMCH, the ligand establishes favourable electrostatic interactions with the surrounding nucleobases (Figure 3.8). The probe shows high mobility around the α and β bonds, in particular the indoline ring, while the chromene moiety is more restrained by the interaction with side nucleobases (A and T). The three independent simulations converge to a similar binding mode for QMC with comparable binding energies, whereas for QMCH two runs show similar conformations but in one simulation the probe establishes the strongest interaction once wrapped by the 3' termination. Both compounds generally establish favourable binding modes (Table 3.2); the protonation increases the gain in energy due to electrostatic interactions, suggesting that the electrostatic interaction might be a key factor in the context of design and development of new SP based G4 binders.

In conclusion, we propose an opening mechanism not involving a direct proton transfer, where instead the presence of the right environment for the protonation offered by G4 in live cells stabilises the QMCH form moving the equilibrium towards the open isomers and leading to the opening process observed for this species in c-Myc G4 environment. This is enhanced by the QMC and QMCH forms being captured by the G4 through stable binding modes. The QMCH-G4 is found to be the most stable interaction as a combination of electrostatic and VdW interaction with the guanine and the side nucleobases in the strand. Full discussion, methods and additional details can be found in the Appendix (section 6.2),

Tab. 3.2: Binding energy (kcal mol⁻¹) at 300K of for three independent simulations for both QMC and QMCH. Energy is computed with MM-ISMSA method along a window of 20 ns of MD.

#	$\Delta E^{electr.}$	ΔE^{VdW}	$\Delta E^{nonpolar}$	$\Delta E^{desolvation}$	ΔG^{bind}
QMC					
1	-0.17 ± 0.29	-46.68 ± 6.07	-2.57 ± 0.32	3.13 ± 0.40 (L) + 11.06 ± 1.29 (R)	-37.24 ± 5.45
2	-0.73 ± 0.15	-52.86 ± 3.11	-2.64 ± 0.13	2-34 ± 0.22 (L) + 10.06 ± 1.01 (R)	43.28 ± 2.64
3	-9.18 ± 0.58	-42.02 ± 2.79	-2.31 ± 0.09	-0.63 ± 0.16 (L) + 13.20 ± 1.01	-40.94 ± 2.09
QMCH					
1	-8.95 ± 0.59	-55.19 ± 3.97	-2.54 ± 0.11	-0.29 ± 0.11 (L) + 11.49 ± 0.93 (R)	-55.48 ± 3.85
2	-8.76 ± 0.53	-47.77 ± 4.30	-2.47 ± 0.17	-0.08 ± 0.44 (L) + 10.94 ± 0.90 (R)	-48.14 ± 4.50
3	-8.63 ± 0.49	-50.05 ± 1.73	-2.41 ± 0.05	-0.22 ± 0.07 (L) + 10.93 ± 0.66 (R)	-50.39 ± 2.03

under the reprinted publication titled "Spiropyran Meets Guanine Quadruplexes: Isomerization Mechanism and DNA Binding Modes of Quinolizidine-Substituted Spiropyran Probes".

Simulation of DNA Damage

In this chapter the simulation of the response of different DNA related systems to light irradiation is reported, responses that might be connected to possible damage of the genetic code due to absorption of light. Starting from a joint experimental/computational work on the adenosine triphosphate (ATP) dianion in the gas phase, the absorption spectrum of c-Myc G4 will be reported and the effect of its binding to QMCH will be evaluated. Finally the effect of light absorption on the nucleobases surrounding the intercalated MCH1 and MCH2 will be investigated.

4.1 Simulation of photooxidation of ATP dianion

The full understanding of the DNA photophysics needs a knowledge of the light induced properties of its components, from which a full picture can be drawn adding more structural complexity. Among the possible DNA building blocks, adenine is also the component of other biologically relevant molecules like ATP, which has a pivot role in the energy transfer and RNA synthesis. The negatively charged phosphate groups make ATP a poly-anion and alter the photophysics of the molecule. In particular, a poly-anion can be oxidated if it absorbs enough energy to promote an electron above a purely electrostatic barrier arising from the poly-anionic nature of the molecule, called repulsive Coulomb barrier (RCB, Figure 4.1.a). Investigating the properties of the RCB of ATP dianion ($[\text{ATP-H}_2]^{2-}$) sheds light on the possible oxidation processes that occur on this molecule after light absorption.

A photoelectron spectrum we obtained after irradiation of $[\text{ATP-H}_2]^{2-}$ at 4.66 eV shows an unexpected signal centred at around 0.55 eV (Figure 4.1.b), which is present with using both, a femto- and a nano-second laser. Unexpected because the RCB is reported to be higher than the excitation energy given¹⁵⁷. The registered signal in the spectrum is assigned to an electron leaving $[\text{ATP-H}_2]^{2-}$ via a tunnelling through the RCB. The hypothesis of the electron tunnelling is supported by taking the spectra for different excitation wavelengths. The signal at 0.55 eV is not affected

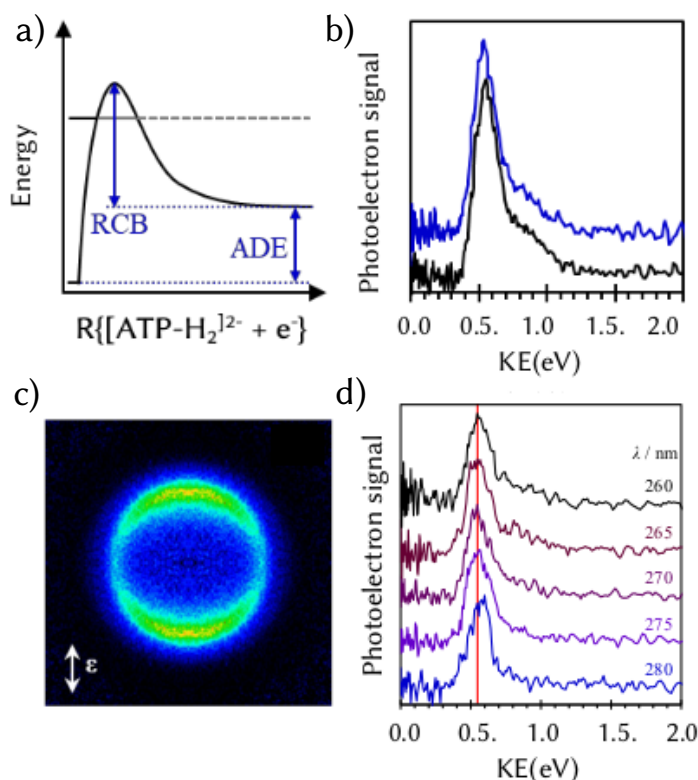


Fig. 4.1: a) Representation of the repulsive Coulomb barrier (RCB) defined with respect of the adiabatic detachment energy and as a function of the distance between ATP dianion and the electron leaving; b) Photoelectron spectra obtained with nanosecond (black line) and femtosecond (blue line) laser pulses at 4.66 eV; c) Photoelectron image recorded with nanosecond laser at 4.66 eV with anisotropy oriented to the polarisation vector ϵ ; d) photoelectron spectra taken with a variable light pulse; the red line shows the unchanged position of the peak at 0.55 eV despite the change of λ of the pulse. Adapted with permission from J. Phys. Chem. Lett. 2020, 11, 19, 8195–820. Copyright 2020 American Chemical Society.

by the change in energy, indicating a wavelength-independent oxidative process. The final suggestion about tunnelling arises from the anisotropy of the actual image recorded. The anisotropy is found in the direction of the polarisation axis of the light and suggests an electron leaving along this direction, which is aligned with the polarisation of the electronic transition, i.e. its transition dipole moment.

In order to confirm the tunnelling hypothesis, to identify the direction of the electron leaving and to give a complete overview over the oxidation process, different computations were necessary. First of all, we confirmed the nature of the absorption to the bright $\pi\pi^*$ state and found that the associated transition dipole moment lies along the plane of the adenine ring by static excited state calculations at ADC(2)/def2-QZVP level of theory (section 2.2.3) at MP2/def2-

TZVP optimised geometry (section 2.2.2). Ground state picosecond-scale *ab initio* MD at the DFT/aug-cc-pVDZ level (section 2.3) shows that the phosphate groups are quite flexible but their relative distributions with respect of the adenine ring is constant. The negative charges located on the phosphates screen the electrons not

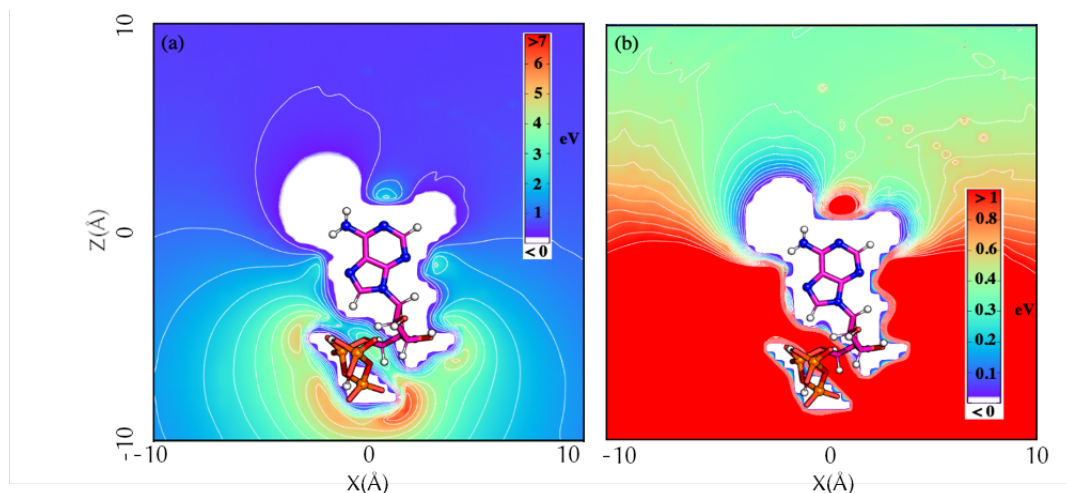


Fig. 4.2: Computed repulsive Coulomb barrier (RCB) in the plane passing through the adenine ring plane and the transition dipole moment of the $\pi\pi^*$ state populated after irradiation at 4.66eV (a). The scale to a maximum of 1 shows the preferential site for electron leaving correspondent to the minimum of the RCB in the proximity of the amino group (b). Adapted with permission from J. Phys. Chem. Lett. 2020, 11, 19, 8195–820. Copyright 2020 American Chemical Society.

allowing them to leave on that direction. The RCB was calculated as the energy of ATP monoanion at the $[\text{ATP-H}_2]^{2-}$ optimised structure in the presence of an electron, whose position was varied along the plane spanned by the transition dipole moment and the adenine ring (Figure 4.2). The calculations show the emitting electron is blocked from the region where the phosphates are located and the electron has a prevalent site for its leaving on the ring side. In particular, the lowest saddle point is calculated to be 0.57 eV, in excellent agreement with the energy of the signal recorded. Additionally, the computational insight allows to determine the preferential site for the electron leaving, showing how this will most easily happen in the proximity of the amino group.

In conclusion, our work shows an important feature of the light-induced behaviour of adenine, with a site specific-oxidation of nucleobases that might be important for a better understanding of electron emission in more complex systems and the overall behaviour and structure of biological relevant poly-anions. Full discussion, methods and additional details can be found in the Appendix (section 6.3), un-

der the reprinted publication titled "Site-Specific Photo-oxidation of the Isolated Adenosine-5'-triphosphate Dianion Determined by Photoelectron Imaging".

4.2 Absorption spectrum of c-Myc Guanine quadruplex and ligand effect

Having explained the binding mode of QMCH to c-Myc G4 (section 3.2), we wanted to evaluate the effect of the presence of the probe on the photophysics of the macromolecule. As shown in section 3.2, we identified the most stable binding mode with the probe stacked over the guanines of the upper tetrad and establishing interaction with the side adenine and thymine, which are present in the sequence of c-Myc G4. Interestingly, we notice in this binding mode a reduced flexibility of the G4. This can be seen by the overlapping of one hundred sampled QM/MM geometries of unbound and bound G4 (Figure 4.3). We wanted to investigate the

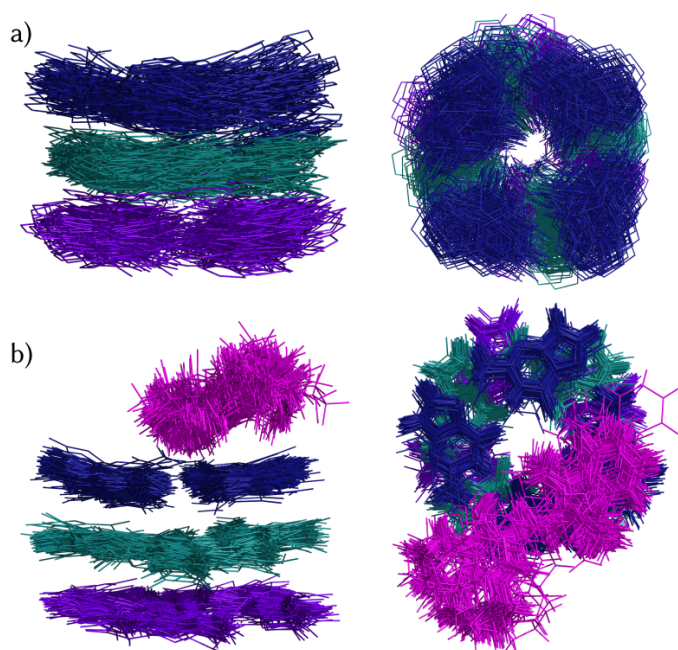


Fig. 4.3: Two views of overlap of 100 sampled QM/MM geometries of the guanines of the c-Myc G4 alone (a) and in the presence of QMCH (b). The ligand induced a conformation rigidity of the macromolecule. Adapted with permission from J. Phys. Chem. Lett. 2020, 11, 23, 10212–10218. Copyright 2020 American Chemical Society.

effect of this enhanced rigidity on the absorption spectrum of c-Myc G4.

First of all, we computed the absorption spectrum of the macromolecule alone, with the QM region including all the twelve guanines composing the three tetrads. The spectrum was obtained from 6000 excited states from an ensemble of 100 geometries. The total absorption consists of two contributions, and it is a combination of local $\pi\pi^*$ excitations, centred at around 5 eV, exciton and mixed exciton/CT states. In our analysis we distinguished exciton and mixed states according to

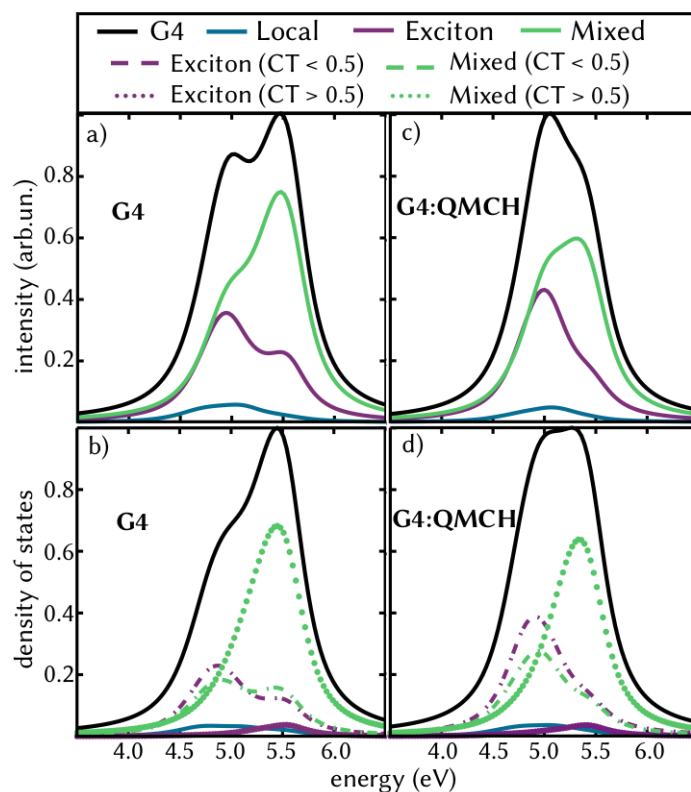


Fig. 4.4: UV absorption spectra of G4 alone (a) and its decomposition according to the excitations contribution (a) and correspondent density of excited states (b) an UV spectrum (c) and density of state (d) in the presence of QMCH. Adapted with permission from J. Phys. Chem. Lett. 2020, 11, 23, 10212–10218. Copyright 2020 American Chemical Society.

the degree of charge transferred in the excitation. In particular, we collected the excitations in two subgroups, depending on whether less (CT number < 0.5) or more (CT number > 0.5) than half of the density is transferred between two or more nucleobases. Combining the UV signal and the density of excited states, we could identify a dominant contribution in the high energy range of the spectrum given by mixed states with high CT number. A smaller density of mixed states with low CT number contributes almost equally to both regions of the spectrum. Also excitonic states show contributions in both regions of the spectrum, but mainly excitonic states with small amounts of CT are observed. The binding of

Tab. 4.1: Binding energy (kcal mol⁻¹) at 300K of for three independent simulations for both QMC and QMCH. Energy is computed with MM-ISMSA method along a window of 20 ns of MD.

excitation	E(eV)	f_{osc}	percentage of state (%)
local excitations within QMCH (first absorption band)	2.4-3.4	~ 1	11
local excitations within QMCH (second absorption band)	4.5-5.9	<0.100	7
pure charge-transfer states (G \rightarrow P)	3.1-3.9	<0.006	16
mixed states (P \rightarrow G)	3.9-4.5	<0.004	3
mixed states (G \rightarrow P)	4.7-6.0	<0.013	3
local/exciton/mixed states within G4	4.3-6.0	<0.400	60

QMCH induces the presence of new states (Table 4.1), like excitations involving only QMCH, mixed states with electron transfer from the probe to the G4 and vice versa and, more interestingly, pure CT states where an electron is transferred from one guanine to the oxidative probe. These states represent guanine-oxidated states and, although they cannot be directly populated from the light irradiation, they might play a crucial role in the photochemistry after absorption leading to the damage of the genetic code. The states involving only the twelve guanines represent 60% of the total ensemble and we can directly evaluate the effect of the ligand in their character and absorption. We analogously find a combination of local, exciton and mixed states, but the total and relative absorption is affected by the presence of the ligand. The maximum of the absorption is shifted from 5.5 to 5.0 eV and the relative height of the two peaks is inverted. The contribution of the exciton is more intense and still almost the full contribution shows CT < 0.5 . Regarding the mixed state we notice a reduction of the states with small CT contribution at high energy and a decrease of the intensity of the signal given by the mixed state with CT > 0.5 . We connect this effect on the UV absorption to the conformational rigidity imposed by the probe, which prevents the mixing of the electronic clouds of the guanine and thus reduce the presence of bright states at high energy and with strong CT character, which represent the main component of states in case of c-Myc G4.

In conclusion, we calculated for the first time the UV absorption properties of c-Myc G4, we identified the presence of potential detrimental CT states and we highlighted the importance and the effect of the conformational flexibility of this DNA structure and the role of the probe interacting with the multi-chromophoric system. Full discussion, methods and additional details can be found in the Appendix (section

6.5), under the reprinted publication titled "Enhanced Rigidity Changes Ultraviolet Absorption: Effect of a Merocyanine Binder on G-Quadruplex Photophysics".

4.3 Excited state characterisation of merocyanine and (poly-dAT)₂ complexes

Following the finding of stable intercalative binding modes for MCH1 and MCH2 (section 3.1) in (poly-dAT)₂, we wanted to investigate the excited states of the intercalative site, the ligand and the four surrounding nucleobases directly interacting with it. Merocyanines strongly absorb light at around 450 nm, at much lower energy than A and T and thus, a pronounced absorption localised on the probe is expected. Moreover, we calculated the standard reduction potential for

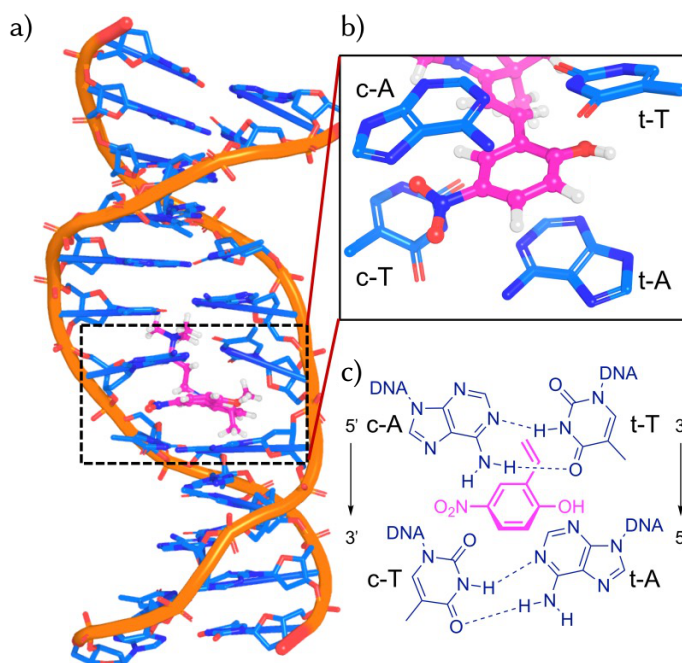


Fig. 4.5: Representative geometry of the MCH1-DNA complex (a) and zoom into the intercalative pocket (b) and on the molecules included in the QM region for the calculation of the excited states (c) with label the strand of belonging, coding (c) and template (t). Adapted from *Phys. Chem. Chem. Phys.*, 2019,21, 17971-17977 with permission from the Royal Society of Chemistry.

MCH1 and the high value (~ 4 V) suggests a potential oxidation of the nucleobases induced by the ligand. Indeed, a merocyanine derivative have been shown being oxidative agents for DNA in cell culture¹⁵⁹. Both ligands, MCH1 and MCH2 were considered, in order to see the effect of the two functional groups and of the total

net charge (+2 for MCH1 and +3 for MCH2) on the excited state distribution and on the possible oxidation of the nucleobases. In order to do that, we included in the QM region, for the calculation of the UV absorption spectra, the ligand and four alternating nucleobases directly interacting with the ligands (Figure 4.5), and doing so we could analyse the excited state involving the ligand and the nucleobases forming the intercalative pocket.

The computed absorption spectra for both MCH1 and MCH2 complexes with DNA consist of the absorption peak of the ligands and show good agreement with the experimental ones, compared for both compounds. More interestingly, the density of excited states show a small amount of bright states that contribute to the total absorption and correspond to the $\pi\pi^*$ state of the merocyanines. The rest

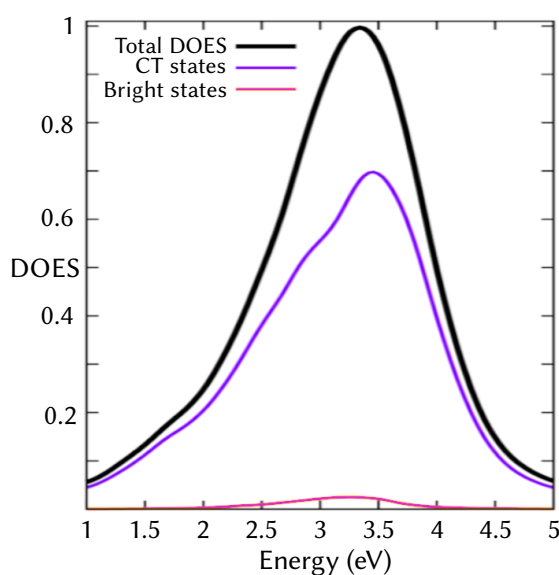


Fig. 4.6: Density of excited state for the MCH1-DNA complex showing a limited amount of bright states (pink line), localised on the ligand and a large amount with CT character (purple line) between the ligand and the surrounding nucleobases. Adapted from Phys. Chem. Chem. Phys., 2019,21, 17971-17977 with permission from the Royal Society of Chemistry.

of the states are CT states, dark but abundant in the energy range of the ligand absorption. Once we investigated the position of the electron holes and excited electrons on these state, we observed how the hole is prevalently located on one of the nucleobases and the excited electron on the probe. These are oxidative states, where the presence of the positively charged probe induces the injection of the electron hole to the double strand. These states cannot be populated directly by light, but once an electron is promoted on the π^* state of the ligand, this can reduce itself transferring the hole to one of the nucleobases. In order to estimate the probability of transferring the electron hole to one nucleobase over the other,

we looked into the hole distribution among the four nucleobases. The ligands interact with both an A and a T in both, the coding (directionality 5' → 3') and the template (directionality 3' → 5') strand, but the two A and T bind asymmetrically the ligand (Figure 4.5). Indeed, MCH1 and MCH2 share the same binding spot and,

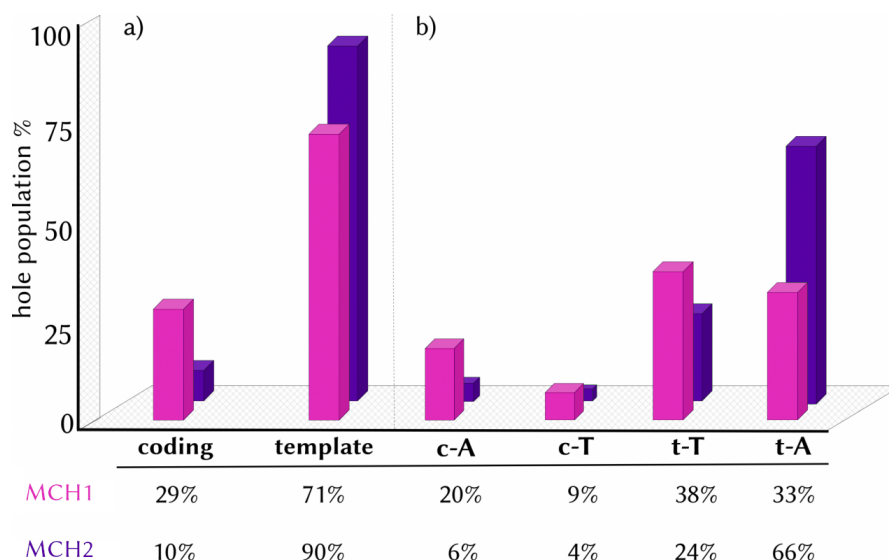


Fig. 4.7: Histogram representation of the electron hole distribution in CT states among the four nucleobases, group by strand of belonging (coding or template) and by single nucleobases (c-A, c-T, t-A and t-T) for both MCH1-DNA (pink) and MCH2-DNA (purple) complexes. Adapted from *Phys. Chem. Chem. Phys.*, 2019,21, 17971-17977 with permission from the Royal Society of Chemistry.

in particular, point an electron donating group (OH) to the template strand and an electron withdrawing group (NO_2 and CN_2H_4^+) to the coding strand, making the binding to the two strands asymmetric. The probe interacts more strongly with the two nucleobases on the coding strand and stabilizes more their molecular orbitals, while the weaker interaction with the template strand makes the hole transfer from its nucleobases to the ligand more favourable. Indeed, 71 and 90 % of the hole population, for the complex with MCH1 and MCH2, respectively, is found being on the nucleobases of the template strand. The higher percentage found for MCH2 can be ascribed to the more selective and stronger interaction established by the ligand with the bases of the two strands, with a calculated non-covalent interaction energy being more than 10 kcal mol^{-1} weaker for the template strand. In particular, MCH2 can establish stronger interaction with the A than with the T of the template strand, leaving this last one as the preferential site for the hole injection in the double strand DNA, as demonstrated by the percentage of the hole population found on this nucleobases (Figure 4.7).

In conclusion, we characterised the excited state of ligand-DNA complexes and proposed merocyanine ligands as potential oxidative agent of DNA, which may be used as a probe of electrochemical properties of the macromolecule as well as to induce damage in the genetic code. Full discussion, methods and additional details can be found in the Appendix (section 6.4), under the reprinted publication titled "Directional and regioselective hole injection of spiropyran photoswitches intercalated into A/T-duplex DNA".

Conclusions and Outlook

In this work different computational methods were applied to study the binding of small ligands to different DNA structures and potential damage induced by light irradiation.

As DNA binders, spiropyran derivatives were chosen, since they were proposed to bind (poly-dAT)₂ DNA as well as, if opportunely modified, G4 after isomerisation to their open merocyanine isomer. However, the mechanism of action and the binding mode were not fully described and lacked the support of computational studies. In this work, the binding mechanism was clarified using umbrella sampling simulations and the most favourable binding modes were evaluated by means of unrestrained full-atom classical molecular dynamics. The binding energy was computed and compared for different possible binding modes. These calculations were run for both binding to A-T rich and G4 structures, since for both structures a binding with merocyanine was demonstrated.

Concerning the simulation of photoinduced damage, two different sources of possible damage of DNA components and structures were investigated. First, the oxidative damage of adenosine triphosphate dianion in the gas phase was studied with a combination of photoelectron spectroscopy and *ab initio* calculations, where computational chemistry helped to identify a prevalent site for the electron emission and oxidation. Second, due to the oxidative nature of the merocyanine ligand, the possible damage induced by merocyanines bound to both (poly-dAT)₂ and G4 was considered by investigating the UV absorption properties of the ligand-DNA complexes. For these studies the development of an ad-hoc computational strategy and a big computational effort were required, to include up to thirteen chromophores, in the case of ligand-G4 complex, and to calculate several thousands of excitations, in the presence of the environment. These calculation allowed to compute the absorption spectra of ligand-DNA complexes and to characterize their excited states and identify the bright states absorbing UV light, the nature of the excitations and the presence of potentially detrimental charge transfer states where the nucleobases donate an electron to the oxidative ligand.

The results obtained, open the doors for further investigation. Starting from the spiropyran-DNA interaction, the most suitable binding modes of merocyanine

to DNA were identified. Atomistic-level models were provided and the most important interactions were highlighted, quantifying the role of electrostatic and Van der Waals interaction. The next step would be to investigate the binding of opportunely modified new derivatives that might enhance the binding and the selectivity for other double strand or other quadruplex structures. Regarding ATP photooxidation, the electron emission by tunneling through a repulsive Coulomb barrier in ATP was identified and follow up studies, possibly based on time-resolved spectroscopy techniques associated with excited state dynamics simulations, might offer additional knowledge of the light-response of adenine and its derivatives. Finally, the excited state characterisation of the ligand-DNA complexes raised important questions about the formation and population of potentially detrimental DNA excited states where the nucleobases result to be oxidated. In this work, as a first step they were identified and quantified, but in order to describe the pathways leading to their potential population, or in general the relaxation of the DNA excited states after light absorption, excited states dynamics simulations are necessary. In order to do that, methods for nonadiabatic dynamics need to be extended to include a QM/MM scheme able to capture the effect of the environment, while still including a large amount of chromophores; only in this way it would be possible to describe accurately the complicated and highly delocalised excitations of the ligand-DNA complexes. All these suggestions are highly stimulating and hopefully, also based on the approaches and results presented in this work, will be a stimulus for further discoveries related to light-control of DNA binding and damage.

Appendix: Reprinted Publications

In this appendix, the publications that compose this work, previously summarised and discussed, are reprinted.

6.1 DNA-binding Mechanism of Spiropyran Photoswitches: The Role of Electrostatics

Davide Avagliano, Pedro A. Sánchez-Murcia, Leticia González

Phys. Chem. Chem. Phys., 2019,21, 8614-8618
<https://doi.org/10.1039/C8CP07508E>

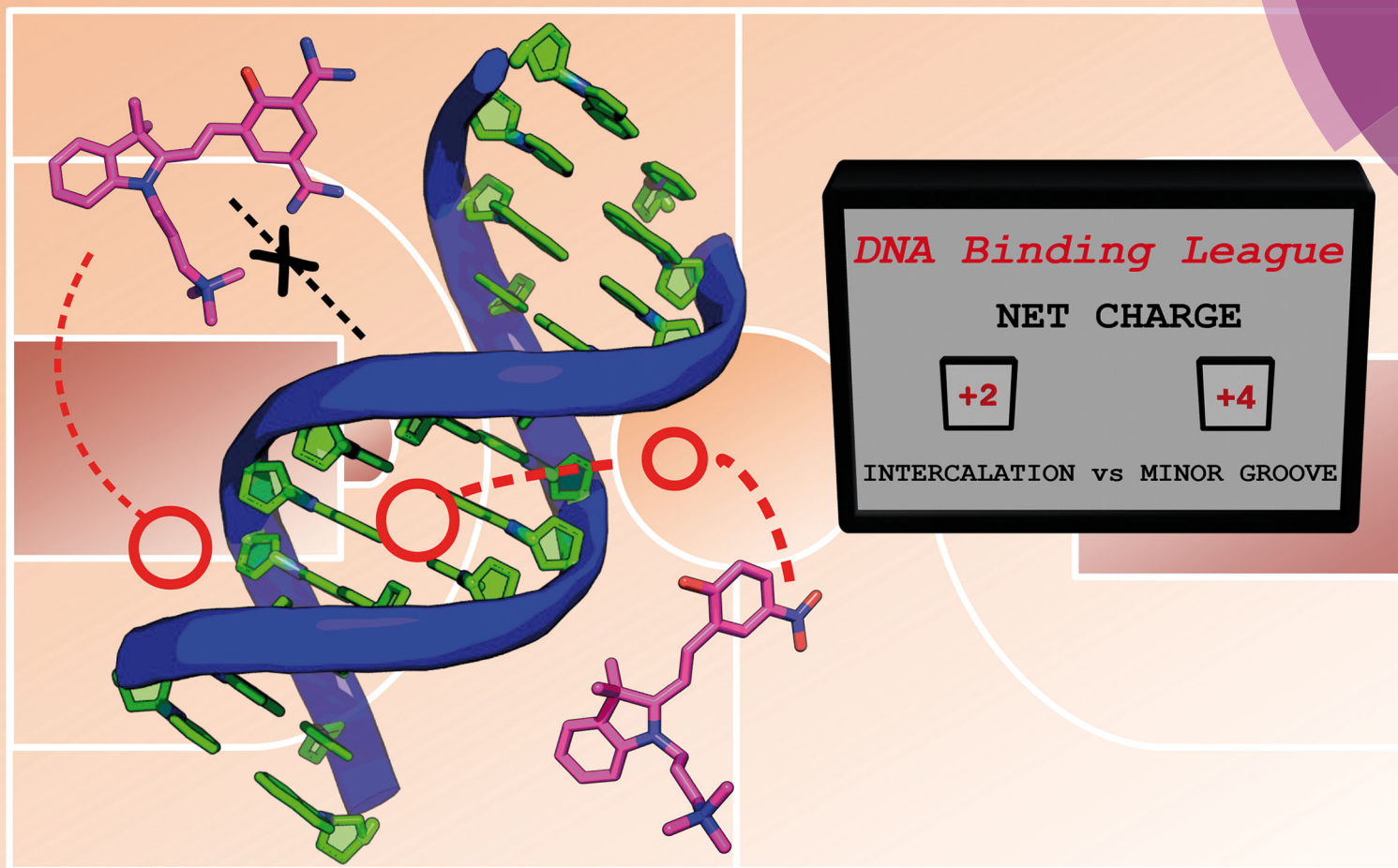
Contributions:

- D. AVAGLIANO: Performed the calculations and the analysis, wrote the first draft of the manuscript and contributed to the final version of the manuscript.
- P.A. SÁNCHEZ-MURCIA: Supervised the methodological development and contributed to the initial and final draft of the manuscript
- L. GONZÁLEZ: conceived the scope of the manuscript, supervised the developments and contributed to the final manuscript

Reproduced from Phys. Chem. Chem. Phys., 2019,21, 8614-8618 with permission from the Royal Society of Chemistry

PCCP

Physical Chemistry Chemical Physics
rsc.li/pccp



ISSN 1463-9076



ROYAL SOCIETY
OF CHEMISTRY

Celebrating
IYPT 2019

COMMUNICATION

Pedro A. Sánchez-Murcia, Leticia González *et al.*
DNA-binding mechanism of spiropyran photoswitches:
the role of electrostatics





Cite this: *Phys. Chem. Chem. Phys.*,
2019, 21, 8614

Received 7th December 2018,
Accepted 15th February 2019

DOI: 10.1039/c8cp07508e

rsc.li/pccp

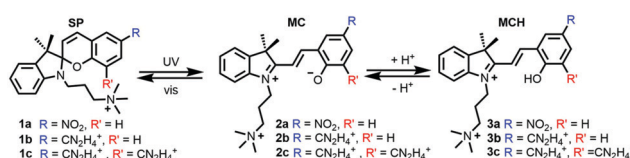
DNA-binding mechanism of spiropyran photoswitches: the role of electrostatics†

Davide Avagliano, Pedro A. Sánchez-Murcia * and Leticia González *

The binding mechanism of the protonated open form of three spiropyran derivatives into a 12-mer (poly-dAT)₂ has been unveiled by means of computational methods. It is found that the electrostatic term in the probe:DNA binding energy, modulates the binding mode, providing new guidelines for the design of spiropyran photoswitches with specific binding modes to DNA.

Molecular photoswitches – molecules that change their chemical configuration upon light absorption – are useful tools to monitor the conformational state of biomolecules, in both temporal and spatial scales.¹ The advantages of these probes are that the implicated species in the photochemical equilibrium highly differ in the binding affinity to the macromolecule upon light absorption, that these species are interconvertible in a reversible manner, and most importantly, that the two species can be selectively accessed by illumination at different wavelengths and/or by thermal activation. Spiropyrans (SP) (Scheme 1) belong to a well-known class of photoswitches, which after light irradiation undergo heterocyclic cleavage to yield the *trans*-merocyanine form (MC).^{2–6}

Spiropyrans have an indoline and a chromene moiety oriented perpendicular to each other and bound together *via* a C_{sp}³ spiro-junction. Upon absorption of UV light, the sigma C–O bond on the spiranic carbon cleaves and builds the MC form. Thereafter, MC can reversibly turn to SP with heat or after irradiation with visible light. Whereas the SP species has moderate lipophilicity, the conjugated coplanar system MC is a zwitterionic species with increased polarity favouring intercalation to polyanions such as DNA *via* π – π stacking interactions. In contrast, the SP form shows no or very little affinity to DNA. The half-time of the MC form increases upon substitution on the phenyl ring with electron withdrawing groups, such as NO₂.^{9–12} For instance, almost half of the population in solution of **1a** (Scheme 1), which contains the scaffold 3',3'-dimethyl-6-nitrospiro[2H-1-benzopyran-2,2'-indoline] (6-nitro-BIPS) with a



Scheme 1 Spirocyclic (SP) compounds **1a**,⁷ **1b**,⁸ the proof of concept **1c** and their corresponding unprotonated (MC, **2**) and protonated (MCH, **3**) merocyanine open forms. Compounds **3** are studied in this work.

trimethylammoniumpropyl group on the indole nitrogen, is converted to **2a** after irradiation at 254 nm.⁷ Despite its efficiency, the binding mode of this family of photoswitches into DNA is still under debate. Experiments with flow-oriented linear dichroism (LD) spectroscopy strongly support that **2a** intercalates into DNA given the dramatic blue-shift of the absorption spectrum of the drug in solution.⁷ A few years later, the same authors found¹³ that the maximum of the absorption spectrum of the protonated form **3a** in buffered aqueous solution is blue-shifted compared to **2a**, in analogy to what was previously observed after binding of **2a** to DNA.⁷ Therefore, it is likely that **2a** gets protonated to form **3a** before binding to DNA.

The latter hypothesis was also supported by the measurement of the DNA binding constants of the amidinium-substituted derivatives **1b**–**3b**.⁸ The protonated **3b** form intercalates stronger into DNA by a factor of *ca.* 50 compared to the unprotonated **2b** one. Again, and in analogy to **3a**, the protonated form **3b** was proposed to be the relevant DNA binder with a similar angle in the LD experiment (*ca.* 70°). Based on this experimental evidence, the authors ascribed the differences in the binding affinities of **3a** and **3b** to the total net positive charge on the photoprobe.⁸ However, the mechanism of the intercalation of **3a** and **3b** remains uncertain since before intercalation these species could bind to different grooves of the DNA.¹⁴ This is, for example, the case of the drug daunomycin (aka. doxorubicin) that can visit up to four states before intercalation, some of them pre-intercalative states,¹⁵ which were identified by molecular dynamics (MD) simulations as local minima bound in the minor groove of DNA.^{16,17}

Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna,
Währinger Strasse 17, 1090, Vienna, Austria. E-mail: leticia.gonzalez@univie.ac.at
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8cp07508e



In this work, we unravel the intercalation modes of spiro-pyran derivatives from the solvent into DNA using MD and umbrella sampling simulations. To this aim, a 12-mer (poly-dAT)₂ has been selected as these compounds are known to bind to AT.⁷ Upon light-driven ring opening and concomitant protonation of the phenoxy moiety (MCH in Scheme 1), **3a** intercalates into (poly-dAT)₂. However, before inserting into DNA, **3a** has to travel from the bulk solvent into the nucleic acid. Potentially, the probe can visit a few binding modes before intercalation, *i.e.* entering DNA by the major (path M) or by the minor (path m) groove, as sketched in Fig. 1a. Experimentally, it is known⁷ from flow-oriented LD that **3a** in the presence of DNA shows one negative peak around 420 nm. A single peak indicates that **3a** has only one binding mode. Furthermore, it was determined that the photoprobe forms an angle of $80 \pm 5^\circ$ with the main Z-axis of DNA, meaning that the transition dipole moment of **3a** is parallel to the plane of the sandwiched nucleobases, and thus, the molecule is intercalated in a sandwich-like way between the DNA base pairs. Unfortunately, and to the best of our knowledge, neither kinetics nor structural data of the intercalation process of any of the derivatives of the SP family have been reported to date. Therefore, we explore here the two possible binding pathways to the major (M) and minor (m) groove of **3a** by means of umbrella sampling MD simulations.

As the reaction coordinate (RC) for binding to both grooves, we define the distance between the center-of-mass of **3a** and the center-of-mass of the four nucleobases in between **3a** inserts.

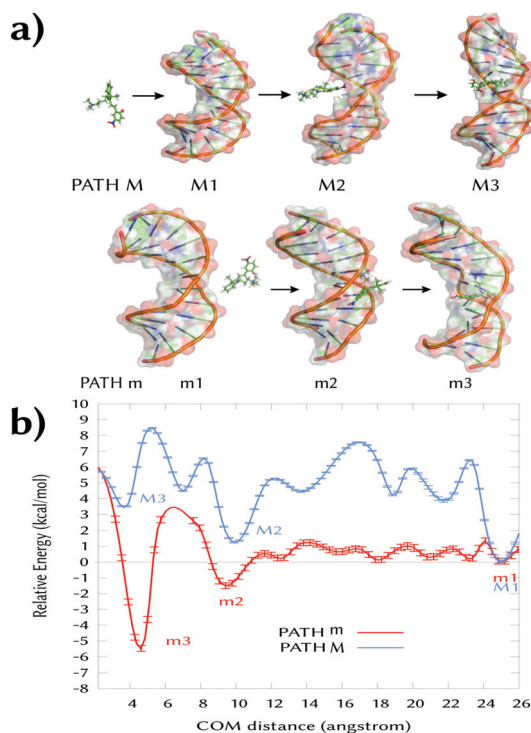


Fig. 1 (a) Intercalation process of **3a** into a 12-mer (poly-dAT)₂ through path M (major groove) and path m (minor groove). (b) Free energy profiles (kcal mol⁻¹) of path M (major) and m (minor) computed with umbrella sampling MD simulations. The bootstrap error values are shown over the plots.

From the initial intercalated geometries (named M3 and m3 in Fig. 1a), we pull out the probe from the DNA (RC ~ 2 Å) into the bulk solvent up to RC = 26 Å (M1 and m1). The whole RC was divided into 50 windows separated by 0.5 Å and the system was allowed to oscillate around the RC for 5 ns per window – the time at which the energy curves were converged (Sections S1 and S2, ESI[†]). The initial guesses for M3 and m3 were obtained by manual docking after structural superimposition of our system with the crystal structure of daunomycin intercalated into DNA (PDB id. 1VTH).¹⁸ Then, the two systems were minimized and equilibrated in explicit solvent. Due to the presence of the two methyl groups on the indoline ring of **3a** (Scheme 1), there was only one possibility in each case for the relative orientation of these groups with respect to the methyl group of the two thymine units, in between the probe intercalates in the states M3 and m3.

Fig. 1b shows the obtained energy profiles for path M (blue line) and m (red line). In both cases, the chemical probe visits several pre-intercalative states before intercalation. However, our calculation clearly predicts that binding to the major groove (pathway M) is unfavorable. When **3a** approaches to the DNA through the major groove it visits only metastable states since all the local minima between 25 and 3 Å are higher in energy compared to the minimum M1. Moreover, the intercalated state M3 is less stable than both the solvated state M1 (RC = 25 Å, $\Delta E > 3$ kcal mol⁻¹) and the local minimum M2 (at ~ 10 Å). Inspection of the M3 geometry reveals that the ligand's rings are coplanar and inserted between four nucleobases parallel to them. In such a conformation, the binding site is highly distorted and the hydrogen bonds between the surrounding base pair are broken (A7:T6' and A5:T8', Fig. S1, ESI[†]). This explains the unfavorable energy for the state M3.

In contrast, the intercalated state m3 is the most thermodynamically stable binding mode. Along the minor groove binding pathway, we identified several metastable pre-intercalative states of **3a** (between RC = 8 Å and RC = 23 Å) that end up in the local minimum m2 at RC = 9 Å. Since the energy of m2 is quite similar to m1 ($\Delta E = -1.5$ kcal mol⁻¹), the binding mode m2 is reversible and **3a** could return to the bulk solvent. However, supplying only *ca.* 5 kcal mol⁻¹, **3a** can overcome the energy barrier between m2 and m3 to reach the latter intercalative state. This extra energy is used to open the space between the nucleobases.

Fig. 2 illustrates the geometry of **3a** in the m2 state. The indoline and the phenol rings are almost co-planar with a small rotation (*ca.* 15° with respect to each other, Fig. S2, ESI[†]) and the molecule is parallel to the DNA backbone. The nitro group is oriented to the same side of the ammoniumpropyl tail and projected outwards to the solvent. In such a disposition, the phenolic hydroxyl group establishes a hydrogen-bond with the carbonyl oxygen of a thymine T6' (dashed line, Fig. 2a). This interaction remains when the system is further simulated during 100 ns without restraints. This observation, together with the orientation of the phenol of **3a** in m2 and m3, could explain the higher experimental binding constant values for the MCH form (**3a**) compared to the MC one (**2a**).⁷ Indeed, as can be seen in Fig. 2b and 3a, the phenoxy group in **2a** would face a negatively-charged region in both m3 and m2 states, respectively.



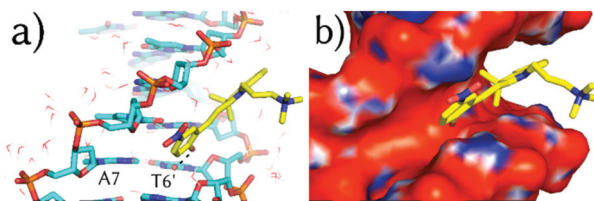


Fig. 2 (a) Representative binding geometry of **3a** (sticks, carbon atoms colored in yellow) in m2 after 100 ns of unrestrained MD simulation. Only the closest water molecules (lines) are shown. (b) The APBS-generated electrostatic potential on the van der Waals surface of the DNA in m2 in the presence of **3a**.

In addition, the computed adaptive Poisson–Boltzmann solver (APBS) electrostatic potential on the surface of the DNA for m2 (Fig. 2b) shows that **3a** binds to a region with high negative potential (red contour) and that the nitro group is oriented to a positive region (blue contour). Thus, **3a** establishes a large number of contacts in the short range (*ca.* 243 in an inner sphere of 4 Å) along the MD simulation, including the interactions between the trimethylammoniumpropyl group of **3a** and the phosphate groups in the DNA.

Once **3a** visits m2, the system requires *ca.* 5 kcal mol^{−1} to make room between the four stacking nucleobases A7:T6' and A5:T8' it binds to. Once in m3, only the phenol ring of **3a** is parallel to the planes defined by the former four nucleobases and almost perpendicular to the main Z-axis of the DNA (Fig. 3a). In m3, **3a** establishes face-to-face π – π stacking interactions between the phenol ring and the base pairs T6:A7' and A5:T8', as observed for other intercalative agents.¹⁹ The distances between the phenol ring and each of the former base pairs are, in fact, below 4 Å. The hydrogen-bonds of the Watson–Crick base pairs T6:A7' and A5:T8' and the surrounding A7:T6' and T8:A5' are kept in the presence of **3a** but the distance between the nucleobases on the same strand T6 and A5 and A7' and T8' is increased from the non-intercalative crystallographic distance of 5 Å (like between A7 and T6) to 7.1 Å and 6.3 Å, respectively.

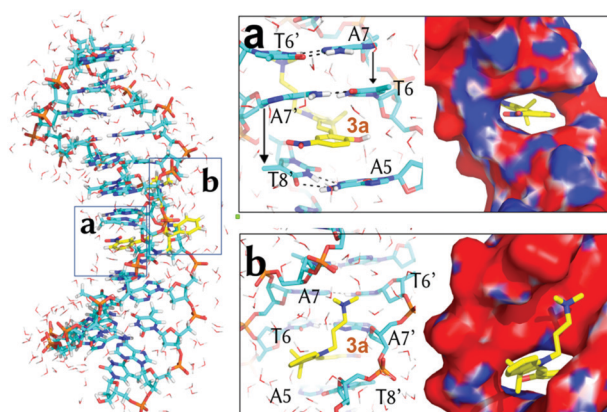


Fig. 3 Representative binding geometry of **3a** in m3 after 100 ns of unrestrained MD simulation. Boxes (a) and (b) zoom-in on the detailed orientation of the phenol and the indoline rings, respectively. Only the closest water molecules (lines) are shown. The APBS-generated electrostatic potential is also shown on the van der Waals surface of the DNA.

Finally, the nitro group of **3a** is again looking at a positive potential region (Fig. 3a) and the ammoniumpropyl tail interacts with a region with negative potential in between the phosphate groups (Fig. 3b).

The intercalation of **3a** increases the angle between the indoline and the chromene moieties (Fig. S2, ESI†) and enlarges the amplitude of this hampered rotation between the two rings. This finding agrees with the larger experimental wavelength shift in the absorption spectrum upon intercalation⁷ than when going from water to chloroform in the absence of DNA, what may indicate a conformational change to a nonplanar structure of **3a** in m3.

The substitution of the nitro group in **3a** by an amidinium group (compound **3b**)⁸ is known to increase the binding constant of the photoprobe to DNA. This effect could be due to the specific interactions of the amidinium group, the increment in the net positive charge in **3b**, or a combination of both factors. In order to discern between these explanations, we simulated the binding states m2 and m3 of **3b** with the (poly-dAT)₂. We directly employed the already calculated binding states m2 and m3 of **3a**, as there is no experimental evidence of different binding geometries or orientations between the nitro and the amidinium substituted MCH (**3a** and **3b**), except for a slight difference in the measured angle in the flow-oriented linear dichroism (LD) spectra.⁸

Our computations show that **3b** in both m2 and m3 states shares the main binding spots found in the nitro derivative (Fig. S3, ESI†) with the amidinium group projected to the solvent. In the m2 binding mode, the amidinium group turned out to have no specific interaction with the DNA backbone or with the nucleobases. Therefore, the first possibility highlighted above cannot explain why **3b** binds stronger to DNA than **3a**. Thus, we computed the binding energies of **3a** and **3b** to DNA using the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) method.²⁰

We considered both m2 and m3 binding modes for each of the compounds. Table 1 collects the total solvated binding energies of **3a** and **3b** to DNA and their individual terms. In both cases, the intercalative state m3 is the most stable state (Fig. S4, ESI†). **3b** shows the largest binding energy value due to a smaller positive (repulsive) electrostatic energy component (ΔE^{elec}) than in **3a**. This term has two contributions: the electrostatic component of the Poisson–Boltzmann solvation energy (ΔE^{PB}) and the electrostatic term in the gas phase (ΔE^{EEL}) (Table S1, ESI†). In the pre-intercalative state m2, ΔE^{elec} strongly decreases between **3a** and **3b** (109%) so that the electrostatic component is so severely reduced in **3b** that it turns negative (attractive by -1.02 kcal mol^{−1}). Therefore, one is left to conclude that the increase of the positive charge on the probe strengthens the binding of **3** to DNA.

With this conclusion in hand, and as a proof of concept, we designed the photoprobe 6,4-diamidinium-MCH (**3c**, Scheme 1), where an extra amidinium group was introduced in the *ortho* position to the hydroxyl group of the phenol ring. This compound – with an extra net charge – should validate our hypothesis that additional electrostatic contributions could

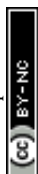


Table 1 Computed MM-PBSA mean binding energies (kcal mol^{−1}) at 300 K of **3a–3c** to (poly-dAT)₂. The errors represent the standard deviations of the mean values. The total binding energy includes the entropy contribution

Compound	Mode	ΔE^{elec}	Non-electrostatic terms		$T\Delta S$	ΔG^{bind}
		$(\Delta E^{\text{EEL}} + \Delta E^{\text{PB}})$	ΔE^{vdW}	ΔE^{nonpol}		
3a	m3	15.81 ± 0.14	−55.17 ± 0.25	−4.32 ± 0.01	−21.73 ± 0.09	−21.96 ± 0.24
3b		10.10 ± 0.40	−55.78 ± 0.23	−4.62 ± 0.40	−20.96 ± 0.17	−29.35 ± 0.33
3c		11.09 ± 0.59	−50.71 ± 0.42	−4.57 ± 0.03	−26.42 ± 0.37	−17.77 ± 0.74
3a	m2	10.82 ± 0.22	−37.72 ± 0.41	−3.23 ± 0.01	−20.97 ± 0.02	−9.29 ± 0.47
3b		−1.02 ± 0.60	−34.16 ± 0.26	−3.22 ± 0.01	−23.88 ± 0.13	−14.52 ± 0.36
3c		−12.07 ± 0.39	−29.91 ± 0.49	−3.19 ± 0.02	−24.71 ± 0.26	−20.46 ± 0.61

then further increase the binding energy to DNA, and thereby, its binding constant. The *ortho* position was selected since it is chemically modifiable²¹ and the orientation of the group does not interfere with the proposed m2 and m3 binding modes. The values for the total binding energy are also shown in Table 1. Surprisingly, we observed that the extra amidinium group of **3c** increases the binding energy only for the m2 mode, and the m3 state is less stable than the corresponding state in **3a** or **3b**. Note that the role of the entropy, although significant, is not the decisive term for the relative stability between states m2 and m3.

Fig. 4 depicts the binding energies for both the m2 and m3 binding modes of all three derivatives **3a–3c**. Whereas in **3a** and **3b** the m2 state is always below in energy compared to m3, in **3c** the pre-intercalative state m2 competes with m3 as the most stable binding mode. This means that, from the thermodynamical point of view, **3c** would not intercalate, but rather bind in the minor groove, as other agents do.²² Our results are supported by the fact that **3c** has similar size and charge compared to other minor groove binders (Fig. S5, ESI†) and also, like most of them, shows pronounced AT selectivity.

These results indicate that care has to be taken in the modulation of the electrostatic component in photoactivatable DNA-targeting spiropyran. Although experiments suggest that MCH species bind to DNA better than MC due to its increased electrostatics (*e.g.* **2b** versus **3b**),⁸ an excess of net positive charge on MCH can stabilize the pre-intercalative state m2 such that

this binding site becomes competitive (*e.g.* **3c** versus **3b**); this is *e.g.* the case in polyamides or diamidines, which are species known to selectively bind AT pairs in the minor groove but not to intercalate.²²

In summary, this work indicates that the modulation of the electrostatic potential in this family of spiropyran can be used as a tool to control their binding mode to DNA (minor groove binder or intercalation). And most importantly, the current assumption that increasing the net positive charge on the spiropyranic photoswitcher always will favor the intercalation process is, at least, not generalizable and may require further experimentation.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

D. A. thanks the H2020-MSCA-ITN-2017 ETN programme LightDyNAMics, under grant agreement no. 765266. P. A. S.-M. thanks the Austrian FWF (Project M 2260). The authors would like to thank Dr J. J. Nogueira for preliminary discussions on spiropyran. The Vienna Scientific Cluster (VSC) is acknowledged for kind allocation of computational resources.

Notes and references

- W. Szymański, J. M. Beierle, H. A. V. Kistemaker, W. A. Velema and B. L. Feringa, *Chem. Rev.*, 2013, **113**, 6114–6178.
- A. Mustafa, *Chem. Rev.*, 1948, **43**, 509–523.
- (a) R. Klajn, *Chem. Soc. Rev.*, 2014, **43**, 148–184; (b) C. L. Fleming, S. Li, M. Grøtli and J. Andréasson, *J. Am. Chem. Soc.*, 2018, **140**, 14069–14072.
- V. I. Minkin, *Chem. Rev.*, 2004, **104**, 2751–2776.
- S. Prager, I. Burghardt and A. Dreuw, *J. Phys. Chem. A*, 2014, **118**, 1339–1349.
- F. Liu and K. Morokuma, *J. Am. Chem. Soc.*, 2013, **135**, 10693–10702.
- J. Andersson, S. Li, P. Lincoln and J. Andréasson, *J. Am. Chem. Soc.*, 2008, **130**, 11836–11837.
- M. Hammarson, J. R. Nilsson, S. Li, P. Lincoln and J. Andréasson, *Chem. – Eur. J.*, 2014, **20**, 15855–15862.

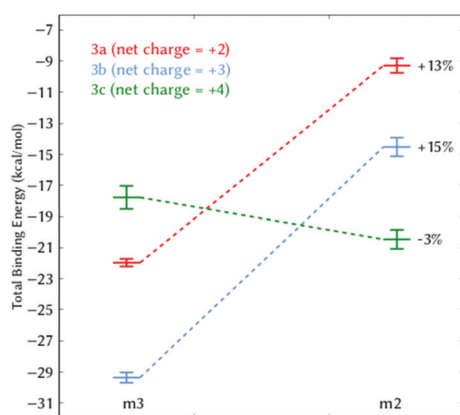
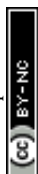


Fig. 4 Total binding energy (kcal mol^{−1}) of **3a–3c** between the binding modes m2 and m3. The percentage indicates the relative energetic change of m2 compared to m3.



- 9 C. Lenoble and R. S. Becker, *J. Phys. Chem.*, 1986, **90**, 62–65.
- 10 M. Gehrtz, C. Braeuchle and J. Voitlaender, *J. Am. Chem. Soc.*, 1982, **104**, 2094–2101.
- 11 Y. Sheng, J. Leszczynski, A. A. García, R. Rosario, D. Gust and J. Springer, *J. Phys. Chem. B*, 2004, **108**, 16233–16243.
- 12 S. Aldridge and C. Jones, *Chem. Soc. Rev.*, 2016, **45**, 763–764.
- 13 M. Hammarson, J. R. Nilsson, S. Li, T. Beke-Somfai and J. Andréasson, *J. Phys. Chem. B*, 2013, **117**, 13561–13571.
- 14 A. A. Almaqwashi, T. Paramanathan, I. Rouzina and M. C. Williams, *Nucleic Acids Res.*, 2016, **44**, 3971–3988.
- 15 V. Rizzo, N. Sacchi and M. Menozzi, *Biochemistry*, 1989, **28**, 274–282.
- 16 A. Mukherjee, R. Lavery, B. Bagchi and J. T. Hynes, *J. Am. Chem. Soc.*, 2008, **130**, 9747–9755.
- 17 M. Wilhelm, A. Mukherjee, B. Bouvier, K. Zakrzewska, J. T. Hynes and R. Lavery, *J. Am. Chem. Soc.*, 2012, **134**, 8588–8596.
- 18 C. M. Nunn, L. Van Meervelt, S. Zhang, M. H. Moore and O. Kennard, *J. Mol. Biol.*, 1991, **222**, 167–177.
- 19 G. Barone, C. F. Guerra, N. Gambino, A. Silvestri, A. Lauria, A. M. Almerico and F. M. Bickelhaupt, *J. Biomol. Struct. Dyn.*, 2008, **26**, 115–129.
- 20 N. Homeyer and H. Gohlke, *Mol. Inf.*, 2012, **31**, 114–122.
- 21 (a) C. Kaiser, T. Halbritter, A. Heckel and J. Wachtveitl, *ChemistrySelect*, 2017, **2**, 4111–4123; (b) H. Görner, *Phys. Chem. Chem. Phys.*, 2001, **3**, 416–423.
- 22 (a) R. R. Tidwell and D. W. Boykin, in *Dicationic DNA Minor Groove Binders as Antimicrobial Agents, in Small Molecule DNA and RNA Binders*, ed. M. Demeunynck, C. Bailly, W. D. Wilson, John Wiley & Sons, Ltd, Chichester, 2002, pp. 414–460; (b) B. Nguyen, D. Hamelberg, C. Bailly, P. Colson, J. Stanek, R. Brun, S. Neidle and W. David Wilson, *Biophys. J.*, 2004, **86**, 1028–1041; (c) D. D. V. H. Xuemei Cai and P. J. Gray Jr., *Cancer Treat. Rev.*, 2009, **35**, 437–450.



Supporting information

DNA-Binding Mechanism of Spiropyran Photoswitches: the Role of Electrostatics

Davide Avagliano, Pedro A. Sánchez-Murcia* and Leticia González*

Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Währinger Straße 17,
A-1090 Vienna (Austria)

Section S1 Molecular dynamics simulation setup

The initial molecular model of the 12-mer dsDNA (poly-dAT)₂ was prepared using the *nab* program implemented in AmberTools14, as described elsewhere.^{1,2} The MCH form of **3a** was inserted into the dsDNA manually and minimized *in vacuum* following a standard protocol using *sander*. The resulting intercalated molecular model **3a**:(poly-dAT)₂ was used as starting geometries for the umbrella sampling MD studies, where both compounds were pulled out to the bulk solvent. Atom-centered ESP charges of **3a**, **3b** and **3c** were computed with *antechamber* (AmberTools17)¹ following the standard procedure for classical MD simulation on their *ab initio* optimized geometries using HF/6-31G* with the Gaussian suite.³ Their GAFF force field parameters were computed using the module *parmchk2*⁴ of Amber17¹ and the ff14SB force field parameter set⁵ was also used to assign the bonded and non-bonded parameters of compounds **3a**, **3b** and **3c** and to describe the nucleotides of the dsDNA. Each MCH:dsDNA complex (e.g. **3a**:dsDNA) was immersed in a cubic box of 30 Å from the solute to the border of the box filled with TIP3P water molecules⁶ and 20 Na⁺ to ensure electroneutrality. In addition, and in order to reproduce the experimental conditions reported by Andersson *et al.*,⁷ a final NaCl concentration of 1 10⁻⁵ M was achieved by addition of Na⁺ and Cl⁻ atoms. *tleap*¹ was used for the setup of all systems. In all cases, periodic boundary conditions were used, and the electrostatic interactions were computed using the Ewald method⁸ with a grid spacing of 1 Å. The cutoff distance for the non-bonded interactions was 10

Å and the SHAKE algorithm⁹ was applied to all bonds involving hydrogens. An integration step of 2.0 fs was defined. Before the production phase, all the simulated systems were prepared following a sequential protocol: First, the system was minimized in three steps, where all the hydrogens, the solvent molecules (waters and counter ions) and, finally, the whole system, were sequentially minimized in 20,000 steps. For the initial 10,000 steps, the steepest descendent algorithm was used; the last 10,000 steps were run using the conjugate gradient algorithm. The resulting minimized solvated geometries were heated from 100 to 300 K in 20 ps using the Langevin thermostat¹⁰ with an integration step of 0.2 fs (collision frequency of 1 ps⁻¹) but keeping the position of all the heavy atoms of the solute restrained with a strong harmonic constant (40 kcal⁻¹ mol⁻¹ Å⁻²). In this step, a random seed was imposed. The Langevin dynamics, as well as the initial velocity for the dynamics are dependent of such random number. After that, the imposed restrains were removed in 6 steps of 20 ps each, where the system was switched from a NVT (40 to 10 kcal⁻¹ mol⁻¹ Å⁻² in 80 ps, constant volume) to a NPT (10 to 0 kcal⁻¹ mol⁻¹ Å⁻² in 40 ps, constant pressure) ensemble. As standard conditions, each of the MCH:dsDNA complexes were simulated at 300 K and constant pressure (1 atm) for three independent MD simulations of 30, 100 or 300 ns with an integration step of 2.0 ps using the *pmemd.cuda* engine of single precision - fixed precision (SPFP)^{11,12} on two GeForce Nvidia GTX 1080Ti GPUs.

Section S2 Umbrella sampling

The intercalation pathways were calculated by umbrella sampling MD simulations using Amber17.¹ The reaction pathway was divided in 50 steps, every 0.5 Å, each one obtained with Steered Molecular Dynamics (SMD) simulations. The SMD simulation was run for 0.1 ns and a harmonic constant of 50 kcal⁻¹ mol⁻¹ Å⁻². For the umbrella sampling MD simulation the force constant was increased to 1000 kcal⁻¹ mol⁻¹ Å⁻² and each of the 50 windows was allowed to oscillate around the anchor position for 5 ns with a

time step of 1 fs. The energy analysis of the potential mean force was performed with the variational free energy profile method.¹³

Section S3 Binding energy analysis

The binding energy analysis was carried out with the program MMPBSA.py implemented in AmberTools17,¹ which allows to calculate the binding free energy with the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA).¹⁴ The entropy contribution, useful to obtain a better description of the order of magnitude of interaction between the two molecules, was also computed with the same program through normal mode analysis.¹⁵

The MM-PBSA analysis provided us: (i) a total electrostatic energy term (ΔE^{elec}), which includes the difference between the ΔE^{EEL} term (non-bonded electrostatic energy + 1,4-electrostatic energy) in gas phase and the electrostatic contribution in the complex probe:DNA to the solvation free energy calculated by Poisson Boltzmann (ΔE^{PB}), (ii) the non-electrostatic terms: a van der Waals term (ΔE^{vdW}) and the nonpolar contribution to the solvation free energy (ΔE^{nonpol}), (iii) an entropy term computed using Normal Mode Analysis using the same frames (ΔS), (iv) and finally, a total solvation free energy term (ΔG^{bind}).

For the binding energy analysis, in each of the MCH:dsDNA complexes we selected a window of 20 ns (400 snapshots) from their trajectories where the root-mean squared distance (RMSD) value (Å) for the complex was constant along the MD simulation. In order to obtain the electrostatic map on the dsDNA surface, we followed the procedure included in the CHARMM software,^{16,17,18} solving numerically the Poisson-Boltzmann equation and obtaining the electrostatic free energy on the macromolecule surface.

Section S4 Trajectories analysis and graphical representation

The 3D representation of the complex was studied with PyMOL 1.8.6.¹⁹ The binding modes are interpreted with CPPTRAJ,²⁰ a program of AmberTools17, which allows to extract important values from the MD trajectories, such as distances between DNA and dye contacts.

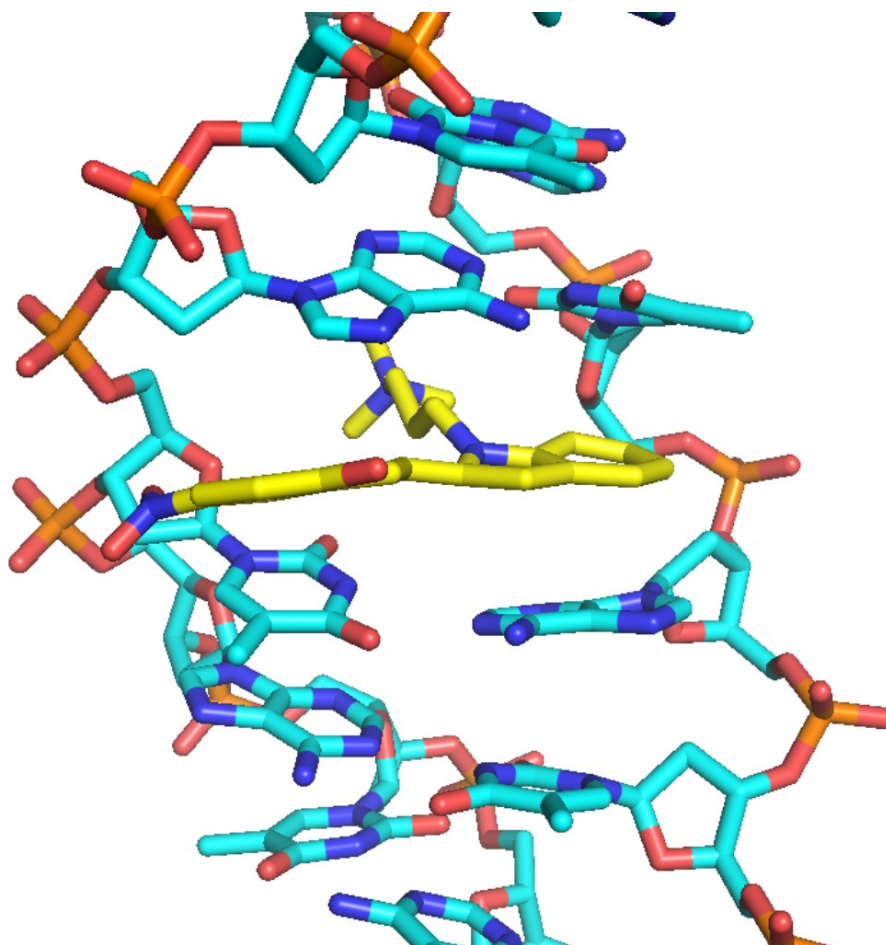


Figure S1: Intercalative geometry of **3a** (sticks, carbon atoms colored in yellow) in the binding mode M3. The intercalation of **3a** from the major groove (path M), introduces a strong perturbation into the dsDNA. As result, the state M3 is metastable.

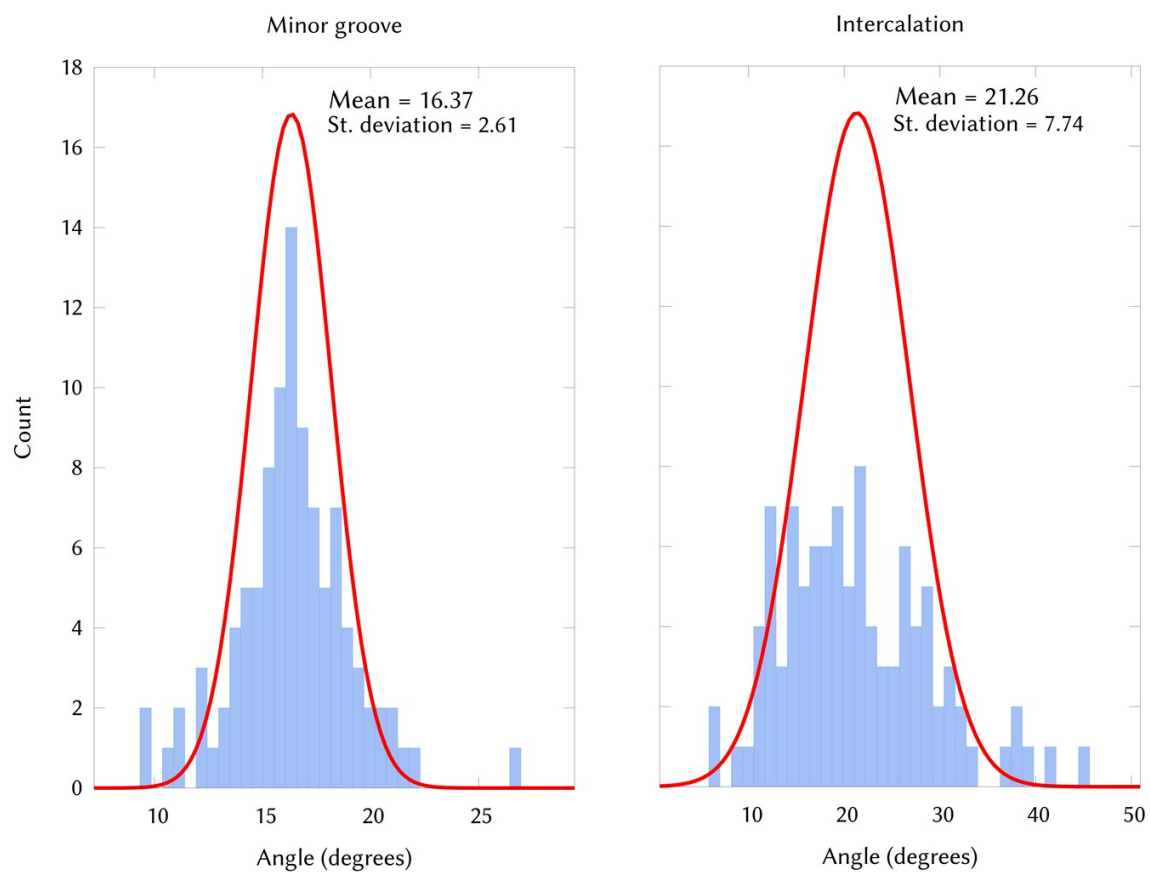


Figure S2: Distribution of the dihedral angle ($^{\circ}$) between the two rings of **3a** in the m2 (left) and m3 (right) states along the 100ns-unrestrained MD simulations.

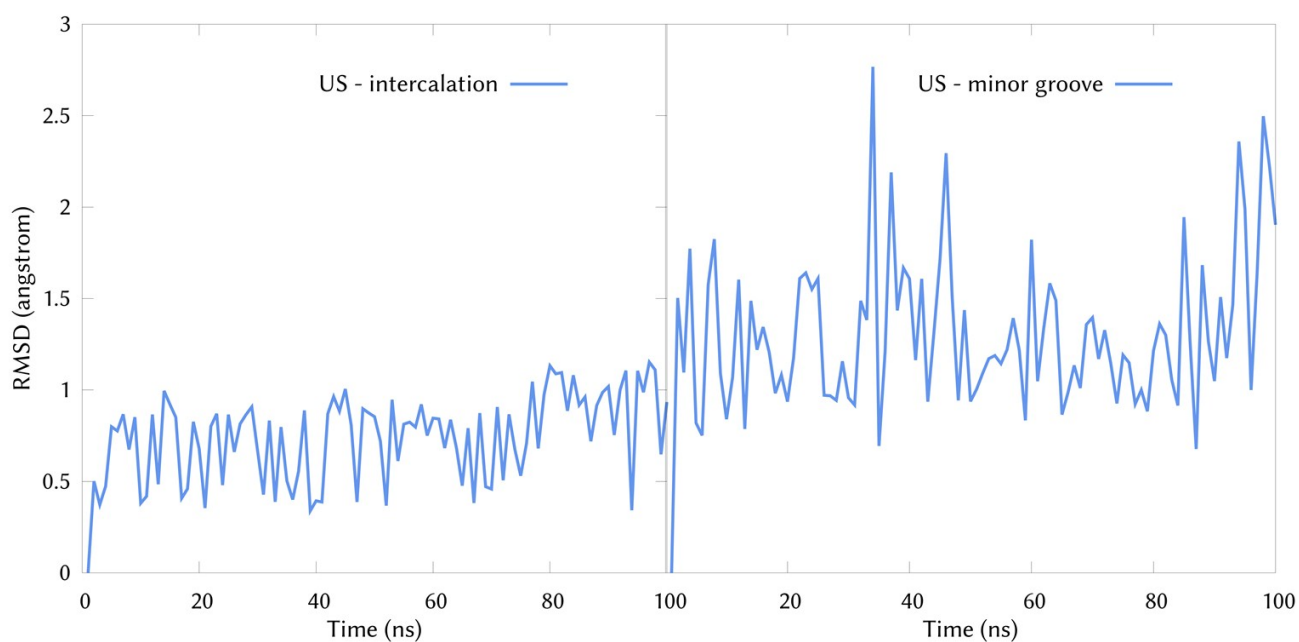


Figure S3: RMSD value (Å) of **3b** along the 100ns-unrestrained MD simulations in the m3 (left) and in the m2 (right) states. The geometries of **3a** in m3 and m2 are taken as reference and only the heavy atoms have been considered. The RMSD values are under 2.0 Å, which indicates that **3b** shares the same binding spots of **3a**.

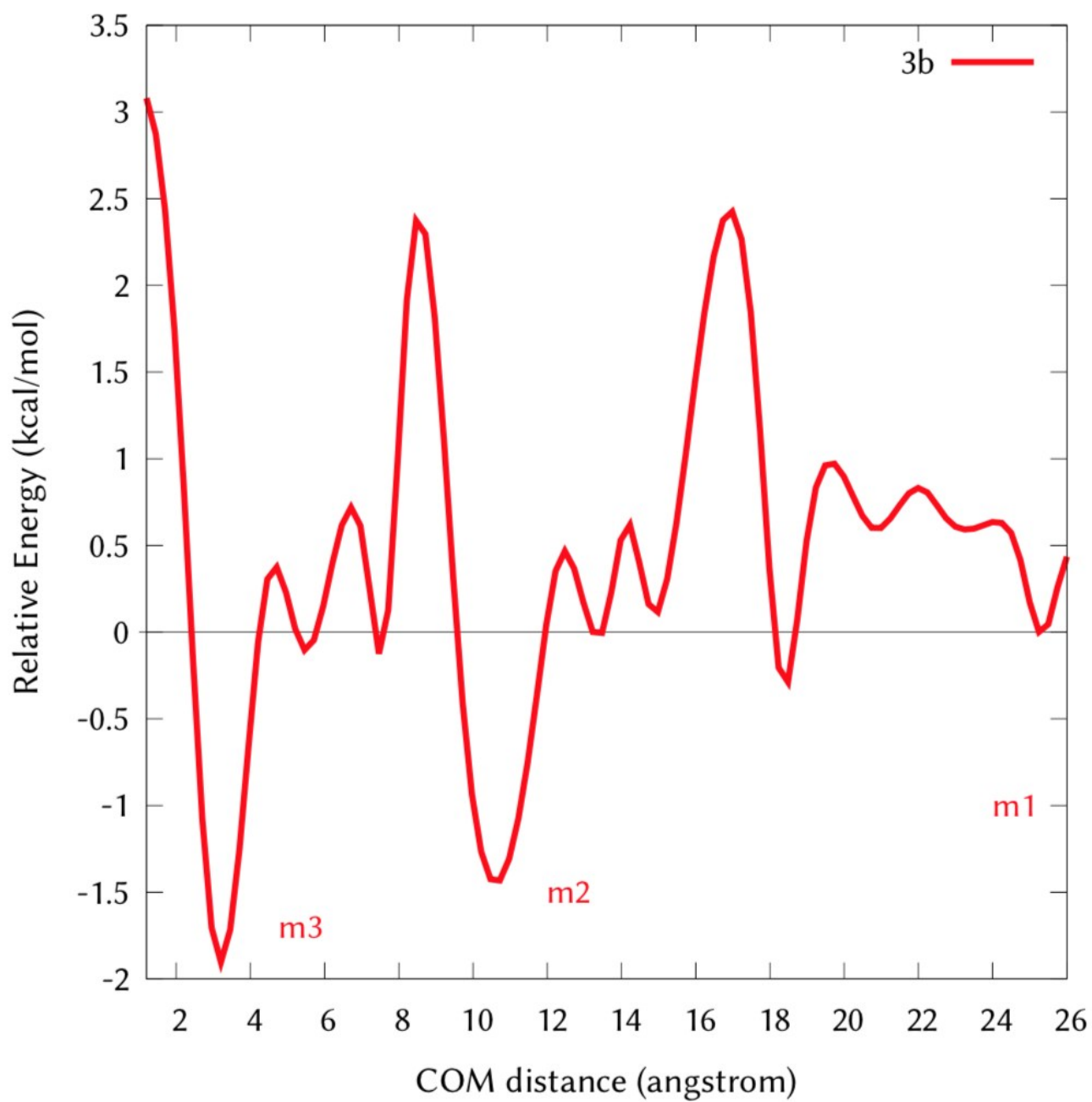


Figure S4: Free energy profiles (kcal mol⁻¹) of path m (minor) of **3b** computed with US MD simulations.

Table S1. Total binding energy (ΔG^{bind} , kcal mol⁻¹) and individual terms used for its calculation.

cmp (charge)	bind. mode	ΔE^{EEL}	ΔE^{PB}	ΔE^{elec} ($\Delta E^{\text{EEL}} +$ ΔE^{PB})	ΔE^{vdW}	$\Delta E^{\text{non-pol}}$	$\Delta E^{\text{vdW+non-pol}}$	ΔG^{MMPBSA}	-TAS	Total ΔG^{bind}
3a (+2)	m3	-1112.69 ± 1.04	1128.5 ± 1.03	15.81 ± 0.14	-55.17 ± 0.19	-4.32 ± 0.01	-59.49 ± 0.19	-43.69 ± 0.32	-21.73 ± 0.09	-21.96 ± 0.24
3b (+3)		-1643.37 ± 2.06	1653.47 ± 2.02	10.1 ± 0.40	-55.78 ± 0.26	-4.62 ± 0.40	-60.4 ± 1.33	-50.31 ± 0.36	-20.96 ± 0.17	-29.35 ± 0.33
3c (+4)		-2155.88 ± 4.39	2126.97 ± 4.35	11.09 ± 0.59	-50.71 ± 0.49	-4.57 ± 0.03	-55.28 ± 0.49	-44.19 ± 0.61	-26.42 ± 0.37	-17.77 ± 0.74
3a (+2)	m2	-997.49 ± 1.09	1008.31 ± 1.17	10.82 ± 0.22	-37.72 ± 0.25	-3.23 ± 0.01	-40.95 ± 0.25	-30.27 ± 0.24	-20.97 ± 0.02	-9.29 ± 0.47
3b (+3)		-1448.37 ± 1.78	1447.35 ± 1.88	-1.02 ± 0.60	34.16 ± 0.23	-3.22 ± 0.01	-37.38 ± 0.23	-38.40 ± 0.34	-23.88 ± 0.13	-14.52 ± 0.36
3c (+4)		-2032.87 ± 7.62	2020.81 ± 7.63	-12.07 ± 0.39	-29.91 ± 0.42	-3.19 ± 0.02	-33.1 ± 0.42	-45.17 ± 0.74	-24.71 ± 0.26	-20.46 ± 0.61

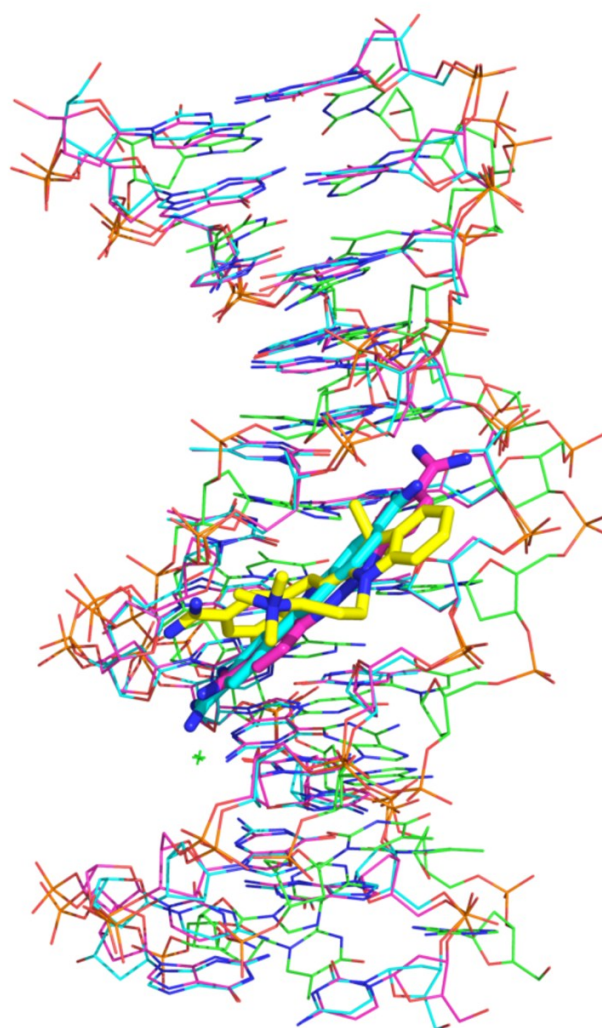


Figure S5: Structural superimposition of the intercalated **3c** (carbon atoms colored in yellow) into 12-mer (poly-dAT)₂ and the crystal structures of minor groove binders berenil (C-atoms colored in blue, PDB id. 1D30) and DAPI (C-atoms colored in magenta, PDB id. 2DBE).^{21,22}

References

1. D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman 2018, *AMBER 2018*, University of California, San Francisco.
2. J. J. Nogueira and L. González, *Biochem.* 2014, **53**, 2391–2412.
3. Gaussian 09, Revision D.01, 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, and D. J. Fox, *Gaussian, Inc.*, Wallingford CT, 2016.
4. J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.* 2006, **25**, 247–260.
5. J. W. Ponder and D. A. Case., *Adv. Prot. Chem.* 2003, **66**, 27-85.
6. W. Junmei, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.* 2005, **25**, 1157–1174.
7. J. Andersson, S. Li, P. Lincoln and J. Andréasson, *J. Am. Chem. Soc.* 2008, **130**, 11836–11837.
8. P. P. Ewald, *Ann. Phys.* 1921, **369**, 253–287.

9. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.* 1977, **23**, 327–341.
10. X. Wu, B. R. Brooks and E. Vanden-Eijnden, *J. Comput. Chem.* 2016, **37**, 595–601.
11. A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.* 2012, **8**, 1542–1555.
12. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.* 2013, **9**, 3878–3888.
13. T. Lee, B. K. Radak, A. Pabis and D. M. York, *J. Chem. Theory Comput.* 2013, **9**, 153–164.
14. N. Homeyer and H. Gohlke, *Mol. Inform.* 2012, **31**, 114–122.
15. B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory Comput.* 2012, **8**, 3314–3321.
16. S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.* 2008, **29**, 1859–1865.
17. S. Jo, M. Vargyas, J. Vasko-Szedlar, B. Roux and W. Im, *Nucleic Acids Res.* 2008, **36**, W270–W275.
18. W. Im, D. Beglov and B. Roux, *Comput. Phys. Commun.* 1998, **111**, 59–75.
19. DeLano, W. L. The PyMOL Molecular Graphics System, version 1.8.6 Schrödinger, LLC. 2015.
20. D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.* 2013, **9**, 3084–3095.
21. Brown D.G., Sanderson M.R., Skelly J.V., Skelly J.V., Jenkins T.C., Brown T., Garman E., Stuart D.I. and Neidle S., *EMBO J.* 1990, **9**, 1329–1334.
22. Larsen T.A., Goodsell D.S., Cascio D., Grzeskowiak K and Dickerson R.E., *J. Biomol. Struct. Dyn.* 1989, **7**, 477–491.

6.2 Spiropyran Meets Guanine Quadruplexes: Isomerization Mechanism and DNA Binding Modes of Quinolizidine-Substituted Spiropyran Probes

Davide Avagliano, Pedro A. Sánchez-Murcia, Leticia González

Chem. Eur. J. 2020, 26, 13039
<https://doi.org/10.1002/chem.202001586>

Contributions:

- D. AVAGLIANO: Conceived the idea of the manuscript, performed the calculations and the analysis, wrote the first draft of the manuscript and contributed to the final version of the manuscript.
- P.A. SÁNCHEZ-MURCIA: Supervised the methodological development and contributed to the analysis and to the initial and final draft of the manuscript
- L. GONZÁLEZ: conceived the scope of the manuscript, supervised the developments and contributed to the final manuscript

Reproduced from Chem. Eur. J. 2020, 26, 13039-13945 with permission of the authors and Wiley-VCH GmbH

■ Spiro Compounds

Spiropyran Meets Guanine Quadruplexes: Isomerization Mechanism and DNA Binding Modes of Quinolizidine-Substituted Spiropyran Probes

Davide Avagliano,^[a] Pedro A. Sánchez-Murcia,^{*,[a]} and Leticia González^{*,[a, b]}

Abstract: The recent delivery of a fluorescent quinolizidine-substituted spiropyran, which is able to switch in vivo and bind to guanine quadruplexes (G4) at physiological pH values, urged us to elucidate its molecular switching and binding mechanism. Combining multiscale dynamical studies and accurate quantum chemical calculations, we show that, both in water and in the G4 environment, the switching of the spiropyran ring is not promoted by an initial protonation event—as expected by the effect of low pH solutions—but that the deprotonated merocyanine form is an

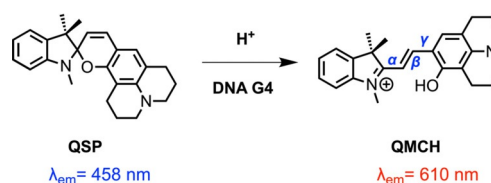
intermediate of the reaction leading to the protonated open species. Additionally, we investigate the binding of both deprotonated and protonated open forms of merocyanine to c-MYC G4s. Both species bind to G4s albeit with different hydrogen-bond patterns and provide distinct rotamers around the exocyclic double bond of the merocyanine forms. Altogether, our study sheds light on the pharmacophoric points for the binding of these probes to DNA, and thereby, contributes to future developments of new G4 binders of the remarkable family of quinolizidine-substituted spiropyrans.

Introduction

Guanine quadruplexes (G4s) are non-canonical secondary structures that can be adopted by particular guanine-rich sequences.^[1] They are constituted by stacked guanine tetrads (G-tetrads), formed by four guanine bases interacting with each other through Hoogsteen hydrogen bonding, and chelating a central metal cation. Depending on the relative orientation of the phosphate backbone, strand direction, and specific base-sequence, these guanine-rich RNA or DNA can fold following different patterns and resulting in a wide diversity of topological families of G4s.^[2,3] G4s have been found to be abundant in cancer-related genes as well as in the telomeres of the chromosomes, and in recent years they have emerged as promising therapeutic targets to silence oncogenes.^[4] For these reasons,

it is important to identify new probes that bind, stabilize, probe, or damage G4s.^[5,6]

Quinolizidine-substituted spiropyran (QSP, see Scheme 1) is a fascinating example of an organic fluorescent dye with the ability to specifically target G4s over single- and double-stranded DNA. QSP can change its fluorescence emission maximum spatiotemporally, switching from its closed spiropyran form to the open protonated merocyanine form (QMCH) in acid media (e.g., inside a lysosome) or upon binding in vivo to G4s of regulatory genes such as c-MYC under physiological pH.^[7] In the QMCH form, the emission is redshifted to 610 nm compared with the QSP form (emission at 458 nm). The evidence that QSP opens to QMCH in acid environment as well as in the presence of c-MYC G4 (at neutral pH values), suggests that the conversion from QSP to QMCH is a proton-mediated process. Although other spiropyran derivatives have been observed to isomerize in the presence of acid as well, they require more severe conditions (e.g., trifluoroacetic acid):^[8] indoline benzo-spiropyran (BIPS) derivatives present lower pK_a values than QSP ($pK_a \approx 5.9$) making pH-mediated opening under physiological conditions inviable. Thus, quinolizidine-spiropyrans constitute a unique family of probes to be used in vivo.



Scheme 1. Quinolizidine-substituted spiropyrans (QSP) isomerizes to the protonated merocyanine (QMCH) under acid conditions or by DNA G4 binding. Bonds for the definition of the QMCH isomers are shown in blue (α , β , γ).^[9]

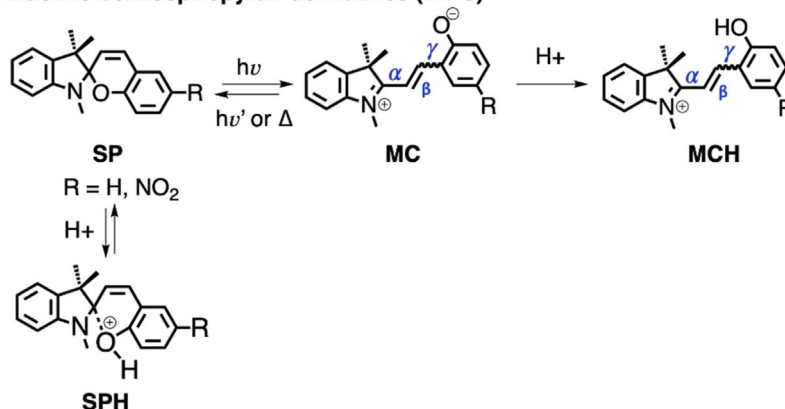
[a] D. Avagliano, Dr. P. A. Sánchez-Murcia, Prof. Dr. L. González
Institute of Theoretical Chemistry, Faculty of Chemistry
University of Vienna
Währinger Straße 17, 1090 Vienna (Austria)
E-mail: pedro.murcia@univie.ac.at
leticia.gonzalez@univie.ac.at

[b] Prof. Dr. L. González
Vienna Research Platform on Accelerating Photoreaction Discovery
University of Vienna
Währinger Straße 17-A, 1090 Vienna (Austria)

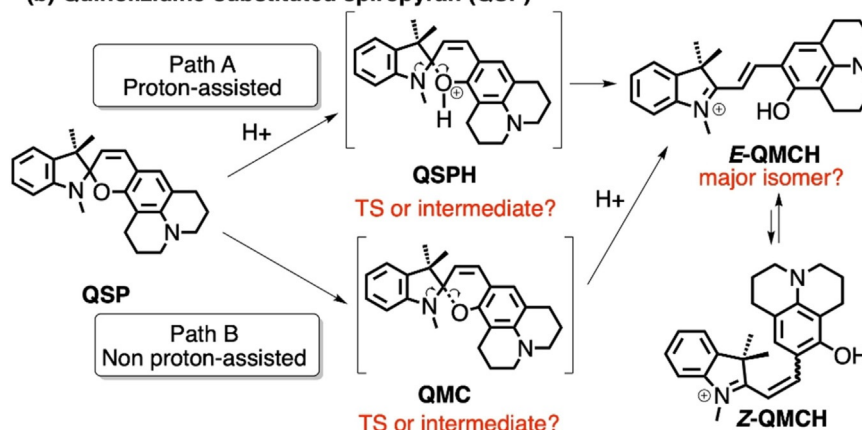
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
<https://doi.org/10.1002/chem.202001586>.

© 2020 The Authors. Published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

(a) Indoline benzospiropyran derivatives (BIPS)



(b) Quinolizidine-substituted spiropyran (QSP)



Scheme 2. (a) The protonation of BIPS SP prevents SP from evolving to MCH. (b) Two possible reaction mechanisms for the QSP–QMCH switching process in solutions with low pH values.

The protonation mechanism of SPs has been extensively discussed. Wojtyk and co-workers^[8] employed ¹H NMR and UV/Vis spectroscopy to show that the protonated species of 6-nitro BIPS is spontaneously generated upon protonation of the oxygen in the presence of increasing concentrations of acid (Scheme 2a). The authors proposed that the SPH species acts as an ‘unreactive sink’, trapping 6-nitro BIPS SP and competing with the SP→MCH opening process. The addition of an extra base to SPH restores the concentration of SP, and in the presence of proper light illumination, promotes the switching to the open MC/MCH species. More recently, Browne and co-workers^[10] showed that under controlled pH conditions and UV irradiation both, MC and MCH *E*-isomers of unsubstituted and 6-nitro BIPS can be formed in solution. Nevertheless, the access to the *E*-isomer of the MCH species required the UV illumination of SP for at least one of the steps. It must be stressed that the *E*-isomer of the exocyclic double bond in MCH (β -bond in Scheme 2a) can present an ensemble of rotamers owing to the relative orientation of the two rings with respect to the exocyclic double bond (α and γ bonds).^[9] The TTC isomer (*E-E-Z* for α , β , and γ bonds, respectively) has been reported to be the most stable form of both MC and MCH species.^[11]

In contrast to the former BIPS derivatives, QSP points to a different switching mechanism where no light is required and where protonation seems to be involved. However, no details in that regard—except for some initial structure-based design campaign^[7]—are known owing to the absence of theoretical calculations. As shown in Scheme 2b, the ring opening from QSP to QMCH requires the breaking of the sigma bond C_{spiro}–O and the protonation of this oxygen. Thus, depending on the order of these two processes, there will be at least two possible reaction mechanisms (paths A and B, Scheme 2b). The aim of the present work is twofold: first, we identify the most probable chemical pathway to activate QSP in solution at low pH values and the nature of its intermediate species. Subsequently, we provide an atomistic molecular model for the binding mode of QSP to the parallel-stranded G4 of the c-MYC promoter,^[12] beyond static molecular docking studies,^[7] by using extensive molecular dynamics simulations up to the μ s-scale. To fulfil these aims, we employ state-of-the-art complementary approaches including quantum mechanical (QM), mixed quantum and classical molecular mechanical (QM/MM), and molecular dynamics (MD) simulations. The results provide a comprehensive picture of QSP chemistry, from solution to its binding to G4 upon thermal opening.

Results and Discussion

First of all, we investigated the two possible reaction mechanisms leading to the formation of QMCH from QSP (recall Scheme 2b). In path A, the protonation of QSP occurs first, with the formation of either a QSPH intermediate or a transition state (TS), before QMCH is generated. In path B, the QSP ring opening and formation of QMC is followed by the subsequent protonation. In contrast to A, in B protonation would not affect the kinetics of the QSP opening, but the thermodynamics of the process.

The reaction mechanism was investigated by using steered QM/MM MD simulations together with two-dimensional umbrella sampling (2D-US) calculations, both in explicit water solution and bound to c-MYC G4 (see computational details in Sections S1–S3 of the Supporting Information). These simulations allowed us to (i) observe the $C_{\text{spiro}}-\text{O}$ bond breaking, (ii) explicitly include the effect of the water or DNA environment in the reaction, and (iii) explore two reaction coordinates independently. The systematic exploration of the potential energy surfaces provided stationary points and the minimum free energy pathway. The two reaction coordinates are defined as the bond breaking between the carbon C_{spiro} and the oxygen O atoms of the spiro-junction (RC_1) and the proton transfer to O from a water molecule (H_{water}) of the solvent (RC_2 ; see Figure 1a). Although Whereas RC_1 is defined as a single distance between two atoms, RC_2 is specified as a linear combination of $O_{\text{water}}-H_{\text{water}}$ and $O_{\text{spiro}}-H_{\text{water}}$ distances. In the QM/MM calculations, the QM region—treated with the Density Functional Tight-Binding (DFTB3) semiempirical level of theory^[13]—contains the probe and a water molecule.

Figure 1b shows the 2D-US free energy profile of the reaction in water, where 600 windows were required to cover the full 2D free energy space (Figures S1 and S2 in the Supporting Information). Through these simulations, four local minima were identified: the initial QSP, two unprotonated open forms of QMC (QMC_1 and QMC_2), and the product QMCH (Figure 1c). The QMCH species is the thermodynamically most stable and, hence, the formation of a previous protonated QSPH species is unfeasible, neither as a TS nor as an intermediate. In contrast, the two minima, corresponding to the unprotonated QMC_1 and QMC_2 intermediates, are thermally accessible through a transition state $TS_{\text{QSP-QMC}}$ lying at 5 kcal mol⁻¹. Importantly, once the system reaches the transition state, it can evolve either to the QMC metastable intermediates or directly to the QMCH species. This means that the pH of the aqueous solution can determine the chemical pathway to get QMCH. These results suggest, therefore, that path B is more probable in solution than path A, as no initial protonation of the QSP species is necessary to reach QMCH.

The two QMC isomers (QMC_1 and QMC_2) are almost degenerate in energy (≈ 4 kcal mol⁻¹) in agreement with the fact that the merocyanine species can adopt different isomers with respect to the double bond (β -bond). The two obtained rotamers correspond to the *E*-isomer of the double bond and differ in their relative orientation of the phenyloxy group with respect to the nitrogen of the indoline (i.e., rotation around the

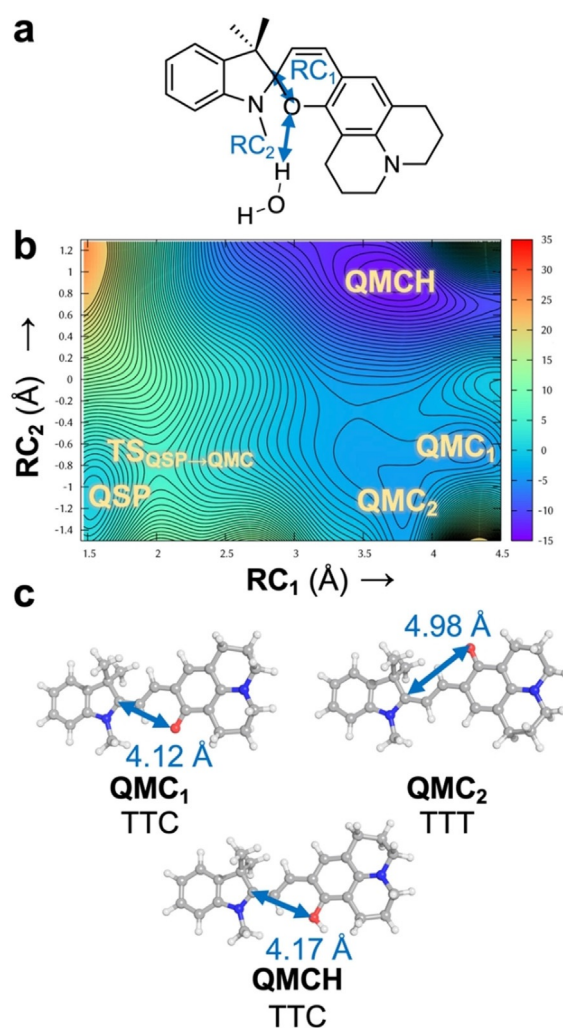
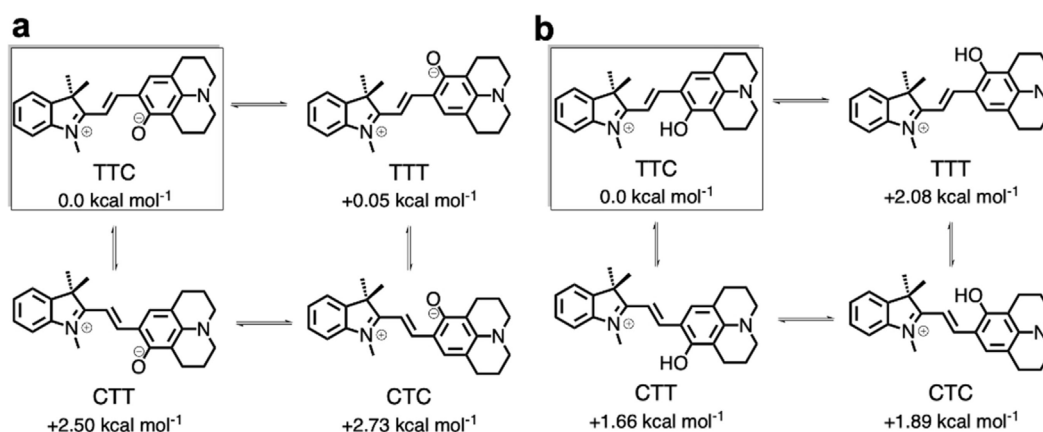


Figure 1. (a) Definition of the reaction coordinates RC_1 and RC_2 (Å) as a distance and a linear combination of distances, respectively. (b) 2D free energy surface (kcal mol⁻¹) for the reaction QSP → QMC/QMCH computed by means of 2D steered and umbrella sampling QM/MM MD simulations in explicit aqueous solvent. The zero energy is set to the QSP minimum and each contour line represents a difference of 0.5 kcal mol⁻¹. (c) Geometries of the local minima identified in the 2D-US energy surface after optimization in implicit solvent (RI-MP2/def2-TZVP@COSMO). The distance for RC_1 is highlighted with a two-headed arrow. TTC and TTT differ in the *cis* → *trans* rotation with respect to the γ bond.

α -bond). To obtain an accurate estimation of the energetics of all stable geometries, four structures, corresponding to possible combinations of rotations around the α/γ -bonds, were optimized at a higher level of theory (Møller–Plesset second-order perturbation theory, using the resolution of identity approximation and a large basis set, RI-MP2/def2-TZVP, Figure 1c, Scheme 3, and Supporting Information) with an implicit description of the aqueous solution (the conductor-like screening COSMO method,^[14] see Section S4 in the Supporting Information). Only *E*-isomers were explored as no *Z*-species were found in our 2D-US studies and the *E*-isomers have been shown experimentally to be thermodynamically more stable for benzoindolic spiropyran.^[8]



Scheme 3. Relative stability of stable rotamers of the *cis-trans E*-isomer with respect to the α/β bond (TTC, TTT, CTT, CTC) of QMC (a) and QMCH (b) species calculated at RI-MP2/def2-TZVP@COSMO level of theory. The zero of the Gibbs energy (kcal mol^{-1} , 300 K) is set to the most stable isomer TTC.

Among the four possible QMC conformations, the TTC and TTT isomers (corresponding to the QMC₁ and QMC₂ geometries obtained in the 2D free energy surface, respectively, Figure 1c) are the energetically lowest minima, with an energy difference of only $0.05 \text{ kcal mol}^{-1}$ between them. Thus, both TTC and TTT isomers are equally stable in water. The computed relative energies agree with the DFT results for the 6-nitro BIPS derivatives. In addition, we also computed the relative energy of the *E*-isomers of QMCH. In this case, the protonation of the phenol oxygen changes slightly the relative stability, resulting in the TTC isomer being $2.08 \text{ kcal mol}^{-1}$ lower in energy than TTT (Scheme 3b). Our RI-MP2 calculations agree well with previous experimental data^[7,10] and with the results of the 2D-US simulation. In summary, QSP would populate—based on the thermodynamics—mainly the TTC isomer of QMC(H) in solution and the protonation would not promote the thermal ring opening.

We then investigated what happens in the G4 context: does the polynucleotide change the reaction mechanism? To answer this question, we studied the reaction in the local environment of G4 with the same methodology. As no structure of QSP bound to G4 is available, we superimposed QSP with the ligand position of a quindoline/*c*-MYC G4 complex (PDB id: 2L7V)^[12] and relaxed the whole system with the MD simulation protocol described in Section S1 (in the Supporting Information). As in solution, the probe and the water molecule closest to the oxygen of the spiropyran were considered in the QM region and the same definition for the reaction coordinates RC_1 and RC_2 was used, as well as the same number of windows (600) and data points (1000 per window, 600 000 total; see Section S3 in the Supporting Information).

The obtained 2D free energy surface is shown in Figure 2a, along its predicted stationary points for QSP, QMC, and QMCH. Accordingly, the pathway leading to QMCH is presumably the same as in water (path B of Scheme 1). The energy barrier of $TS_{QSP \rightarrow QMC}$, as well as the relative energies of the QMCH and QMC minima are comparable to those in water. However, in the presence of G4, only the TTC isomer of the QMC form is identified. Further, a new local minimum, Z-QMCH ($RC_1 = 3.1 \text{ \AA}$,

$RC_2 = 1.2 \text{ \AA}$) is found in the 2D free energy surface (Figure 2a). This isomer corresponds to the Z-isomer identified previously by Browne et al.^[10] for the 6-nitro BIPS (SPH, Scheme 2a), where the oxygen of the spiropyran is protonated and the $C_{\text{spiro}}\text{--O}$ bond broken. In the gas-phase optimized geometry (RI-MP2/def2-TZVP), the distance between the C_{spiro} and the O atoms is 2.71 \AA . In this geometry, the two rings around the exocyclic double bond (β -bond) are not co-planar, with a dihe-

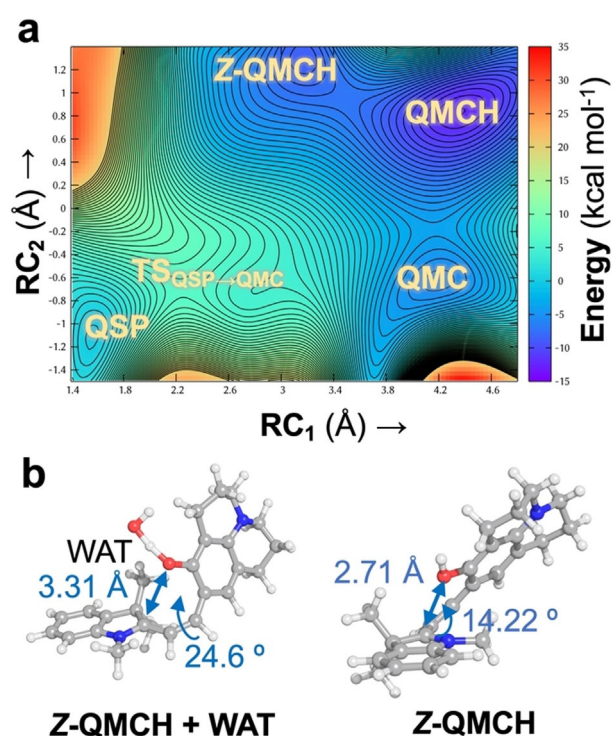


Figure 2. (a) Two-dimensional free energy surface (kcal mol^{-1}) for the opening $QSP \rightarrow QMC/QMCH$ when bound to G4. The zero energy is set to the QSP minimum and each contour line represents a difference of $0.5 \text{ kcal mol}^{-1}$. (b) Identified geometry (Z-QMCH + WAT) and optimized geometry of the Z-QMCH intermediate at RI-MP2/def2-TZVP level of theory in the gas phase.

dral angle of 14° . The steric repulsion of the two methyl groups on C3 of the indoline ring with the oxygen atom of Z-QMCH prevents the co-planarity of both π -systems. This angle is slightly larger in the geometry obtained in our 2D-US exploration (24° , Z-QMCH+WAT, Figure 2). According to these results, Z-QMCH would then be accessible through thermal equilibrium from the more stable QMCH, whereas in the case of 6-BIPS derivatives it requires UV-light irradiation. Interestingly, Z-QMCH was not found in our previous 2D-US exploration in water. The electrostatic environment exerted by the G4 (polyanion, phosphate groups negatively charged) may be responsible for the relative stabilization of this positively charged Z-isomer intermediate.

In summary, our results indicate that the thermal opening of QSP to form QMCH is not a proton-assisted process neither in water nor bound to G4. As soon as QSP opens, it forms the *E*-QMC species, which, in the presence of protons in solution (i.e., aqueous solution at physiological pH), evolves to the more stable QMCH species. In the DNA context, QSP follows the same reaction mechanism. However, the electrostatics of the environment allows the isomerization between Z-QMCH and *E*-QMCH, the latter species being the most stable one. In both cases, the pH value of the solution will control the QMC/QMCH ratio. At low pH values or physiological pH, the equilibrium will be shifted to QMCH. At high pH values, it is expected to have a representative population of QMC. The role of the DNA may be the same as the one of an excess of protons: as G4 binds more strongly to QMCH species over QSP or QMC, it will shift the equilibrium towards the QMCH:G4 complex. This chemical equilibrium is different from other SP derivatives. The quinolizidine substitution increases the thermodynamic stability of the QMC with respect to the closed QSP isomer, owing to its structural and chemical nature (i.e., tertiary amine group vs. $-\text{NO}_2$ group in 6-nitro BIPS), which was the result of a structure-based G4 probe design.^[6] These features make the QSP an excellent candidate as a G4 probing system.

To validate the former hypothesis, we performed all-atom MD simulations, binding the three species QSP, QMC, and QMCH to the parallel-stranded DNA G4 of the promoter c-MYC. To this end, the three species QSP, QMC, and QMCH were positioned on the surface of the external G-tetrads at the 3'-end of the DNA by superimposition of each of the compounds with one of the quindoline molecules present in the solution structure of a 2:1 quindoline-c-MYC G4 (template structure: PDB id: 2L7V).^[12] Then, three independent 300 ns MD simulations were carried out for each system for a total of $0.9\ \mu\text{s}$ each, to efficiently explore the conformational space without restraints (Section S1 in the Supporting Information). As found for other spiropyran forms when binding to DNA,^[15] the QSP species does not show a stable binding mode to G4; indeed, after a few ns, QSP moves into the bulk solvent (data not shown). In contrast, QMC and QMCH bind to G4 owing to their planar extended π system, stacking with the guanine rings of upper G-tetrads at the 3'-terminus. Figure 3a shows the structural superposition of the final geometries obtained from the three independent MD trajectories for QMC:G4 and QMCH:G4 complexes. Although for QMCH all three trajectories ended up

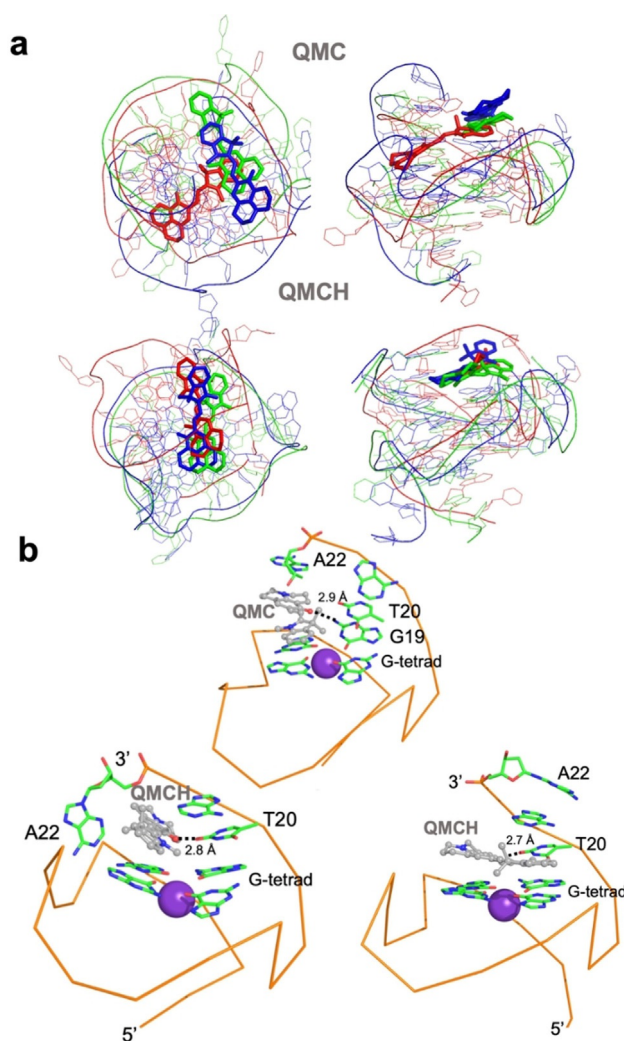


Figure 3. (a) Superimposition of the final geometries of the three independent fully relaxed MD simulations for both QMC and QMCH. Top and lateral views are provided. The geometries of each of the three independent MD simulation are colored in red, blue, and green, respectively, and the probe is shown as sticks. (b) Detail of the binding mode of QMC and QMCH to G4 obtained from one of the MD simulations. Only the surrounding nucleobases and the upper G-tetrad + K^+ ion are shown for simplicity. The backbone is shown as a ribbon. Two solutions are shown for QMCH on the bottom as an example of the mobility of A22 and show the flipping of the 3'-terminal A22 when QMCH binds, peculiar of the QMCH:G4 complex.

in a similar binding mode, in QMC we found more than a single solution (Figure S3 in the Supporting Information). In the most stable binding mode of QMC (Table 1, QMC simulation 2), the probe interacts through its negatively charged oxygen with the exocyclic amine group of G19 (dashed line Figure 3b and Figure S4 in the Supporting Information). However, in the other two QMC simulations, this interaction is missing and the binding with the macromolecule is mainly through π -stacking interactions with the surrounding guanines (simulations 1 and 3, figure not shown). In contrast, the hydroxyl group of QMCH is hydrogen bonded with the carbonyl oxygen on C-2 of T20 (Figure 3b). The 3'-terminus, where A22 is located, shows high mobility along the simulation time, but the most stable complex (Table 1, simulation 1) is formed when

Table 1. Binding energy (kcal mol⁻¹) of QMCH and QMC to c-MYC G4 and its components along the MD simulations computed with MM-ISMSA.^[18]

Simulation	vdW	Electrostatics	Energy components ^[a, b]			Total
			L desolv.	R desolv.	Nonpolar	
QMCH						
1	−55.19 ± 3.97	−8.95 ± 0.59	−0.29 ± 0.11	11.49 ± 0.93	−2.54 ± 0.11	−55.48 ± 3.85
2	−47.77 ± 4.30	−8.76 ± 0.53	−0.08 ± 0.44	10.94 ± 0.90	−2.47 ± 0.17	−48.14 ± 4.50
3	−50.05 ± 1.73	−8.63 ± 0.49	−0.22 ± 0.07	10.93 ± 0.66	−2.41 ± 0.05	−50.39 ± 2.03
QMC						
1	−48.68 ± 6.07	−0.17 ± 0.29	3.13 ± 0.40	11.06 ± 1.29	−2.57 ± 0.32	−37.24 ± 5.45
2	−52.86 ± 3.11	−0.73 ± 0.15	2.34 ± 0.22	10.06 ± 1.01	−2.64 ± 0.13	−43.28 ± 2.64
3	−42.02 ± 2.79	−9.18 ± 0.58	−0.63 ± 0.16	13.20 ± 1.01	−2.31 ± 0.09	−40.94 ± 2.09
[a] vdW=van der Waals, L desolv.=ligand desolvation, R desolv.=receptor desolvation. [b] A window of 20 ns was analyzed for each simulation.						

[a] vdW = van der Waals, L desolv. = ligand desolvation, R desolv. = receptor desolvation. [b] A window of 20 ns was analyzed for each simulation.

the 3'-terminus folds around the probe as shown in Figure 3b. This kind of interaction has been already reported for similar G4 probes.^[16,17]

The indoline ring shows higher mobility than the chromene-quindoline one along the MD simulations. This property can be measured by monitoring the evolution of the dihedral angle around the α -bond (Figure 4) along the simulation time, as this bond connects the indoline ring to the exocyclic double bond. QMC simulations show a dihedral angle between -50 and 50 degrees around the α -bond, which corresponds

to a *cis* conformation whereas for QMCH mostly *trans* α -bond conformers are found. In contrast, the dihedral angle controlling the γ -bond is fixed as *cis* in QMCH but varies to *cis* and *trans* in QMC. This means that the major conformer of QMCH is TTC and both CTC and TTC isomers dominate QMC. The hydrogen bond between the quinolizidine hydroxyl group of QMCH (absent in QMC) and G4 (see Figure 3c) explains the 'rigidity' of the quinolizidine ring with a *cis* conformation of the γ -bond, and thereby, the preference for the TTC isomer.

Finally, we analyzed the binding energies of QMC and QMCH along the three different MD simulations by using the MM-ISMSA method^[18] (Table 1 and Section S5 in the Supporting Information).

In all the simulations, QMCH binds more strongly to G4 than QMC ($\Delta E_{\text{QMC/QMCH}} \approx 10$ kcal mol⁻¹). Looking at the energy decomposition, the van der Waals component (vdW) is the largest term in both complexes. Both species present comparable values for this term, but the electrostatics make a difference, with an extra stabilization of approximately 8 kcal mol⁻¹ in QMCH. In contrast to QMC, QMCH is positively charged, and as a general trend in other merocyanine species, this favors its interaction with DNA.^[15] These binding energies agree with the data obtained from the 2D free energy surface shown in Figure 2, that is, QMCH is the absolute minimum, supporting the initial hypothesis of QMCH species being the strongest binder to G4.

Conclusion

In summary, we unraveled the molecular mechanism governing the binding of the fluorescent quinolizidine-substituted spiropyran QSP to c-MYC G4. We described the QSP ring opening to the QMC/QMCH open forms and characterized the dynamical binding of these merocyanine isomers to the macro-molecule. The calculations show that the opening of QSP is not a proton-assisted process, and the QMC species is an intermediate, both in solution and in G4. In the latter case, G4 would trap the QMCH form, shifting the chemical equilibrium $\text{QSP} \rightleftharpoons \text{QMC/QMCH}$ towards the protonated form. A similar effect can be expected at low pH value solutions. Extensive MD simulations on the μ s-scale with QMC and QMCH bound to c-MYC G4 support that QMCH is the strongest G4 binder, showing a major rotamer around the *E*-double β -bond, which

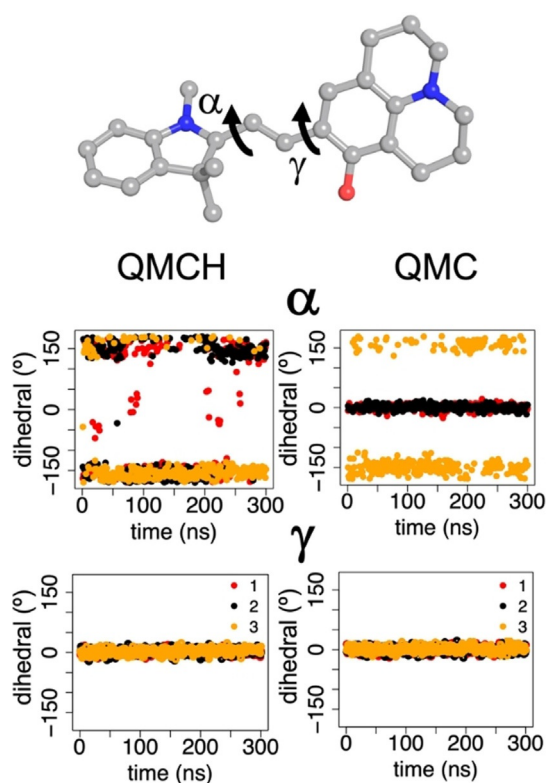


Figure 4. Time evolution of dihedral α and γ (degree) along the three independent 300 ns MD simulations (red, black, and yellow points, respectively) of QMCH and QMC bound to G4. As an illustration, the two dihedral angles are shown on the structure of the TTC isomer (-180° , 180°) of QMC. For each of the dihedral angles, values of approximately 0° correspond to a *cis* conformation and values close to $\pm 180^\circ$ to a *trans* rotamer. In all cases, only the *trans* isomer of the β -double bond is considered.

share a common binding site. QMC, in contrast, presents higher degrees of rotations and not a common binding mode. These results shed light on the structural features that govern the binding of QSP to G4s and open up a great opportunity to use the found structural models to design novel SP derivatives with enhanced G4 binding activity.

Acknowledgments

D.A. thanks the funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 765266 (LightDyNAMics). P.A.S.-M. thanks the Austrian FWF, Project M 2260. The authors thank the Vienna Scientific Cluster (VSC) for allocation of generous computational resources and Dr. Boris Maryasin for helpful discussions on the calculation of organic molecules reactivity.

Conflict of interest

The authors declare no conflict of interest.

Keywords: guanine-quadruplexes • QM/MM • spiropyran • umbrella sampling

- [1] J. T. Davis, *Angew. Chem. Int. Ed.* **2004**, *43*, 668–698; *Angew. Chem.* **2004**, *116*, 684–716.
- [2] R. Hänsel-Hertsch, M. Di Antonio, S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 279–284.

- [3] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 5402–5415.
- [4] P. Armas, A. David, N. B. Calcaterra, *Transcription* **2017**, *8*, 21–25.
- [5] D. Monchaud, M. P. Teulade-Fichou, *Org. Biomol. Chem.* **2008**, *6*, 627–636.
- [6] M. P. O'Hagan, J. C. Morales, M. C. Galan, *Eur. J. Org. Chem.* **2019**, 4995–5017.
- [7] J. Li, X. Yin, B. Li, X. Li, Y. Pan, J. Li, Y. Guo, *Anal. Chem.* **2019**, *91*, 5354–5361.
- [8] J. T. C. Wojtyk, A. Wasey, N.-N. Xiao, P. M. Kazmaier, S. Hoz, C. Yu, R. P. Lemieux, E. Buncel, *J. Phys. Chem. A* **2007**, *111*, 2511–2516.
- [9] N. P. Ernstring, B. Dick, T. Arthen-Engel, *Pure Appl. Chem.* **1990**, *62*, 1483–1488.
- [10] L. Kortekaas, L. Chen, D. Jacquemin, W. R. Browne, *J. Phys. Chem. B* **2018**, *122*, 6423–6430.
- [11] M. Hammarson, J. R. Nilsson, S. Li, T. Beke-Somfai, J. Andréasson, *J. Phys. Chem. B* **2013**, *117*, 13561–13571.
- [12] J. Dai, M. Carver, L. H. Hurley, D. Yang, *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680.
- [13] M. Gaus, Q. Cui, M. Elstner, *J. Chem. Theory Comput.* **2011**, *7*, 931–948.
- [14] A. Klamt, *J. Phys. Chem.* **1995**, *99*, 2224–2235.
- [15] D. Avagliano, P. A. Sánchez-Murcia, L. González, *Phys. Chem. Chem. Phys.* **2019**, *21*, 8614–8618.
- [16] A. Terenzi, H. Gattuso, A. Spinello, B. K. Keppler, C. Chipot, F. Dehez, G. Barone, A. Monari, *Antioxidants* **2019**, *8*, 472.
- [17] Q. Zhai, C. Gao, J. Ding, Y. Zhang, B. Islam, W. Lan, H. Hou, H. Deng, J. Li, Z. Hu, H. I. Mohamed, S. Xu, C. Cao, S. M. Haider, D. Wei, *Nucleic Acids Res.* **2019**, *47*, 2190–2204.
- [18] J. Klett, A. Núñez-Salgado, H. G. Dos Santos, Á. Cortés-Cabrera, A. Perona, R. Gil-Redondo, D. Abia, F. Gago, A. Morreale, *J. Chem. Theory Comput.* **2012**, *8*, 3395–3408.

Manuscript received: April 1, 2020

Revised manuscript received: May 4, 2020

Accepted manuscript online: May 5, 2020

Version of record online: September 17, 2020

Chemistry–A European Journal

Supporting Information

Spiropyran Meets Guanine Quadruplexes: Isomerization Mechanism and DNA Binding Modes of Quinolizidine-Substituted Spiropyran Probes

Davide Avagliano,^[a] Pedro A. Sánchez-Murcia,^{*,[a]} and Leticia González^{*,[a, b]}

Table of contents

Section S1	System setup and classical molecular dynamics simulations
Section S2	Two-dimensional umbrella sampling (2D-US) QM/MM MD simulation
Section S3	Umbrella Sampling partition scheme
Section S4	Geometry optimization by means of Møller-Plesset second order perturbation theory calculations
Section S5	Binding energy calculation
Section S6	Cartesian coordinates of RI-MP2/def2-TZVP optimized geometries
References	

Section S1: System setup and classical molecular dynamics simulations

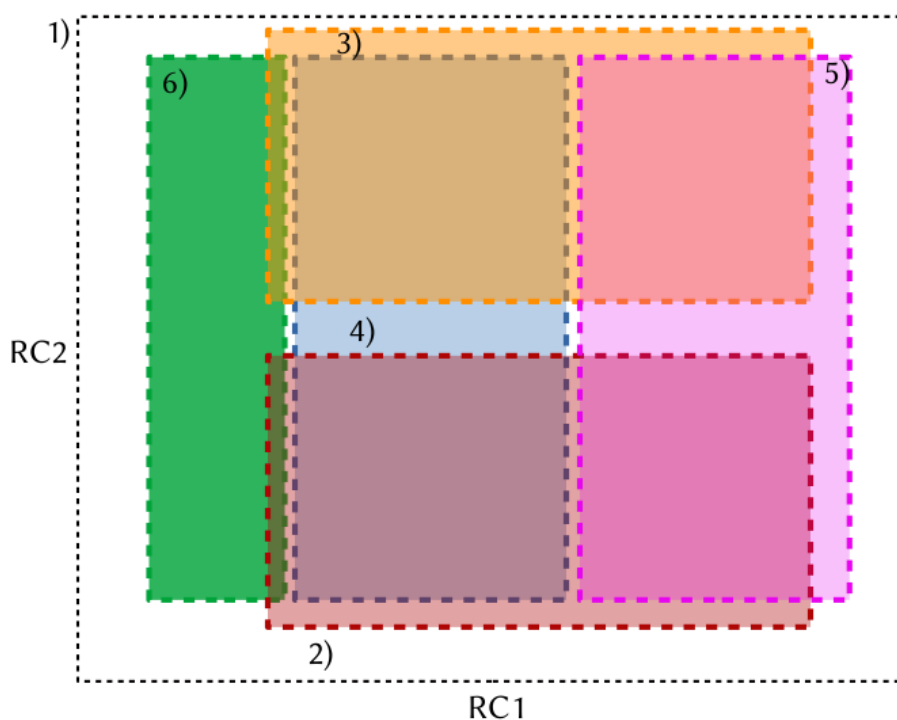
The initial cartesian coordinates of the complexes between the QSP, QMC and QMCH forms and the DNA guanine quadruplex (G4) of the promotor c-MYC were obtained by manual superimposition of the julolidine moiety of QSP with the quindoline ligand at the 3'-flank in the NMR-solved structure quindoline:DNA G4 deposited in the Protein Data Bank (PDB id. 2L7V). The force field parameters for the molecular dynamics (MD) for the three forms of the probe were generated following the general procedure described elsewhere^[1] (geometry optimization and atom-centered ESP charge computed at the HF/6-31G* level) and their atom types were described as GAFF atom types. The Amber force field ff14SB^[2] was used for the description of the system. The QSP:DNA G4 complex was solvated with a box of ~12500 TIP3P water molecules with a distance from the solute to the border of 20 Å. 18 Na⁺ (QSP, QMC) or 17 Na⁺ (QMCH) atoms were added to ensure the electroneutrality of the system. Before running the QM/MM (quantum mechanical/molecular mechanics) MD studies, the system was initially minimized in three steps where the hydrogen atoms, the solvent molecules and all the system were sequentially minimized. After that, the system was heated from 100 to 300 K in 20 ps using a time step of 0.2 fs and the weak-coupling thermostat. For this step the position of the all the heavy atoms on the ligand and DNA G4 had to be restrained (harmonic constant = 40 kcal⁻¹ mol⁻¹ Å⁻²). Finally, the restrains were gradually removed in 6 steps (6 x 20 ps) and the system switched from an NVT ensemble (steps 1-4) to an NPT ensemble (steps 5 and 6). In all the simulations, periodic boundary conditions were used, and the electrostatic interaction were treated with the Ewald method (grid space of 1 Å and cutoff distance of 10 Å). The ligand:DNA G4 complexes were further simulated for 3 x 300 ns. A random seed was placed in the heating step for the generation of the three independent simulations. The SHAKE algorithm was used to treat the bonds where hydrogen atoms were involved.

Section S2: Two-dimensional umbrella sampling (2D-US) QM/MM MD simulation

The equilibrated snapshots after heating of QSP in solution or QSP:G4 QSP obtained in Section S1 were used as initial geometries for the umbrella sampling (US) studies after a short MD minimization (see Section S1). QSP and the closest water molecule to the spiropyran oxygen were treated at the density functional tight binding method version 3 (DFTB3)^[3] whereas the rest of the system (c-MYC G4, solvent molecules and counter ions) were treated with force fields. A cut off of 10 Å was defined for the electrostatic interactions between the QM and MM. Two reaction coordinates (RC) were defined for the QSP ring opening: the distance between C and O atoms of the spiro-junction of the probe (O-C bond cleavage, RC₁) and the linear combination of distances of the distances O_{water}-H_{water} and H_{water}-O_{spiro} (proton transfer, RC₂). The reaction pathway was explored by means of 2ps-steered QM/MM MD dynamics imposing a harmonic restraint (50 kcal⁻¹ mol⁻¹ Å⁻²) for both RC₁ and RC₂. Six different reaction pathways with RC-values (1.2 < RC₁ < 6) Å and (-1.2 < RC₂ < 1.2) were defined to efficiently explore the energy surface. 50 points per trajectory along the chemical pathway were printed. Then, these snapshots were used as initial geometries for the 2D-US. Each of the windows in the 2D-US QM/MD MD study was allowed to oscillate for 5 ps around the reaction coordinate value (harmonic constant = 100 kcal⁻¹ mol⁻¹ Å⁻² and a time step of 0.1 fs). The 2D free energy landscape was constructed using the maximum likelihood principle implemented in the variational Free Energy Profile (vFEP) method.^[4]

Section S3: Umbrella Sampling partition scheme

We partitioned the (RC_1 , RC_2) space in up to six sub-regions, as shown in Figure S1. Region 1 represents the exploration of all the range of both two coordinates from the respective starting ($RC_1=1.2$ Å, $RC_2=-1.2$ Å) and final ($RC_1=6$ Å, $RC_2=1.2$) values of interest. Region 2 is the opening ($RC_1=1.2:6.0$ Å) with proton assistance ($RC_2=-1.2:0$ Å). Region 3 is the opening ($RC_1=1.2:6.0$ Å) mediated by the proton transfer ($RC_2=0:1.2$ Å). Regions 4 and 5 represent the effect of the protonation ($RC_2=-1.2:1.2$ Å) in the first and the second part of the ring opening process, respectively. Finally, region 6 was explored to sample the possible protonation of the close formed before the ring opening occurs.



Restraints: 1) $RC_1 = 1.2 : 6.0$; $RC_2 = -1.2 : 1.2$
2) $RC_1 = 1.2 : 6.0$; $RC_2 = -1.2 : 0.0$
3) $RC_1 = 1.2 : 6.0$; $RC_2 = 0.0 : 1.2$
4) $RC_1 = 1.0 : 3.0$; $RC_2 = -1.2 : 1.2$
5) $RC_1 = 3.0 : 6.0$; $RC_2 = -1.2 : 1.2$
6) $RC_1 = 1.0 : 2.0$; $RC_2 = -1.2 : 1.2$

Figure S1: Schematic representation of reaction coordinates partitioning for the 2D-QM/MM-US calculation.

In total, we got 600 umbrella sampling windows (~600000 data points) that were used to calculate the free energy profile. The use of a large number of data is mandatory to get a converged surface. As example, Figure S2 shows how the FEP surface changes when the number of data increases.

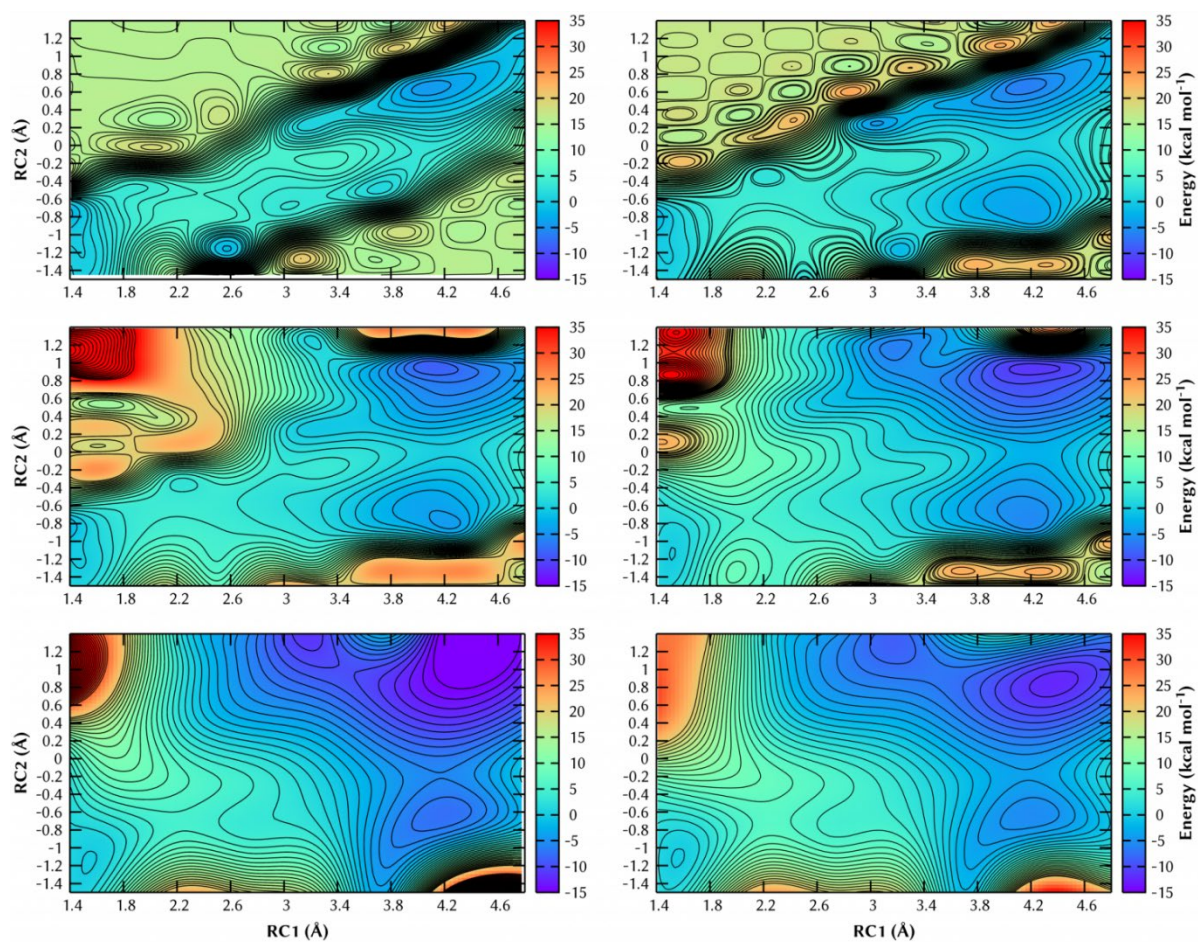


Figure S2. Free energy profiles (kcal mol^{-1}) obtained with increasing number of input data (from left to right, from top to bottom: 50, 100, 150, 200, 250 and 300 windows). The 0 of the energy is referred to the QSP minimum and each contour line represents a difference of 1 kcal mol^{-1} .

Section S4: Geometry optimization by means of Møller-Plesset second order perturbation theory calculations

Ground state geometry optimizations of the open forms of QSP in their different forms (unprotonated (QMC) and protonated (QMCH) merocyanine, and the protonated spiropyran (QSPH)) were performed with Møller-Plesset second order perturbation theory using the resolution of identity approximation (RI-MP2) and the def2-TZVP basis set. The zero-point energy correction to the Gibbs free energies at 298 K were included by means of frequency calculations. The four possible QMC isomers exocyclic double bond and the possible relative orientations of the indoline N and O atoms were taken into account in the calculations. The Z-isomers of the double bond were not considered because are known to be higher in energy. All the quantum chemical calculations were run using the TURBOMOLE suite, version 7.2.^[5] The optimized geometries are reported below in section S6 in xyz format (compound name//number of atom//electronic energy//atomic coordinates)

Section S5: Binding energy calculation

The binding free energies were calculated using the program MM-ISMSA.^[6] For each of the simulations, the last 50 ns were considered. Before running the energy analysis, the snapshots every 10 ns were cold down in 1 ns from 300 to 100 K by means of unrestrained MD simulations. MM-ISMSA estimates solvent-corrected binding energies as well as their per-residue decomposition into van der Waals, Coulombic, apolar, and desolvation ligand/receptor contributions, using a sigmoidal, distance-dependent function for the dielectric.

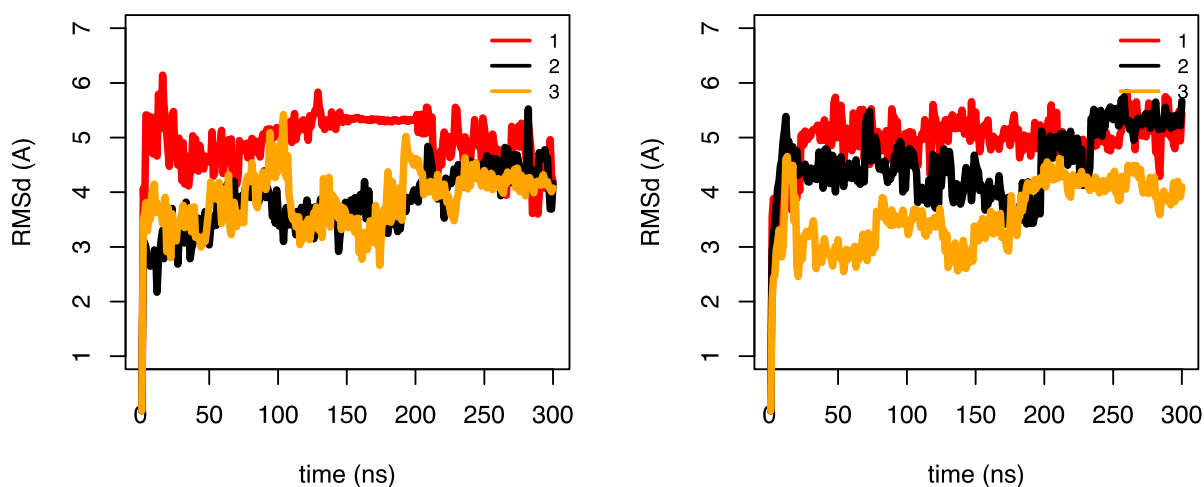


Figure S3. Root-mean squared deviation (RMSd, Å) of the complexes QMCH:G4 (left) and MC:G4 (right) along the 3 independent 300-ns MD simulations.

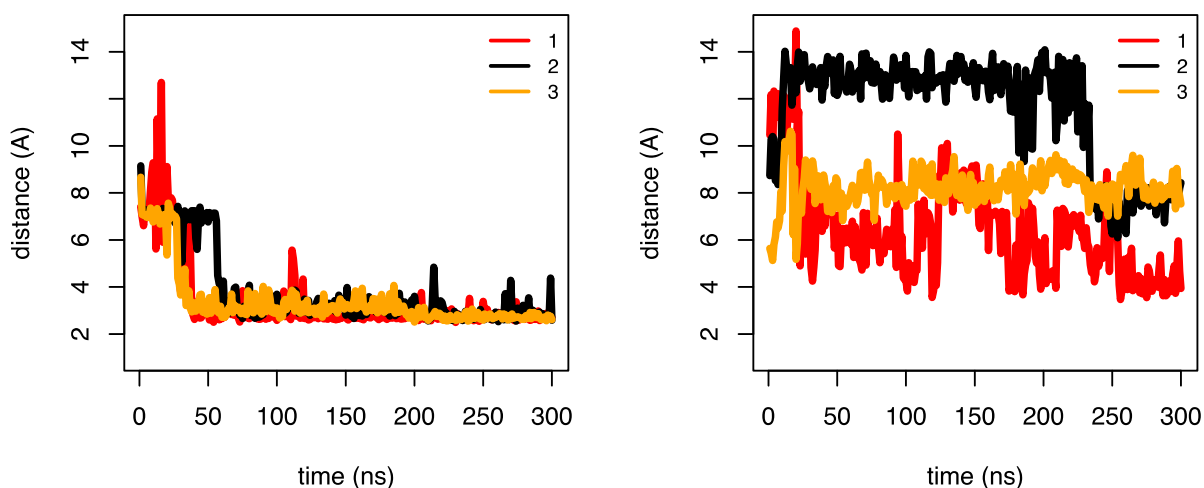


Figure S4. Distance (Å) between the spiropyran oxygen of QMCH (left) and QMC (right) and the O2 of T21 and the NH₂ group of G20, respectively, along the 3 independent 300-ns MD simulations. Whereas in QMCH the three simulations ended up in the same solution, in the case of QMC this interaction is only present in the first simulation (red line).

Section S6: Cartesian coordinates of RI-MP2/def2-TZVP optimized geometries

QSP

56

ELECTRONIC ENERGY = -1151.79461 a.u.

C	0.104629	0.890035	-0.234721
C	0.543649	1.303201	-1.496828
C	1.316771	2.464803	-1.566587
C	1.648369	3.199048	-0.435042
C	0.169483	0.497655	-2.632372
C	-0.408356	-0.707299	-2.454231
C	-0.643976	-1.247392	-1.084924
O	-0.724721	-0.185886	-0.102248
N	0.386433	-2.200901	-0.691987
C	-0.155313	-2.960512	0.349494
C	-1.553281	-2.903565	0.306134
C	-1.954668	-2.064594	-0.879901
C	0.505853	-3.693876	1.329472
C	-0.272810	-4.393882	2.258067
C	-1.665600	-4.348173	2.215210
C	-2.317246	-3.598051	1.228067
H	-3.401679	-3.552179	1.198217
H	-2.244757	-4.892097	2.952495
H	0.219913	-4.971218	3.032751
H	1.588107	-3.727150	1.382006
H	-0.650303	-1.343108	-3.296189
H	0.377928	0.866749	-3.632937
C	-3.170897	-1.181888	-0.653102
H	-3.070525	-0.588907	0.254637
H	-3.311342	-0.502677	-1.498790
H	-4.065343	-1.804973	-0.570032
C	-2.201210	-3.005370	-2.064864
H	-2.990882	-3.710745	-1.796613
H	-2.527803	-2.453074	-2.949808
H	-1.301524	-3.575014	-2.308295
C	0.418383	1.584855	0.930871
C	1.206545	2.751758	0.834021
C	-0.090332	1.059693	2.246169
C	0.642868	1.700847	3.414148
C	0.769422	3.192924	3.174689
N	1.579958	3.431638	1.994521
C	2.452103	4.467802	-0.576422
C	3.085264	4.870933	0.745672
C	2.034537	4.800053	1.836982
H	0.007694	-0.030072	2.262519
H	-1.165504	1.261174	2.322841
H	0.113948	1.511020	4.351270
H	1.648591	1.281518	3.506769
H	1.263762	3.683696	4.017032
H	-0.232676	3.643818	3.067160
H	1.195519	5.475438	1.593510
H	2.446760	5.121185	2.797127
H	3.899280	4.187249	1.002317
H	3.498995	5.880628	0.686508
H	1.795774	5.275852	-0.920595
H	3.214034	4.334160	-1.349386
C	1.755285	-1.739389	-0.573117
H	2.424403	-2.601107	-0.538274
H	1.912978	-1.131037	0.327896
H	2.008386	-1.144845	-1.451382
H	1.664125	2.808248	-2.538927

QSPH

57

ELECTRONIC ENERGY = -1152.17919 a.u.

C	-0.076271	1.166795	0.318707
C	0.318961	1.335251	-1.027784
C	1.184016	2.422972	-1.290573
C	1.674458	3.260185	-0.315789
C	-0.155197	0.597000	-2.156748
C	-0.745703	-0.636736	-2.280418
C	-0.836993	-1.695598	-1.345657
O	-0.991100	0.185401	0.596971
N	0.097482	-2.049786	-0.480270
C	-0.346563	-3.113304	0.341421
C	-1.618016	-3.491842	-0.082741
C	-2.016591	-2.631135	-1.245568
C	0.323815	-3.721176	1.393840
C	-0.346570	-4.763364	2.034749
C	-1.621076	-5.165649	1.622136
C	-2.269311	-4.537138	0.555943
H	-3.257016	-4.859894	0.244464
H	-2.112918	-5.979586	2.141542
H	0.130124	-5.269611	2.865522
H	1.309070	-3.408379	1.717990
H	-1.238627	-0.835713	-3.228592
H	-0.080926	1.150033	-3.092567
C	-3.309632	-1.851061	-0.991670
H	-3.231321	-1.248718	-0.086305
H	-3.525770	-1.188178	-1.832192
H	-4.140708	-2.550447	-0.879830
C	-2.129281	-3.469796	-2.530485
H	-2.935758	-4.196040	-2.411758
H	-2.366259	-2.837175	-3.388198
H	-1.203063	-4.012021	-2.730829
C	0.404082	1.981310	1.341205
C	1.290665	3.045093	1.038676
C	-0.037238	1.721025	2.759861
C	0.908233	2.374640	3.757250
C	1.160693	3.810562	3.348254
N	1.768561	3.851966	2.029169
C	2.574175	4.412220	-0.673388
C	3.443189	4.791051	0.513940
C	2.552841	5.039202	1.714132
H	-0.067751	0.642094	2.953265
H	-1.053555	2.105981	2.918210
H	0.486775	2.338996	4.763502
H	1.861538	1.840056	3.773549
H	1.848726	4.294694	4.043695
H	0.221898	4.383829	3.358057
H	1.889290	5.895161	1.524074
H	3.149214	5.276433	2.597967
H	4.143580	3.983484	0.744690
H	4.026513	5.688858	0.301390
H	1.962359	5.274147	-0.962773
H	3.181115	4.147867	-1.542086
C	1.436660	-1.503453	-0.389649
H	2.120565	-2.317221	-0.152646
H	1.482132	-0.734372	0.385227
H	1.698470	-1.058364	-1.348039
H	1.474419	2.603600	-2.322681
H	-1.357405	0.352850	1.478839

QMC_CTC

56

ELECTRONIC ENERGY = -1151.75953 a.u.

C	0.944291	-0.650938	-5.257444
C	1.464379	-0.136006	-3.973223
C	2.421472	0.926098	-4.002734
C	2.857117	1.505538	-5.154501
C	1.072707	-0.623555	-2.737786
C	0.093372	-1.611531	-2.502337
C	-0.304569	-2.154409	-1.304921
O	0.152668	-1.617080	-5.328777
N	0.277243	-2.061079	-0.067770
C	-0.495984	-2.713404	0.897886
C	-1.600157	-3.308847	0.283128
C	-1.548725	-3.024655	-1.192191
C	-0.276528	-2.807465	2.268330
C	-1.203278	-3.536756	3.018347
C	-2.305619	-4.143841	2.415840
C	-2.510575	-4.031783	1.036272
H	-3.368952	-4.502935	0.567570
H	-3.009427	-4.700714	3.023279
H	-1.062114	-3.625742	4.089541
H	0.564186	-2.325062	2.752689
H	-0.440715	-1.955570	-3.380305
H	1.533386	-0.129333	-1.885348
C	-2.791348	-2.268718	-1.666167
H	-2.920842	-1.343724	-1.100671
H	-2.701660	-2.018447	-2.725752
H	-3.681759	-2.887901	-1.528420
C	-1.365954	-4.317577	-1.996047
H	-2.238395	-4.963446	-1.865151
H	-1.253257	-4.091817	-3.059140
H	-0.479793	-4.859578	-1.659291
C	1.380721	0.025569	-6.455794
C	2.299890	1.061699	-6.419705
C	0.773066	-0.453829	-7.742168
C	0.946617	0.576956	-8.845169
C	2.376202	1.083138	-8.845686
N	2.680210	1.716983	-7.575578
C	3.936130	2.553528	-5.125311
C	3.839094	3.466773	-6.334680
C	3.827371	2.604808	-7.580155
H	-0.281742	-0.677416	-7.564480
H	1.228134	-1.407916	-8.033408
H	0.702394	0.152694	-9.822286
H	0.279632	1.426868	-8.674432
H	2.530653	1.827920	-9.630741
H	3.070483	0.248595	-9.036394
H	4.771261	2.037322	-7.648312
H	3.755671	3.221850	-8.479864
H	2.917195	4.054258	-6.297395
H	4.683041	4.160270	-6.371786
H	4.918377	2.063525	-5.140320
H	3.876771	3.114579	-4.189050
C	1.594927	-1.540491	0.213958
H	1.996615	-2.059408	1.084520
H	1.579306	-0.466595	0.418190
H	2.243047	-1.733165	-0.641733
H	2.829799	1.271647	-3.053854

QMC_CTT

56

ELECTRONIC ENERGY = -1151.75881 a.u.

C	1.702926	-2.490082	-0.434958
C	1.029066	-1.182978	-0.226862
C	1.827132	-0.013960	-0.042091
C	3.187458	-0.052641	-0.056046
C	-0.346894	-1.195466	-0.220930
C	-1.204801	-0.091246	-0.006800
C	-2.576313	-0.067498	-0.036754
O	1.058134	-3.545160	-0.577652
N	-3.447779	-1.046309	-0.443879
C	-4.771636	-0.654120	-0.228300
C	-4.796514	0.646729	0.279994
C	-3.383716	1.144282	0.415215
C	-5.940891	-1.373837	-0.454159
C	-7.152813	-0.740293	-0.166044
C	-7.190484	0.562203	0.332110
C	-6.003075	1.266533	0.559512
H	-6.030681	2.279104	0.950230
H	-8.145151	1.026745	0.549230
H	-8.079983	-1.277992	-0.329061
H	-5.929203	-2.392831	-0.821764
H	-0.753408	0.859538	0.258860
H	-0.763642	-2.187968	-0.352031
C	-3.061536	1.515167	1.863580
H	-3.232358	0.665158	2.526858
H	-2.017587	1.823095	1.959609
H	-3.696507	2.343434	2.189150
C	-3.136347	2.335374	-0.516737
H	-3.773286	3.175533	-0.227150
H	-2.093331	2.656929	-0.460791
H	-3.362619	2.069548	-1.551279
C	3.147242	-2.483059	-0.485998
C	3.869275	-1.315432	-0.313611
C	3.811196	-3.798527	-0.770324
C	5.236871	-3.594310	-1.254806
C	5.935705	-2.586927	-0.361684
N	5.246728	-1.309730	-0.413875
C	3.986599	1.184884	0.247712
C	5.353835	1.126554	-0.409703
C	6.025710	-0.162832	0.013191
H	3.209691	-4.336683	-1.506956
H	3.797510	-4.422618	0.130994
H	5.791555	-4.535865	-1.256886
H	5.234716	-3.206220	-2.277385
H	6.965392	-2.418868	-0.687464
H	5.970460	-2.962918	0.673555
H	6.171166	-0.164368	1.106882
H	7.014467	-0.254882	-0.444153
H	5.255830	1.140442	-1.499032
H	5.969098	1.979533	-0.112564
H	4.129187	1.264325	1.333439
H	3.423574	2.068672	-0.063224
C	-3.101148	-2.264052	-1.143837
H	-3.961127	-2.580498	-1.733128
H	-2.821831	-3.068375	-0.458964
H	-2.266574	-2.069402	-1.817845
H	1.343711	0.942168	0.142604

QMC_TTC

56

ELECTRONIC ENERGY = -1151.76334 a.u.

C	1.873780	-1.556998	-0.617503
C	1.261269	-0.282917	-0.185501
C	2.118649	0.830471	0.079525
C	3.471552	0.779906	-0.058468
C	-0.101485	-0.105872	-0.016283
C	-1.101620	-1.077401	-0.215045
C	-2.449883	-0.870914	-0.040525
O	1.205112	-2.592601	-0.835296
N	-3.357409	-1.872918	-0.260131
C	-4.661862	-1.446519	-0.016687
C	-4.632423	-0.105878	0.377029
C	-3.203969	0.378790	0.393593
C	-5.856595	-2.152706	-0.117939
C	-7.036494	-1.470428	0.188379
C	-7.019722	-0.131766	0.581630
C	-5.808335	0.561151	0.678449
H	-5.794415	1.602588	0.984840
H	-7.950967	0.371686	0.813178
H	-7.982116	-1.995927	0.118097
H	-5.889889	-3.192411	-0.420341
H	-0.751122	-2.052908	-0.526863
H	-0.400353	0.891397	0.298508
C	-2.818220	0.811056	1.812991
H	-2.956664	-0.014923	2.513640
H	-1.777371	1.132469	1.864229
H	-3.454249	1.641756	2.130298
C	-3.031781	1.523213	-0.612084
H	-3.669913	2.363601	-0.326034
H	-1.998511	1.869982	-0.650281
H	-3.321097	1.197229	-1.613270
C	3.305925	-1.562888	-0.798691
C	4.083849	-0.446288	-0.536804
C	3.900317	-2.843667	-1.309185
C	5.289489	-2.612460	-1.880117
C	6.093276	-1.751232	-0.925114
N	5.448331	-0.461951	-0.755743
C	4.326324	1.955823	0.329032
C	5.639439	1.947304	-0.433583
C	6.294037	0.595752	-0.235752
H	3.224200	-3.264391	-2.057490
H	3.936354	-3.582939	-0.500147
H	5.807024	-3.559607	-2.051918
H	5.220203	-2.091424	-2.839328
H	7.099559	-1.568087	-1.310747
H	6.196112	-2.262013	0.046187
H	6.512941	0.442000	0.834765
H	7.245780	0.542092	-0.771187
H	5.462807	2.106486	-1.501251
H	6.305554	2.737323	-0.077592
H	4.549729	1.904156	1.402579
H	3.770080	2.882573	0.166071
C	-2.974791	-3.199478	-0.684332
H	-3.867566	-3.805994	-0.813198
H	-2.325567	-3.665957	0.060927
H	-2.431789	-3.152446	-1.631448
H	1.660486	1.754099	0.430932

QMC_TTT

56

ELECTRONIC ENERGY = -1151.76134 a.u

C	1.975316	-2.027077	-5.432329
C	2.161133	-1.117652	-4.273006
C	2.795915	0.144740	-4.477595
C	3.213629	0.557556	-5.705139
C	1.704701	-1.570369	-3.057718
C	1.749123	-0.891432	-1.818778
C	1.276419	-1.414754	-0.641851
O	1.475660	-3.159514	-5.304492
N	1.335818	-0.730367	0.547571
C	0.789042	-1.473230	1.593335
C	0.349626	-2.701029	1.092467
C	0.625462	-2.769927	-0.389467
C	0.655996	-1.141063	2.937911
C	0.063797	-2.085862	3.779798
C	-0.378322	-3.315766	3.291866
C	-0.236262	-3.630202	1.936219
H	-0.580265	-4.586810	1.555370
H	-0.834185	-4.029492	3.967925
H	-0.052599	-1.853843	4.832387
H	0.991643	-0.191070	3.336235
H	2.175700	0.105524	-1.789075
H	1.269613	-2.562673	-3.104170
C	-0.695165	-2.924122	-1.154649
H	-1.363885	-2.089578	-0.933144
H	-0.534603	-2.963862	-2.232494
H	-1.188105	-3.849812	-0.846286
C	1.583799	-3.930649	-0.685120
H	1.123835	-4.871338	-0.371185
H	1.816182	-4.000376	-1.748323
H	2.517359	-3.804549	-0.132575
C	2.381982	-1.529856	-6.727208
C	2.969581	-0.285505	-6.869772
C	2.105586	-2.422131	-7.901615
C	2.131395	-1.632833	-9.199741
C	3.355255	-0.737162	-9.223290
N	3.310355	0.202077	-8.116512
C	3.973655	1.846390	-5.857211
C	3.782218	2.431096	-7.245211
C	4.157550	1.371052	-8.259170
H	1.139655	-2.908691	-7.745752
H	2.841230	-3.234507	-7.931295
H	2.142483	-2.298471	-10.066343
H	1.240433	-1.002828	-9.274444
H	3.399645	-0.154746	-10.147169
H	4.270047	-1.349335	-9.170753
H	5.222814	1.107756	-8.143687
H	4.025866	1.743528	-9.278599
H	2.738736	2.721518	-7.397618
H	4.406288	3.316912	-7.387793
H	5.044345	1.653619	-5.708914
H	3.666512	2.546098	-5.075619
C	1.897581	0.590346	0.671381
H	1.849756	0.906317	1.710407
H	1.340797	1.305039	0.058599
H	2.942979	0.592859	0.350179
H	2.988877	0.792054	-3.625540

QMCH_CTC

57

ELECTRONIC ENERGY = -1152.18104 a.u.

C	1.863283	0.903793	-0.255247
C	1.189884	-0.328983	-0.061439
C	2.005156	-1.466537	0.152933
C	3.376678	-1.424780	0.166777
C	-0.212095	-0.522322	-0.069566
C	-1.219668	0.416387	-0.154068
C	-2.599890	0.150718	-0.114993
O	1.091884	2.001724	-0.494088
N	-3.237525	-0.979746	-0.422069
C	-4.630393	-0.856961	-0.194079
C	-4.896305	0.441363	0.232361
C	-3.606126	1.203950	0.297317
C	-5.614870	-1.822763	-0.352991
C	-6.923385	-1.430168	-0.068134
C	-7.213700	-0.129212	0.352377
C	-6.201483	0.821402	0.506757
H	-6.436840	1.827812	0.836339
H	-8.240631	0.142434	0.565894
H	-7.726535	-2.149610	-0.173053
H	-5.396592	-2.836856	-0.664708
H	-0.955342	1.464111	-0.157043
H	-0.500936	-1.559179	0.081918
C	-3.308651	1.705527	1.714529
H	-3.309061	0.880940	2.429948
H	-2.336142	2.201038	1.750981
H	-4.074233	2.423807	2.015220
C	-3.596008	2.364551	-0.708700
H	-4.376300	3.079275	-0.438978
H	-2.636161	2.884938	-0.693740
H	-3.790231	2.008153	-1.722177
C	3.246335	1.000760	-0.211002
C	4.029646	-0.169783	-0.015657
C	3.901912	2.348454	-0.376606
C	5.326402	2.333202	0.156333
C	6.058017	1.131067	-0.401613
N	5.385096	-0.094518	0.000586
C	4.187996	-2.677740	0.350757
C	5.536651	-2.341652	0.963668
C	6.213464	-1.290434	0.108676
H	3.336081	3.114189	0.167108
H	3.908460	2.639283	-1.434614
H	5.848666	3.250808	-0.120156
H	5.316736	2.267757	1.247237
H	7.078903	1.086446	-0.018776
H	6.117080	1.189902	-1.497586
H	6.427180	-1.691977	-0.891622
H	7.164732	-0.984808	0.549590
H	5.405715	-1.954533	1.977954
H	6.173796	-3.226034	1.022327
H	4.343671	-3.159362	-0.621374
H	3.633682	-3.384864	0.971842
C	-2.667542	-2.166944	-1.036838
H	-3.417041	-2.600077	-1.697164
H	-2.383958	-2.902392	-0.282297
H	-1.794288	-1.876356	-1.619458
H	1.670522	2.762485	-0.651452
H	1.513308	-2.424685	0.303860

QMCH_CTT

57

ELECTRONIC ENERGY = -1152.18103 a.u.

C	1.714737	-2.379056	-0.406075
C	1.025814	-1.143904	-0.292951
C	1.815722	0.022510	-0.185532
C	3.187227	-0.009528	-0.161931
C	-0.386758	-1.143417	-0.292095
C	-1.202713	-0.043461	-0.130738
C	-2.606534	-0.033656	-0.103451
O	0.939452	-3.494582	-0.507655
N	-3.451463	-0.955929	-0.570020
C	-4.794680	-0.602398	-0.292239
C	-4.805150	0.638162	0.339414
C	-3.391242	1.114288	0.497918
C	-5.949333	-1.319574	-0.575172
C	-7.157434	-0.729161	-0.201567
C	-7.190337	0.520203	0.423753
C	-6.011902	1.217581	0.701223
H	-6.046968	2.185150	1.190648
H	-8.145408	0.950020	0.701297
H	-8.085143	-1.252823	-0.398609
H	-5.933756	-2.294856	-1.045875
H	-0.750622	0.919280	0.079773
H	-0.841422	-2.122788	-0.371843
C	-3.020534	1.305078	1.972738
H	-3.187402	0.387484	2.540049
H	-1.972244	1.594018	2.074693
H	-3.638362	2.095190	2.404592
C	-3.147683	2.402602	-0.300936
H	-3.780754	3.197851	0.098105
H	-2.107308	2.724281	-0.218473
H	-3.392785	2.263537	-1.355629
C	3.098710	-2.456340	-0.426543
C	3.860839	-1.266319	-0.282538
C	3.773629	-3.792631	-0.602419
C	5.213352	-3.619135	-1.060706
C	5.900781	-2.598751	-0.177862
N	5.215748	-1.317109	-0.265362
C	3.985090	1.253361	0.009779
C	5.351850	1.101299	-0.634930
C	6.029035	-0.123583	-0.056936
H	3.238153	-4.390235	-1.349029
H	3.752932	-4.354975	0.339347
H	5.747420	-4.569966	-1.017969
H	5.237171	-3.266844	-2.095035
H	6.932664	-2.442080	-0.495870
H	5.923479	-2.943940	0.865046
H	6.221035	0.016657	1.015677
H	6.991330	-0.300678	-0.542009
H	5.249108	0.983218	-1.717217
H	5.974058	1.979144	-0.451223
H	4.112498	1.462011	1.078352
H	3.434122	2.094798	-0.416290
C	-3.115420	-2.127684	-1.360855
H	-3.954234	-2.345644	-2.018942
H	-2.922186	-2.991075	-0.721968
H	-2.232786	-1.908315	-1.960741
H	1.507832	-4.277546	-0.518773
H	1.329776	0.990260	-0.105394

QMCH_TTC

57

ELECTRONIC ENERGY = -1152.18407 a.u.

C	1.962457	-1.569754	-0.658688
C	1.269662	-0.400011	-0.253345
C	2.064435	0.740486	0.020511
C	3.433111	0.750295	-0.067857
C	-0.128558	-0.268108	-0.095865
C	-1.122521	-1.206221	-0.283326
C	-2.485816	-0.934807	-0.086032
O	1.214029	-2.680983	-0.911906
N	-3.417153	-1.870031	-0.303273
C	-4.713391	-1.398127	-0.003805
C	-4.606576	-0.072487	0.406926
C	-3.159889	0.342236	0.384601
C	-5.926055	-2.071499	-0.077148
C	-7.065207	-1.348735	0.277846
C	-6.979031	-0.014869	0.686068
C	-5.745720	0.637314	0.755568
H	-5.688518	1.671991	1.076747
H	-7.883244	0.518133	0.954805
H	-8.033022	-1.833794	0.236595
H	-6.006544	-3.106822	-0.384025
H	-0.842401	-2.203441	-0.586365
H	-0.432678	0.726587	0.213414
C	-2.714157	0.733312	1.800924
H	-2.863434	-0.095994	2.495281
H	-1.664387	1.024967	1.832179
H	-3.315242	1.578087	2.143906
C	-2.968023	1.488607	-0.619784
H	-3.573003	2.342065	-0.306364
H	-1.928074	1.809727	-0.679046
H	-3.294448	1.186612	-1.617003
C	3.341503	-1.594518	-0.803148
C	4.104757	-0.436359	-0.488132
C	4.015232	-2.852686	-1.289714
C	5.410819	-2.557580	-1.817511
C	6.159732	-1.712179	-0.809506
N	5.457331	-0.456843	-0.590285
C	4.224341	1.978279	0.289102
C	5.541233	1.988943	-0.467877
C	6.270658	0.689384	-0.197595
H	3.428881	-3.308533	-2.096390
H	4.077448	-3.587806	-0.477445
H	5.953813	-3.485636	-2.004759
H	5.345671	-2.010145	-2.761183
H	7.159854	-1.470310	-1.172632
H	6.273434	-2.253742	0.140012
H	6.535915	0.614467	0.865968
H	7.198251	0.637877	-0.771688
H	5.359939	2.086459	-1.541859
H	6.167676	2.826889	-0.156569
H	4.427086	1.982720	1.366166
H	3.632922	2.870537	0.072304
C	-3.141997	-3.219122	-0.771971
H	-4.073921	-3.674050	-1.094547
H	-2.703004	-3.815606	0.029518
H	-2.457332	-3.180189	-1.618313
H	1.557404	1.649874	0.334448
H	1.808547	-3.410861	-1.140286

QMCH_TTT

57

ELECTRONIC ENERGY = -1152.17833 a.u.

C	1.736422	-1.857495	-5.534949
C	1.939353	-1.086344	-4.353393
C	2.668238	0.122313	-4.499258
C	3.220930	0.516240	-5.687970
C	1.492264	-1.554425	-3.104206
C	1.532169	-0.872002	-1.900911
C	1.140184	-1.418542	-0.674429
O	1.020102	-3.017881	-5.526563
N	1.176942	-0.703995	0.459787
C	0.782173	-1.472946	1.574670
C	0.451474	-2.748936	1.126825
C	0.644466	-2.821730	-0.364620
C	0.711068	-1.101554	2.911303
C	0.278308	-2.076407	3.810671
C	-0.063549	-3.360599	3.378787
C	0.020378	-3.709785	2.028796
H	-0.244190	-4.710181	1.702732
H	-0.395653	-4.094687	4.103118
H	0.208729	-1.830100	4.863375
H	0.978865	-0.112815	3.262365
H	1.893262	0.149369	-1.895408
H	1.105329	-2.567066	-3.070411
C	-0.705655	-3.108966	-1.038135
H	-1.435954	-2.337292	-0.786780
H	-0.618349	-3.160372	-2.123765
H	-1.086119	-4.069012	-0.683112
C	1.695734	-3.888488	-0.700161
H	1.347494	-4.859890	-0.342683
H	1.874877	-3.964174	-1.773186
H	2.643961	-3.663788	-0.207796
C	2.265747	-1.491039	-6.758339
C	3.031133	-0.297735	-6.854601
C	2.002639	-2.345848	-7.967552
C	2.192475	-1.542087	-9.242096
C	3.523016	-0.820600	-9.188437
N	3.569310	0.074404	-8.035211
C	4.055357	1.762068	-5.780679
C	3.958480	2.344146	-7.180305
C	4.375024	1.284770	-8.178656
H	0.993304	-2.755127	-7.905238
H	2.682941	-3.204826	-7.961502
H	2.163114	-2.193284	-10.117511
H	1.392129	-0.804753	-9.348711
H	3.677380	-0.210022	-10.079212
H	4.350531	-1.539528	-9.129775
H	5.437308	1.036497	-8.053565
H	4.240173	1.641399	-9.201841
H	2.931998	2.658678	-7.388403
H	4.604969	3.216607	-7.291451
H	5.101946	1.517944	-5.563859
H	3.730137	2.479856	-5.024476
C	1.592765	0.685506	0.545110
H	1.346785	1.066192	1.531898
H	1.059979	1.276132	-0.200128
H	2.669026	0.770559	0.383603
H	2.852149	0.740883	-3.626349
H	0.416137	-3.019672	-4.771944

References

- [1] D. Avagliano, P. A. Sánchez-Murcia, L. González, *Phys. Chem. Chem. Phys.* **2019**, *21*, 8614–8618.
- [2] J. W. Ponder, D. A. Case, *Adv. Prot. Chem.* **2003**, *66*, 27–85.
- [3] M. Gaus, Q. Cui, M. Elstner, *J. Chem. Theory Comput.* **2011**, *7*, 931–948.
- [4] T. S. Lee, B. K. Radak, M. Huang, K. Y. Wong, D. M. York, *J. Chem. Theory Comput.* **2014**, *10*, 24–34.
- [5] TURBOMOLE GmbH, “TURBOMOLE V7.0 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>,” can be found under <http://www.turbomole.com>
- [6] J. Klett, A. Núñez-Salgado, H. G. Dos Santos, Á. Cortés-Cabrera, A. Perona, R. Gil-Redondo, D. Abia, F. Gago, A. Morreale, *J. Chem. Theory Comput.* **2012**, *8*, 3395–3408.

6.3 Site-Specific Photo-oxidation of the Isolated Adenosine-5'-triphosphate Dianion Determined by Photoelectron Imaging

Marie E. Castellani, Davide Avagliano, Leticia González, Jan R. R. Verlet

J. Phys. Chem. Lett. 2020, 11, 19, 8195–8201
<https://doi.org/10.1021/acs.jpcllett.0c02089>

Contributions:

- M.E CASTELLANI: Performed the experiments, wrote the initial draft of the manuscript and contributed to the final version of the manuscript
- D. AVAGLIANO: Contributed to the experiments, performed the calculations and contribute to the final manuscript.
- L. GONZÁLEZ: supervised the developments and contributed to the final manuscript
- J.R.R. VERLET: conceived the scope of the manuscript, supervised the developments and contributed to the final manuscript

Reprinted with permission from J. Phys. Chem. Lett. 2020, 11, 19, 8195–820.
Copyright 2020 American Chemical Society.

Site-Specific Photo-oxidation of the Isolated Adenosine-5'-triphosphate Dianion Determined by Photoelectron Imaging

Maria Elena Castellani,[§] Davide Avagliano,[§] Leticia González,^{*} and Jan R. R. Verlet^{*}



Cite This: *J. Phys. Chem. Lett.* 2020, 11, 8195–8201



Read Online

ACCESS |



Metrics & More



Article Recommendations



Supporting Information

ABSTRACT: Photoelectron imaging of the isolated adenosine-5'-triphosphate dianion excited to the $^1\pi\pi^*$ states reveals that electron emission is predominantly parallel to the polarization axis of the light and arises from subpicosecond electron tunneling through the repulsive Coulomb barrier (RCB). The computed RCB shows that the most probable electron emission site is on the amino group of adenine. This is consistent with the photoelectron imaging: excitation to the $^1\pi\pi^*$ states leads to an aligned ensemble distributed predominantly parallel to the long axis of adenine; the subsequent electron tunneling site is along this axis; and the negatively charged phosphate groups guide the outgoing electron mostly along this axis at long range. Imaging of electron tunneling from polyanions combined with computational chemistry may offer a general route for probing the intrinsic photo-oxidation site and dynamics as well as the overall structure of complex isolated species.

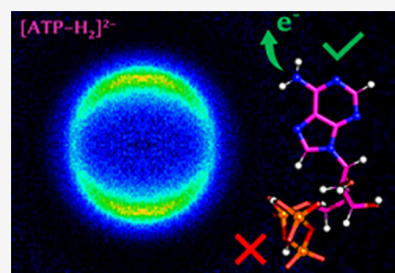


Photo-oxidation is an important source of biological damage, especially in DNA and its derivatives. All of the nucleobases possess bright $^1\pi\pi^*$ states that can readily absorb ultraviolet light. Adenine in particular is a common nucleobase, as it is present not only as one of the four DNA bases but also in adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide, which are central to cell energy distribution and metabolism, respectively.^{1,2} To gain an understanding of the intrinsic photoresponse of such molecules, gas-phase spectroscopy in conjunction with electronic structure theory has played an important role.^{3,4} For example, gas-phase environments have been used to probe the decay dynamics of photoexcited nucleobases,^{5–8} nucleotides⁹ and oligonucleotides;¹⁰ the ionization potentials of the nucleobases in nucleotides;^{9,11} their geometric structures;^{12–16} and their photodissociation products.^{17,18} The phosphate groups render most nucleotide derivatives negatively charged, and therefore, multiple phosphates lead to polyanions. While ubiquitous in the condensed phase, a polyanion of charge Q can also exist in the gas phase, but with the unique property that oxidation leads to a repulsive Coulomb barrier (RCB) between the departing electron and the remaining molecule with overall charge $Q + 1$.^{19–22} Here we exploit the unique properties of the RCB to determine the location of photo-oxidation in isolated doubly deprotonated adenosine-5'-triphosphate dianions, $[\text{ATP-H}_2]^{2-}$, and as a crude probe of its molecular structure.

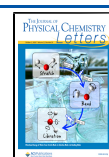
The electronic RCB arises from a balance between short-range attraction and long-range repulsion and has been extensively studied both theoretically^{20,21,23} and experimentally.^{22,24} The latter has been conveniently enabled by photoelectron (PE) spectroscopy of the polyanion.^{19,22,25–27}

Consider a dianion, as in the case of $[\text{ATP-H}_2]^{2-}$. In the absence of photoexcited states, the RCB height shown in Figure 1a can be directly measured from an electron kinetic energy (eKE) cutoff in the PE spectrum below which a PE cannot be emitted. If an excited state of the dianion lies below the RCB but higher than the adiabatic detachment energy (ADE) of the dianion, then the excited state will be a resonance and will be metastable with respect to electron tunneling through the RCB. This scenario is shown in Figure 1a. The tunneling lifetime can be directly measured using time-resolved PE spectroscopy.²⁸ Moreover, the RCB is not isotropic and depends on the relative location of the charged sites with respect to the departing electron. This anisotropy can be clearly seen in the computation of the RCB and in principle can be probed experimentally through PE imaging.²⁹ However, to do so requires a connection between the laboratory and molecular frames of reference, as the experiment probes only the laboratory-frame PE angular distribution. Such a connection can be most conveniently attained through photoexcitation of a chromophore with a known transition dipole moment, as has been shown for a model dianion.³⁰ While PE spectroscopy of polyanions has been predominantly an academic curiosity, there have been some PE spectroscopic measurements on polyanionic DNA fragments.^{17,31–39} In

Received: July 8, 2020

Accepted: September 4, 2020

Published: September 4, 2020



ACS Publications

© 2020 American Chemical Society

8195

<https://dx.doi.org/10.1021/acs.jpclett.0c02089>
J. Phys. Chem. Lett. 2020, 11, 8195–8201

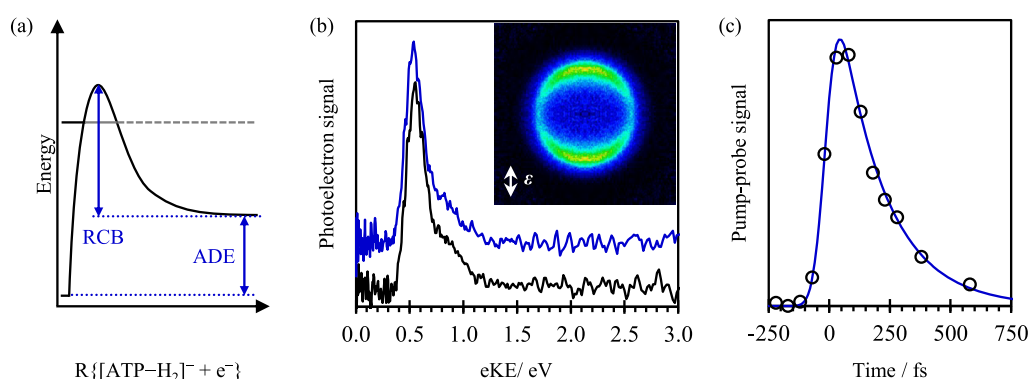


Figure 1. (a) Schematic of the repulsive Coulomb barrier (RCB), indicating its height as a function of the distance between $[\text{ATP-H}_2]^{2-}$ and the free electron, and the adiabatic detachment energy (ADE). (b) Photoelectron spectra of $[\text{ATP-H}_2]^{2-}$ taken at 4.66 eV with femtosecond (blue) and nanosecond (black) light pulses. The inset is a photoelectron image of $[\text{ATP-H}_2]^{2-}$ taken at 4.66 eV with nanosecond light pulses, with the polarization vector, ϵ , indicated. (c) Integrated pump-probe photoelectron signal following excitation to $^1\pi\pi^*$ states.

particular, Schinle et al.³⁴ performed a study of adenosine diphosphate, $[\text{ADP-H}_2]^{2-}$, and $[\text{ATP-H}_2]^{2-}$. For the latter, they observed no PE emission for excitation at 4.66 eV, although the photodetachment action spectrum by Cercola et al.¹⁷ clearly suggests that such emission should be present. Here we observe this emission, show that it arises from resonance tunneling, and, with the aid of electronic structure calculations, reveal that photo-oxidation leads to electron emission from the amino group of adenine.

Figure 1b shows the PE spectrum of mass-selected $[\text{ATP-H}_2]^{2-}$ taken with 4.66 eV light pulses derived from either a nanosecond laser (~ 5 ns pulse duration) or a femtosecond laser (~ 80 fs pulse duration). The PE spectra using the different light sources are essentially indistinguishable and show a dominant detachment feature that peaks at $\text{eKE} = 0.55$ eV with a slight shoulder at higher eKE . The PE spectra were obtained from their respective PE images, an example of which is shown in the Figure 1b inset. The PE feature at $\text{eKE} = 0.55$ eV is correlated with the intense anisotropic inner ring. The anisotropy of this feature peaks parallel to the polarization axis of the light, ϵ . The PE angular distribution, $I(\theta) \propto 1 + \frac{1}{2}\beta_2[3\cos^2(\theta) - 1]$, where θ is the angle between ϵ and the velocity vector of the emitted electron, can be quantified using the anisotropy parameter β_2 .^{40–42} For the dominant feature at $\text{eKE} = 0.55$ eV, $\beta_2 = +0.69 \pm 0.07$.

Schinle et al.³⁴ determined the adiabatic and vertical detachment energies of $[\text{ATP-H}_2]^{2-}$ to be 3.35 ± 0.11 and 4.01 ± 0.08 eV, respectively, and estimated the RCB to have a height of ~ 1.9 eV. Accordingly, the PE spectra in Figure 1b would arise from resonance tunneling. We verified this using time-resolved PE spectroscopy, which probed the initially excited state pumped at 4.66 eV with a 1.55 eV probe photon. The probe detaches a fraction of the excited-state population generating the electron signal at higher eKE . Concomitantly, the peak at 0.55 eV is depleted as less population is available to tunnel. The integrated time-resolved PE signal over the energy range $0.9 < \text{eKE} < 2.0$ eV, which probes the excited-state population as a function of the pump-probe delay time, t , is shown in Figure 1c, together with a fit to the data. The measured excited state lifetime is 190 ± 30 fs. This decay depends on the electron tunneling lifetime, τ_{tun} , as well as any other decay processes that may compete, such as internal conversion, which is known to be fast in adenine following excitation to the $^1\pi\pi^*$ states.^{5,6} Unfortunately, the signal-to-

noise ratio in the time-resolved PE spectra was insufficient to allow further data analysis. However, Cercola et al.¹⁷ measured the relative quantum yields of electron emission and photodissociation. Assuming a simple competitive first-order kinetic model might then suggest that τ_{tun} is ~ 2.5 times the observed lifetime, allowing us to estimate τ_{tun} as ~ 475 fs.

Figure 2 shows the computed RCB for $[\text{ATP-H}_2]^{2-} + e^-$ in the xz plane containing the adenine ring system, as shown. Details of the calculation of the RCB are given in section S1.2 in the Supporting Information, as well as a larger region of the plane (Figure S2). The loss of a π electron from adenine results in an RCB that is highest near the phosphate groups, where most of the negative charge resides. However, the potential energy is also highly structured on adenine itself. The lowest-energy saddle point of the RCB can be found at the amino group, as is more clearly shown in Figure 2b, and has a computed barrier height of 0.57 eV. Although this is different from the 1.9 eV RCB estimated by Schinle et al.,³⁴ caution should be applied regarding the previous PE spectrum. First, the PE spectrum by Schinle et al. was taken at a photon energy of 6.42 eV, which is likely to be resonant with excited states of not just the nucleobase but also the sugar and phosphate. It has been shown that the presence of resonances can lead to significant complications in terms of assigning the RCB from the PE spectrum.²² Second, at a photon energy of 6.42 eV, electron emission (direct or by tunneling) can arise from several sites of $[\text{ATP-H}_2]^{2-}$, and it is not known a priori which one results in the strongest signal.¹¹ In the present experiment, the photon energy of 4.66 eV is resonant only with excited states localized on the nucleobase, ensuring that the electron is removed exclusively from the nucleobase.⁴³ In a separate experiment, we focused the femtosecond 4.66 eV light in the interaction region to obtain the two-photon resonance-enhanced PE spectrum shown in Figure 3a. This shows an onset of the eKE signal at ~ 5.2 eV, which yields the ADE associated with the nucleobase: $\text{ADE}_A = 2 \times 4.66 - 5.2 = 4.1$ eV. However, the highest occupied molecular orbital is also on the nucleobase, so this measured ADE_A is probably equivalent to the ADE of $[\text{ATP-H}_2]^{2-}$. The computed RCB is consistent with our measurements. Given $\tau_{\text{tun}} < 1$ ps, the resonance leading to the peak at $\text{eKE} = 0.55$ eV will be near the lowest RCB saddle point (as previously seen for the fluorescein dianion),²⁸ which is energetically close to the calculated RCB height of 0.57 eV. The RCB for a plane perpendicular to the

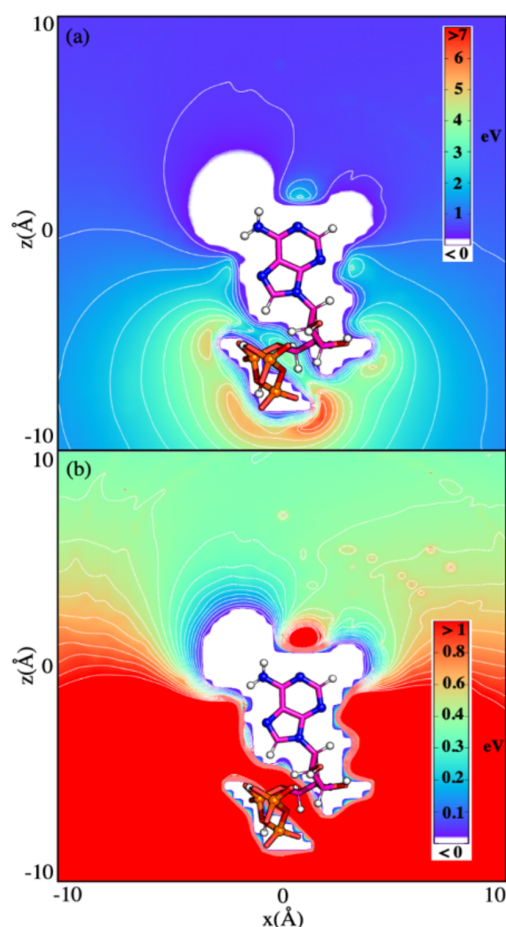


Figure 2. (a) Repulsive Coulomb barrier (RCB) for $[\text{ATP-H}_2]^{2-} + e^-$ in the plane passing through the calculated transition dipole moment of the excitation to the $^1\pi\pi^*$ state and including the adenine ring; each contour line represents an increment of 1 eV. (b) Map scaled to a maximum RCB of 1 eV; each contour line represents an increment of 0.05 eV. The RCB maps were computed at the MP2/def2-SVP level of theory.

nucleobase was also computed (see Figure S3), but the emission barrier along the yz plane was found to be higher than that along the xz plane. While the signal at 0.55 eV is consistent with tunneling through the RCB, it may also arise from direct detachment or autodetachment from the $^1\pi\pi^*$ states. We explored these possibilities by measuring the PE spectra over a range of photon energies from 4.43 to 4.77 eV (280 to 260 nm) as shown in Figure 3b. The PE peak at $\text{eKE} = 0.55$ eV does not shift despite the change in $h\nu$ by 0.34 eV, which should be readily observable. This observation excludes direct detachment. Autodetachment could still be consistent, but in that case, the electron simply tunnels through a different potential barrier and does not affect the following arguments concerning the angular distributions. Finally, we briefly comment on the possibility that the peak at 0.55 eV arises from thermionic emission (i.e., statistical electron emission from the ground state of the dianion following internal conversion from the $^1\pi\pi^*$ states). After all, some ground-state recovery does take place, as shown by Cercola et al.¹⁷ By delaying an electronic gate on the detector, we showed that all of the electron emission takes place within 100 ns, thus excluding thermionic emission, which proceeds on a much

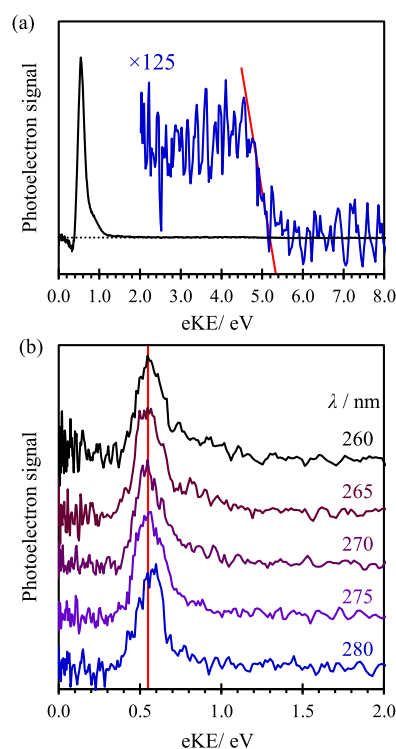


Figure 3. PE spectra of $[\text{ATP-H}_2]^{2-}$ taken with (a) focused 4.66 eV femtosecond light pulses and (b) variable-wavelength nanosecond light pulses. In (a), a resonance-enhanced two-photon photoelectron spectrum is obtained. The red line shows the high-eKE edge of the photoelectron spectrum, which intersects the zero signal (dashed line) at ~ 5.2 eV. In (b) the red line shows the position of 0.55 eV and highlights that the change in wavelength does not lead to a change in photoelectron spectrum of the tunneling peak.

longer time scale. Moreover, the photoelectron spectrum and angular distribution would be inconsistent with such a process.

An electron produced by tunneling will subsequently move along a trajectory determined by the RCB. Given the RCB in Figure 2a, intuition suggests that such a trajectory will avoid the high potential associated with the phosphate groups. The PE image inset in Figure 1b in principle contains this directional information. The resonance initially excited on the nucleobase involves the optically bright states of adenine, the $^1\pi\pi^*$ states.^{44–46} Our calculations show that the transition with the highest oscillator strength is of this character (for details, see the Supporting Information). Upon excitation with linearly polarized light of polarization ϵ , only those $[\text{ATP-H}_2]^{2-}$ molecules within a $\cos^2 \varphi$ distribution will be excited, where φ is the angle between ϵ and the transition dipole moment (TDM). In Figure 2, $[\text{ATP-H}_2]^{2-}$ has been aligned in such a way that the vertical axis (z) is aligned with the TDM of the brightest transition. Hence, from Figure 2 we anticipate that the PE will be emitted predominantly parallel to ϵ . This is consistent with the PE image in Figure 1b and its positive β_2 value.

In principle, one could simulate the emission using classical molecular dynamics (MD) of an electron under the influence of the RCB.⁴⁷ However, we refrained from performing such a simulation here because there are several factors that would undermine such a quantitative prediction and potentially even the qualitative reasoning provided thus far. First, the ions are

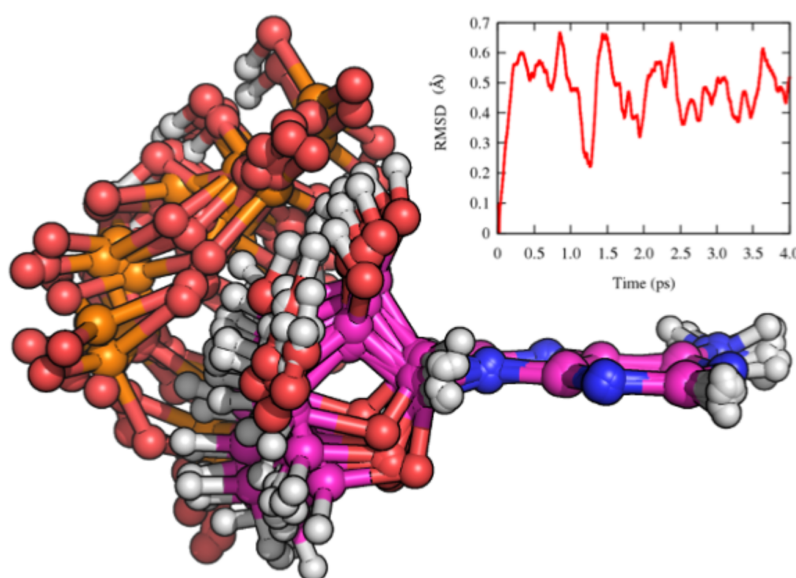


Figure 4. Superimposition of 10 equidistant (every 100 fs) $[\text{ATP-H}_2]^{2-}$ geometries along the first picosecond of the ab initio molecular dynamics simulation. The inset shows the root-mean-square deviation (RMSD) calculated along the 4 ps ground-state dynamical trajectory.

thermalized at ~ 300 K, and therefore, rotational motion of $[\text{ATP-H}_2]^{2-}$ could blur the alignment.³⁰ On the basis of the computed rotational constants, the rotational dephasing time⁴⁸ of $[\text{ATP-H}_2]^{2-}$ at 300 K is 5 ps, which exceeds τ_{tun} by at least 1 order of magnitude. Hence, rotational motion will not be a significant factor. Second, at ~ 300 K a number of isomers could be present in the gas-phase sample. Infrared multiphoton dissociation experiments and calculations show that $[\text{ATP-H}_2]^{2-}$ is predominantly $\alpha\beta$ deprotonated and that it forms an intramolecular hydrogen bond between the β phosphate and a neighboring oxygen atom.^{15,16,34} Third, even if a single dominant isomer were present, significant thermal motion could be associated with the phosphates, which could lead to dramatic changes in the RCB. Figure 4 presents the results of a ground-state ab initio molecular dynamics simulation over 4 ps (for details, see section S1.3). This shows the conformational flexibility of $[\text{ATP-H}_2]^{2-}$ around the minimum-energy structure. The inset shows the root-mean-square deviation (RMSD) calculated along the simulation time. While the phosphates clearly move, their locations relative to the adenine remain quite similar, and therefore, this motion does not qualitatively affect the RCB or the angular distribution of the emitted electrons. Fourth, as demonstrated by the results of Cercola et al.,¹⁷ excited-state dynamics competes with tunneling. On the ground-state potential energy surface, $[\text{ATP-H}_2]^{2-}$ dissociates rather than losing an electron,¹⁷ so this cannot contribute to the PE signal. However, substantial nuclear motion clearly takes place in the $^1\pi\pi^*$ states, and $^1n\pi^*$ states may also be involved in the decay mechanism.^{46,49–51} The nuclear dynamics likely does not have a major impact on the photoelectron angular distribution because the positions of the phosphates relative to the nucleobase do not change significantly over the time scale of the dynamics (the lifetime is 190 fs). The electronic dynamics, on the other hand, will alter the relevant molecular orbitals that should be considered. For example, in the case of a $^1n\pi^*$ state, the orbital from which the electron is removed will differ, as will the final states in $[\text{ATP-H}_2]^-$. The former could be taken into account by calculating the RCB for loss of an electron from the nonbonded orbital

located on the amino nitrogen atom. The consequence of this should be reflected in a different final eKE of the emitted electron. It should be noted that Figures 1b and 3b show a shoulder to higher eKE, which also has a slightly reduced anisotropy ($\beta_2 = +0.48 \pm 0.33$, where the much larger error reflects the reduced signal intensity). This feature could be consistent with emission from the $^1n\pi^*$ state, although there would also be a different RCB associated with such a process and thus a different lifetime, which was not seen in the experiment; we stress that the data are not of sufficient quality at present to be assertive on this conclusion. Alternatively, the shoulder could also arise from hot bands and/or different isomers or from direct and/or autodetachment. It is interesting to notice that in Figure 3b, the shoulder appears to become more prominent as $h\nu$ increases. This observation is very reminiscent of the behavior in the fluorescein dianion,²⁸ where direct detachment contributes once the lowest saddle point of the RCB is surpassed.

The above discussion shows that MD simulations would provide only limited added insight. In any case, such simulations would only be feasible if a high level of approximation (e.g., no conformational space investigation, low-dimensional analytical potentials, low accuracy) were considered, which in turn would lower the significance of any quantitative results. Despite the complications highlighted above, we can conclude that following photoexcitation to the $^1\pi\pi^*$ states, resonance tunneling in $[\text{ATP-H}_2]^{2-}$ is qualitatively consistent with emission from the amino group of adenine. Moreover, it provides a crude structural measure of isolated $[\text{ATP-H}_2]^{2-}$ by showing that the negative charges are located on phosphate groups and that these are located around the sugar. This latter observation is in agreement with IR multiphoton dissociation spectra of $[\text{ATP-H}_2]^{2-}$ and computational work.^{15,16,34} Tunneling detachment from the amino groups is expected to result in a β_2 that approaches +2. The reduced value observed here likely arises from the fact that the phosphate lies out of the xz plane (Figure 4), which imposes a Coulombic force on the outgoing electron along the

y axis, thus reducing the overall anisotropy observed experimentally.

The above observations of electron tunneling following photoexcitation of adenine in a polyanionic system potentially offer some insight into the photoemission observed in larger DNA complexes studied by the Gabelica and Dugourd groups.^{18,31–33} Specifically, they studied polyanions of 6- and 20-mer single strands and 12 base pair double strands^{31,32} as well as a four-tetrad G-quadruplex.³³ A general conclusion of their work is that only the purine nucleobases (adenine and guanine) appear to show enhanced electron emission compared with dissociation: for a homobase 6-mer single strand, this was observed over the $1\pi\pi^*$ absorption band regardless of charge state (from -2 to -4).¹⁸ The Kappes group also measured the PE spectra at 4.66 eV for the -3 charge state of the homobase 6-mer single strands.³⁷ However, these are difficult to interpret without knowledge of which features arise from direct detachment and which arise from tunneling. It should be noted that even if a resonance is above the RCB saddle point, it may still be subject to emission by tunneling, as shown for the fluorescein dianion²⁸ and the bisdisulizole tetraanion.²⁷ Very recently, Daly et al. showed that electron circular dichroism can be observed for chiral DNA strands.³⁹ Here we have shown that tunneling may be efficient in DNA fragments and that PE imaging may offer some insight into the emission process.

Our method provides a new route to understanding the photo-oxidation sites and overall structures of polyanions. Looking ahead, reducing the internal temperature of $[\text{ATP}-\text{H}_2]^{2-}$ using cryogenic ion traps will provide a more controlled environment^{52,53} and may enable quantitative determination of the emission process. A computational perspective on the excited-state dynamics of the molecule may in the end allow a complete picture of the photo-oxidation of this molecule to be obtained.

EXPERIMENTAL AND COMPUTATIONAL DETAILS

The details of the TR-PEI instrument can be found elsewhere.^{54,55} An ~ 1 mM methanolic solution of adenosine-5'-triphosphate disodium salt (Sigma-Aldrich), was pushed through a syringe and introduced into the first vacuum region through a capillary. A potential gradient led the ions through a series of ring-electrode ion guides until they reached a pulsed ion trap. Ions were ejected and focused into a collinear Wiley–McLaren time-of-flight mass spectrometer⁵⁶ at 10 Hz. At the focus of the mass spectrometer, a mass-selected ion packet was irradiated with laser pulses either from a commercial Ti:sapphire laser (producing femtosecond pulses) or a Nd:YAG-pumped optical parametric oscillator (producing nanosecond pulses). For time-resolved measurements, the third harmonic of the fundamental at 1.55 eV was used as a pump pulse and the fundamental as a probe pulse. The time resolution of the experiment was ~ 100 fs. Following the interaction of the light with the ion packet, emitted photoelectrons were collected and imaged by a perpendicular velocity map imaging arrangement.⁵⁷ A 300 ns electronic gate was applied to the microchannel plates so that only signal over this window was collected. PE spectra were obtained from the PE images using polar onion peeling.⁵⁸ The PE images were calibrated using the well-known PE image of I^- . The spectral resolution was approximately $\sim 5\%$ of the kinetic energy.

$[\text{ATP}-\text{H}_2]^{2-}$ was optimized with Møller–Plesset second-order perturbation theory using the resolution of identity

approximation (RI-MP2)⁵⁹ and the def2-TZVP basis set.⁶⁰ The ground-state energy was then refined with the def2-QZVP⁶⁰ basis set. The first five excitation energies were obtained with the algebraic diagrammatic construction scheme for the polarization propagator at second order (ADC(2))⁶¹ method and the same def2-QZVP⁶⁰ basis set. The RCB was calculated using the local static approximation model,⁶² in which the energy of the monoanion plus a point-charge electron (in the geometry of $[\text{ATP}-\text{H}_2]^{2-}$) was calculated with an interval of 0.5 Å. Two planes (*xz* and *yz*) that pass through the transition dipole moment vector connecting the S_0 and S_2 states were considered. RCB calculations were performed at the MP2/def2-SVP level of theory. Ground-state ab initio molecular dynamics calculations were carried out at the B3LYP^{63,64}/aug-cc-pVDZ⁶⁵ level of theory starting from the $[\text{ATP}-\text{H}_2]^{2-}$ minimum-energy geometry. Further computational details are given in the Supporting Information.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcllett.0c02089>.

Computational details and results of calculations on ground-state optimized structure and excited states, the repulsive Coulomb barrier, and ab initio molecular dynamics (PDF)

AUTHOR INFORMATION

Corresponding Authors

Leticia González – Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, 1090 Vienna, Austria;

orcid.org/0000-0001-5112-794X;

Email: leticia.gonzalez@univie.ac.at

Jan R. R. Verlet – Department of Chemistry, Durham

University, Durham DH1 3LE, United Kingdom; orcid.org/

0000-0002-9480-432X; Email: j.r.r.verlet@durham.ac.uk

Authors

Maria Elena Castellani – Department of Chemistry, Durham University, Durham DH1 3LE, United Kingdom

Davide Avagliano – Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, 1090 Vienna, Austria

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpcllett.0c02089>

Author Contributions

[§]M.E.C. and D.A. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement 765266 (Light-DyNAMics). D.A. and L.G. thank the University of Vienna and the Vienna Scientific Cluster (VSC) for allocation of computational resources.

REFERENCES

(1) Bruice, P. Y. *Essential Organic Chemistry*, 3rd ed.; Pearson: Upper Saddle River, NJ, 2016.

- (2) McMurry, J.; Begley, T. P. *The Organic Chemistry of Biological Pathways*, 2nd ed.; Roberts and Company Publishers: Greenwood Village, CO, 2016.
- (3) Stavros, V. G.; Verlet, J. R. R. Gas-Phase Femtosecond Particle Spectroscopy: A Bottom-Up Approach to Nucleotide Dynamics. *Annu. Rev. Phys. Chem.* **2016**, *67* (1), 211–232.
- (4) Weber, J. M.; Marcum, J.; Nielsen, S. B. UV Photophysics of DNA and RNA Nucleotides In Vacuo: Dissociation Channels, Time Scales, and Electronic Spectra. In *Photophysics of Ionic Biochromophores*; Brøndsted Nielsen, S., Wyer, J. A., Eds.; Springer: Berlin, 2013; pp 181–207.
- (5) Satzger, H.; Townsend, D.; Zgierski, M. Z.; Patchkovskii, S.; Ullrich, S.; Stolow, A. Primary processes underlying the photostability of isolated DNA bases: Adenine. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (27), 10196–10201.
- (6) Bisgaard, C. Z.; Satzger, H.; Ullrich, S.; Stolow, A. Excited-State Dynamics of Isolated DNA Bases: A Case Study of Adenine. *ChemPhysChem* **2009**, *10* (1), 101–110.
- (7) Wells, K. L.; Hadden, D. J.; Nix, M. G. D.; Stavros, V. G. Competing $\pi\sigma^*$ States in the Photodissociation of Adenine. *J. Phys. Chem. Lett.* **2010**, *1* (6), 993–996.
- (8) Mai, S.; Richter, M.; Marquetand, P.; González, L. Excitation of Nucleobases from a Computational Perspective II: Dynamics. In *Photoinduced Phenomena in Nucleic Acids I: Nucleobases in the Gas Phase and in Solvents*; Barbatti, M., Borin, A. C., Ullrich, S., Eds.; Springer: Cham, Switzerland, 2015; pp 99–153.
- (9) Chatterley, A. S.; West, C. W.; Stavros, V. G.; Verlet, J. R. R. Time-resolved photoelectron imaging of the isolated deprotonated nucleotides. *Chem. Sci.* **2014**, *5* (10), 3963–3975.
- (10) Chatterley, A. S.; West, C. W.; Roberts, G. M.; Stavros, V. G.; Verlet, J. R. R. Mapping the Ultrafast Dynamics of Adenine onto Its Nucleotide and Oligonucleotides by Time-Resolved Photoelectron Imaging. *J. Phys. Chem. Lett.* **2014**, *5* (5), 843–848.
- (11) Yang, X.; Wang, X.-B.; Vorpägel, E. R.; Wang, L.-S. Direct experimental observation of the low ionization potentials of guanine in free oligonucleotides by using photoelectron spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (51), 17588–17592.
- (12) Parker, A. W.; Quinn, S. J. Infrared Spectroscopy of DNA. In *Encyclopedia of Biophysics*; Roberts, G. C. K., Ed.; Springer: Berlin, 2013; pp 1065–1074.
- (13) Keane, P. M.; Baptista, F. R.; Gurung, S. P.; Devereux, S. J.; Sazanovich, I. V.; Towrie, M.; Brazier, J. A.; Cardin, C. J.; Kelly, J. M.; Quinn, S. J. Long-Lived Excited-State Dynamics of i-Motif Structures Probed by Time-Resolved Infrared Spectroscopy. *ChemPhysChem* **2016**, *17* (9), 1281–1287.
- (14) Hall, J. P.; Poynton, F. E.; Keane, P. M.; Gurung, S. P.; Brazier, J. A.; Cardin, D. J.; Winter, G.; Gunnlaugsson, T.; Sazanovich, I. V.; Towrie, M.; Cardin, C. J.; Kelly, J. M.; Quinn, S. J. Monitoring one-electron photo-oxidation of guanine in DNA crystals using ultrafast infrared spectroscopy. *Nat. Chem.* **2015**, *7* (12), 961–967.
- (15) Ligare, M. R.; Rijs, A. M.; Berden, G.; Kabeláč, M.; Nachtigallova, D.; Oomens, J.; de Vries, M. S. Resonant Infrared Multiple Photon Dissociation Spectroscopy of Anionic Nucleotide Monophosphate Clusters. *J. Phys. Chem. B* **2015**, *119* (25), 7894–7901.
- (16) van Outersterp, R. E.; Martens, J.; Berden, G.; Steill, J. D.; Oomens, J.; Rijs, A. M. Structural characterization of nucleotide 5'-triphosphates by infrared ion spectroscopy and theoretical studies. *Phys. Chem. Chem. Phys.* **2018**, *20* (44), 28319–28330.
- (17) Cercola, R.; Matthews, E.; Dessent, C. E. H. Photoexcitation of Adenosine 5'-Triphosphate Anions in Vacuo: Probing the Influence of Charge State on the UV Photophysics of Adenine. *J. Phys. Chem. B* **2017**, *121* (22), 5553–5561.
- (18) Daly, S.; Porrini, M.; Rosu, F.; Gabelica, V. Electronic spectroscopy of isolated DNA polyanions. *Faraday Discuss.* **2019**, *217* (0), 361–382.
- (19) Wang, X.-B.; Ding, C.-F.; Wang, L.-S. Photodetachment Spectroscopy of a Doubly Charged Anion: Direct Observation of the Repulsive Coulomb Barrier. *Phys. Rev. Lett.* **1998**, *81*, 3351–3354.
- (20) Simons, J. Molecular Anions. *J. Phys. Chem. A* **2008**, *112* (29), 6401–6511.
- (21) Dreuw, A.; Cederbaum, L. S. Multiply Charged Anions in the Gas Phase. *Chem. Rev.* **2002**, *102* (1), 181–200.
- (22) Verlet, J. R. R.; Horke, D. A.; Chatterley, A. S. Excited states of multiply-charged anions probed by photoelectron imaging: riding the repulsive Coulomb barrier. *Phys. Chem. Chem. Phys.* **2014**, *16* (29), 15043–15052.
- (23) Scheller, M. K.; Compton, R. N.; Cederbaum, L. S. Gas-Phase Multiply Charged Anions. *Science* **1995**, *270* (5239), 1160–1166.
- (24) Wang, L.-S.; Wang, X.-B. Probing Free Multiply Charged Anions Using Photodetachment Photoelectron Spectroscopy. *J. Phys. Chem. A* **2000**, *104* (10), 1978–1990.
- (25) Wang, X.-B.; Wang, L.-S. Observation of negative electron-binding energy in a molecule. *Nature* **1999**, *400* (6741), 245–248.
- (26) Wang, X.-B.; Woo, H.-K.; Huang, X.; Kappes, M. M.; Wang, L.-S. Direct Experimental Probe of the On-Site Coulomb Repulsion in the Doubly Charged Fullerene Anion C_{70}^{2-} . *Phys. Rev. Lett.* **2006**, *96* (14), 143002.
- (27) Dau, P. D.; Liu, H.-T.; Yang, J.-P.; Winghart, M.-O.; Wolf, T. J. A.; Unterreiner, A.-N.; Weis, P.; Miao, Y.-R.; Ning, C.-G.; Kappes, M. M.; Wang, L.-S. Resonant tunneling through the repulsive Coulomb barrier of a quadruply charged molecular anion. *Phys. Rev. A: At., Mol., Opt. Phys.* **2012**, *85* (6), No. 064503.
- (28) Horke, D. A.; Chatterley, A. S.; Verlet, J. R. Effect of internal energy on the repulsive Coulomb barrier of polyanions. *Phys. Rev. Lett.* **2012**, *108* (8), No. 083003.
- (29) Xing, X.-P.; Wang, X.-B.; Wang, L.-S. Photoelectron imaging of multiply charged anions: Effects of intramolecular Coulomb repulsion and photoelectron kinetic energies on photoelectron angular distributions. *J. Chem. Phys.* **2009**, *130* (7), No. 074301.
- (30) Horke, D. A.; Chatterley, A. S.; Verlet, J. R. R. Femtosecond Photoelectron Imaging of Aligned Polyanions: Probing Molecular Dynamics through the Electron–Anion Coulomb Repulsion. *J. Phys. Chem. Lett.* **2012**, *3* (7), 834–838.
- (31) Gabelica, V.; Tabarin, T.; Antoine, R.; Rosu, F.; Compagnon, I.; Broyer, M.; De Pauw, E.; Dugourd, P. Electron photodetachment dissociation of DNA polyanions in a quadrupole ion trap mass spectrometer. *Anal. Chem.* **2006**, *78* (18), 6564–72.
- (32) Gabelica, V.; Rosu, F.; Tabarin, T.; Kinet, C.; Antoine, R.; Broyer, M.; De Pauw, E.; Dugourd, P. Base-dependent electron photodetachment from negatively charged DNA strands upon 260-nm laser irradiation. *J. Am. Chem. Soc.* **2007**, *129* (15), 4706–13.
- (33) Rosu, F.; Gabelica, V.; De Pauw, E.; Antoine, R.; Broyer, M.; Dugourd, P. UV spectroscopy of DNA duplex and quadruplex structures in the gas phase. *J. Phys. Chem. A* **2012**, *116* (22), 5383–91.
- (34) Schinle, F.; Crider, P. E.; Vonderach, M.; Weis, P.; Hampe, O.; Kappes, M. M. Spectroscopic and theoretical investigations of adenosine 5'-diphosphate and adenosine 5'-triphosphate dianions in the gas phase. *Phys. Chem. Chem. Phys.* **2013**, *15* (18), 6640–50.
- (35) Burke, R. M.; Dessent, C. E. H. Effect of Cation Complexation on the Structure of a Conformationally Flexible Multiply Charged Anion: Stabilization of Excess Charge in the Na⁺-Adenosine 5'-Triphosphate Dianion Ion-Pair Complex. *J. Phys. Chem. A* **2009**, *113* (12), 2683–2692.
- (36) Burke, R. M.; Pearce, J. K.; Boxford, W. E.; Bruckmann, A.; Dessent, C. E. H. Stabilization of Excess Charge in Isolated Adenosine 5'-Triphosphate and Adenosine 5'-Diphosphate Multiply and Singly Charged Anions. *J. Phys. Chem. A* **2005**, *109* (43), 9775–9785.
- (37) Vonderach, M.; Ehrler, O. T.; Matheis, K.; Weis, P.; Kappes, M. M. Isomer-Selected Photoelectron Spectroscopy of Isolated DNA Oligonucleotides: Phosphate and Nucleobase Deprotonation at High Negative Charge States. *J. Am. Chem. Soc.* **2012**, *134* (18), 7830–7841.
- (38) Weber, J. M.; Ioffe, I. N.; Berndt, K. M.; Löffler, D.; Friedrich, J.; Ehrler, O. T.; Danell, A. S.; Parks, J. H.; Kappes, M. M. Photoelectron Spectroscopy of Isolated Multiply Negatively Charged Oligonucleotides. *J. Am. Chem. Soc.* **2004**, *126* (27), 8585–8589.

- (39) Daly, S.; Rosu, F.; Gabelica, V. Mass-resolved electronic circular dichroism ion spectroscopy. *Science* **2020**, 368 (6498), 1465–1468.
- (40) Reid, K. L. Photoelectron angular distributions. *Annu. Rev. Phys. Chem.* **2003**, 54, 397–424.
- (41) Cooper, J.; Zare, R. N. Angular Distribution of Photoelectrons. *J. Chem. Phys.* **1968**, 48 (2), 942–943.
- (42) Cooper, J.; Zare, R. N. Erratum: Angular Distribution of Photoelectrons. *J. Chem. Phys.* **1968**, 49 (9), 4252–4252.
- (43) Chatterley, A. S.; Johns, A. S.; Stavros, V. G.; Verlet, J. R. R. Base-Specific Ionization of Deprotonated Nucleotides by Resonance Enhanced Two-Photon Detachment. *J. Phys. Chem. A* **2013**, 117 (25), 5299–5305.
- (44) Santoro, F.; Improta, R.; Fahleson, T.; Kauczor, J.; Norman, P.; Coriani, S. Relative Stability of the La and Lb Excited States in Adenine and Guanine: Direct Evidence from TD-DFT Calculations of MCD Spectra. *J. Phys. Chem. Lett.* **2014**, 5 (11), 1806–1811.
- (45) Serrano-Andrés, L.; Merchán, M.; Borin, A. C. A Three-State Model for the Photophysics of Adenine. *Chem. - Eur. J.* **2006**, 12 (25), 6559–6571.
- (46) Conti, I.; Garavelli, M.; Orlandi, G. Deciphering Low Energy Deactivation Channels in Adenine. *J. Am. Chem. Soc.* **2009**, 131 (44), 16108–16118.
- (47) West, C. W.; Bull, J. N.; Woods, D. A.; Verlet, J. R. R. Photoelectron imaging as a probe of the repulsive Coulomb barrier in the photodetachment of antimony tartrate dianions. *Chem. Phys. Lett.* **2016**, 645, 138–143.
- (48) Blokhin, A. P.; Gelin, M. F.; Khoroshilov, E. V.; Kryukov, I. V.; Sharkov, A. V. Dynamics of optically induced anisotropy in an ensemble of asymmetric top molecules in the gas phase. *Opt. Spectrosc.* **2003**, 95 (3), 346–352.
- (49) Conti, I.; Nenov, A.; Höfinger, S.; Flavio Altavilla, S.; Rivalta, I.; Dumont, E.; Orlandi, G.; Garavelli, M. Excited state evolution of DNA stacked adenines resolved at the CASPT2//CASSCF/Amber level: from the bright to the excimer state and back. *Phys. Chem. Chem. Phys.* **2015**, 17 (11), 7291–7302.
- (50) Conti, I.; Di Donato, E.; Negri, F.; Orlandi, G. Revealing Excited State Interactions by Quantum-Chemical Modeling of Vibronic Activities: The R2PI Spectrum of Adenine. *J. Phys. Chem. A* **2009**, 113 (52), 15265–15275.
- (51) Picconi, D.; Avila Ferrer, F. J.; Improta, R.; Lami, A.; Santoro, F. Quantum-classical effective-modes dynamics of the $\pi\pi^* \rightarrow n\pi^*$ decay in 9H-adenine. A quadratic vibronic coupling model. *Faraday Discuss.* **2013**, 163 (0), 223–242.
- (52) Wang, X. B.; Wang, L. S. Development of a low-temperature photoelectron spectroscopy instrument using an electrospray ion source and a cryogenically controlled ion trap. *Rev. Sci. Instrum.* **2008**, 79 (7), No. 073108.
- (53) Garand, E. Spectroscopy of Reactive Complexes and Solvated Clusters: A Bottom-Up Approach Using Cryogenic Ion Traps. *J. Phys. Chem. A* **2018**, 122 (32), 6479–6490.
- (54) Lecointre, J.; Roberts, G. M.; Horke, D. A.; Verlet, J. R. R. Ultrafast Relaxation Dynamics Observed Through Time-Resolved Photoelectron Angular Distributions. *J. Phys. Chem. A* **2010**, 114 (42), 11216–11224.
- (55) Stanley, L. H.; Anstöter, C. S.; Verlet, J. R. R. Resonances of the anthracenyl anion probed by frequency-resolved photoelectron imaging of collision-induced dissociated anthracene carboxylic acid. *Chem. Sci.* **2017**, 8 (4), 3054–3061.
- (56) Wiley, W. C.; McLaren, I. H. Time-of-Flight Mass Spectrometer with Improved Resolution. *Rev. Sci. Instrum.* **1955**, 26 (12), 1150–1157.
- (57) Horke, D. A.; Roberts, G. M.; Lecointre, J.; Verlet, J. R. R. Velocity-map imaging at low extraction fields. *Rev. Sci. Instrum.* **2012**, 83 (6), No. 063101.
- (58) Roberts, G. M.; Nixon, J. L.; Lecointre, J.; Wrede, E.; Verlet, J. R. R. Toward real-time charged-particle image reconstruction using polar onion-peeling. *Rev. Sci. Instrum.* **2009**, 80 (5), No. 053104.
- (59) Weigend, F.; Häser, M. RI-MP2: first derivatives and global consistency. *Theor. Chem. Acc.* **1997**, 97 (1), 331–340.
- (60) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, 8 (9), 1057–1065.
- (61) Dreuw, A.; Wormit, M. The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, 5 (1), 82–95.
- (62) Dreuw, A.; Cederbaum, L. S. Erratum: Nature of the repulsive Coulomb barrier in multiply charged negative ions [Phys. Rev. A 63, 012501 (2000)]. *Phys. Rev. A: At., Mol., Opt. Phys.* **2001**, 63 (4), No. 049904.
- (63) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, 98 (7), 5648–5652.
- (64) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, 37 (2), 785–789.
- (65) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, 96 (9), 6796–6806.

SUPPORTING INFORMATION

Site-Specific Photo-Oxidation of the Isolated Adenosine-5'-

Triphosphate Dianion Determined by Photoelectron Imaging

Maria Elena Castellani*, Davide Avagliano†, Leticia González†, and Jan R. R. Verlet*

**Department of Chemistry, Durham University, DH1 3LE Durham, United Kingdom.*

†Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Währinger

Str. 17 1090 Vienna, Austria.

Contents

Section S.1 Computational details

- S.1.1** [ATP-H₂]²⁻ ground state optimization and excited states calculation
- S.1.2** Repulsive Coulomb Barrier calculations
- S.1.3** Ab initio molecular dynamics

References

Section S.1 Computational details

Section S.1.1 [ATP-H₂]²⁻ ground state optimization and excited states calculation

Methods

An initial conformational space sampling of [ATP-H₂]²⁻ was done using the OPLS_2005 force field [1] and MACROMODEL 11.5. [2] with the Monte Carlo method. The resulting minimum energy conformer was optimized with the Møller-Plesset second order perturbation theory using the resolution of identity approximation (RI-MP2)[3] and the def2-TZVP [4] basis set. The ground state energy was then redefined at the def2-QZVP level of theory to be compatible with the five lowest-lying excitation energies calculated with the Algebraic Diagrammatic Construction scheme for the polarization propagator at the second order (ADC(2)) [5] method and def2-QZVP[4] basis set. Scaled opposite spin was used with a scaling factor of 1.3 [6]. The energies are reported in Table S1. All the quantum chemical calculations were performed with the TURBOMOLE suite, v. 7.0 [7].

Data

Table S1: energy and oscillator strength (fos) for the first five singlet excited states of [ATP-H₂]²⁻

State	Energy (eV)	fos
S ₁	5.048	0.055
S ₂	5.330	0.212
S ₃	5.538	0.001

S ₄	6.072	0.003
S ₅	6.269	0.099

Table S2: relevant information about S₂ transition: orbitals and transition dipole moment

Occupied orbital	Virtual orbital	coefficient
130	133	0.7033
TDM component	Left/right moment	Transition moment
x	0.545	0.298
y	0.071	0.005
z	1.149	1.320

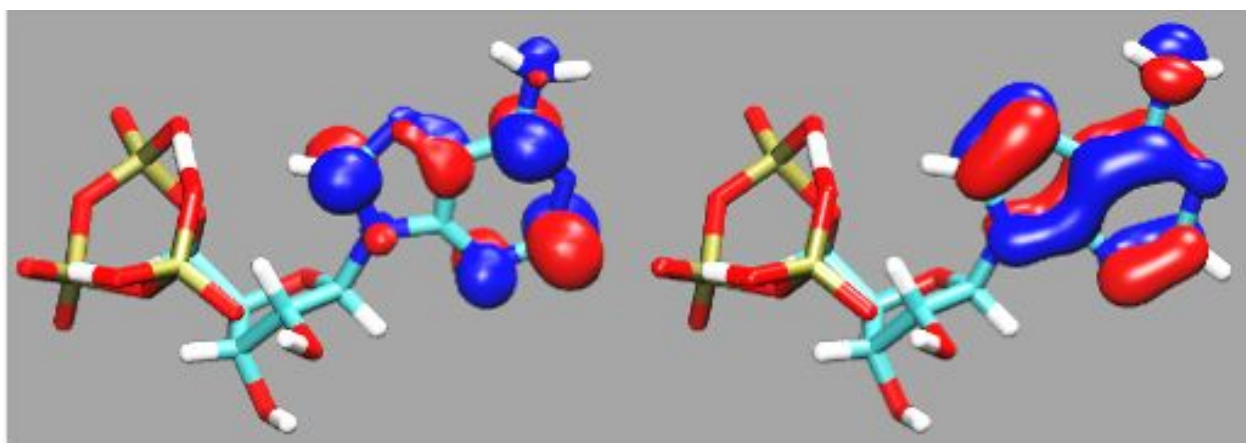


Figure S1: natural transition orbitals for the S₂ ¹ππ* state (isovalue = 0.046)

Cartesian coordinate of [ATP-H₂]²⁻ optimized geometry at RI-MP2/def2-TZVP level of theory in the xyz format (Å).

45

[ATP-H₂]²⁻

O	-0.3181870	-1.2406154	-7.3746486
C	-1.4599304	-2.0884473	-7.4606841
H	-1.1733859	-3.1065930	-7.1805698
H	-1.8715451	-2.0758802	-8.4756371
C	-2.5003129	-1.5540728	-6.4974069
H	-3.3837408	-2.2008788	-6.5354468
O	-1.9892416	-1.5840671	-5.1409899
C	-1.8431705	-0.2561629	-4.6714926
H	-2.6348058	-0.0030107	-3.9594463
N	-0.6039278	-0.1513380	-3.9258059
C	0.6816484	-0.1759440	-4.4077936
H	0.8899289	-0.2572101	-5.4664305
N	1.6065407	-0.1038108	-3.4566580
C	0.8737721	-0.0300463	-2.3029117
C	1.2519355	0.0981189	-0.9627561
N	2.5785931	0.0993505	-0.5940390
H	2.7532239	0.5610243	0.2865291
H	3.2055120	0.3512720	-1.3464607
N	0.3126745	0.1785309	-0.0182617
C	-0.9792366	0.1208615	-0.4121159
N	-1.4739853	0.0025322	-1.6419745
C	-0.5024709	-0.0607779	-2.5623005
C	-2.8959137	-0.1153181	-6.7603570
H	-2.8337256	0.1443776	-7.8175679
C	-1.9087309	0.6598212	-5.8897669
H	-0.9384563	0.6936984	-6.3806660
O	-2.3587640	1.9433955	-5.5450522
H	-1.8747739	2.5573965	-6.1757128
O	-4.2096896	0.1105214	-6.2429546
P	0.8861805	-1.3158286	-8.4439069
O	1.3797817	-2.6848327	-8.6731501
P	-0.9089972	0.3840356	-10.1429143
O	-2.1511015	-0.4176477	-10.2574986
P	-0.1406236	2.5675946	-8.3276243
O	1.2177706	2.0520405	-7.9002494
H	-4.1779658	1.0489607	-5.9830598
O	0.3260587	-0.6498065	-9.7715730

O	-1.0199776	1.2671310	-8.7803469
O	-0.0363970	3.4001405	-9.6807404
O	-0.9477792	3.3117375	-7.3114358
O	-0.3873211	1.2754942	-11.2274697
O	1.8913781	-0.3116061	-7.8236158
H	1.5960120	0.7166469	-7.8838971
H	-0.0563997	2.7338907	-10.4288747
H	-1.7081754	0.1851238	0.3904611

Section S.1.2 Repulsive Coulomb Barrier (RCB) calculations

Methods

The RCB maps were calculated using the Local Static Approximation model [8] employing the [ATP-H₂]²⁻ minimum energy geometry and calculating the energy for the monoanion plus an electron, located as a point charge along the two planes of interest, with an interval of 0.5 Å. These two planes pass through the transition dipole moment of the S₂ state. The plane showed in the main text (xz) is the one passing through the adenine ring, while in figure S1 is reported the RCB map along the plane orthogonal to it (yz). The coordinates of the planes (Å), which refer to the optimized geometry mentioned above, are given in the following (bottom left BL, top left TL, top right TR and bottom right BR). All the calculations are performed at MP2/def2-SVP[4] level of theory with the TURBOMOLE suite, v. 7.0 [7].

Data

XZ plane coordinates:

BL: (-16.69, 0.38, 27.28); TL: (-24.83, 0.24, -8.81); TR: (18.48, -0.36, -21.01);
BR: (26.62, -0.24, 15.09)

YZ plane coordinates:

BL: (-3.43, -6.14, -14.82); TL: (1.41, -6.07, 6.64); TR: (1.52, 6.93, 6.55); BR: (-3.26, 6.85, -14.90);

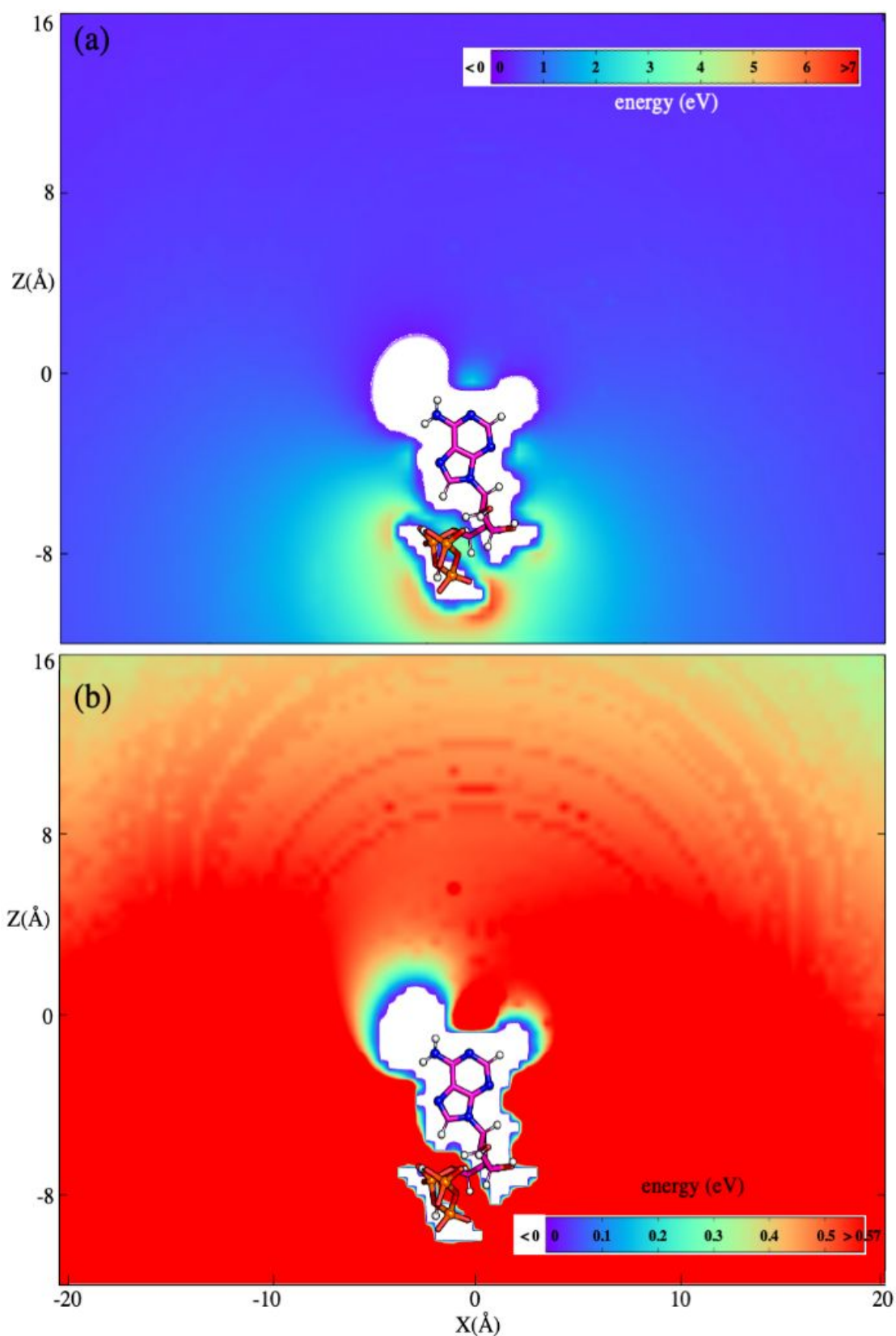


Figure S.2 Extended RCB map for the plane xz. Comparing with fig. 2 of the main text, this wider investigation of the RCB does not show any long range effect on the barrier.

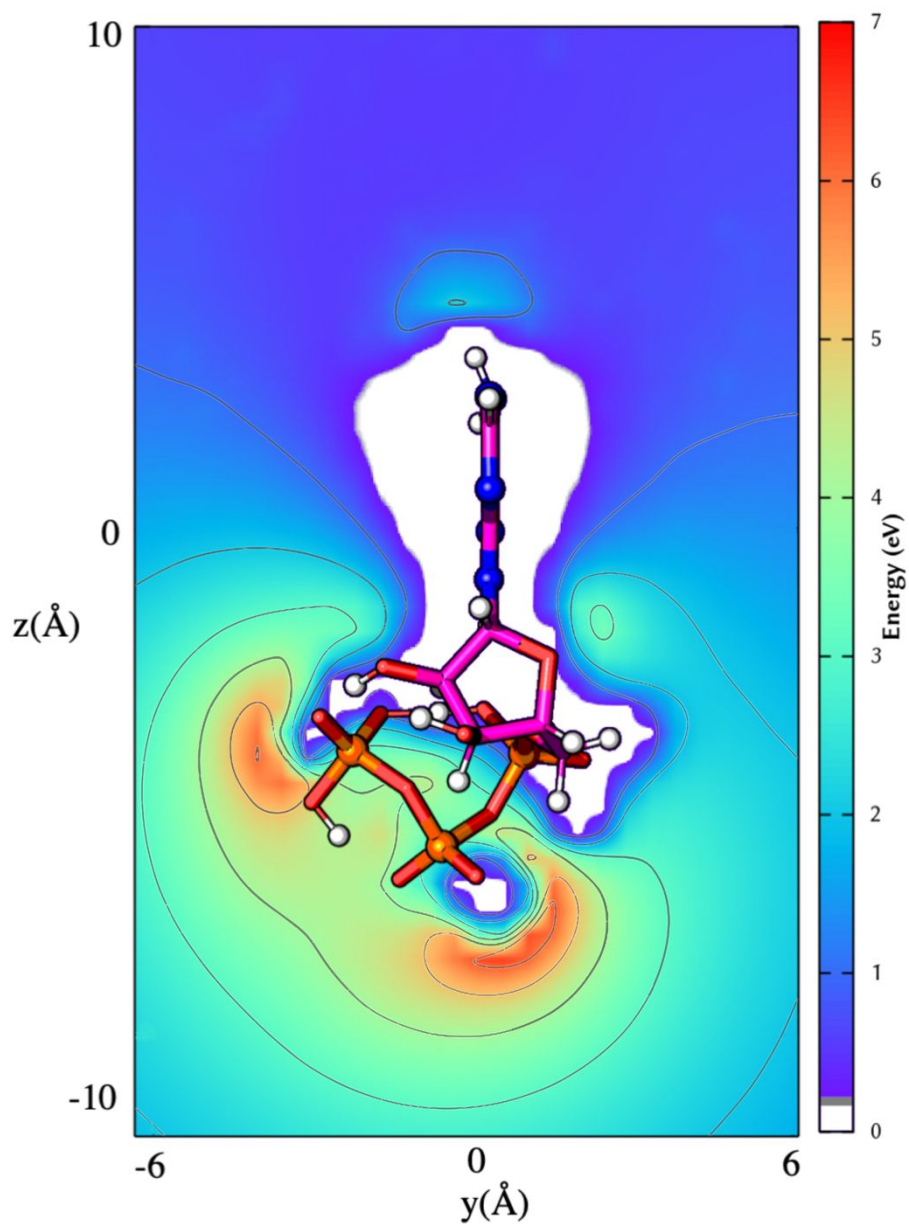


Figure S.3 RCB map for the plane yz. Each contour line represents an increment of 1 eV. The calculated high of the RCB in this plane is 0.65 eV

Section S.1.3 Ab initio molecular dynamics

Methods

Picosecond long *ab initio* molecular dynamics was run, starting from the [ATP–H₂]²⁻ minimized geometry, at B3LYP[9]/aug-cc-pVDZ[10] level of theory. The simulation time was 4ps, time step 0.2 fs, and T=300K using Langevin thermostat. The dynamics was running with Terachem suite [11].

References

- 1) Jay L. Banks and George A. Kaminski “Parametrizing a polarizable force field from *ab initio* data. I. The fluctuating point charge model”, *J. Chem. Phys.*, 1999, 110, 741
- 2) MacroModel, Schrödinger, LLC, New York, NY, 2018.
- 3) Weigend, F.; Häser, M., “RI-MP2: first derivatives and global consistency”. *Theor. Chem. Acc.* 1997, 97(1), 331-340
- 4) Weigend, F., “Accurate Coulomb-fitting basis sets for H to Rn”. *Phys. Chem. Chem. Phys.* 2006, 8(9), 1057-1065.
- 5) Dreuw, A. and Wormit, M “The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states.” *WIREs Comput. Mol. Sci.*, 2015, 5: 82-956)
- 6) Rohini C. Lochan, Yousung Jung, Martin Head-Gordon, “Scaled Opposite Spin Second Order Møller–Plesset Theory with Improved Physical Description of Long-Range Dispersion Interactions” *J. Phys. Chem. A* 2005, 109, 33, 7598–7605
- 7) TURBOMOLE GmbH, “TURBOMOLE V7.0 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH”, 1989-2007, TURBOMOLE GmbH, since 2007

8) Dreuw, A.; Cederbaum, L. S., Erratum: Nature of the repulsive Coulomb barrier in multiply charged negative ions [Phys. Rev. A 63, 012501 (2000)]. *Phys. Rev. A* 2001, *63*(4), 049904

9) a) Becke, A. D., "Density-functional thermochemistry. III. The role of exact exchange" *J. Chem. Phys.* 1993, *98* (7), 5648-5652. b) Lee, C.; Yang, W.; Parr, R. G., "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density." *Phys. Rev. B* 1988, *37*(2), 785-789.

10) Kendall, R. A.; Jr., T. H. D.; Harrison, R. J., "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions." *J. Chem. Phys.* 1992, *96* (9), 6796-6806

11) a) I. S. Ufimtsev and T. J. Martinez. "Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients and First Principles Molecular Dynamics", *Journal of Chemical Theory and Computation*, 5, 2009, 2619-2628 b) TeraChem, v. 1.9, PetaChem LLC, 2015

6.4 Enhanced Rigidity Changes Ultraviolet Absorption: Effect of a Merocyanine Binder on G-Quadruplex Photophysics

Davide Avagliano, Sara Tkaczyk, Pedro A. Sánchez-Murcia, Leticia González

J. Phys. Chem. Lett. 2020, 11, 23, 10212–10218
<https://doi.org/10.1021/acs.jpcllett.0c03070>

Contributions:

- D. AVAGLIANO: Conceived the idea of the manuscript, performed the calculations and the analysis, wrote the first draft of the manuscript and contributed to the final version of the manuscript.
- S. TKACZYK: Performed part of the calculations and of the analysis
- P.A. SÁNCHEZ-MURCIA: Supervised the development and contributed to the initial and final draft of the manuscript
- L. GONZÁLEZ: conceived the scope of the manuscript, supervised the developments and contributed to the final manuscript

Reprinted with permission from J. Phys. Chem. Lett. 2020, 11, 23, 10212–10218. Copyright 2020 American Chemical Society. Further permission should be directed to ACS.

Enhanced Rigidity Changes Ultraviolet Absorption: Effect of a Merocyanine Binder on G-Quadruplex Photophysics

Davide Avagliano, Sara Tkaczyk, Pedro A. Sánchez-Murcia,* and Leticia González*

Cite This: *J. Phys. Chem. Lett.* 2020, 11, 10212–10218

Read Online

ACCESS |



Metrics & More

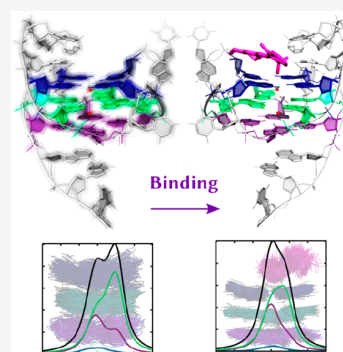


Article Recommendations



Supporting Information

ABSTRACT: The urge to discover selective fluorescent binders to G-quadruplexes (G4s) for rapid diagnosis must be linked to understand the effect that those have on the DNA photophysics. Herein, we report on the electronic excited states of a bound merocyanine dye to c-Myc G4 using extensive multiscale quantum mechanics/molecular mechanics calculations. We find that the absorption spectra of c-Myc G4, both without and with the intercalated dye, are mainly composed of exciton states and mixed local/charge-transfer states. The presence of merocyanine hardly affects the energy range of the guanine absorption or the number of guanines excited. However, it triggers a substantial amount (16%) of detrimental pure charge-transfer states involving oxidized guanines. We identify the rigidity introduced by the probe in G4, reducing the overlap among guanines, as the one responsible for the changes in the exciton and charge-transfer states, ultimately leading to a redshift of the absorption maximum.



G-Quadruplexes (G4s) are noncanonical secondary structures formed in nucleic acids where groups of four guanines interact via Hoogsteen base-pairing to form square structures, tetrads, that stack and are stabilized by a central metal cation,^{1,2} see Scheme 1. The formation of G4 motifs appears throughout the human genome and evinces essential functions in transcription, replication, stability, epigenetic regulation, as well as in cancer formation.^{3,4} G4s are present in the promoter regions of oncogenes, like in the protooncogene c-Myc,⁵ which regulates several elongation factors in cellular transcription. They also appear in viruses⁶ and have been discussed in the context of the pathogenicity of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) currently ravaging the world.⁷

While in the presence of UV radiation DNA/RNA nucleobases are able to efficiently dissipate the gained energy into heat into the environment,⁸ G4s can be damaged by the generation of guanine radical cations,^{9–13} thereby potentially affecting regulation of the transcription by c-Myc.¹⁴ Conversely, there is a strong drive to exploit the fluorescence of G4s in order to develop biomarkers for rapid non-destructive diagnosis.¹⁵ In this sense, small-molecule fluorescent probes are highly desirable tools to develop real-time diagnostics and also to monitor photo-oxidative lesions,⁶ as well as to understand the photophysics and photochemistry of G4 motifs themselves.^{16,17} However, the binding of chemical probes can affect the intrinsic chemical and physical properties of the G4, in both its electronic ground and excited states. Thus, only if the nature of the perturbation (structural, chemical, physical) induced by the probe is clearly identified is it possible to understand the native properties developed in G4 after light

excitation. In other words, it might be possible to connect the difference in the absorption after binding, with the perturbation induced by the fluorescent probe.

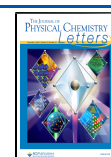
The number of available fluorescent binders with preference for quadruplex over duplex structures is very reduced¹⁸ and studies rationalizing the effect of the G4-binding are very scarce,¹⁹ hampering the design of G4-fluorophores with optimal properties.^{16,20} Most theoretical studies on the absorption of G4 have focused on unprobed G4s so far.^{21–24} There exist calculations of UV/vis spectra of telomeric G4 structures,^{23,24} but the spectroscopic properties of G4–probe complexes are only used as a G4 diagnostic tool to assess whether the probe binds.²⁵ Molecular dynamics studies have been carried out with G4 binders^{26–28} with the aim to monitor the spatiotemporal status of G4s by means of fluorescent probes and shed light into the biological role of these DNA structures. In this Letter, we follow a different approach. We study the photophysics of a G4 structure, in particular c-Myc G4, in the presence and absence of a spiropyran probe in order to quantify the effect of the binder on the nature of the electronic excited states of the G4 and to correlate the differences found with the structural effects induced by the G4 binder.

Spiroyrans belong to a class of organic photochromic molecules that, depending on UV and visible light, can

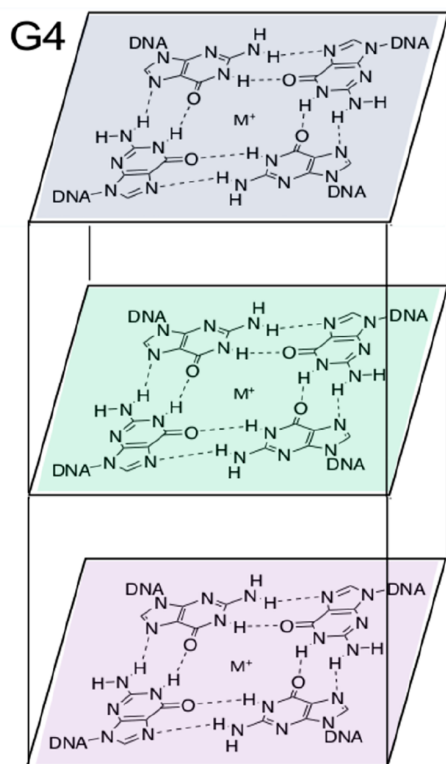
Received: October 8, 2020

Accepted: November 11, 2020

Published: November 18, 2020

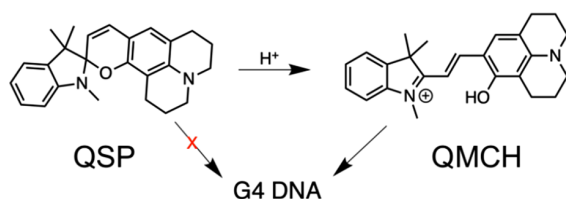


Scheme 1. Schematic Representation of a G-Quadruplex (G4) with Three Tetrads ($M^+ = K^+$ in the present study)



reversibly convert into a ring-opened merocyanine form. Intriguingly, the recently developed spiropyran decorated with a quinolizidine moiety (QSP, Scheme 2) does not bind

Scheme 2. Spiropyran (QSP)–Merocyanine (QMCH) Chemical Equilibrium, Indicating that QSP Does Not Bind G4 DNA but that QMCH Does



DNA, but once it isomerizes to its open and protonated merocyanine form (QMCH), it binds strongly with c-Myc G4 *in vivo*.²⁹ The closed QSP form emits at 458 nm and after QSP isomerizes to QMCH in the presence of c-Myc G4, its emission is drastically shifted to 610 nm. This visible change from blue to red allows the *in vivo* detection of G4 DNAs by the QSP/QMCH system. We recently investigated the QSP → QMCH isomerization reaction mechanism and the most probable binding mode of QMCH to G4, showing that QMCH “rigidifies” G4, reducing its conformational flexibility.³⁰ In this binding mode, the probe is stacked to the upper tetrad of the G4 pocket, at the 3′-end, interacting mainly with the four guanines via π – π stacking, but also with the side nucleobases with non-covalent interactions³⁰ (see Figure 1c). Without the probe, the guanines of G4 present high mobility, as shown in Figure 1a, but the presence of QMCH reduces the

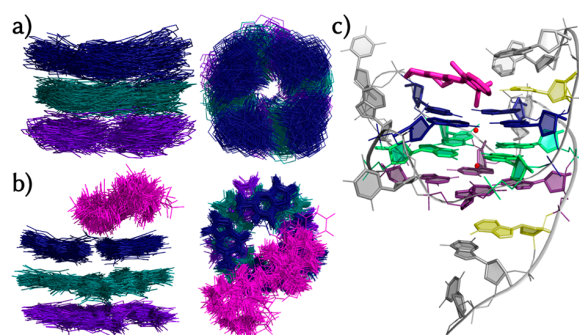


Figure 1. Side and top views of 100 overlapped geometries sampled with QM/MM-MD simulations of c-Myc G4 alone (a) and in the presence of merocyanine (b). Guanines are colored according to the tetrad they belong to (blue, green, and purple), with the probe in magenta. The presence of the probe produces an enhanced rigidity of the macromolecule, and the less flexibility reduces guanines overlapping. (c) A three-dimensional view of c-Myc G4 probed with merocyanine. Guanines included in the QM region for the calculation of the UV spectrum are colored by tetrads in blue, green, and purple. The additional frontier thymine and adenine nucleobases included in the QM region, but excluded in the excited states analysis, are colored in yellow. Cations are symbolized by red dots.

mobility of G4 so that the guanines stack more compactly and have less degree of movement—notice the unoccupied space between bases in Figure 1b, c. We hypothesize that the change on the mobility of c-Myc G4 may alter the overlap between the guanine electron densities, influencing the nature of its excited states, which is in line with recent experiments that suggest³¹ that restrained conformational changes are more important than the nature of the central cation or the folding topology in governing the excitation deactivation. In order to unveil this possibility, we characterize the nature of the electronic excited states of c-Myc G4 in the presence and absence of the fluorescent merocyanine probe QMCH.

Classical and mixed quantum mechanics/molecular mechanics (QM/MM) molecular dynamics were used to sample the conformational space of a c-Myc G4-folded 22-mer single-stranded DNA chain and provided 100 initial conditions on which the lowest 60 excited states are calculated with time-dependent density functional theory (TD-DFT).³² Here, an electrostatic embedding QM/MM scheme was employed, where the MM part was represented as point charges and the QM region was calculated with the CAM-B3LYP³³ functional with a def2-SVP³⁴ basis set. The QM region includes the 12 guanines involved in the three tetrads, the QMCH probe (see Figure 1c), and two additional frontier nucleobases to avoid spurious effects (see the Supporting Information for further details). In summary, our ad-hoc protocol relies on extensive sampling to reproduce an experimental-like ensemble of geometries and on a quantitative analysis of the electronic excitations. The high number of excitations calculated (6000 for the whole ensemble of geometries) allows for statistical analysis of the electronic effect induced by absorption of light in the G4. Ultimately, the applicability of the same protocol to both the unprobed and probed system gives the possibility of a direct comparison of the electronic excitations of the G4 in the absence and presence of the external probe.

A simple way of visualizing electronic excitations is to consider that when light absorption promotes an excited electron (E) to an upper electronic state, an electron hole (H)

is created at the initial location of the electron. Depending on where H and E are located, the electronic excitations can be classified (see Figure 2) as (a) monomer-like or local

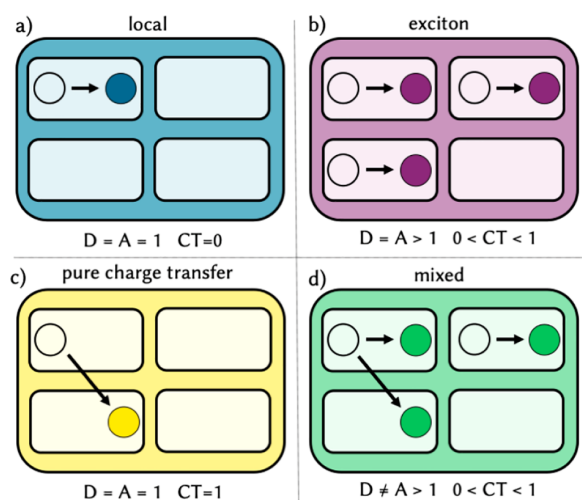


Figure 2. Schematic representation of the excited states formed in a G-quadruplex. Each rectangle represents a single guanine in a tetrad. Empty and full circles denote an electron hole and an excited electron, respectively. The number of electron donor (D) and acceptor (A) fragments involved in each type of excited state and their charge-transfer number (CT) is indicated for each type of electronic state.

excitations, where both H (empty circles) and E (full circles) are located on the same nucleobase; (b) exciton states, when more than one local excitation is present in more than one nucleobase; and (c) charge-transfer states, where the H and E are on different nucleobases.³⁵ As implemented in the wavefunction analysis software TheoDORE,³⁶ we additionally employ two parameters to discriminate between these states (also in Figure 2): the charge-transfer number (CT) and the number of donor (D) and acceptor (A) units. In a monomer, or local excitation, the H and E are fully localized on the same guanine unit, and thus, there is only one donor and one acceptor ($D = A = 1$) with no charge-transfer component ($CT = 0$, Figure 2a). In an exciton, the local excitations take place on different guanines, so that the number of D and A are still the same but larger than one ($D = A > 1$, Figure 2b).

The analysis of these descriptors in the c-Myc G4 shows that in the exciton states, the involved guanines also interact with each other. This implies that a fraction of the total density transfers from one monomer to another. Thus, CT can be anything in the range 0.1 to 0.9 (Figure 2b). In contrast, pure charge-transfer states have $CT > 0.9$ and the H and E are separated on different D and A units, but the number of D and A is the same; they are labelled with $D = A = 1$ (Figure 2c). These three scenarios usually coexist after light absorption in flexible multi-chromophoric systems, as was found by studying the fluorescent behavior of human telomeric G4 DNA.²³ Here, the H and E can be localized on a different number of D and A units ($D \neq A$) and different CT ($0.1 < CT < 0.9$) are possible. Accordingly, we label these states as mixed local/charge-transfer states (Figure 2d). In this case, some guanines are responsible of a local excitation while others induce electron density transfer between nucleobases, leading to different number of D and A participating units. Thus, these states are a combination of local, exciton, and charge-transfer states. As we

will show below, this rich mosaic of excited states with diverse charge-transfer values and diverse spatial localizations will contribute differently to the UV spectrum of c-Myc G4 with or without the fluorescent probe.

The absorption spectrum of c-Myc G4 alone (without probe), obtained from 6000 excited states calculated from an ensemble of 100 geometries, is shown in Figure 3a (black line).

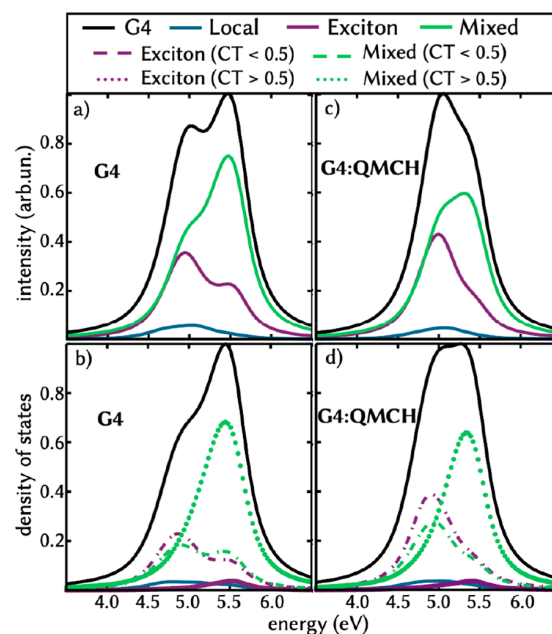


Figure 3. (a) UV absorption spectra of c-Myc G4 (black) and decomposition according to different excitations contributions. (b) Corresponding density of states decomposed according to the nature of excitations, also attending to the amount of charge transfer. (c) UV absorption spectrum of c-Myc G4 in the presence of QMCH (black line) and decomposition according to different excitations contributions. (d) Corresponding density of states decomposed according to the nature of excitations, also attending to the amount of charge transfer.

It displays two peaks centered at 4.8 and 5.4 eV, respectively. We additionally deconvoluted the spectrum according to the contributions given by the monomer local excitations (blue line), exciton states (violet line), and mixed states (green line). Noteworthy, pure charge-transfer states ($D = A = 1$, $CT > 0.9$) do not exist in our ensemble of vertical excitations because a small contribution of local excitations is always present throughout all the excitations. The contribution of monomer or local excitations alone is small; it peaks at circa 5.0 eV and corresponds to local guanine $\pi \rightarrow \pi^*$ excitations.³⁷ Thus, the relevant electronic excitations underlying both peaks are excitonic and/or mixed states excitations but with an important difference: whilst the excitonic excited states are significant at lower energies, the mixed states dominate the spectrum at high energies.

We can compare the computed absorption spectrum with a convoluted spectrum of the density of excited states (Figure 3b), which includes all excited states regardless of their brightness. We decompose it also into local, exciton, and mixed excitations. Additionally, we discriminate exciton and mixed states according to their amount of charge-transfer contributions ($CT < 0.5$ and $CT > 0.5$). As observed in the absorption

spectrum, exciton states prevail at low energies, but they have mostly low charge transfer ($CT < 0.5$, dashed line) while only a little amount of exciton states with high charge transfer ($CT > 0.5$, dotted line) is found at high energies. The mixed states, which dominate the density of states spectrum, have significant charge-transfer character at high energy (dotted line), while those with small charge transfer (dashed line) are equally distributed behind the two peaks.

We can therefore conclude that the two peaks observed in the absorption spectrum of c-Myc G4 correspond to excited states that differ in their excitation length and in their amount of charge transfer. The first one at 4.8 eV, less intense, is dominated by local, exciton and mixed states with low charge-transfer character. The second one at 5.4 eV is mostly composed of mixed states with strong charge-transfer character between different guanines. In the molecular orbital picture it means that a high fraction of density transfer ($CT > 0.5$) leads to a blue-shift in the absorption peak with respect of the local $\pi \rightarrow \pi^*$ guanine absorption band.³⁷ That signifies that the G4 acts as a H-like molecular aggregate, shifting the maximum of the absorption to the blue, with respect to the single local guanine excitation, once the guanines are compacted in the tetrads. States with small charge-transfer character can be found in any region of the spectrum, although they are most relevant at low energies. In spite of the large amount of charge-transfer character contributing to the high energy peak at 5.4 eV, something that is known to lead to dark excited states in the DNA context,³⁸ in c-Myc G4 this peak is the brightest thanks to the mixed local/charge-transfer character of the excitations.

Figure 3c clearly shows a different absorption of the guanine in the tetrads once the probe is bound. In particular, the H-like behavior vanishes, with a maximum of the absorption shifted to 5.0 eV, in agreement to the c-Myc G4 peak observed experimentally in the presence of QMCH.²⁹ In order to analyze the effect of the probe in the G4 absorption, we first focus on the electronic effect induced by QMCH. Due to its extended conjugated system, QMCH is, as other cyanine systems, highly prompted to absorb UV/vis light. Indeed, in the absorption spectrum of the complex G4:QMCH, we found two bands belonging to QMCH (Table 1 and Figures S1 and S2), i.e., the first, with a strong oscillator strength ($f_{osc} \sim 1$) centered at 2.8 eV, and the second, between 4.5–5.9 eV, weaker ($f_{osc} < 0.1$). The first absorption band represents 11% of the total excitations in the complex and appears in a region

that does not overlap with G4; the second one takes 7%. Importantly, in contrast to unprobed G4, there is now a relevant population of pure charge-transfer states (16% of the total excitations, Table 1). The location of H and E reveals that these pure charge-transfer states correspond to guanine-oxidized states, as the H is fully localized on the guanines (G^+) and E on the probe (P^-). These states are dark ($f_{osc} < 0.006$) and lie between the QMCH and G4 absorption bands (3.1–3.9 eV, Table 1), and they exist due to oxidative nature of the positively charged merocyanine.^{30,39} Although these states cannot be directly populated by absorption of light, if they are accessed non-radiatively after excitation, they would lead to an oxidative damage of the genetic code.⁴⁰ We also see that the H can be differently delocalized on nucleobases according to the electronic nature of the ligand, as shown for different merocyanine derivatives intercalated in duplex A/T.³⁹ Additionally, it has been extensively shown that guanine radicals, as a precursor of oxidative damage of the genetic code, can be generated by absorption of low energy UV light.⁴¹ Since G4 is a structure with propensity to form such experimentally observed guanine photo-oxidation, this would be enhanced by the interaction with the probe. On the other hand, the possible induced damage of the genetic code could have important consequences for applications in photoinduced therapy. Therefore, we believe that despite challenges, an investigation of the dynamics of such process will be of high interest in the future.

The presence of the QMCH also leads to the appearance of few mixed guanine/probe states, where the electron density can be both transferred from the guanine to the probe ($G \rightarrow P$) and vice versa ($P \rightarrow G$). These states show a very small oscillator strength and represent a small percentage of the total amount of excitation (3% each). Altogether, from an electronic point of view, merocyanine does not affect particularly the energy range of the guanine absorption because the largest group of excitations (60%) is represented by excited states involving only guanines (local/exciton/mixed states within G4, Table 1), whose absorption energy range is unshifted upon binding. In the following, we shall analyze in detail this region of the QMCH:G4 electronic absorption spectrum in order to discern the effect of the probe on the character of the $G \rightarrow G$ excitations. Example excitations of the G4:QMCH complex can be found in Figure S3.

We now analyze whether the number of guanines that participate in the electronic excitations changes, in the absence or presence of the probe, and whether it leads to differences in the two spectra. To this purpose, we employ the electron delocalization length (DEL) descriptor,³⁸ which indicates over how many guanines an excitation is delocalized (see Supporting Information). We focus on the exciton and mixed states, as they are the relevant states in the absorption spectrum of c-Myc G4. Figure 4 displays the number of guanines involved in each of the states of G4 (Figure 4a) and of G4:QMCH (Figure 4b). Unexcitingly, the differences are negligible, meaning that QMCH does not affect the DEL distributions, neither for exciton nor for mixed states. However, much more interesting is to see that the number of excited guanines is very different in the exciton and in the mixed states. While in the former, half of the population is found in two guanines ($D = A = 2$), the mixed states are mostly delocalized in three guanine units (ca. 30 %) but can reach up to nine different guanines, i.e., at least three tetrads can be simultaneously involved in one excitation. Excluding the role of

Table 1. Excitations Found in the Electronic Absorption Spectrum of the Complex G4:QMCH^a

excitation	<i>E</i> (eV)	<i>f</i> _{osc}	percentage of state (%)
local excitations within probe (P, first absorption band)	2.4–3.4	~1	11
local excitations within probe (P, second absorption band)	4.5–5.9	<0.100	7
pure charge-transfer states ($G \rightarrow P$)	3.1–3.9	<0.006	16
mixed states ($P \rightarrow G$)	3.9–4.5	<0.004	3
mixed states ($G \rightarrow P$)	4.7–6.0	<0.013	3
local/exciton/mixed excitations within G4 (G)	4.3–6.0	<0.400	60

^aEnergy range of absorption (*E*, eV), oscillator strength (*f*_{osc}), and their percentage with respect of the total number of excitation (%). P is the probe (QMCH) and G is guanine.

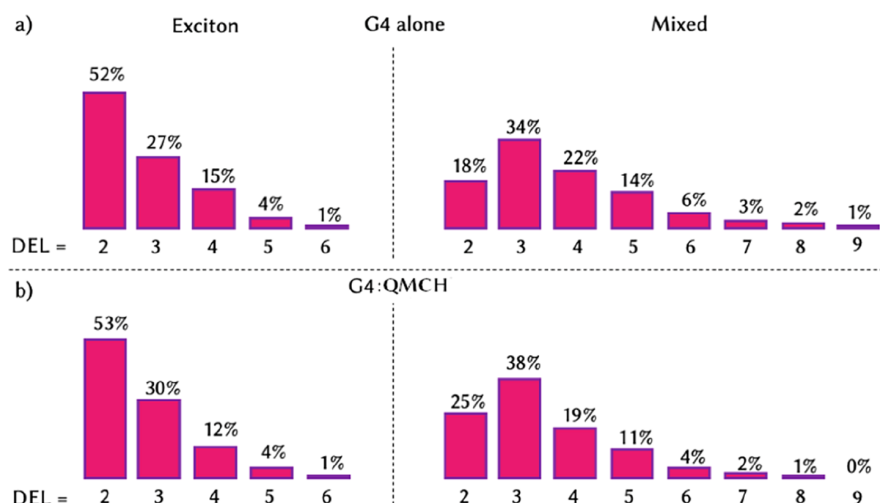


Figure 4. Percentage of exciton (left column) and mixed (right column) states with different delocalization lengths (DEL) for G4 (a) and G4:QMCH (b). DEL indicates the number of guanines involved in the excitation.

the DEL, we can conclude that it is the rigidity imposed by the probe (recall Figure 1a, b) that is responsible for the different absorptions at 4.5 and 5.5 eV of the two systems.

Figure 3c shows the computed UV absorption spectrum of G4 in the presence of QMCH (black line). It appears in the same energy range as G4 (Figure 3a), but there is an inversion on the relative intensities of the two absorption peaks. The absolute absorption maximum (~ 5.0 eV) is now the lowest-energy peak, and it is more intense than the one still centered at 5.4 eV. The low-energy region though has a higher density of states (black line, Figure 3d). In detail, we see that the number of exciton and mixed states with CT < 0.5 increases at low energies (dashed lines), while at high energy, mixed and exciton states with CT > 0.5 show comparable density of states (dotted lines) as without the probe. In the same region, the net number of exciton and mixed states with CT < 0.5 decreases (dashed lines), leading to an absolute higher number of exciton and mixed states in the low energy peak. The presence of QMCH affects mainly the oscillator strength of the mixed states at high energy, reducing their brightness, and promotes the formation of exciton and mixed states of small charge-transfer character in the low energy region. The intrinsic flexibility of the G4 promotes bright excited states at high energies with a strong charge-transfer character (mainly mixed states) and, thereby, with a high transfer of electron density among the involved fragments. Nevertheless, this situation changes upon binding of the probe. There is a reduction in the absorbance of the high energy peak due to the rigidity imposed by the dye, which stabilizes excitonic and mixed states with small charge transfer, thereby red-shifting the main absorption peak to 5.0 eV. The presence of the probe induces an external perturbation, which in turn induces a different absorption of the DNA. Knowing the nature of the perturbation, in this case the reduced mobility of the guanines, we are able to trace back the properties responsible of the UV/vis absorption of the system. The overlapping of numerous chromophores leads to the presence of mixed states, combining local, exciton, and charge-transfer states, and with a strong coupling between exciton and charge-transfer states. We showed how this overlapping is necessary for the population of mixed states with high charge transfer, responsible for the absorption at 5.4

eV of the unprobed G4 and the aggregate-like behaviour. Once the probe is bound, the absorption is shifted to the red, missing this requirement for the main absorption at those energies.

In conclusion, we have investigated for the first time the UV/vis absorption spectrum of c-Myc G4. We have characterized the absorption spectrum in terms of local, exciton, and mixed local/charge-transfer states and evaluated the impact of binding of a merocyanine binder on the photophysics of c-Myc G4. The probe does not affect either the energy range of G4 guanines absorption or the extent of the delocalization of the excited states, but its oxidative nature induces the formation of guanine oxidative states. Accordingly, the binding changes remarkably the photophysics of c-Myc G4 in the UV region. The probe induces an enhanced conformational rigidity on G4, altering the yield of exciton and mixed states absorption, ultimately leading to a global redshift in the G4 absorption maximum. These findings point out the importance of the structural flexibility in the photophysics of G4 DNA structures.

The characterization of the excitations involving the G4 binder will be helpful to functionalize fluorescent probes with optimally tuned photophysical properties. For instance, if the probe is modified so that the population of mixed/charge-transfer states is promoted, this species could evolve via non-emissive pathways and potentially react with G4. On the contrary, if functionalization promotes the population of electronic states that can relax to lower energy states, then fluorescence will be reinforced. Identifying these mixed states is thus interesting to rationalize the effect on the fluorescence of the probe. This could be an attractive avenue to explore in the future, both theoretically and experimentally. Further, our results can contribute to understand the photochemistry of related G4-binders as well as motivate the study of the temporal evolution of these excited states from both computational and experimental points of view.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpclett.0c03070>.

Extended computational details, additional UV spectra, and natural transition orbitals (PDF)

AUTHOR INFORMATION

Corresponding Authors

Pedro A. Sánchez-Murcia – Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, A-1180 Vienna, Austria; orcid.org/0000-0001-8415-870X; Email: pedro.murcia@medunigraz.at

Leticia González – Institute of Theoretical Chemistry, Faculty of Chemistry and Vienna Research Platform on Accelerating Photoreaction Discovery, University of Vienna, A-1180 Vienna, Austria; orcid.org/0000-0001-5112-794X; Email: leticia.gonzalez@univie.ac.at

Authors

Davide Avagliano – Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, A-1180 Vienna, Austria

Sara Tkaczyk – Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, A-1180 Vienna, Austria

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpclett.0c03070>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are thankful for funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 765266 (LightDyNAMics). The Vienna Scientific Cluster (VSC) is gratefully acknowledged for its generous allocation of computational resources. Basile F.E. Curchod and Lea M. Ibele are thanked for stimulating discussions.

REFERENCES

- (1) Williamson, J. R. G-Quartet Structures in Telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703–730.
- (2) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34* (19), 5402–5415.
- (3) Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S. The Regulation and Functions of DNA and RNA G-Quadruplexes. *Nat. Rev. Mol. Cell Biol.* **2020**, *21* (8), 459–474.
- (4) Guilbaud, G.; Murat, P.; Recolin, B.; Campbell, B. C.; Maiter, A.; Sale, J. E.; Balasubramanian, S. Local Epigenetic Reprogramming Induced by G-Quadruplex Ligands. *Nat. Chem.* **2017**, *9* (11), 1110–1117.
- (5) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (18), 11593–11598.
- (6) Ruggiero, E.; Richter, S. N. G-Quadruplexes and G-Quadruplex Ligands: Targets and Tools in Antiviral Therapy. *Nucleic Acids Res.* **2018**, *46* (7), 3270–3283.
- (7) Hognon, C.; Miclot, T.; García-Iriepa, C.; Francés-Monerris, A.; Grandemange, S.; Terenzi, A.; Marazzi, M.; Barone, G.; Monari, A. Role of RNA Guanine Quadruplexes in Favoring the Dimerization of SARS Unique Domain in Coronaviruses. *J. Phys. Chem. Lett.* **2020**, *11* (14), 5661–5667.
- (8) Barbatti, M.; Barone, V.; Borin, A. C.; Crespo-Hernández, C. E.; Vries, M. S. de.; Giussani, A.; González, L.; Hochlaf, M.; Improta, R.; Mai, S.; et al. *Photoinduced Phenomena in Nucleic Acids I*; Barbatti, M., Ullrich, S., Borin, A. C., Eds.; Springer: Heidelberg, 2015.
- (9) Balanikas, E.; Banyasz, A.; Baldacchino, G.; Markovitsi, D. Populations and Dynamics of Guanine Radicals in DNA Strands: Direct versus Indirect Generation. *Molecules* **2019**, *24* (13), 2347.
- (10) Banyasz, A.; Balanikas, E.; Martínez-Fernández, L.; Baldacchino, G.; Douki, T.; Improta, R.; Markovitsi, D. Radicals Generated in Tetramolecular Guanine Quadruplexes by Photoionization: Spectral and Dynamical Features. *J. Phys. Chem. B* **2019**, *123* (23), 4950–4957.
- (11) Hall, J. P.; Poynton, F. E.; Keane, P. M.; Gurung, S. P.; Brazier, J. A.; Cardin, D. J.; Winter, G.; Gunnlaugsson, T.; Sazanovich, I. V.; Towrie, M.; Cardin, C. J.; Kelly, J. M.; Quinn, S. J. Monitoring One-Electron Photo-Oxidation of Guanine in DNA Crystals Using Ultrafast Infrared Spectroscopy. *Nat. Chem.* **2015**, *7* (12), 961–967.
- (12) Wu, L.; Liu, K.; Jie, J.; Song, D.; Su, H. Direct Observation of Guanine Radical Cation Deprotonation in G-Quadruplex DNA. *J. Am. Chem. Soc.* **2015**, *137* (1), 259–266.
- (13) Banyasz, A.; Martínez-Fernández, L.; Balty, C.; Perron, M.; Douki, T.; Improta, R.; Markovitsi, D. Absorption of Low-Energy UV Radiation by Human Telomere G-Quadruplexes Generates Long-Lived Guanine Radical Cations. *J. Am. Chem. Soc.* **2017**, *139* (30), 10561–10568.
- (14) Fleming, A. M.; Burrows, C. J. Interplay of Guanine Oxidation and G-Quadruplex Folding in Gene Promoters. *J. Am. Chem. Soc.* **2020**, *142* (3), 1115–1136.
- (15) Kozitsina, A. N.; Svalova, T. S.; Malysheva, N. N.; Okhokhonin, A. V.; Vidrevich, M. B.; Brainina, K. Z. Sensors Based on Bio and Biomimetic Receptors in Medical Diagnostic, Environment, and Food Analysis. *Biosensors* **2018**, *8* (2), 35.
- (16) Li, Q.; Xiang, J. F.; Yang, Q. F.; Sun, H. X.; Guan, A. J.; Tang, Y. L. G4LDB: A Database for Discovering and Studying G-Quadruplex Ligands. *Nucleic Acids Res.* **2013**, *41* (D1), D1115–D1123.
- (17) Manna, S.; Srivatsan, S. G. Fluorescence-Based Tools to Probe G-Quadruplexes in Cell-Free and Cellular Environments. *RSC Adv.* **2018**, *8* (45), 25673–25694.
- (18) Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. Solution Structure of a 2:1 Quindoline-c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J. Am. Chem. Soc.* **2011**, *133* (44), 17673–17680.
- (19) Deiana, M.; Mettra, B.; Martínez-Fernández, L.; Mazur, L. M.; Pawlik, K.; Andraud, C.; Samoc, M.; Improta, R.; Monnerneau, C.; Matczyszyn, K. Specific Recognition of G-Quadruplexes over Duplex-DNA by a Macromolecular NIR Two-Photon Fluorescent Probe. *J. Phys. Chem. Lett.* **2017**, *8* (23), 5915–5920.
- (20) Di Antonio, M.; Ponjavic, A.; Radzevičius, A.; Ranasinghe, R. T.; Catalano, M.; Zhang, X.; Shen, J.; Needham, L. M.; Lee, S. F.; Klennerman, D.; Balasubramanian, S. Single-Molecule Visualization of DNA G-Quadruplex Formation in Live Cells. *Nat. Chem.* **2020**, *12* (9), 832–837.
- (21) Improta, R. Quantum Mechanical Calculations Unveil the Structure and Properties of the Absorbing and Emitting Excited Electronic States of Guanine Quadruplex. *Chem. - Eur. J.* **2014**, *20* (26), 8106–8115.
- (22) Saravanan, V.; Rajamani, A.; Subramaniam, V.; Ramasamy, S. Interaction of (G4)₂ and (X4)₂ DNA Quadruplexes with Cu⁺, Ag⁺ and Au⁺ Metal Cations: A Quantum Chemical Calculation on Structural, Energetic and Electronic Properties. *Struct. Chem.* **2020**, *31* (1), 465–484.
- (23) Martínez-Fernández, L.; Changenet, P.; Banyasz, A.; Gustavsson, T.; Markovitsi, D.; Improta, R. Comprehensive Study of Guanine Excited State Relaxation and Photoreactivity in G-Quadruplexes. *J. Phys. Chem. Lett.* **2019**, *10* (21), 6873–6877.
- (24) Martínez-Fernández, L.; Esposito, L.; Improta, R. Studying the Excited Electronic States of Guanine Rich DNA Quadruplexes by Quantum Mechanical Methods: Main Achievements and Perspectives. *Photochem. Photobiol. Sci.* **2020**, *19* (4), 436–444.

- (25) Terenzi, A.; Gattuso, H.; Spinello, A.; Keppler, B. K.; Chipot, C.; Dehez, F.; Barone, G.; Monari, A. Targeting G-Quadruplexes with Organic Dyes: Chelerythrine–DNA Binding Elucidated by Combining Molecular Modeling and Optical Spectroscopy. *Antioxidants* **2019**, *8* (10), 472.
- (26) Zhai, Q.; Gao, C.; Ding, J.; Zhang, Y.; Islam, B.; Lan, W.; Hou, H.; Deng, H.; Li, J.; Hu, Z.; Mohamed, H. I.; Xu, S.; Cao, C.; Haider, S. M.; Wei, D. Selective Recognition of C-MYC Pu22 G-Quadruplex by a Fluorescent Probe. *Nucleic Acids Res.* **2019**, *47* (5), 2190–2204.
- (27) Mulholland, K.; Siddiquei, F.; Wu, C. Binding Modes and Pathway of RHPS4 to Human Telomeric G-Quadruplex and Duplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent. *Phys. Chem. Chem. Phys.* **2017**, *19* (28), 18685–18694.
- (28) Li, M. H.; Luo, Q.; Xue, X. G.; Li, Z. S. Molecular Dynamics Studies of the 3D Structure and Planar Ligand Binding of a Quadruplex Dimer. *J. Mol. Model.* **2011**, *17* (3), 515–526.
- (29) Li, J.; Yin, X.; Li, B.; Li, X.; Pan, Y.; Li, J.; Guo, Y. Spiropyran in Situ Switching: A Real-Time Fluorescence Strategy for Tracking DNA G-Quadruplexes in Live Cells. *Anal. Chem.* **2019**, *91* (8), 5354–5361.
- (30) Avagliano, D.; Sánchez-Murcia, P. A.; González, L. Spiropyran Meets Guanine Quadruplexes: Isomerization Mechanism and DNA Binding Modes of Quinolizidine-substituted Spiropyran Probes. *Chem. - Eur. J.* **2020**, *26*, 13039–13045.
- (31) Ma, C.; Chan, R. C. T.; Chan, C. T. L.; Wong, A. K. W.; Kwok, W. M. Real-Time Monitoring Excitation Dynamics of Human Telomeric Guanine Quadruplexes: Effect of Folding Topology, Metal Cation, and Confinement by Nanocavity Water Pool. *J. Phys. Chem. Lett.* **2019**, *10* (24), 7577–7585.
- (32) Casida, M. E.; Huix-Rotlant, M. Progress in Time-Dependent Density-Functional Theory. *Annu. Rev. Phys. Chem.* **2012**, *63*, 287–323.
- (33) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393* (1–3), 51–57.
- (34) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305.
- (35) Ibele, L. M.; Sánchez-Murcia, P. A.; Mai, S.; Nogueira, J. J.; González, L. Excimer Intermediates En Route to Long-Lived Charge-Transfer States in Single-Stranded Adenine DNA as Revealed by Nonadiabatic Dynamics. *J. Phys. Chem. Lett.* **2020**, *11*, 7483–7488.
- (36) Plasser, F. TheoDORE: A Toolbox for a Detailed and Automated Analysis of Electronic Excited State Computations. *J. Chem. Phys.* **2020**, *152* (8), 084108.
- (37) Chaugnet-Barret, P.; Hua, Y.; Markovitsi, D. Electronic Excitations in Guanine Quadruplexes. In *Photoinduced Phenomena in Nucleic Acids II*; Springer: Berlin, 2014; pp 183–202.
- (38) Nogueira, J. J.; Plasser, F.; González, L. Electronic Delocalization, Charge Transfer and Hypochromism in the UV Absorption Spectrum of Polyadenine Unravelling by Multiscale Computations and Quantitative Wavefunction Analysis. *Chem. Sci.* **2017**, *8* (8), 5682–5691.
- (39) Avagliano, D.; Sánchez-Murcia, P. A.; González, L. Directional and Regioselective Hole Injection of Spiropyran Photoswitches Intercalated into A/T-Duplex DNA. *Phys. Chem. Chem. Phys.* **2019**, *21* (32), 17971–17977.
- (40) Kanvah, S.; Joseph, J.; Schuster, G. B.; Barnett, R. N.; Cleveland, C. L.; Landman, U. Z. I. Oxidation of DNA: Damage to Nucleobases. *Acc. Chem. Res.* **2010**, *43* (2), 280–287.
- (41) Balanikas, E.; Banyasz, A.; Douki, T.; Baldacchino, G.; Markovitsi, D. Guanine Radicals Induced in DNA by Low-Energy Photoionization. *Acc. Chem. Res.* **2020**, *53* (8), 1511–1519.

SUPPORTING INFORMATION

Enhanced Rigidity Changes Ultraviolet Absorption: Effect of a Merocyanine Binder on G-Quadruplex Photophysics

Davide Avagliano,^[a] Sara Tkaczyk,^[a] Pedro A. Sánchez-Murcia^{[a]‡*} and Leticia González^{[a,b]*}

^[a]Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Währinger Straße 17, A-1180 Vienna, Austria. E-mail: leticia.gonzalez@univie.ac.at

^[b]Vienna Research Platform on Accelerating Photoreaction Discovery, University of Vienna, Währinger Straße 17, A-1180 Vienna, Austria.

[‡]Present address: Division of Physiological Chemistry, Otto-Loewi Research Center, Medical University Graz, Neue Stiftingtalstraße 6/III A-8010 Graz, Austria. E-mail: pedro.murcia@medunigraz.at

Contents:

	Page:
Section S1	Computational details
Section S2	Additional UV spectra
Section S3	Natural Transition Orbitals
References	

Section S1 Computational Details

We consider the NMR-solved structure of the major quadruplex, formed in the promoter region of the human c-MYC oncogene (sequence 5'-d(*TP*GP*AP*GP*GP*GP*TP*GP*GP*GP*TP*AP*GP*GP*GP* TP*GP*GP*GP*TP*AP*A)-3', PDB id. 2L7V¹) by a 22-mer single stranded DNA. The two potassium ions of NMR-solved structure were preserved in our simulations.

The initial conditions for the excited states calculation were obtained through a multiscale computational approach based on classical and mixed quantum/classical molecular dynamics (QM/MM). Details about the classical molecular dynamics (MM) can be found elsewhere.² From the MM trajectory, 100 equidistant geometries were used to run 100 independent QM/MM-MD Born-Oppenheimer simulations. The density functional tight binding method version 3 (DFTB3)³ is used in the QM region, as implemented in the self-built module in AMBER.⁴ The time step was 0.1 fs, the temperature was kept constant at 300 K with Langevin thermostat,⁵ and the electrostatic cutoff for the QM-MM interaction was set 10 Å. Periodical boundary conditions were kept and the cutoff for the non-bonding interaction calculated in the direct space was 10 Å. The part of the system treated with QM consisted of all the guanine molecules of the quadruplex, two additional nucleobases on the edges, and - in case of the G4:probe complex- the probe as well. The rest of the system was treated at the force field level, as reported in Ref. 2. The inclusion of the two side nucleobases (adenine and thymine), was essential to avoid spurious limit effects on the description of the QM region, so that the excitations involving these fragments are systematically excluded for that reason. The QM/MM trajectories were propagated for 1 ps and the last geometries were used as initial conditions for the excited states calculations. All the QM/MM dynamics calculations were run with AMBER18.⁴

The electronic excitation energies of the first 60 singlet excited states were calculated for each of the 100 initial geometries using time dependent density functional theory (TD-DFT).^{6,7} As before, an electrostatic embedding QM/MM scheme was employed, where the MM part was represented as point charges and the QM region is represented with the long-range Hartree Fock exchange correlation corrected functional CAM-B3LYP⁸ with the Grimme's dispersion correction D3⁹ and a def2-SVP

basis set.¹⁰ The calculations were done with the GPU-based code TeraChem1.93^{11,12} via the AMBER18 interface.

The UV/vis spectra were convoluted as a sum of Gaussian functions:

$$\sigma(E) = \sum_g^{geom} \sum_s^{state} f_{gs} \exp(-4\ln(2)(E - E_{gs})^2 (FWMH)^{-2}) \quad (S1)$$

where f_{gs} is the oscillator strength in the ground state, E and E_{gs} are the energies in the excited state and in the ground state, respectively, and FWMH is the full width at half maximum (set as 0.2 eV). To characterize charge transfer proceses, the system was divided in fragments, every guanine representing a fragment, and the probe, if applicable, also being a fragment. We computed charge transfer numbers (CT) using the Lödwin-like population analysis,

$$\Omega_{AB} = \sum_{\mu \in A} \sum_{\nu \in B} (S^{1/2} D^{OI} S^{1/2})_{\mu\nu}^2 \quad (S2)$$

where A and B represents two different fragments, μ and ν indicates atomic orbitals, S is the overlap matrix, and D^{OI} is the one-particle transition density matrix (D^{OI}).¹³ If we consider the states I and J and the orbitals α and β , one element of the one-particle transition density matrix is given by:

$$D_{\alpha\beta}^{OI} = \langle \Psi^I | \hat{a}_{\alpha}^{\dagger} \hat{a}_{\beta} | \Psi^J \rangle \quad (S3)$$

where $\hat{a}_{\alpha}^{\dagger}$ and \hat{a}_{β} are respectively the creation and annihilation operators. We then calculate delocalization lengths (DEL) as total delocalization of the electron hole (H), excited electron (E) and total excitation among fragments, as implemented in the wavefunction analysis software TheoDORE.¹⁴

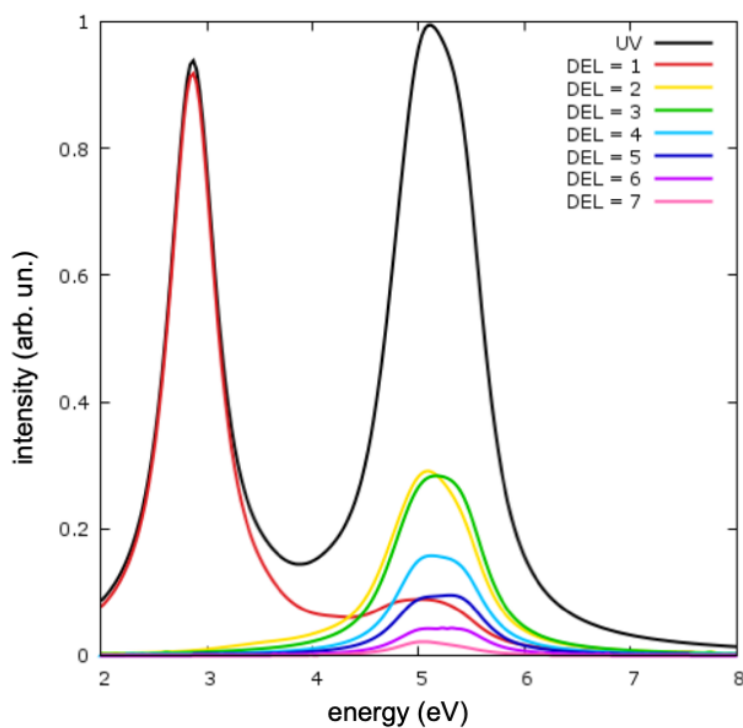


Figure S 1: total absorption spectrum for QMCH:G4 complex and individual decomposition for different delocalization lengths (DEL). The band centered at 2.88 eV is given by the absorption of the probe (DEL=1), while the higher energy band is given by the excitations localized up to 7 guanines.

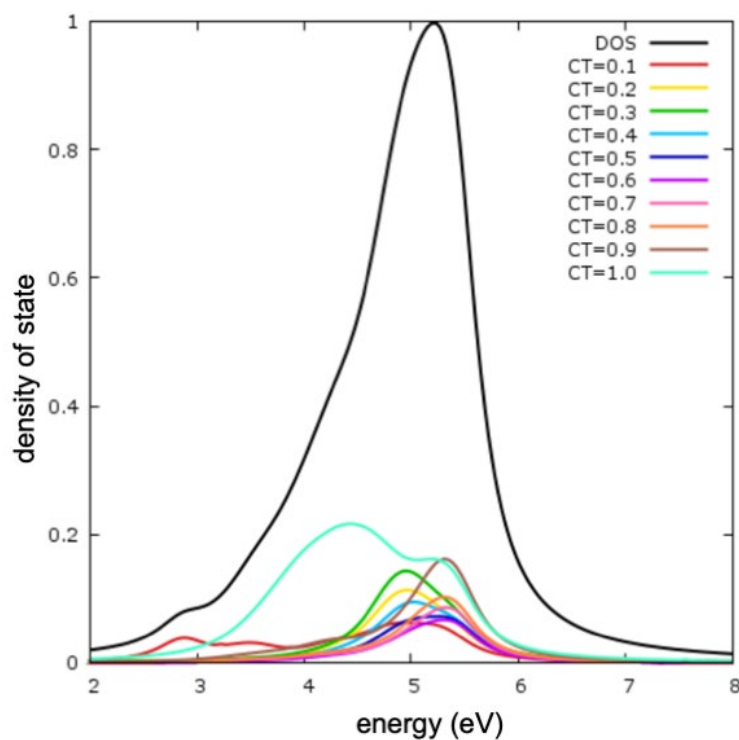


Figure S 2: density of state (DOS) for QMCH:G4 complex and individual decomposition for different charge transfer numbers (CT). The states with CT=1.0 represent G \rightarrow P pure charge transfer states. The states around 5 eV represents state with different CT numbers, both exciton and mixed.

The Natural Transition Orbitals (NTO) are calculated as single value decomposition of the transition density matrix as implemented in TheoDORE.¹⁴ Relevant orbitals are reported in Figure S3.

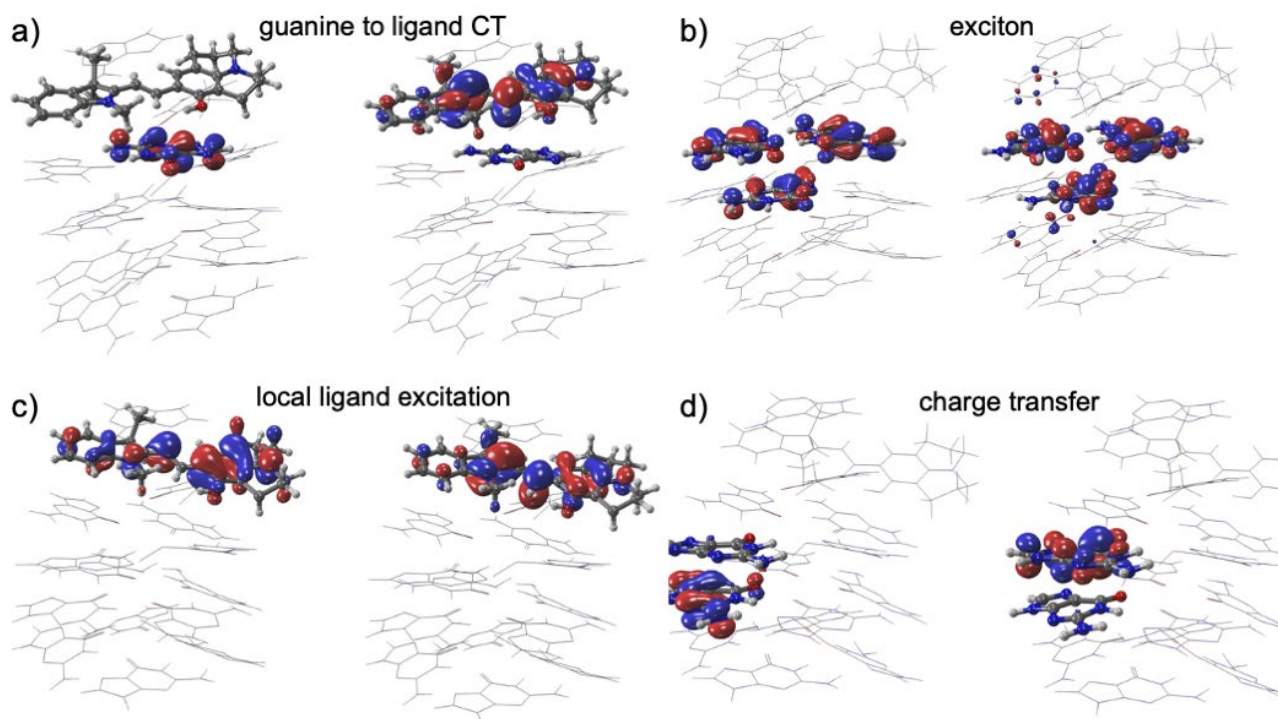


Figure S 3: Representative NTOs for the excitation of different nature found in QMCH:G4 complex

References

- (1) Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. Solution Structure of a 2:1 Quindoline-c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J. Am. Chem. Soc.* **2011**, *133* (44), 17673–17680.
- (2) Avagliano, D.; Sánchez-Murcia, P. A.; González, L. Spiropyran Meets Guanine Quadruplexes: Isomerization Mechanism and DNA Binding Modes of Quinolizidine-substituted Spiropyran Probes. *Chem. – A Eur. J.* **2020**, *26*, 13039–13045.
- (3) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7* (4), 931–948.
- (4) D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T, D. M. Y. and P. A. K. AMBER 2020. University of California: San Francisco 2020.
- (5) M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*; Oxford university press: New York, 1991.
- (6) Casida, M. E.; Huix-Rotllant, M. Progress in Time-Dependent Density-Functional Theory. *Annu. Rev. Phys. Chem.* **2012**, *63*, 287–323.
- (7) Runge, E.; Gross, E. K. I. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.
- (8) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393* (1–3), 51–57.

- (9) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104.
- (10) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305.
- (11) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Generating Efficient Quantum Chemistry Codes for Novel Architectures. *J. Chem. Theory Comput.* **2013**, *9* (1), 213–221.
- (12) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5* (10), 2619–2628.
- (13) Kimber, P.; Plasser, F. Toward an Understanding of Electronic Excitation Energies beyond the Molecular Orbital Picture. *Phys. Chem. Chem. Phys.* **2020**, *22* (11), 6058–6080.
- (14) Plasser, F. TheoDORE. A Package for Theoretical Density, Orbital Relaxation and Exciton Analysis, [Http://Theodore-Qc.Sourceforge.Net](http://Theodore-Qc.Sourceforge.Net). URL: <http://theodore-qc.sourceforge.net>. F. Plasser 2020.

6.5 Directional and regioselective hole injection of spiropyran photoswitches intercalated into A/T-duplex DNA

Davide Avagliano, Pedro A. Sánchez-Murcia, Leticia González

Phys. Chem. Chem. Phys., 2019,21, 17971-17977
<https://doi.org/10.1039/C9CP03398J>

Contributions:

- D. AVAGLIANO: Performed the calculations and the analysis, wrote the first draft of the manuscript and contributed to the final version of the manuscript.
- P.A. SÁNCHEZ-MURCIA: Supervised the development and contributed to the initial and final draft of the manuscript
- L. GONZÁLEZ: conceived the scope of the manuscript, supervised the developments and contributed to the final manuscript

Reproduced from Phys. Chem. Chem. Phys., 2019,21, 17971-17977 with permission from the Royal Society of Chemistry



Cite this: *Phys. Chem. Chem. Phys.*,
2019, 21, 17971

Received 16th June 2019,
Accepted 25th July 2019

DOI: 10.1039/c9cp03398j

rsc.li/pccp

Directional and regioselective hole injection of spiropyran photoswitches intercalated into A/T-duplex DNA†

Davide Avagliano,^{id} Pedro A. Sánchez-Murcia^{id}* and Leticia González^{id}*

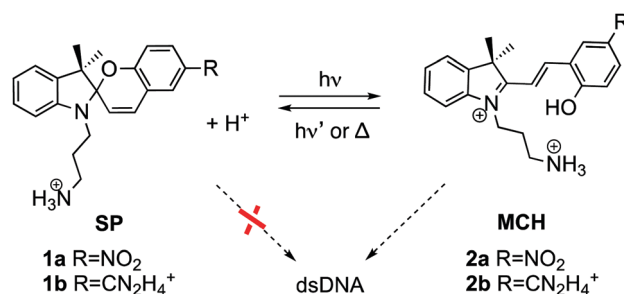
The electron-hole injection from a family of spiropyran photoswitches into A/T-duplex DNA has been investigated at the molecular level for the first time. Multiscale computations coupled with automatized quantitative wavefunction analysis reveal a pronounced directionality and regioselectivity towards the template strand of the duplex DNA. Our findings suggest that this directional and regioselective photoinduced electron-hole transfer could thus be exploited to tailor the charge transport processes in DNA in specific applications.

Light-driven charge transport processes through DNA play a central role in photodamage¹ and are the key to design DNA-based molecular wires.^{2–4} Upon illumination, one electron is excited from a donor unit to an acceptor species, generating a hole – a positive charge – in DNA.⁵ This electron-hole can efficiently migrate long molecular distances through the DNA helix.^{2,6} Given its importance for a wide range of applications, understanding the dynamics of electron transport in DNA has been a subject of study for decades.^{7–9} For instance, it is known that the migration of the electron-hole strongly depends on stacking¹⁰ and on the energies of the involved nucleobases.¹¹ However, despite several experimental¹² and theoretical¹³ setups that have been used to investigate the electron-hole injection and migration in DNA, the characterization of the initial electronic excited states is in most cases unknown.

The donor and acceptor units can be covalently bound^{14–16} or intercalated in DNA,^{17–21} such as the spiropyran (SP) photoswitches. After light irradiation, SPs undergo heterocyclic cleavage to yield the open merocyanine (MC) form.^{22,23} Actually, the SP form (**1** in Scheme 1) does not bind DNA but the protonated open MCH form (**2**) does.^{24,25} While the photocleavage of SP in solution has been widely investigated,^{26–31} and it is known that the protonated open MCH form is able to oxidize DNA nucleobases in cell culture,³² the excited states of these photoswitches intercalated in DNA have never been investigated. This is the subject of this work. We address here the central question of how electron-hole injection operates from

merocyanine derivatives to a duplex A/T DNA strand. We use atomistic multiscale calculations coupled with quantitative wavefunction analysis to model explicitly the hole injection from two MCH derivatives (**2a** and **2b**) into 12-mer (poly-dAT)₂, for which these compounds show selectivity.²⁴

The intercalation mechanism of the nitro (**2a**) and amidinium (**2b**) derivatives into a 12-mer (poly-dAT)₂ is described elsewhere.²⁵ An ensemble of geometries of **2a**:DNA and **2b**:DNA obtained by unrestrained classical and Born Oppenheimer molecular dynamics simulations in an explicit solvent is used to ensure an efficient sampling of the environment and of the vibrational space of the chromophore.^{34,35} A total of 4000 excited states per complex were calculated within a quantum mechanics/molecular mechanics (QM/MM) framework, where the chromophore and the first four surrounding nucleobases (121 atoms for **2a**:DNA and 125 atoms for **2b**:DNA, see Fig. 1a and b) are considered quantum mechanically. Further computational details can be found in the Computational details section.



Scheme 1 Spiropyran (SP) and protonated merocyanine (MCH) derivatives studied in this work bearing a nitro (**2a**)²⁴ or an amidinium (**2b**)³³ group. Only MCH species, and not SP, bind to dsDNA.

Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna,
Währinger Str. 17, A-1090 Vienna, Austria. E-mail: pedro.murcia@univie.ac.at,
leticia.gonzalez@univie.ac.at

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9cp03398j



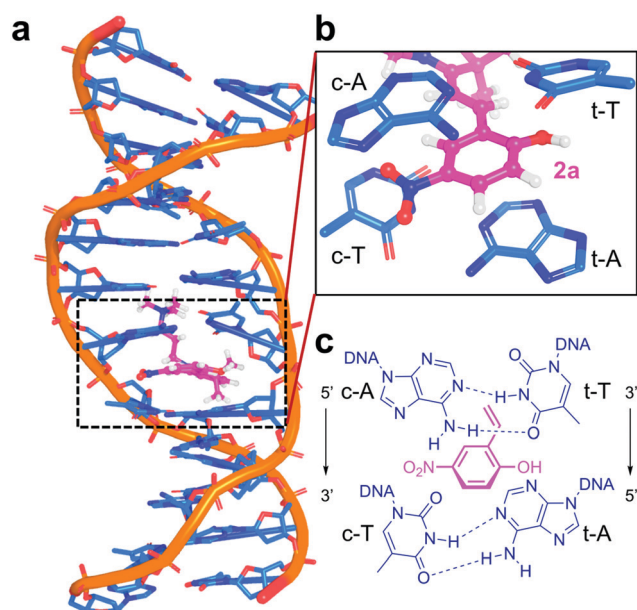


Fig. 1 Complex **2a**:DNA. Color code in 3D representation: P atoms of the DNA backbone in orange, C atoms of nucleobases and **2a** in light blue and pink, respectively, N atoms in dark blue and O atoms in red. (a) Full view, (b) zoomed-in on the intercalative pocket considered quantum mechanically, and (c) 2D projection of panel b.

Our theoretical protocol is validated by the good agreement of the computed absorption spectra (Fig. S1, ESI†) of **2a** in water and **2b** intercalated in DNA with available experimental data.^{33,36} The spectra consist of two absorption bands. The lower-energy band corresponds to the intramolecular excitation from a π orbital (HOMO) to a π^* molecular orbital (LUMO), with absorption maxima at 382 and 398 nm, respectively. The peaks are slightly blue-shifted (<30 nm) with respect to the experimental ones, as usual at this level of theory.³⁴ The brightest states of **2a** and **2b** are the same in solution as bound to the DNA. However, upon DNA intercalation, the brightest state is surrounded by dark states with strong charge transfer (CT) character involving nucleobases, see Table 1 for the excited states of **2a**:DNA and Table S1 (ESI†) for **2b**:DNA.

The CT character, defined by the CT number³⁷ from 0 to 1, measures the intermolecular light-driven electron transfer

between the donor and the acceptor. To define CT numbers, the system is split into five fragments, the chromophore (e.g. **2a**) and the four interacting nucleobases: adenine (c-A) and thymine (c-T) of the coding strand (5' → 3') and the corresponding ones of the template strand (t-A and t-T, 3' → 5'), see Fig. 1c.

The total density of excited states (DOES) of the lowest absorption band of the complex **2a**:DNA (black line of Fig. 2a) calculated from the ensemble of the structures and classified according to their CT character clearly shows that there are a few absorbing states (red line) embedded by many dark states with strong CT character (blue line). Fig. 2b and c illustrate the natural transition orbitals for the brightest state and one representative CT state, respectively. The brightest state corresponds to an intramolecular excitation; the CT state is an intermolecular excitation from the probe to t-A.

The fact that the brightest state of **2a** in solution and in the duplex DNA is the same, indicates that upon intercalation an electron of **2a** is excited creating a hole on the HOMO. This orbital is surrounded by the electron-rich HOMOs of the nucleobases. Therefore, one electron from the orbitals of the nucleobase can relax to the half-occupied HOMO of **2a**, transferring the hole from **2a** to one of the stacked nucleobases of the dsDNA. The hypothesis that **2a** acts as a photooxidant probe of the surrounding nucleobases is supported by its substantial reduction potential $E^0(\mathbf{2a})$ (calculated value 3.94 V, see Computational details) and that the related merocyanine 540 derivative is able to oxidize the DNA nucleobases in cell culture.³²

Finding this high density of CT states for the minimum geometry of the **2a**:DNA complex urged us to investigate the excited electron and electron-hole populations of these CT states within the first ten excited states of the ensemble of geometries (1000 excited states per complex, see Computational details). We found that upon light absorption, 97% of the population of the excited electron is localized on **2a**. Complementarily, 70% of the hole population is found in only one nucleobase and 24% is delocalized in two different fragments. In principle, the electron hole could have been transferred from the probe to any of the four nucleobases since **2a** intercalates in the middle of the site with similar distance to all the nucleobases (Table S2, ESI†). However, the four nucleobases are not chemically equivalent, as they are located in different strands and they are affected by the asymmetrical binding of **2a**, recall Fig. 1b and c. As a consequence, the majority of the electron hole is localized on the template strand (3' → 5'), with 38% and 33% on t-T and t-A, respectively, and 29% on the coding strand (Fig. 3, pink bars). The strand selectivity is ascribed to the presence of the NO₂ group, a strong electron-withdrawing group (EWG), oriented to the coding strand and a hydroxyl group (OH), an electron-donating group (EDG), projected to the template strand. Due to its electronic character, the NO₂ group prevents the hole injection in the nucleobases around it. Moreover, since the differences between the electron-hole populations between t-A and t-T are small, **2a** promotes the electron-hole injection into the template strand, in a static picture, in both directions (5' ↔ 3').

The strand selectivity increases in **2b**:DNA, as the template/coding ratio is 90/10 (Fig. 3a, green bars), compared to **2a** (71/29).

Table 1 Energy in eV, oscillator strength (f), CT number and electron/hole decomposition (in parentheses the fragment of e/h localization) for the first ten excited singlet states of the minimum energy geometry of **2a**:DNA

State	Energy	f	CT number	Electron hole population	Excited electron population
S ₁	2.40	0.004	0.996	0.969 (t-A)	0.953 (2a)
S ₂	2.51	0.004	0.997	0.985 (t-T)	0.958 (2a)
S ₃	2.72	0.001	0.996	0.977 (t-A)	0.956 (2a)
S ₄	2.99	0.001	0.996	0.905 (t-A)	0.925 (2a)
S ₅	3.07	0.000	0.998	0.990 (t-T)	0.964 (2a)
S ₆	3.31	0.004	0.997	0.900 (t-A)	0.927 (2a)
S ₇	3.53	0.696	0.087	0.846 (2a)	0.874 (2a)
S ₈	3.89	0.006	0.987	0.958 (t-T)	0.950 (2a)
S ₉	3.94	0.009	0.928	0.709 (c-A)	0.887 (2a)
S ₁₀	3.97	0.009	0.774	0.503 (c-T)	0.885 (2a)



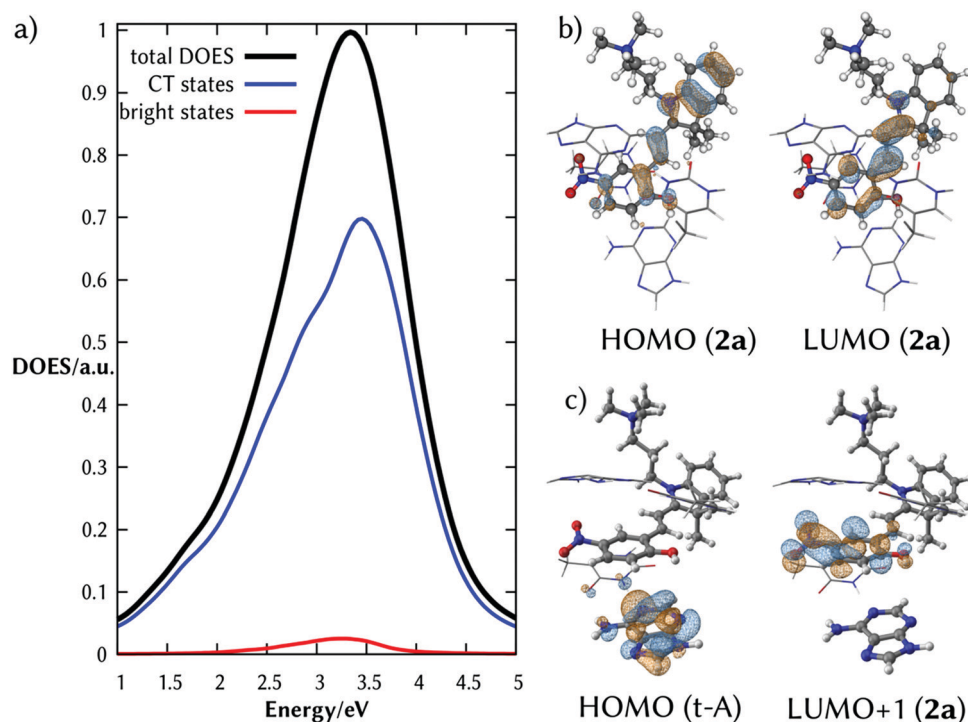


Fig. 2 (a) Total density of excited states (DOES, black line), DOES of excited states with CT > 0.7 (blue line) and DOES of the bright states (red line, $f > 0.1$) of the complex **2a**:DNA upon light absorption. (b) Natural transition orbitals (NTOs) involved in the intramolecular $\pi\pi^*$ excitation in **2a** of the minimum energy structure. They correspond to the HOMO to the LUMO of **2a**. (c) NTOs involved in a representative CT state, which correspond to the HOMO of a nucleobase (t-A) and the LUMO+1 of **2a**.

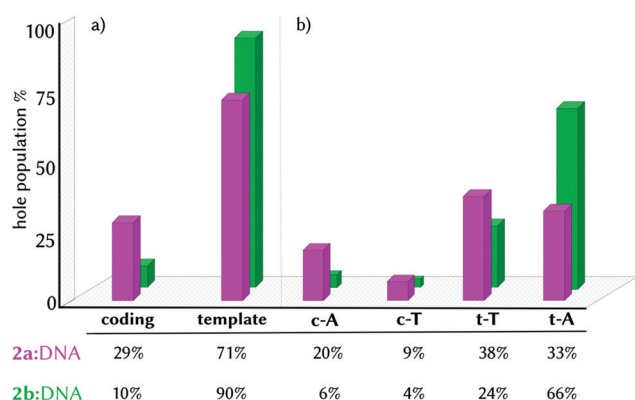


Fig. 3 Histogram representation of electron-hole localization percentage in the CT states group by strand (a) and by single nucleobases (b) in the complexes **2a**:DNA (pink) and **2b**:DNA (green).

Of greater importance is the fact that in the complex **2b**:DNA, the electron-hole injection goes preferentially into t-A (66%) instead of t-T (24%). This observation suggests that the electron-hole injection into the template strand propagates unidirectionally ($3' \rightarrow 5'$) in the presence of **2b**.

A possible explanation for such discrimination between t-T and t-A could be found in geometrical differences between the two complexes, which would promote the electron-hole injection unidirectionally in the case of **2b**. In our previous work²⁵ and in Fig. S3 (ESI[†]), we showed that the two complexes shared the same intercalative binding site and the same average

distance differences between the two nucleobases of the template strand and the probe for both complexes (Table S2, ESI[†]). These observations led us to exclude geometrical reasons as a discriminant for the regioselectivity and to investigate other possible reasons. The origin of the structural and electronic differences that regulate the electron-hole injection directionality into DNA can be traced back to the interactions between the π -systems of the chromophore and those of the neighbouring nucleobases. These interactions can be assessed in the minimum energy geometries with two descriptors. One is the non-covalent interaction energy between the probe and the nucleobases of each of the strands or π - π stacking interactions (Table 2), calculated *via* the Grimme's dispersion energy correction ($D3$).^{38,39} The other descriptor is the energy difference between the HOMO of the probe and the HOMO of each of the nucleobases forming the binding site (Table 3 and Fig. S2, ESI[†]).¹¹

Table 2 Computed dispersion correction energy ($D3$, kcal mol⁻¹) for the interaction between **2a** and **2b** and the coding and template strands, respectively, in their minimum energy geometries. $\Delta D3$ is the dispersion energy difference between the two strands ($\Delta D3 = D3_{\text{coding}} - D3_{\text{template}}$)

Complex	$D3$ dispersion energy (kcal mol ⁻¹)		$\Delta D3$
	Coding strand ($3' \rightarrow 5'$)	Template strand ($5' \rightarrow 3'$)	
2a:DNA	-75.49	-63.04	-12.44
2b:DNA	-79.59	-72.28	-6.81

Table 3 Difference energies ($\Delta\epsilon$, eV) between the HOMOs of the probe (**2a** and **2b**) and the surrounding nucleobases on the template (t-A, t-T) and the coding strand (c-A, c-T)

Nucleobase	$\Delta\epsilon^{\text{HOMO}}$ (eV)	
	2a :DNA	2b :DNA
t-A	+1.82	+0.96
t-T	+1.72	+0.24
c-A	−0.10	−1.24
c-T	−0.37	−2.16

The presence of the strong nitro EWG (**2a**) or amidinium (**2b**) EWGs exerts a strong interaction with the nucleobases of the coding strand ($-75.49 \text{ kcal mol}^{-1}$ for **2a** and $-79.59 \text{ kcal mol}^{-1}$ for **2b**, Table 2). As a consequence, the probe induced an energy split of the HOMOs of the nucleobases. In particular, the nucleobases of the coding strand are lower in energy than the HOMO of the probes (Table 3). In contrast, the interaction with the template strand is weaker ($-63.04 \text{ kcal mol}^{-1}$ for **2a** and $-72.28 \text{ kcal mol}^{-1}$ for **2b**), and thus, the HOMOs of t-A and t-T are lying higher in energy than the HOMOs of the probes (Table 3). This explains why most of the electron-hole population is found on the template strand. Upon light absorption by the photoprobe, one electron from the HOMO of the template strand relaxes in energy and occupies the HOMO of the probe, injecting the hole into the template strand. Remarkably, the derivative **2a** shows a stronger interaction with the coding strand than with the template one ($\Delta E = -12.44 \text{ kcal mol}^{-1}$, Table 2), while in **2b**, this difference is half ($\Delta E = -6.81 \text{ kcal mol}^{-1}$). This means that the HOMOs of t-A and t-T, although higher in energy, are more stabilized by **2b** due to favourable π - π stacking interactions. In addition, whereas in **2a** the HOMOs of t-A and t-T are close in energy ($\Delta\epsilon^{\text{HOMO}} = +1.82$ and $+1.72$, respectively), in **2b** t-T is much more stabilized than t-A. This is why most of the total electron-hole population (66%) is found on t-A (Fig. 3b, green bar)—the nucleobase with the higher HOMO level.

We are now in the position to propose a mechanistic model for the electron-hole injection in dsDNA by MCH derivatives containing EWGs, such as **2a** and **2b** (Fig. 4). It is the combination of π - π stacking interactions and the presence of an EWG that stabilizes

the HOMOs of the nucleobases differently at the binding site (Fig. 4a). Upon irradiation by UV light, the brightest excited state is populated, which can decay to one of the lower-lying dark CT states. That is, one electron from the π -system of the probe (HOMO) is promoted to an excited state with π^* character fully localized on the probe, creating a hole within the intercalated photoprobe (circle, Fig. 4b). The proximity in the energy of the HOMOs of t-A and t-T allows the migration of one electron of the nucleobases to MCH. The probe oxidizes thus the neighbouring nucleobases, injecting the hole into the DNA (arrow, Fig. 4c) and triggering hole migration through the double strand. This process is directional because the probe oxidizes the DNA, injecting an electron hole; is asymmetric because the binding mode of the open merocyanin species projecting the EWG to the coding strand promotes the hole injection into the template strand; and is regioselective because the nature of this EWG affects the energy levels of the HOMOs of the nucleobases of the template strand. As an example, in **2b** the hole injection happens mainly into t-A, allowing a $5' \rightarrow 3'$ electron-hole propagation.

In conclusion, we have portrayed how chemical modification of spiropyrans can modulate the directionality of the hole transport in DNA, arguably offering many application prospects. The quantitative direct observation of the CT states between a photooxidant and nucleobases, where the ligand is intercalated, has no precedent in the study of spiropyran photoswitches and evidences the importance of characterizing their excited states in order to prevent or enhance the photoinduced process involving DNA. Our findings open new questions on how the temporal evolution of these excited states is influenced by this selectivity and its biological implications. To answer these questions, further theoretical and experimental studies on the photo-dynamics of the injected electron hole are necessary.

Computational details

MD simulations

The initial structures of both probes **2a** and **2b** intercalated into a 12-mer dsDNA (poly-dAT)₂ were obtained from umbrella sampling MD simulation studies, as described elsewhere.²⁵

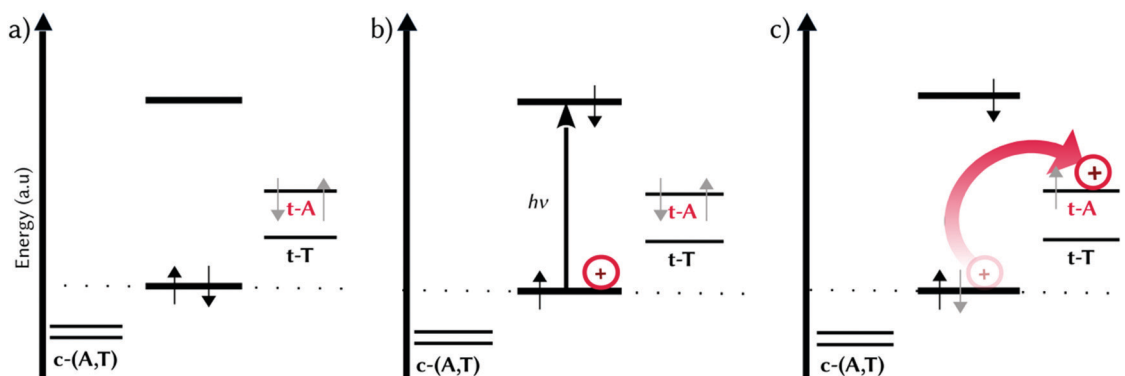


Fig. 4 Schematic mechanistic model proposed for the electron-hole injection into dsDNA by **2b**. Each line represents the frontier orbitals (HOMO and LUMO of the probe, HOMOs of the nucleobases) localized on a single fragment. (a) HOMO splitting induced by π - π stacking interactions and an electron-withdrawing group; (b) UV light absorption and hole creation on the probe; (c) hole injection from the probe to the t-A nucleobase.



In all cases, each complex was immersed in a cubic box of 30 Å from the solute to the border of the box filled with TIP3P water molecules⁴⁰ 20 (2a) and 19 (2b) Na⁺ to ensure electroneutrality. To reproduce the experimental conditions reported by Andersson *et al.*,²⁴ a final NaCl concentration of 1×10^{-5} M was achieved by the addition of Na⁺ and Cl⁻ atoms. Each of the systems was simulated for 100 ns without any restraint following the protocol described in ref. 25

QM/MM MD simulations

A total of 100 equidistant snapshots from the former MD simulations were selected to carry out QM/MM MD simulations using the sander program implemented in the AMBER17 suite.⁴¹ The system was partitioned in two regions: the QM and the MM region. Each of the probes (2a or 2b) and the four nucleobases around them (after cutting the glycosidic bond) were included in the QM region (121 and 125 atoms, respectively). The rest of the atoms were treated classically using the ff14SB⁴² force field and the TIP3P⁴⁰ model for the water molecules. The QM region was treated with density-functional tight-binding (version 3, DFTB3),⁴³ the semiempirical method of DFT, which is internally provided within the AMBER17 suite. The interaction term between the QM and the MM regions was calculated using the electrostatic embedding scheme.⁴⁴ In the MM part, periodic boundary conditions were used and the electrostatic interactions were computed using the Ewald method⁴⁵ with a grid spacing of 1 Å. The cutoff distance for the non-bonded interactions was 10 Å and the SHAKE algorithm⁴⁶ was applied to all bonds involving hydrogen atoms. An integration step of 2.0 fs was defined. In the QM region, the PME and SHAKE algorithms were deactivated and a cutoff of 10 Å was defined for the interaction between the two regions. Each of the 100 QM/MM MD trajectories was propagated for 1 ps, with a time step of 0.1 fs at 300 K and 1 atm. In this way, the QM/MM MD sampling included the quantum mechanical effects in the phase space sampling, obtaining more accurate initial conditions for the excited state calculations.

Static TD-DFT vertical excitations

The final geometries from the QM/MM MD simulations were used to compute the first 40 singlet states using time-dependent density functional theory (TD-DFT). As above, the QM region included the probe and the four surrounding nucleobases (c-T, c-A, t-A and t-T), in this case treated with the long-range corrected functional CAM-B3LYP⁴⁷ and the def2-svp⁴⁸ basis set. Grimme's dispersion correction D3 was considered.^{38,39} The rest of the system was printed as MM point charges and they were included in the Hamiltonian by means of electrostatic embedding. These calculations were performed with the TeraChem code^{49,50} on GeForce Nvidia GTX 1080Ti GPUs.

Wavefunction analysis

The quantitative wavefunction analysis was performed in a second step after calculating the vertical excitations and the corresponding orbitals in TeraChem. Such an analysis was possible using the TheoDOR software.^{37,51} Detailed information

can be found on the documentation of TheoDOR;⁵¹ here we only summarize the features employed.

Vis/UV spectra

The Vis/UV spectra were convoluted as a sum of Gaussian functions (eqn (1)):

$$\sigma(E) = \sum_g \sum_s^{\text{geom state}} f_{gs} \exp(-4 \ln(2)(E - E_{gs})^2 (\text{FWMH})^{-2}) \quad (1)$$

where f_{gs} is the oscillator strength in the ground state, E and E_{gs} are the energies in the excited state and in the ground state, respectively, and FWMH is the full width at half maximum. A value of 0.5 eV was used for FWMH.

Transition density matrix, charge transfer numbers and natural transition orbitals

TheoDOR relies on the transition density matrix (D^{OI}) analysis,⁵² computed as shown in eqn (2). Briefly, considering the states I and J and the orbitals α and β , one element of the one-particle transition density matrix is given by:

$$D_{\alpha\beta}^{OI} = \langle \Psi^I | \hat{a}_\alpha^\dagger \hat{a}_\beta | \Psi^J \rangle \quad (2)$$

where \hat{a}_α^\dagger and \hat{a}_β are the creation and annihilation operators, respectively.

For the charge transfer (CT) analysis, the system was divided into five fragments: the probe and each of the four nucleobases. The CT numbers were computed using the Mulliken-like population analysis:

$$\Omega_{AB} = \sum_{\mu \in A} \sum_{\nu \in A} (D^{OI} S)_{\mu\nu} (S D^{OI})_{\mu\nu} \quad (3)$$

where A and B are two different fragments, μ and ν are atomic orbitals, D^{OI} is the transition density matrix and S is the overlap matrix, both matrices expressed in atomic orbital basis.

Natural transition orbitals (NTOs)⁵³

NTO is built through a singular value decomposition of the D^{OI} given by:

$$D^{On} = UXV^\dagger \quad (4)$$

where U is the hole orbital coefficients matrix, V is the particle orbital coefficients matrix and X is the diagonal matrix of the transition amplitudes.

Redox potential calculation

The standard redox potential of the probe in a DNA environment is calculated using the Born–Haber cycle⁵⁴ and the Nernst equation, based on the Gibbs free energy difference in gas and solvated environments for both the reduced and oxidized species. The oxidized (closed-shell, singlet, net charge = +2) and the reduced (open shell, doublet, net charge = +1) species were optimized at the CAM-B3LYP/def2-SVP level of theory. With this, we obtained the value of G_{gas} . Then, the geometries of both species were solvated with an acetonitrile shell ($\epsilon = 35.688$), which is known to reproduce a more similar DNA-like environment



than water,⁵⁵ and the free energies (G_{solv}°) were calculated with the polarizable continuum model (PCM).⁵⁶ The standard redox potential E° was then calculated from the Gibbs free energy change $\Delta G_{\text{red(solv)}}^{\circ}$ as shown in eqn (5) and (6):

$$\Delta G_{\text{red(solv)}}^{\circ, \text{redox}} = \Delta G_{\text{(g)}}^{\circ, \text{redox}} + \Delta G_{\text{(solv)}}^{\circ, \text{(red)}} - \Delta G_{\text{(solv)}}^{\circ, \text{(ox)}} \quad (5)$$

$$E^{\circ} = \frac{\Delta G_{\text{red(solv)}}^{\circ, \text{redox}}}{-nF} \quad (6)$$

These calculations are performed with Gaussian 09, version D.01.⁵⁷

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

D. A. acknowledges funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765266 (LightDyNAMics). P. A. S.-M. would like to thank the Austrian FWF, Project M 2260. The authors thank the Vienna Scientific Cluster (VSC) for the allocation of computational resources.

References

- 1 D. B. Hall, R. E. Holmlin and J. K. Barton, *Nature*, 1996, **382**, 731–735.
- 2 J. D. Slinker, N. B. Muren, S. E. Renfrew and J. K. Barton, *Nat. Chem.*, 2011, **3**, 228.
- 3 C. H. Wohlgamuth, M. A. McWilliams and J. D. Slinker, *Anal. Chem.*, 2013, **85**, 8634–8640.
- 4 A. R. Chandrasekaran, *ChemBioChem*, 2017, **18**, 1886–1887.
- 5 H. Wagenknecht, *Charge Transfer in DNA: From Mechanism to Application*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2005.
- 6 B. Giese, *Acc. Chem. Res.*, 2000, **33**, 631–636.
- 7 K. Kawai and T. Majima, *Acc. Chem. Res.*, 2013, **46**, 2616–2625.
- 8 F. D. Lewis, R. M. Young and M. R. Wasielewski, *Acc. Chem. Res.*, 2018, **51**, 1746–1754.
- 9 J. C. Genereux and J. K. Barton, *Chem. Rev.*, 2010, **110**, 1642–1662.
- 10 A. J. A. Aquino, D. Nachtigallova, P. Hobza, D. G. Truhlar, C. Hättig and H. Lischka, *J. Comput. Chem.*, 2011, **32**, 1217–1227.
- 11 K. Kawai, M. Hayashi and T. Majima, *J. Am. Chem. Soc.*, 2012, **134**, 4806–4811.
- 12 I. Kratochvílová, T. Todorciuc, K. Král, H. Němec, M. Bunčák, J. Šebera, S. Zálšíš, Z. Vokáčová, V. Sychrovský, L. Bednářová, P. Mojžeš and B. Schneider, *J. Phys. Chem. B*, 2010, **114**, 5196–5205.
- 13 D. Rawtani, B. Kuntmal and Y. Agrawal, *Front. Life Sci.*, 2016, **9**, 214–225.
- 14 N. Renaud, M. A. Harris, A. P. N. Singh, Y. A. Berlin, M. A. Ratner, M. R. Wasielewski, F. D. Lewis and F. C. Grozema, *Nat. Chem.*, 2016, **8**, 1015.
- 15 F. D. Lewis, X. Liu, Y. Wu and R. L. Letsinger, *J. Am. Chem. Soc.*, 1999, **121**, 9905–9906.
- 16 K. E. Brown, A. P. N. Singh, Y.-L. Wu, A. K. Mishra, J. Zhou, F. D. Lewis, R. M. Young and M. R. Wasielewski, *J. Am. Chem. Soc.*, 2017, **139**, 12084–12092.
- 17 A. Jacques, A. Kirsch-De Mesmaeker and B. Elias, *Inorg. Chem.*, 2014, **53**, 1507–1512.
- 18 C. Murphy, M. Arkin, Y. Jenkins, N. Ghatlia, S. Bossmann, N. Turro and J. Barton, *Science*, 1993, **262**, 1025–1029.
- 19 G. B. Schuster, *Acc. Chem. Res.*, 2000, **33**, 253–260.
- 20 F. Shao, K. Augustyn and J. K. Barton, *J. Am. Chem. Soc.*, 2005, **127**, 17445–17452.
- 21 F. Shao and J. K. Barton, *J. Am. Chem. Soc.*, 2007, **129**, 14733–14738.
- 22 C. L. Fleming, S. Li, M. Grötlí and J. Andréasson, *J. Am. Chem. Soc.*, 2018, **140**, 14069–14072.
- 23 J. Kohl-Landgraf, M. Braun, C. Özçoban, D. P. N. Gonçalves, A. Heckel and J. Wachtveitl, *J. Am. Chem. Soc.*, 2012, **134**, 14070–14077.
- 24 J. Andersson, S. Li, P. Lincoln and J. Andréasson, *J. Am. Chem. Soc.*, 2008, **130**, 11836–11837.
- 25 D. Avagliano, P. A. Sánchez-Murcia and L. González, *Phys. Chem. Chem. Phys.*, 2019, **21**, 8601–8966.
- 26 I. Gómez, M. Reguero and M. A. Robb, *J. Phys. Chem. A*, 2006, **110**, 3986–3991.
- 27 A. Eilmes, *J. Phys. Chem. A*, 2013, **117**, 2629–2635.
- 28 Y. Kalisky, T. E. Orlowski and D. J. Williams, *J. Phys. Chem.*, 1983, **87**, 5333–5338.
- 29 C. J. Wohl and D. Kuciauskas, *J. Phys. Chem. B*, 2005, **109**, 22186–22191.
- 30 N. A. Murugan, S. Chakrabarti and H. Ågren, *J. Phys. Chem. B*, 2011, **115**, 4025–4032.
- 31 R. A. Rogers, A. R. Rodier, J. A. Stanley, N. A. Douglas, X. Li and W. J. Brittain, *Chem. Commun.*, 2014, **50**, 3424–3426.
- 32 M. Dušínská, A. Mateášik, D. J. Chorvát, I. Lajdová, V. Spustová and A. Chorvátová, *Neoplasma*, 1999, **46**, 69–70.
- 33 M. Hammarson, J. R. Nilsson, S. Li, P. Lincoln and J. Andréasson, *Chem. – Eur. J.*, 2014, **20**, 15855–15862.
- 34 J. J. Nogueira, F. Plasser and L. González, *Chem. Sci.*, 2017, **8**, 5682–5691.
- 35 J. J. Nogueira and L. González, *Annu. Rev. Phys. Chem.*, 2018, **69**, 473–497.
- 36 M. Hammarson, J. R. Nilsson, S. Li, T. Beke-Somfai and J. Andréasson, *J. Phys. Chem. B*, 2013, **117**, 13561–13571.
- 37 F. Plasser and H. Lischka, *J. Chem. Theory Comput.*, 2012, **8**, 2777–2789.
- 38 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.
- 39 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 40 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.



- 41 I. Y. B.-S. D. A. Case, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER*, University of California, 2018.
- 42 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
- 43 M. Gaus, Q. Cui and M. Elstner, *J. Chem. Theory Comput.*, 2011, **7**, 931–948.
- 44 A. W. Götz, M. A. Clark and R. C. Walker, *J. Comput. Chem.*, 2014, **35**, 95–108.
- 45 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
- 46 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 47 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.
- 48 A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571–2577.
- 49 I. S. Ufimtsev and T. J. Martinez, *J. Chem. Theory Comput.*, 2009, **5**, 2619–2628.
- 50 *TeraChem*, v. 1.9, PetaChem LLC, 2015.
- 51 F. Plasser, *TheoDORE, A package for Theoretical Density, Orbital Relaxation and Exciton analysis*, 2019.
- 52 F. Plasser, M. Wormit and A. Dreuw, *J. Chem. Phys.*, 2014, **141**, 024106.
- 53 F. Plasser, *J. Chem. Phys.*, 2016, **144**, 194107.
- 54 D. F. C. Morris and E. L. Short, *Nature*, 1969, **224**, 950–952.
- 55 C. A. M. Seidel, A. Schulz and M. H. M. Sauer, *J. Phys. Chem.*, 1996, **100**, 5541–5553.
- 56 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3094.
- 57 G. W. T. M. J. Frisch, 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 and D. J. Fox, *Gaussian 09, Revision D.01*, Gaussian, Inc., Wallingford CT, 2016.



SUPPORTING INFORMATION

Directional and Regioselective Hole injection in A/T-rich DNA Induced by Intercalated Spiropyran Photoswitches

Davide Avagliano, Pedro A. Sánchez-Murcia* and Leticia González*

Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Währinger Straße
17, A-1090 Vienna (Austria)

* leticia.gonzalez@univie.ac.at

* pedro.murcia@univie.ac.at

Contents

Figures S1-S2	Experimental vs computed UV/Vis spectra
Figure S3	Representation of the HOMOs of 2a bound to dsDNA and the surrounding nucleobases
Figure S4	Superimposition of complex 2a :DNA and 2b :DNA in their minimum energy geometries
Table S1	Analysis of the excited singlet states of the complex 2b :dsDNA
Table S2	Center of mass distances
Table S3	HOMOs energy of A-T without intercalative probe
References	

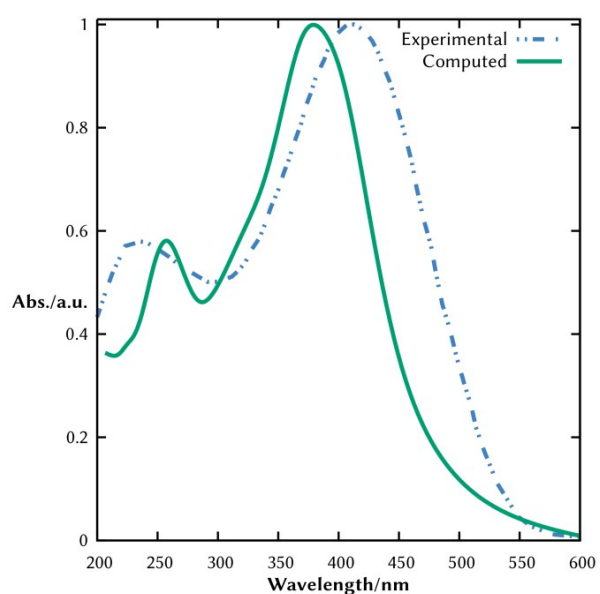


Figure S1. Comparison between computed and experimental spectra of **2a** in water.¹

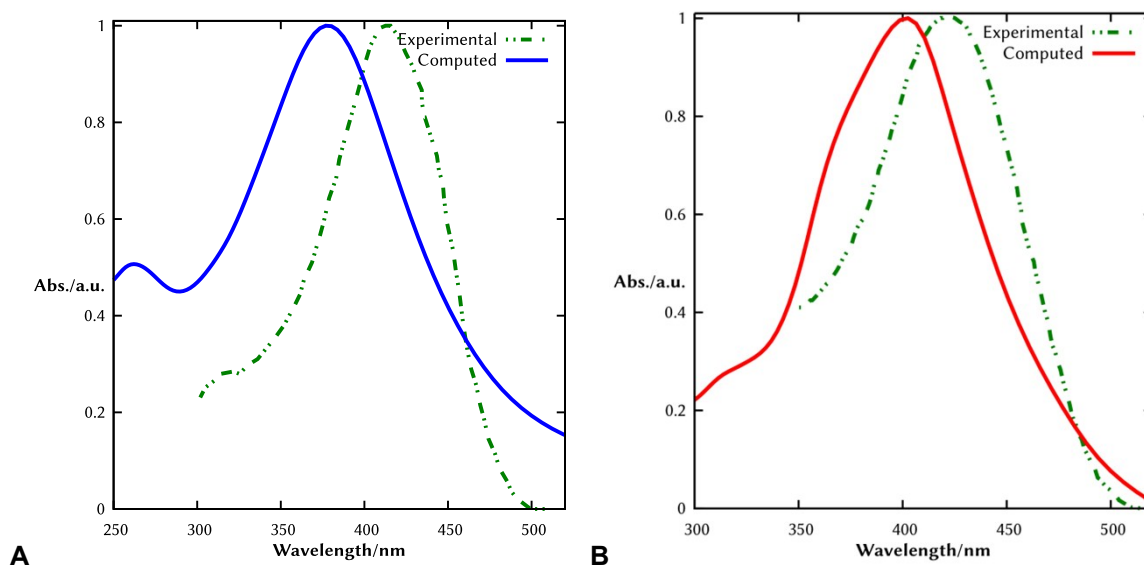


Figure S2 Comparison between computed and experimental spectra of **2b** in water² (**A**) and DNA environment² (**B**). We reproduced the lower energy absorption band in the limit of error of our method and the small red shift induced by DNA embedding.

Table S1. Energy (eV), oscillator strength (f), CT number and excited electron/hole population (in parenthesis the fragment of e/h localization) for the first five excited singlet states of the minimum energy geometry of **2b**:DNA.

State	Energy	f	CT number	electron hole	excited electron
S₁	2.79	0.012	0.980	0.956 (t-A)	0.897 (2a)
S₂	3.20	0.002	0.997	0.956 (t-A)	0.896 (c-T)
S₃	3.24	0.907	0.057	0.877 (2b)	0.980 (2b)
S₄	3.67	0.009	0.991	0.943 (t-T)	0.924 (2a)
S₅	3.71	0.004	0.982	0.935 (t-A)	0.868 (2a)

Table S2. Average distances (Å) between the center of mass (COM) of probes **2a** and **2b** and the COM of the four surrounding nucleobases on the coding (c-T, c-A) and on the template strand (t-T, t-A) along the QM/MM MD simulations. Distances computed with the *cpptraj* program of AMBER suite.

	distance between COMs (Å)			
	c-A	c-T	t-A	t-T
2a :DNA	4.12	5.16	4.13	4.97
2b :DNA	4.31	5.83	4.24	5.09

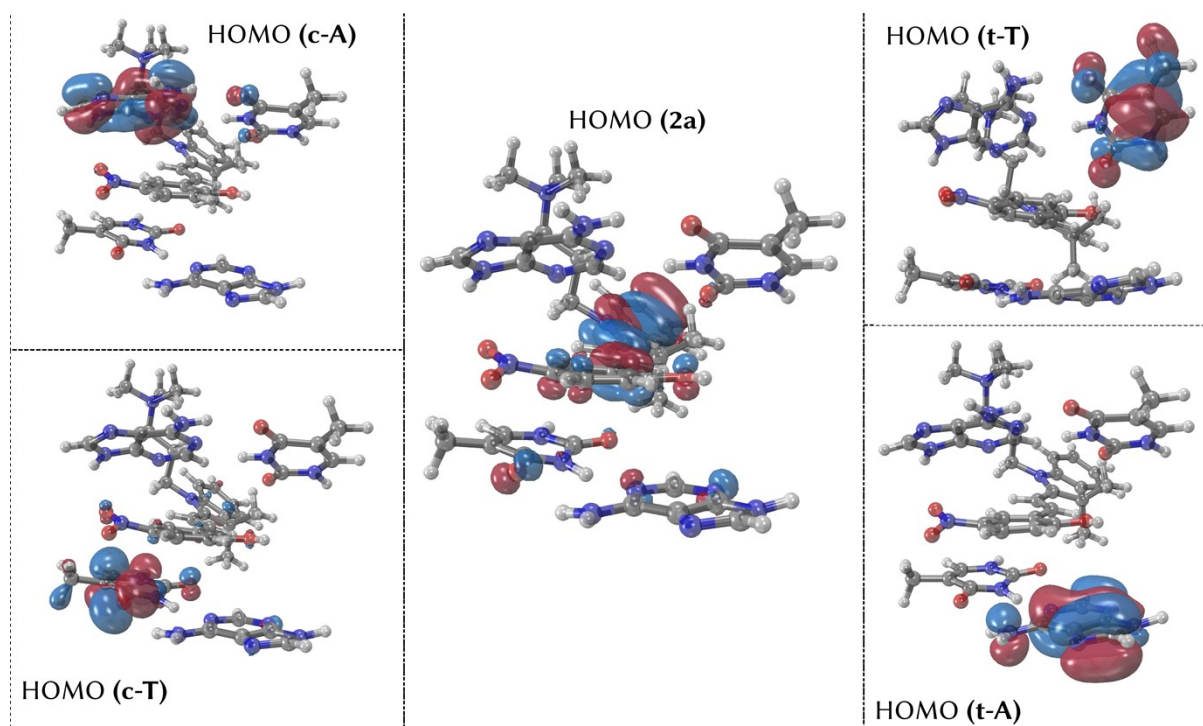


Figure S3. HOMO of the photoprobe **2a** and the four surrounding nucleobases c-A, c-T (left, coding strand) and t-T and t-A (right, template strand) for the minimum energy geometry of the complex **2a**:DNA.

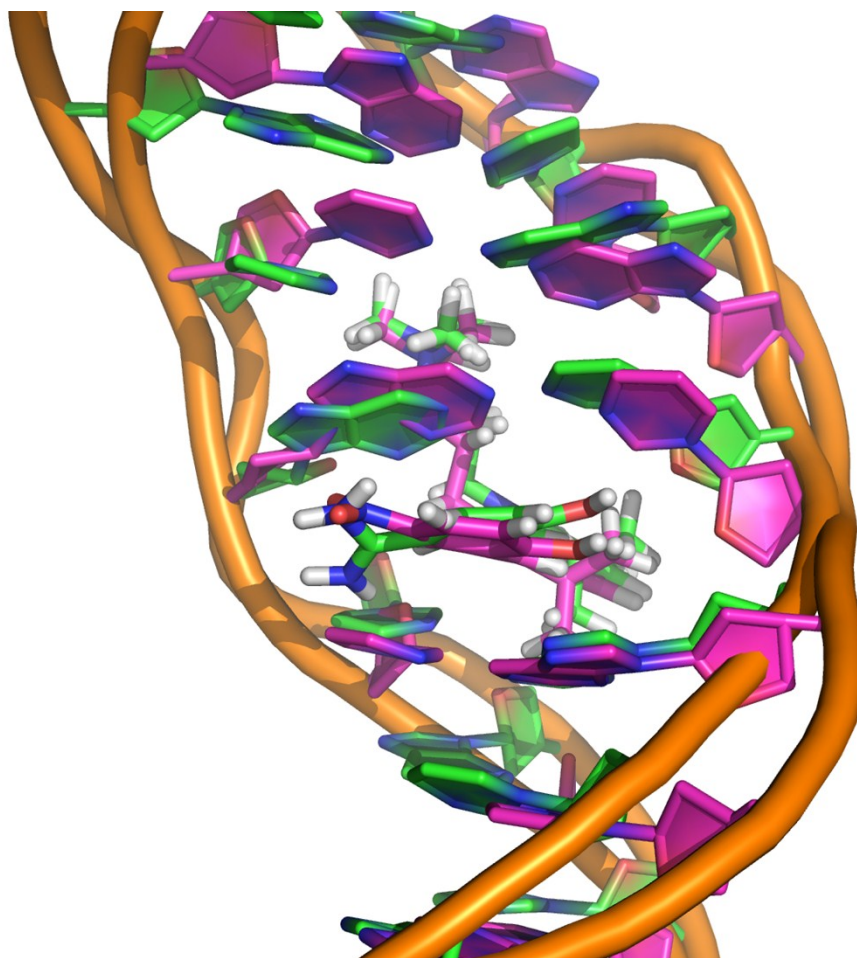


Figure S4. Superimposition of complexes **2a**:DNA (C-atoms colored in pink) and **2b**:DNA (C-atoms colored in green) in their minimum energy geometries.

Table S3. HOMO energies (eV) and electronic delocalization on the c-A, t-A, c-T and t-T of the A-T nucleobases of the DNA duplex in the absence of the intercalative probes **2a** and **2b**. The representative structures were obtained by means of MD simulations and the values were calculated at the CAM-B3LYP/def2-SVP level.

Orbital	$\Delta\epsilon$ (eV)	Delocalization
HOMO	0.00	c-A, t-A
HOMO-1	-0.34	t-A, c-A
HOMO-2	-0.76	t-T, c-T
HOMO-3	-0.78	c-T, t-T

References

1. M. Hammarson, J. R. Nilsson, S. Li, T. Beke-Somfai and J. Andréasson, *J. Phys. Chem. B*, 2013, 117, 13561-13571.
2. M. Hammarson, J. R. Nilsson, S. Li, P. Lincoln and J. Andréasson, *Chem. Eur. J.*, 2014, 20, 15855-15862.

Bibliography

- ¹R. E. Franklin and R. G. Gosling, “Molecular Configuration in Sodium Thymonucleate”, *Nature* **171**, 740–741 (1953).
- ²R. E. Franklin and R. G. Gosling, “The structure of sodium thymonucleate fibres. I. The influence of water content”, *Acta Crystallographica* **6**, 673–677 (1953).
- ³R. E. Franklin and R. G. Gosling, “The structure of sodium thymonucleate fibres. II. The cylindrically symmetrical Patterson function”, *Acta Crystallographica* **6**, 678–685 (1953).
- ⁴R. E. Franklin and R. G. Gosling, “Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate”, *Nature* **172**, 156–157 (1953).
- ⁵Y. Ke, C. Castro, and J. H. Choi, “Structural dna nanotechnology: artificial nanostructures for biomedical research”, *Annual Review of Biomedical Engineering* **20**, 375–401 (2018).
- ⁶T. B. Singh, N. S. Sariciftci, and J. G. Grote, “Bio-organic optoelectronic devices using dna”, in *Organic electronics*, edited by T. Grasser, G. Meller, and L. Li (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010), pp. 73–112.
- ⁷Y. Chen, B. Groves, R. Muscat, and G. Seelig, “DNA nanotechnology from the test tube to the cell”, *Nature Nanotech* **10**, 74–760 (2015).
- ⁸M. Bochman, K. Paeschke, and V. Zakian, “DNA secondary structures: stability and function of G-quadruplex structures”, *Nature Reviews genetics* **417**, 876–880 (2002).
- ⁹S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, and S. Neidle, “Quadruplex DNA: sequence, topology and structure”, *Nucleic Acids Research* **34**, 5402–5415 (2006).
- ¹⁰G. Prkinson, M. Lee, and S. Neidle, “Crystal structure of parallel quadruplexes from human telomeric DNA”, *Nature* **417**, 876–880 (2002).
- ¹¹T. Simonsson, M. Kubista, and P. Pecinka, “DNA tetraplex formation in the control region of c-myc”, *Nucleic Acids Research* **26**, 1167–1172 (1998).
- ¹²E. Ruggiero and S. N. Richter, “G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy”, *Nucleic Acids Research* **46**, 3270–3283 (2018).
- ¹³J. B. Chaires, “Drug–dna interactions”, *Current Opinion in Structural Biology* **8**, 314–320 (1998).
- ¹⁴P. L. Hamilton and D. P. Arya, “Natural product dna major groove binders”, *Nat. Prod. Rep.* **29**, 134–143 (2012).
- ¹⁵P. G. Baraldi, A. Bovero, F. Fruttarolo, D. Preti, M. A. Tabrizi, M. G. Pavani, and R. Romagnoli, “Dna minor groove binders as potential antitumor and antimicrobial agents”, *Medicinal Research Reviews* **24**, 475–528 (2004).

- ¹⁶W. D. Wilson and R. L. Jones, "Intercalating drugs: dna binding and molecular pharmacology", in, Vol. 18, edited by S. Garattini, A. Goldin, F. Hawking, I. Kopin, and R. Schnitzer, *Advances in Pharmacology* (Academic Press, 1981), pp. 177–222.
- ¹⁷K. Wójcik, M. Zarębski, A. Cossarizza, and J. W. Dobrucki, "Daunomycin, an antitumor dna intercalator, influences histone–dna interactions", *Cancer Biology & Therapy* **14**, 823–832 (2013).
- ¹⁸O. Tacar, P. Sriamornsak, and C. R. Dass, "Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems", *Journal of Pharmacy and Pharmacology* **65**, 157–170 (2012).
- ¹⁹H.-K. Liu and P. J. Sadler, "Metal complexes as dna intercalators", *Accounts of Chemical Research* **44**, 349–359 (2011).
- ²⁰G. Bischoff and S. Hoffmann, "Dna-binding of drugs used in medicinal therapies", *Current Medicinal Chemistry* **9**, 321–348 (2002).
- ²¹F. HW. and S. C., "Electrical conduction through DNA molecules", *Nature* **398**, 407–410 (1999).
- ²²T. T. Williams and J. K. Barton, "Charge transport in dna", in *Small molecule dna and rna binders* (John Wiley & Sons, Ltd, 2002) Chap. 7, pp. 146–177.
- ²³W. Bochu, T. Jun, and Z. Liancai, "Selective binding of small molecules to dna: application and perspectives", *Colloids and Surfaces B: Biointerfaces* **79**, 1–4 (2010).
- ²⁴P. Alberti, M. Hoarau, L. Guittat, M. Takasugi, P. B. Arimondo, L. Lacroix, M. Mills, M.-P. Teulade-Fichou, J.-P. Vigneron, J.-M. Lehn, P. Mailliet, and J.-L. Mergny, "Triplex- versus quadruplex-specific ligands and telomerase inhibition", in *Small molecule dna and binders* (John Wiley & Sons, Ltd, 2002) Chap. 12, pp. 315–336.
- ²⁵E. M. Arpa, M. M. Brister, S. J. Hoehn, C. E. Crespo-Hernández, and I. Corral, "On the origin of the photostability of dna and rna monomers: excited state relaxation mechanism of the pyrimidine chromophore", *The Journal of Physical Chemistry Letters* **11**, 5156–5161 (2020).
- ²⁶C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández, and B. Kohler, "Dna excited-state dynamics: from single bases to the double helix", *Annual Review of Physical Chemistry* **60**, 217–239 (2009).
- ²⁷C. E. Crespo-Hernández, B. Cohen, P. M. Hare, and B. Kohler, "Ultrafast excited-state dynamics in nucleic acids", *Chemical Reviews* **104**, 1977–2020 (2004).
- ²⁸W. J. Schreier, P. Gilch, and W. Zinth, "Early events of dna photodamage", *Annual Review of Physical Chemistry* **66**, 497–519 (2015).
- ²⁹S. Kanvah, J. Joseph, G. B. Schuster, R. N. Barnett, C. L. Cleveland, and U. Landman, "Oxidation of dna: damage to nucleobases", *Accounts of Chemical Research* **43**, 280–287 (2010).
- ³⁰W. J. Schreier, T. E. Schrader, F. O. Koller, P. Gilch, C. E. Crespo-Hernández, V. N. Swaminathan, T. Carell, W. Zinth, and B. Kohler, "Thymine dimerization in dna is an ultrafast photoreaction", *Science* **315**, 625–629 (2007).
- ³¹V. Gabelica and F. Rosu, "Gas-phase spectroscopy of nucleic acids", in *Nucleic acids in the gas phase*, edited by V. Gabelica (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014), pp. 103–130.

- ³²C.-H. C., C. B., and K. B., “Base stacking controls excited-state dynamics in A·T DNA”, *Nature* **436**, 1141–1144 (2005).
- ³³W. Y. Liang, “Excitons”, *Physics Education* **5**, 226–228 (1970).
- ³⁴R. Borrego-Varillas, G. Cerullo, and D. Markovitsi, “Exciton trapping dynamics in dna multimers”, *The Journal of Physical Chemistry Letters* **10**, 1639–1643 (2019).
- ³⁵E. M. Conwell, P. M. McLaughlin, and S. M. Bloch, “Charge-transfer excitons in dna”, *The Journal of Physical Chemistry B* **112**, 2268–2272 (2008).
- ³⁶B. Bouvier, J.-P. Dognon, R. Lavery, D. Markovitsi, P. Millié, D. Onidas, and K. Zakrzewska, “Influence of conformational dynamics on the exciton states of dna oligomers”, *The Journal of Physical Chemistry B* **107**, 13512–13522 (2003).
- ³⁷J. J. Nogueira, S. Roßbach, C. Ochsenfeld, and L. González, “Effect of dna environment on electronically excited states of methylene blue evaluated by a three-layered qm/qm/mm oniom scheme”, *Journal of Chemical Theory and Computation* **14**, 4298–4308 (2018).
- ³⁸J. J. Nogueira, F. Plasser, and L. González, “Electronic delocalization, charge transfer and hypochromism in the uv absorption spectrum of polyadenine unravelled by multiscale computations and quantitative wavefunction analysis”, *Chemical Science* **8**, 5682–5691 (2017).
- ³⁹L. M. Ibele, P. A. Sánchez-Murcia, S. Mai, J. J. Nogueira, and L. González, “Excimer intermediates en route to long-lived charge-transfer states in single-stranded adenine dna as revealed by nonadiabatic dynamics”, *The Journal of Physical Chemistry Letters* **11**, 7483–7488 (2020).
- ⁴⁰H. Wagenknecht, “Charge transfer in dna”, (2005).
- ⁴¹B. Lukyanov and M. Lukyanova, “Spiropyrans: synthesis, properties, and application. (review).”, *Chemistry of Heterocyclic Compounds* **41**, 281–311 (2005).
- ⁴²R. Klajn, “Spiropyran-based dynamic materials”, *Chemical Society Reviews* **43**, 148–184 (2014).
- ⁴³V. I. Minkin, “Photoswitchable molecular systems based on spiropyrans and spirooxazines”, in *Molecular switches* (John Wiley & Sons, Ltd, 2011) Chap. 2, pp. 37–80.
- ⁴⁴L. Kortekaas and W. R. Browne, “The evolution of spiropyran: fundamentals and progress of an extraordinarily versatile photochrome”, *Chemical Society Reviews* **48**, 3406–3424 (2019).
- ⁴⁵W. Szymański, J. M. Beierle, H. A. V. Kistemaker, W. A. Velema, and B. L. Feringa, “Reversible photocontrol of biological systems by the incorporation of molecular photoswitches”, *Chemical Reviews* **113**, 6114–6178 (2013).
- ⁴⁶J. Andersson, S. Li, P. Lincoln, and J. Andréasson, “Photoswitched dna-binding of a photochromic spiropyran”, *Journal of the American Chemical Society* **130**, 11836–11837 (2008).
- ⁴⁷A. K. Chibisov and H. Görner, “Photoprocesses in spiropyran-derived merocyanines”, *The Journal of Physical Chemistry A* **101**, 4305–4312 (1997).
- ⁴⁸C. Brieke and A. Heckel, “Spiropyran photoswitches in the context of dna: synthesis and photochromic properties”, *Chemistry – A European Journal* **19**, 15726–15734 (2013).

- ⁴⁹M. Hammarson, J. R. Nilsson, S. Li, P. Lincoln, and J. Andréasson, "Dna-binding properties of amidine-substituted spiropyran photoswitches", *Chemistry – A European Journal* **20**, 15855–15862 (2014).
- ⁵⁰J. Li, X. Yin, B. Li, X. Li, Y. Pan, J. Li, and Y. Guo, "Spiropyran in situ switching: a real-time fluorescence strategy for tracking dna g-quadruplexes in live cells", *Analytical Chemistry* **91**, 5354–5361 (2019).
- ⁵¹K. Jänich, *Linear algebra* (Springer-Verlag, 1994).
- ⁵²A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*, First (Dover Publications, Inc., Mineola, 1996).
- ⁵³J. J. Sakurai and J. Napolitano, *Modern quantum mechanics*, 2nd ed. (Cambridge University Press, 2017).
- ⁵⁴F. Jensen, *Introduction to computational chemistry* (Wiley, 2006).
- ⁵⁵E. Schrödinger, "An undulatory theory of the mechanics of atoms and molecules", *Phys. Rev.* **28**, 1049–1070 (1926).
- ⁵⁶M. Born and R. Oppenheimer, "Zur quantentheorie der molekeln", *Annalen der Physik* **389**, 457–484 (1927).
- ⁵⁷J. C. Slater, "The theory of complex spectra", *Phys. Rev.* **34**, 1293–1322 (1929).
- ⁵⁸P.-O. Löwdin, "Correlation problem in many-electron quantum mechanics i. review of different approaches and discussion of some current ideas", in *Advances in chemical physics* (John Wiley & Sons, Ltd, 1958), pp. 207–322.
- ⁵⁹M. Born and V. Fock, "Beweis des adiabatsatzes", *Zeitschrift für Physik* **51**, 165–180 (1928).
- ⁶⁰P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka, and R. Shepard, "Multiconfiguration self-consistent field and multireference configuration interaction methods and applications", *Chemical Reviews* **112**, 108–181 (2012).
- ⁶¹D. I. Lyakh, M. Musiał, V. F. Lotrich, and R. J. Bartlett, "Multireference nature of chemistry: the coupled-cluster view", *Chemical Reviews* **112**, PMID: 22220988, 182–243 (2012).
- ⁶²C. Møller and M. S. Plesset, "Note on an approximation treatment for many-electron systems", *Phys. Rev.* **46**, 618–622 (1934).
- ⁶³G. Li Manni, K. Guthrie, D. Ma, and W. Dobrautz, "Foundation of multi-configurational quantum chemistry", in *Quantum chemistry and dynamics of excited states* (John Wiley & Sons, Ltd, 2020) Chap. 6, pp. 133–203.
- ⁶⁴J. Olsen, "The casscf method: a perspective and commentary", *International Journal of Quantum Chemistry* **111**, 3267–3272.
- ⁶⁵P.-Å. Malmqvist, A. P. Rendell, and B. O. Roos, "The restricted active space self-consistent-field method, implemented with a split graph unitary group approach", *The Journal of Physical Chemistry* **94**, 5477–5482 (1990).

- ⁶⁶D. Roca-Sanjuán, F. Aquilante, and R. Lindh, “Multiconfiguration second-order perturbation theory approach to strong electron correlation in chemistry and photochemistry”, *WIREs Computational Molecular Science* **2**, 585–603.
- ⁶⁷J. Oddershede, P. Jørgensen, and D. L. Yeager, “Polarization propagator methods in atomic and molecular calculations”, *Computer Physics Reports* **2**, 33–92 (1984).
- ⁶⁸A. Dreuw and M. Wormit, “The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states”, *WIREs Computational Molecular Science* **5**, 82–95.
- ⁶⁹A. Dreuw, “The algebraic-diagrammatic construction scheme for the polarization propagator”, in *Quantum chemistry and dynamics of excited states* (John Wiley & Sons, Ltd) Chap. 5, pp. 109–131.
- ⁷⁰J. Schirmer, “Beyond the random-phase approximation: a new approximation scheme for the polarization propagator”, *Phys. Rev. A* **26**, 2395–2416 (1982).
- ⁷¹P. H. P. Harbach, M. Wormit, and A. Dreuw, “The third-order algebraic diagrammatic construction method (adc(3)) for the polarization propagator for closed-shell molecules: efficient implementation and benchmarking”, *The Journal of Chemical Physics* **141**, 064113 (2014).
- ⁷²R. Sarkar, M. Boggio-Pasqua, P.-F. Loos, and D. Jacquemin, “Benchmarking td-dft and wave function methods for oscillator strengths and excited-state dipole moments”, *Journal of Chemical Theory and Computation* **17**, 1117–1132 (2021).
- ⁷³P.-F. Loos, F. Lipparini, M. Boggio-Pasqua, A. Scemama, and D. Jacquemin, “A mountaineering strategy to excited states: highly accurate energies and benchmarks for medium sized molecules”, *Journal of Chemical Theory and Computation* **16**, 1711–1741 (2020).
- ⁷⁴“Electron density and hole functions”, in *A chemist’s guide to density functional theory* (John Wiley & Sons, Ltd) Chap. 2, pp. 19–28.
- ⁷⁵“The hohenberg-kohn theorems”, in *A chemist’s guide to density functional theory* (John Wiley & Sons, Ltd) Chap. 4, pp. 33–40.
- ⁷⁶“The kohn-sham approach”, in *A chemist’s guide to density functional theory* (John Wiley & Sons, Ltd) Chap. 5, pp. 41–64.
- ⁷⁷P. Hohenberg and W. Kohn, “Inhomogeneous electron gas”, *Phys. Rev.* **136**, B864–B871 (1964).
- ⁷⁸L. H. Thomas, “The calculation of atomic fields”, *Mathematical Proceedings of the Cambridge Philosophical Society* **23**, 542–548 (1927).
- ⁷⁹W. Kohn and L. J. Sham, “Self-consistent equations including exchange and correlation effects”, *Phys. Rev.* **140**, A1133–A1138 (1965).
- ⁸⁰A. J. Cohen, P. Mori-Sánchez, and W. Yang, “Challenges for density functional theory”, *Chemical Reviews* **112**, 289–320 (2012).
- ⁸¹N. T. Maitra, “Perspective: fundamental aspects of time-dependent density functional theory”, *The Journal of Chemical Physics* **144**, 220901 (2016).

- ⁸²M. Huix-Rotllant, N. Ferré, and M. Barbatti, “Time-dependent density functional theory”, in *Quantum chemistry and dynamics of excited states* (John Wiley & Sons, Ltd) Chap. 2, pp. 13–46.
- ⁸³E. Runge and E. K. U. Gross, “Density-functional theory for time-dependent systems”, *Phys. Rev. Lett.* **52**, 997–1000 (1984).
- ⁸⁴G. Vignale, “Real-time resolution of the causality paradox of time-dependent density-functional theory”, *Phys. Rev. A* **77**, 062511 (2008).
- ⁸⁵N. T. Maitra, K. Burke, and C. Woodward, “Memory in time-dependent density functional theory”, *Phys. Rev. Lett.* **89**, 023002 (2002).
- ⁸⁶Y. Kurzweil and R. Baer, “Time-dependent exchange-correlation current density functionals with memory”, *The Journal of Chemical Physics* **121**, 8731–8741 (2004).
- ⁸⁷M. Marques and E. Gross, “Time-dependent density functional theory”, *Annual Review of Physical Chemistry* **55**, 427–455 (2004).
- ⁸⁸K. Yabana and G. F. Bertsch, “Time-dependent local-density approximation in real time”, *Phys. Rev. B* **54**, 4484–4487 (1996).
- ⁸⁹M. E. Casida, “Time-dependent density-functional theory for molecules and molecular solids”, *Journal of Molecular Structure: THEOCHEM* **914**, Time-dependent density-functional theory for molecules and molecular solids, 3–18 (2009).
- ⁹⁰S. Hirata and M. Head-Gordon, “Time-dependent density functional theory within the tamm–dancoff approximation”, *Chemical Physics Letters* **314**, 291–299 (1999).
- ⁹¹N. T. Maitra, F. Zhang, R. J. Cave, and K. Burke, “Double excitations within time-dependent density functional theory linear response”, *The Journal of Chemical Physics* **120**, 5932–5937 (2004).
- ⁹²N. T. Maitra, “Charge transfer in time-dependent density functional theory”, *Journal of Physics: Condensed Matter* **29**, 423001 (2017).
- ⁹³A. Dreuw and M. Head-Gordon, “Failure of time-dependent density functional theory for long-range charge-transfer excited states: the zincbacteriochlorophyllinbacteriochlorophyllin and bacteriochlorophyllspheroidene complexes”, *Journal of the American Chemical Society* **126**, 4007–4016 (2004).
- ⁹⁴D. J. Tozer and N. C. Handy, “On the determination of excitation energies using density functional theory”, *Phys. Chem. Chem. Phys.* **2**, 2117–2121 (2000).
- ⁹⁵T. Yanai, D. P. Tew, and N. C. Handy, “A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp)”, *Chemical Physics Letters* **393**, 51–57 (2004).
- ⁹⁶L. Kronik, T. Stein, S. Refaely-Abramson, and R. Baer, “Excitation gaps of finite-sized systems from optimally tuned range-separated hybrid functionals”, *Journal of Chemical Theory and Computation* **8**, 1515–1531 (2012).
- ⁹⁷B. G. Levine, C. Ko, J. Quenneville, and T. J. Martínez, “Conical intersections and double excitations in time-dependent density functional theory”, *Molecular Physics* **104**, 1039–1051 (2006).

- ⁹⁸D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, and R. Kaschner, “Construction of tight-binding-like potentials on the basis of density-functional theory: application to carbon”, *Phys. Rev. B* **51**, 12947–12957 (1995).
- ⁹⁹G. d. M. Seabra, R. C. Walker, M. Elstner, D. A. Case, and A. E. Roitberg, “Implementation of the scc-dftb method for hybrid qm/mm simulations within the amber molecular dynamics package”, *The Journal of Physical Chemistry A* **111**, 5655–5664 (2007).
- ¹⁰⁰M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, “Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties”, *Phys. Rev. B* **58**, 7260–7268 (1998).
- ¹⁰¹M. Gaus, Q. Cui, and M. Elstner, “Dftb3: extension of the self-consistent-charge density-functional tight-binding method (scc-dftb)”, *Journal of Chemical Theory and Computation* **7**, 931–948 (2011).
- ¹⁰²González, M.A., “Force fields and molecular dynamics simulations”, *JDN* **12**, 169–200 (2011).
- ¹⁰³Z. Jin, C. Yang, F. Cao, F. Li, Z. Jing, L. Chen, Z. Shen, L. Xin, S. Tong, and H. Sun, “Hierarchical atom type definitions and extensible all-atom force fields”, *Journal of Computational Chemistry* **37**, 653–664 (2016).
- ¹⁰⁴J. E. Lennard-Jones, “Cohesion”, *Proceedings of the Physical Society* **43**, 461–482 (1931).
- ¹⁰⁵J. D. Jackson, in *Classical electrodynamics* (John Wiley & Sons, Ltd, 1998).
- ¹⁰⁶J. A. Harrison, J. D. Schall, S. Maskey, P. T. Mikulski, M. T. Knippenberg, and B. H. Morrow, “Review of force fields and intermolecular potentials used in atomistic computational materials research”, *Applied Physics Reviews* **5**, 031104 (2018).
- ¹⁰⁷J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, “Development and testing of a general amber force field”, *Journal of Computational Chemistry* **25**, 1157–1174 (2004).
- ¹⁰⁸W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, “A second generation force field for the simulation of proteins, nucleic acids, and organic molecules”, *Journal of the American Chemical Society* **117**, 5179–5197 (1995).
- ¹⁰⁹A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczero, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczero, D. Yin, and M. Karplus, “All-atom empirical potential for molecular modeling and dynamics studies of proteins”, *The Journal of Physical Chemistry B* **102**, 3586–3616 (1998).
- ¹¹⁰T. E. Cheatham III and D. A. Case, “Twenty-five years of nucleic acid simulations”, *Biopolymers* **99**, 969–977 (2013).
- ¹¹¹M. Zgarbová, F. J. Luque, J. Šponer, T. E. Cheatham, M. Otyepka, and P. Jurečka, “Toward improved description of dna backbone: revisiting epsilon and zeta torsion force field parameters”, *Journal of Chemical Theory and Computation* **9**, 2339–2354 (2013).

- ¹¹²R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka, and T. E. Cheatham, "Assessing the current state of amber force field modifications for dna", *Journal of Chemical Theory and Computation* **12**, 4114–4127 (2016).
- ¹¹³M. Krepl, M. Zgarbová, P. Stadlbauer, M. Otyepka, P. Banáš, J. Koča, T. E. Cheatham, P. Jurečka, and J. Šponer, "Reference simulations of noncanonical nucleic acids with different variants of the amber force field: quadruplex dna, quadruplex rna, and z-dna", *Journal of Chemical Theory and Computation* **8**, 2506–2520 (2012).
- ¹¹⁴M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo, and P. Jurečka, "Refinement of the sugar–phosphate backbone torsion beta for amber force fields improves the description of z- and b-dna", *Journal of Chemical Theory and Computation* **11**, 5723–5736 (2015).
- ¹¹⁵I. Ivani, P. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. Gelpí, C. González, M. Vendruscolo, C. Laughton, S. Harris, D. Case, and M. Orozco, "Parmbsc1: a refined force field for dna simulations", *Nature Methods* **13**, 55–58 (2016).
- ¹¹⁶W. L. Jorgensen, "Quantum and statistical mechanical studies of liquids. 10. transferable intermolecular potential functions for water, alcohols, and ethers. application to liquid water", *Journal of the American Chemical Society* **103**, 335–340 (1981).
- ¹¹⁷W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water", *The Journal of Chemical Physics* **79**, 926–935 (1983).
- ¹¹⁸P. Ren and J. W. Ponder, "Polarizable atomic multipole water model for molecular mechanics simulation", *The Journal of Physical Chemistry B* **107**, 5933–5947 (2003).
- ¹¹⁹H. M. Senn and W. Thiel, "Qm/mm methods for biomolecular systems", *Angewandte Chemie International Edition* **48**, 1198–1229 (2009).
- ¹²⁰A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme", *Journal of Molecular Biology* **103**, 227–249 (1976).
- ¹²¹M. J. Field, P. A. Bash, and M. Karplus, "A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations", *Journal of Computational Chemistry* **11**, 700–733.
- ¹²²M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, "Oniom: a multilayered integrated mo + mm method for geometry optimizations and single point energy predictions. a test for dielsalder reactions and pt(p(t-bu)₃)₂ + h₂ oxidative addition", *The Journal of Physical Chemistry* **100**, 19357–19363 (1996).
- ¹²³T. H. Rod and U. Ryde, "Accurate qm/mm free energy calculations of enzyme reactions: methylation by catechol o-methyltransferase", *Journal of Chemical Theory and Computation* **1**, 1240–1251 (2005).
- ¹²⁴H. Yu and W. F. van Gunsteren, "Accounting for polarization in molecular simulation", *Computer Physics Communications* **172**, 69–85 (2005).

- ¹²⁵A. Warshel, M. Kato, and A. V. Pisliakov, "Polarizable force fields: history, test cases, and prospects", *Journal of Chemical Theory and Computation* **3**, 2034–2045 (2007).
- ¹²⁶D. Bakowies and W. Thiel, "Hybrid models for combined quantum mechanical and molecular mechanical approaches", *The Journal of Physical Chemistry* **100**, 10580–10594 (1996).
- ¹²⁷R. B. Murphy, D. M. Philipp, and R. A. Friesner, "A mixed quantum mechanics/molecular mechanics (qm/mm) method for large-scale modeling of chemistry in protein environments", *Journal of Computational Chemistry* **21**, 1442–1457 (2000).
- ¹²⁸G. A. Cisneros, S. N.-I. Tholander, O. Parisel, T. A. Darden, D. Elking, L. Perera, and J.-P. Piquemal, "Simple formulas for improved point-charge electrostatics in classical force fields and hybrid quantum mechanical/molecular mechanical embedding", *International Journal of Quantum Chemistry* **108**, 1905–1912 (2008).
- ¹²⁹Y. Zhang, H. Lin, and D. G. Truhlar, "Self-consistent polarization of the boundary in the redistributed charge and dipole scheme for combined quantum-mechanical and molecular-mechanical calculations", *Journal of Chemical Theory and Computation* **3**, 1378–1398 (2007).
- ¹³⁰U. C. Singh and P. A. Kollman, "A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the $\text{CH}_3\text{Cl} + \text{Cl}$ exchange reaction and gas phase protonation of polyethers", *Journal of Computational Chemistry* **7**, 718–730 (1986).
- ¹³¹M. A. Cuendet and W. F. van Gunsteren, "On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm", *The Journal of Chemical Physics* **127**, 184102 (2007).
- ¹³²L. Verlet, "Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules", *Phys. Rev.* **159**, 98–103 (1967).
- ¹³³B. L., "Vorlesungen über gastheorie", Leipzig, J. A. Barth (1896).
- ¹³⁴U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh ewald method", *The Journal of Chemical Physics* **103**, 8577–8593 (1995).
- ¹³⁵P. P. Ewald, "Die berechnung optischer und elektrostatischer gitterpotentiale", *Annalen der Physik* **369**, 253–287 (1921).
- ¹³⁶J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes", *Journal of Computational Physics* **23**, 327–341 (1977).
- ¹³⁷H. C. Andersen, "Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations", *Journal of Computational Physics* **52**, 24–34 (1983).
- ¹³⁸M. R. Hestens and E. Stiefel, "Methods of conjugate gradients for solving linear systems", *Journal of Research of the National Bureau of Standards* **49**, 409–436 (1952).
- ¹³⁹H. B. Curry, "The method of steepest descent for non-linear minimization problems", *Quarterly of Applied Mathematics* **2**, 258–261 (1944).
- ¹⁴⁰J. Liu, D. Li, and X. Liu, "A simple and accurate algorithm for path integral molecular dynamics with the langevin thermostat", *The Journal of Chemical Physics* **145**, 024103 (2016).

- ¹⁴¹H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath", *The Journal of Chemical Physics* **81**, 3684–3690 (1984).
- ¹⁴²H. C. Andersen, "Molecular dynamics simulations at constant pressure and/or temperature", *The Journal of Chemical Physics* **72**, 2384–2393 (1980).
- ¹⁴³S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble", *Molecular Physics* **52**, 255–268 (1984).
- ¹⁴⁴N. Homeyer and H. Gohlke, "Free energy calculations by the molecular mechanics poissonboltzmann surface area method", *Molecular Informatics* **31**, 114–122 (2012).
- ¹⁴⁵J. Klett, A. Núñez-Salgado, H. G. Dos Santos, Cortés-Cabrera, A. Perona, R. Gil-Redondo, D. Abia, F. Gago, and A. Morreale, "Mm-ismsa: an ultrafast and accurate scoring function for protein–protein docking", *Journal of Chemical Theory and Computation* **8**, 3395–3408 (2012).
- ¹⁴⁶B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg, "Mmpbsa.py: an efficient program for end-state free energy calculations", *Journal of Chemical Theory and Computation* **8**, 3314–3321 (2012).
- ¹⁴⁷R. Qi, G. Wei, B. Ma, and R. Nussinov, "Replica exchange molecular dynamics: a practical application protocol with solutions to common problems and a peptide aggregation and self-assembly example", in *Peptide self-assembly: methods and protocols*, edited by B. L. Nilsson and T. M. Doran (Springer New York, New York, NY, 2018), pp. 101–119.
- ¹⁴⁸A. Laio and M. Parrinello, "Escaping free-energy minima", *Proceedings of the National Academy of Sciences* **99**, 12562–12566 (2002).
- ¹⁴⁹A. F. Voter, "Hyperdynamics: accelerated molecular dynamics of infrequent events", *Phys. Rev. Lett.* **78**, 3908–3911 (1997).
- ¹⁵⁰J. Kästner, "Umbrella sampling", *WIREs Computational Molecular Science* **1**, 932–942 (2011).
- ¹⁵¹T. Mülders, P. Krüger, W. Swegat, and J. Schlitter, "Free energy as the potential of mean constraint force", *The Journal of Chemical Physics* **104**, 4869–4870 (1996).
- ¹⁵²M. Souaille and B. Roux, "Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations", *Computer Physics Communications* **135**, 40–57 (2001).
- ¹⁵³T.-S. Lee, B. K. Radak, M. Huang, K.-Y. Wong, and D. M. York, "Roadmaps through free energy landscapes calculated using the multidimensional vfep approach", *Journal of Chemical Theory and Computation* **10**, 24–34 (2014).
- ¹⁵⁴T.-S. Lee, B. K. Radak, A. Pabis, and D. M. York, "A new maximum likelihood approach for free energy profile construction from molecular simulations", *Journal of Chemical Theory and Computation* **9**, 153–164 (2013).
- ¹⁵⁵B. Isralewitz, M. Gao, and K. Schulten, "Steered molecular dynamics and mechanical functions of proteins", *Current Opinion in Structural Biology* **11**, 224–230 (2001).

- ¹⁵⁶L. Kortekaas, J. Chen, D. Jacquemin, and W. R. Browne, “Proton-stabilized photochemically reversible e/z isomerization of spiropyrans”, *The Journal of Physical Chemistry B* **122**, 6423–6430 (2018).
- ¹⁵⁷F. Schinle, P. E. Crider, M. Vonderach, P. Weis, O. Hampe, and M. M. Kappes, “Spectroscopic and theoretical investigations of adenosine 5-diphosphate and adenosine 5-triphosphate dianions in the gas phase”, *Physical Chemistry Chemical Physics* **15**, 6640–6650 (2013).
- ¹⁵⁸F. Plasser, “Theodore: a toolbox for a detailed and automated analysis of electronic excited state computations”, *The Journal of Chemical Physics* **152**, 084108 (2020).
- ¹⁵⁹M Dušínská, A Mateášik, D. J. Chorvát, I Lajdová, V Spustová, and A. Chorvtová, “Merocyanine 540 induces specific base oxidation in human lymphocyte dna”, *NEOPLASMA* **46**, 69–70 (1999).

Declaration of Authenticity

I declare that this thesis is my original work. All the calculations and the thesis itself, including all the figures, are result of my own work, with the help of the advisor Professors and researchers during the three years of the research project. The only external contributions are cited and their references, personally consulted, reported.

Vienna, March 2021

Davide Avagliano

CV DAVIDE AVAGLIANO

Born: 12/06/1992, Naples, Italy
Email address: davide.avagliano@univie.ac.at

CV updated 29/03/2021

WORK EXPERIENCES

JUNE 2018 – MAY 2021 **INSTITUTE OF THEORETICAL CHEMISTRY**, UNIVERSITY OF VIENNA, VIENNA, AUSTRIA
PH.D. RESEARCH PROJECT (Prof. L. González)

Research internships included:

- *October 2019: **Durham University***, Durham, United Kingdom (Prof. J.R.R. Verlet)
- *January 2020: **University of Bologna***, Bologna, Italy (Prof. M. Garavelli)

FEBRUARY 2017 – JULY 2017 **INSTITUTE OF THEORETICAL CHEMISTRY**, UNIVERSITY OF VIENNA, VIENNA, AUSTRIA
MASTER RESEARCH PROJECT (Prof. E. Bodo, Prof. L. González)

OCTOBER 2014 – NOVEMBER 2014 **LABORATORY OF ENVIRONMENTAL CATALYSIS**, SAPIENZA UNIVERSITY OF
ROME, ROME, ITALY
BACHELOR'S DEGREE INTERNSHIP (Dr. D. Pietrogiaconi)

EDUCATION

JUNE 2018 - MAY 2021 **UNIVERSITY OF VIENNA**, Vienna, Austria, *Ph.D. student in Chemistry*

OCTOBER 2015 – DECEMBER 2017 **SAPIENZA UNIVERSITY OF ROME**, Rome, Italy, *Master of Science in Chemistry*

OCTOBER 2011 – MARCH 2015 **SAPIENZA UNIVERSITY OF ROME**, Rome, Italy, *Bachelor's Degree in Chemistry*,

SEPTEMBER 2006 – JULY 2011 **LICEO GINNASIO STATALE "TORQUATO TASSO"**, Rome, Italy, *Secondary school diploma in classical studies*

SCHOLARSHIPS

JUNE 2018 – MAY 2021 **MSCA-ETN LIGHTDYNAMICS FELLOWSHIP**, 2018-2021, UNIVERSITY OF VIENNA
Early Stage Researcher fellowship funded by the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement No765866

FEBRUARY 2017 - JULY 2017 **ERASMUS+ SCHOLARSHIP**, AY 2016/17, SAPIENZA UNIVERSITY OF ROME
Scholarship for the international mobility of University students

- 8) **D. Avagliano**, M. Bonfanti, M. Garavelli, L. González,
“QM/MM Nonadiabatic Dynamics: The SHARC/COBRAMM Approach” (*in preparation*), 2021
- 7) M. Castellani, **D. Avagliano**, J. Verlet,
“Ultrafast Dynamics of the Isolated Adenosine-5'-triphosphate Dianion Probed By Time-Resolved Photoelectron Imaging” (*in revision*), 2021
- 6) **D. Avagliano**; S. Tkaczyk, P. A. Sánchez-Murcia, L. González,
“Enhanced Rigidity Changes Ultraviolet Absorption: Effect of a Merocyanine Binder on G-Quadruplex Photophysics” *J. Phys. Chem. Lett.*, 11, 10212-10218, 2020
- 5) M. Castellani*, **D. Avagliano***; L. González, J.R.R. Verlet
“Site-specific Photo-oxidation of the Isolated Adenosine-5'-triphosphate Dianion Determined by Photoelectron Imaging” *J. Phys. Chem. Lett.*, 11, 8195–8201, 2020, *equal contribution
- 4) **D. Avagliano**, P. A. Sánchez-Murcia, L. González
“Spiropyran Meets Guanine Quadruplexes: Isomerization Mechanism and DNA Binding Modes of Quinolizidine-Substituted Spiropyran Probes” *Chem.: Eur. J.*, 26, 13039-13045, 2020
- 3) P. Heim, S. Mai, B. Thaler, S. Cesnik, **D. Avagliano**, D. Bella-Velidou, W. Ernst, L. González, M. Koch
“Revealing ultrafast population transfer between nearly degenerated electronic states” *J. Phys. Chem. Lett.* 11, 1443-1449, 2020
- 2) **D. Avagliano**, P.A. Sánchez-Murcia, L. González,
“Directional and Regioselective Hole Injection of Spiropyran Photoswitches Intercalated in A/T-duplex DNA” *Phys. Chem. Chem. Phys.* 21, 17971-17977, 2019
- 1) **D. Avagliano**, P.A. Sánchez-Murcia, L. González,
“DNA-Binding Mechanism of Spiropyran Photoswitches: The Role of Electrostatics” *Phys. Chem. Chem. Phys.* 21, 8614-8618, 2019

PRESENTATIONS

- 5) “SHARC/COBRAMM approach for nonadiabatic dynamics: Methods, key steps and applications”
Invited Talk: Bologna (Italy), Computational Photochemistry Webinar Series (Online event), 28.01.2021
- 4) “Modelling and Disentangling Exciton-Excimer Coupling in c-MYC Guanine Quadruplex by Ligand Interaction”
Oral contribution: London (United Kingdom), Modelling Photoinduced Processes in Molecular Systems, 06.02.2020 - 07.02.2020
- 3) “Excited states characterization of ligand-DNA complexes”
Invited Talk: Durham (United Kingdom), Prof. Verlet and Prof. Curchod group seminar, 08.10.2019
- 2) “Hole Injection in DNA induced by Spiropyran Photoswitches: a Regioselective Oxidative Process”
Flash Communication: València (Spain), 3rd Workshop on DNA Damage and Repair (DNADNR), 25.09.2019 - 26.09.2019
- 1) “Drastic effect of the electrostatic energy on the binding mode of spiropyran photoprobe to dsDNA”
Flash Communication: Madrid (Spain), 6th-European Chemical Biology Symposium/LS-EuChemS, 03.04.2019 - 05.04.2019