

University of Warsaw
Faculty of Physics

Łukasz Milewski

Record book number: 394521

Exploring the excited state dynamics of molecular probes for cancer biomarkers: a multiscale modeling study

Master's thesis
in the field of "Applications of physics in biology and medicine"

The thesis was written under the supervision of
dr hab. Silvio Osella
Centre for New Technology
and prof. dr hab. Bogdan Lesyng
Faculty of Physics

Warsaw, September, 2022

Summary

Cyclooxygenases are a family of enzymes consisting of the isoenzymes COX-1 and COX -2, which are important for the biosynthesis of prostaglandins in inflammatory reactions. In addition, COX-2 is overexpressed in all stages of cancer and therefore can be used to identify cancer cells through the use of molecular fluorescent probes. The aim of this study is to develop a novel computational method to calculate the fluorescence properties of the ANQ-IMC-6 probe, which consists of the fluorophore acenaphtho-1,2-b-quinoxaline (ANQ) and an indomethacin inhibitor (IMC) when embedded in a COX-2 environment. The geometry of the probe was optimized in the excited state and molecular dynamics simulation for the probe inside active site of COX-2 was performed using modified CHARMM force field. On the extracted frames from MD simulations, excited-state ab-initio molecular dynamics computation were performed. As a result, the evolution of the fluorescence spectra of the probe in time were obtained. For each considered time, the molecular orbitals underneath the electronic transition were calculated and the spectra analyzed. As main result, we observe that the evolution in time allows for a change in electronic transition from a quenched HOMO-2 → LUMO to an allowed HOMO → LUMO transition. This, in turn, translates into a strong fluorescence of the ANQ-IMC-6 probe due to the presence of the surrounding biological environment.

Keywords

biophysics, cancer, molecular probe, fluorescence, cyclooxygenase, anq-imc-6, molecular dynamics, computational multiscale approach, ab initio computations

Title of the thesis in Polish language

Badanie dynamiki stanu wzbudzonego sond molekularnych jako biomarkerów raka: modelowanie wieloskalowe

Contents

1. Introduction	5
1.1. Molecular probes	5
1.2. Fluorescence principles	6
1.2.1. Nature of light	6
1.2.2. Luminescence and Jabłoński diagram.	6
1.2.3. Fluorescence quenching	8
1.3. Cyclooxygenases and cancer	11
1.4. Subject of the study	15
1.4.1. ANQ-ICM-6 as a fluorescent probe	15
2. Methods	19
2.1. Molecular dynamics	19
2.1.1. Basic principles	19
2.1.2. Approximations	19
2.1.3. Unit cell and periodic boundary conditions	20
2.1.4. Cut-off restrictions	20
2.1.5. Global MD procedure	21
2.1.6. Computed variables	21
2.1.7. Integrating algorithms	22
2.1.8. Conserving the temperature and pressure	22
2.1.9. Force field	24
2.2. Density Functional Theory	25
2.2.1. Schrödinger equation	25
2.2.2. Hohenberg-Kohn Theorems	26
2.2.3. Time Dependent DFT	27
2.2.4. Runge-Gross Theorem	27
2.2.5. Functionals and Basis Sets	27
2.2.6. Configuration Interaction Singles (CIS)	28
2.2.7. Equation-of-Motion Couple Cluster Singles and Doubles (EOM-CCSD)	30
2.3. Ab initio molecular dynamics	31
3. Results	33
3.1. Optimization in the excited state	33
3.2. MD	34
3.3. AIMD Validation	39
3.4. Ab initio MD	46
Conclusions	55
Bibliography	57

Chapter 1

Introduction

1.1. Molecular probes

A molecular probe, also known as a fluorescent chemosensor, is defined as any compound possessing a reactive site that binds and a fluorophore¹. Such a reactive site may be an inhibitor that binds specifically to a given protein or enzyme. After the probe binds to the target molecule, the fluorophore part changes its response and we obtain physical measurable quantities, such as fluorescence intensity or emission wavelength, altered by the environment.

Fluorescence imaging techniques are not invasive and can be very sensitive. Detection of fluorescence probes is easy using optical instruments. Such a probe can be easily used in cells of living organisms and can act specifically with a chosen biomolecule, such as previously mentioned proteins². Several fluorescence mechanisms, such as photoinduced electron transfer (PET)³, intramolecular charge transfer (ICT) and fluorescence resonance energy transfer (FRET) can be considered⁴. I will describe these mechanisms in the following chapter "Fluorescence principles". Regardless of the acting mechanisms, the principles of design of molecular probes remain the same: we need something that can interact with a chosen biomarker and that can give us a change of emission wavelength or emission intensity - decrease (quenching) or increase (enhancement) when inserted in a biological environment⁵. Generally, two major strategies are considered, either based on the receptor (i.e. coordination interaction with ions or binding to the active site of an enzyme) or on the reaction that might occur (i.e. enzymatic reaction altering the structure of the fluorescent probe)⁵. This thesis is focused on fluorescent probes which bind to a specific class of proteins. The nature of the probe is twofold; on one hand, a strong binding moiety is needed, acting as an inhibitor. Inhibitors are molecules that can block enzymatic reactions when the inhibitor is bound in the enzyme's active site when they possess a fitting size and shape, which can give very high specificity. On the other hand, the probe must be fluorescent, in order to see to what extent the binding process to the protein alters the intrinsic optical properties of the probe. The combination of the inhibitor and the fluorophore across a bridge (or linker) forms a complete molecular probe for fluorescence imaging.

1.2. Fluorescence principles

1.2.1. Nature of light

As described by quantum mechanics⁶, light has a dual nature: it can be described as a particle as well as a wave. When described using particle theory, these corpuscles are massless photons. Each photon carries the same defined amount of energy. Light rays consist thus of streams of particles. This view also properly describes interactions with the matter. Yet, the propagation of light is better described by the wave theory. In the terms of the wave light theory, light is described as an electromagnetic wave governed by the Maxwell equations. The wave thus consists of an electric and magnetic field oscillating perpendicular to each other and propagating along the direction of the wave (see the image in Fig. 1.1).

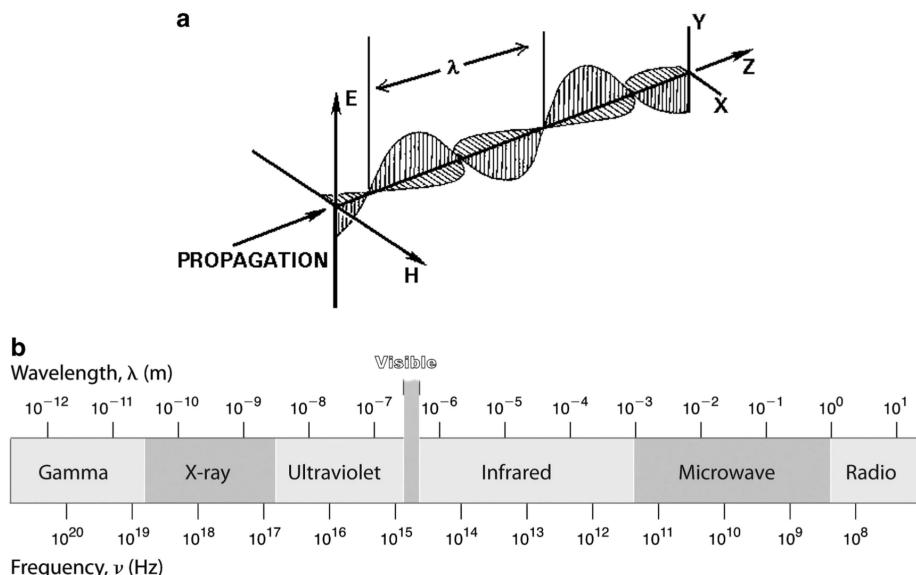


Figure 1.1: a) Sketch of electromagnetic wave with electric field E magnetic field H and wavelength λ . b)electromagnetic spectrum, the visible part of the spectrum is visible light. Image from Sliney, 2016⁶.

The visible light we see is only a small part of the whole visible electromagnetic radiation spectrum. Depending on the wavelength of the wave we distinguish different colours. The nature of the light is key to understanding all spectroscopy imaging techniques based on using light and associated processes. For example, we can use the phenomenon of fluorescence to distinguish different parts of the cell in microscopy or get information about protein structures using absorption spectra analysis.

1.2.2. Luminescence and Jabłoński diagram.

According to quantum mechanics⁶, electronic systems have defined fixed energy values. For example, an atom has energy corresponding to a region where we can find electrons in an atom or molecule (an orbital), as the electron position is described by a certain probability. Depending on the orbital, an electron has a fixed energy state. The lowest stable energy level of the system is the ground state. When the energy of an atom or molecule is higher than its ground state, an excited state is defined⁷. It can be described as that extra energy causing an electron to change its allowed orbital. We call this process excitation. After some time, the excited electron has to go back to the ground state, emitting quanta of light. This phenomenon is called luminescence⁷. Depending on the type of transition, it can be fluorescence or phosphorescence. Fluorescence occurs when

we have a transition of an electron from excited singlet states (when the excited electron is paired by opposite spin to the second ground-state electron)⁸. Phosphorescence however occurs when we have triplet excited states (when the excited electron has the same spin as the electron in the ground state and the straight-forward transition to the ground state is forbidden, see the image in Fig. 1.2)⁸. The difference between the two mechanisms can be also based on the time of emission - lifetime τ of the fluorophore. Fluorescence (lifetime near 1 nanosecond) is a rapid process, while phosphorescence can be long (lasts from milliseconds to seconds or sometimes even longer) because the conversion of the triplet state into the singlet state to eventually pair spin in the ground state is needed.

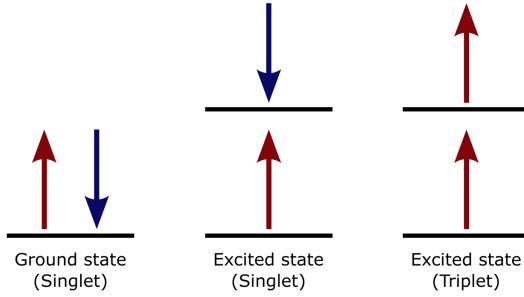


Figure 1.2: Sketch explaining singlet and triplet states. Source: PSIBERG⁹.

Such processes can be described by using the Jabłoński diagram proposed in 1935¹⁰ by Polish Professor Aleksander Jabłoński (Fig. 1.3). Looking at the diagram, we can differentiate processes as absorption, fluorescence and phosphorescence. When a photon is absorbed by a system, depending on the wavelength of the light causing that effect, we observe the excitation of an electron to a given energy state. After some time the state relaxes by electron going back to the ground state, which is accompanied by emission of the light. The system can also relax by non-radiative processes such as internal conversion. Alternatively, we can have inter-system crossing causing the system to go from the singlet state to the triplet state leading to the possibility of phosphorescent emission.

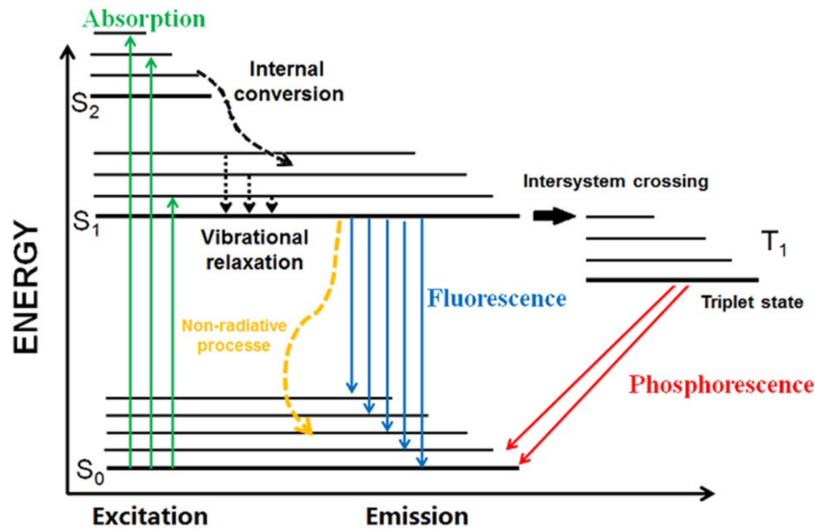


Figure 1.3: Jabłoński diagram, from Berardozzi, et al, 2021¹¹.

1.2.3. Fluorescence quenching

Different physical processes can cause decrease of the intensity of the fluorescence, a phenomenon known as fluorescence quenching. Among the variety of different processes, one of the most important is the Förster (also called Fluorescence) Resonance Energy Transfer (FRET). It is a physical process which transfer energy in non-radiate way between an excited fluorophore as a donor and another fluorophore as an acceptor. It is done by intermolecular long-range dipole-dipole coupling, if they are in close proximity (less than 10 nm). Mechanism of this transfer of energy can be described as "treating an excited fluorophore as an oscillating dipole that can undergo an energy exchange with a second dipole having a similar resonance frequency."¹². This process doesn't depend on the surrounding solvent and only depends on the distance between coupled chromophores. In some situations, the donor do not need to be a fluorophore, so FRET can also occur without emission of photon. As it is caused by resonance, the fluorescence spectrum of the donor and absorption spectrum of the acceptor overlaps, which is shown in Fig. 1.4.

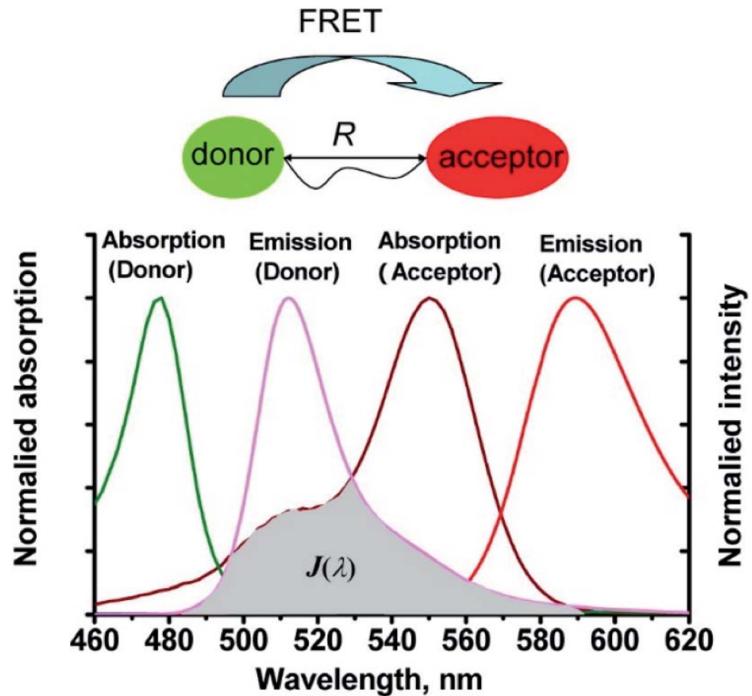


Figure 1.4: The diagram showing the Förster resonance energy transfer. R denotes the distance between them. Below donor-acceptor pair spectrum for the FRET phenomenon. Spectral overlap is marked by the grey region. Image from Yuan et al., 2013¹³.

In the Fig. 1.5, the same Jabłoński diagram is presented, but concerning only the FRET transition.

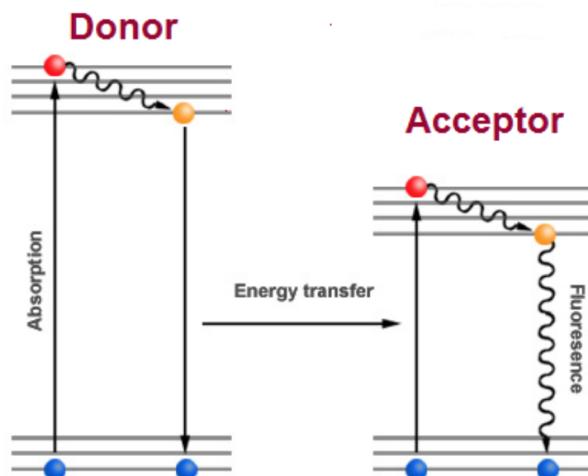


Figure 1.5: Jabłoński diagram for FRET phenomenon. Image from Hussain et al., 2009¹².

Förster transfer of energy can cause quenching of the fluorescence of a molecule. Fluorescence quenching can be also caused by electron transfer between the excited fluorophore (donor) and receptor ligand (acceptor) capable of redox activity. This process is called photoinduced electron transfer (PET).

Excitation of the fluorophore leads to the electron transfer from highest occupied molecular orbital (HOMO) to lowest unoccupied molecular orbital (LUMO). If the energy of some linked receptor to the fluorophore is higher than the HOMO of the fluorophore, but lower than the LUMO, than the electron can be transferred, which "locks" the LUMO in the state where the electron cannot relax back to the HOMO and thus, the fluorescence is quenched¹⁴. The diagram explaining the PET phenomenon is shown in Fig.1.6.

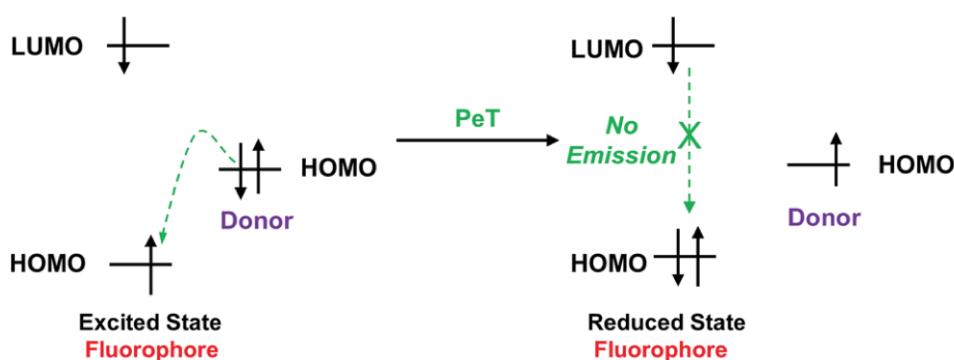


Figure 1.6: Diagram showing the mechanism of photoinduced electron transfer based on the arrangement of molecular orbitals and their occupancy states. Image from Tian et al., 2021¹⁴.

The mechanism of PET can be also used for the rational design of a molecular probes; when the receptors binds to the target, it can lower its HOMO below the HOMO of the fluorophore, blocking the PET process and allowing fluorescence¹⁴. This is called an 'on-off' mechanism of the probe. Scheme of the mechanism is shown in Fig. 1.7.

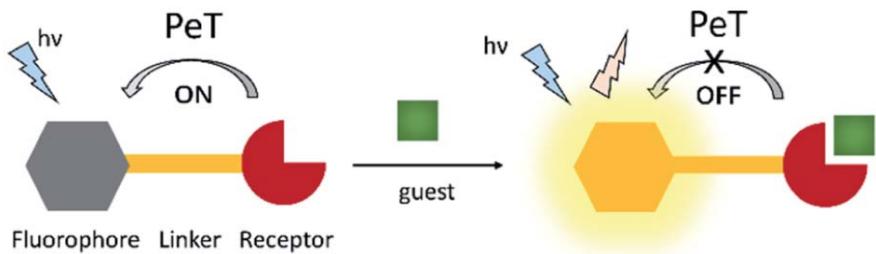


Figure 1.7: Diagram showing the PET mechanism for an on-off molecular probe. "Guest" here can be, for example, the active site of an enzyme. Image from Tian et al., 2021¹⁴.

The third major quenching mechanism involved into fluorescence probe design, is intramolecular charge transfer (ICT). When the fluorophore made of electron donor (D), electron acceptor (A) and π -conjugated linker interact with an analyte, the electron density is altered in some of the parts which cause electron transfer inside the molecule electron density cloud¹⁴. Such a probe will exhibit 'turn-on' mechanism, where we see a shift of the fluorescence wavelength after the interaction with the environment.

In the Fig. 1.8, the schematic comparison of the FRET, PET and ICT is shown.

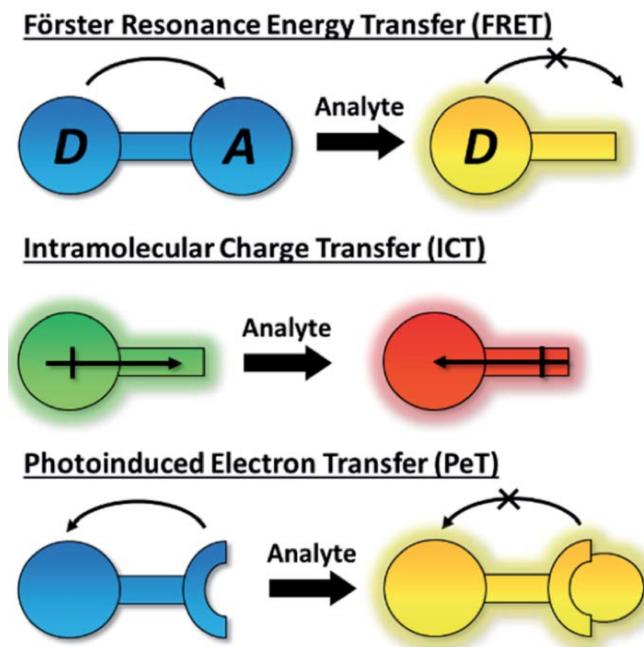


Figure 1.8: The diagram showing schematic representation of selected fluorescence probe mechanisms. Image from Tian et al., 2021¹⁴.

1.3. Cyclooxygenases and cancer

Cyclooxygenases (also called prostaglandin endoperoxide synthases (E.C. 1.14.99.1) or prostaglandin G/H synthases) are enzymes important for the biosynthesis of prostaglandins in inflammatory responses¹⁵. The family of cyclooxygenases consists of three isoforms: COX-1, COX-2, and COX-3, each of size close to 600 amino acids with similar sequence¹⁶. First purified in 1976 by Hemler et al.¹⁷, cyclooxygenases have been used as an inhibition target for treating pain and controlling inflammation for many years. As prostaglandins are involved in pain sensitivity and fever¹⁸, blocking synthesising them by cyclooxygenases is key to the activity of so-called nonsteroidal anti-inflammatory drugs (NSAIDs). Several examples of such drugs making use of this inhibition are aspirin (mechanism described in 1971¹⁹), ibuprofen, rofecoxib and indomethacin²⁰ (see Fig. 1.9).

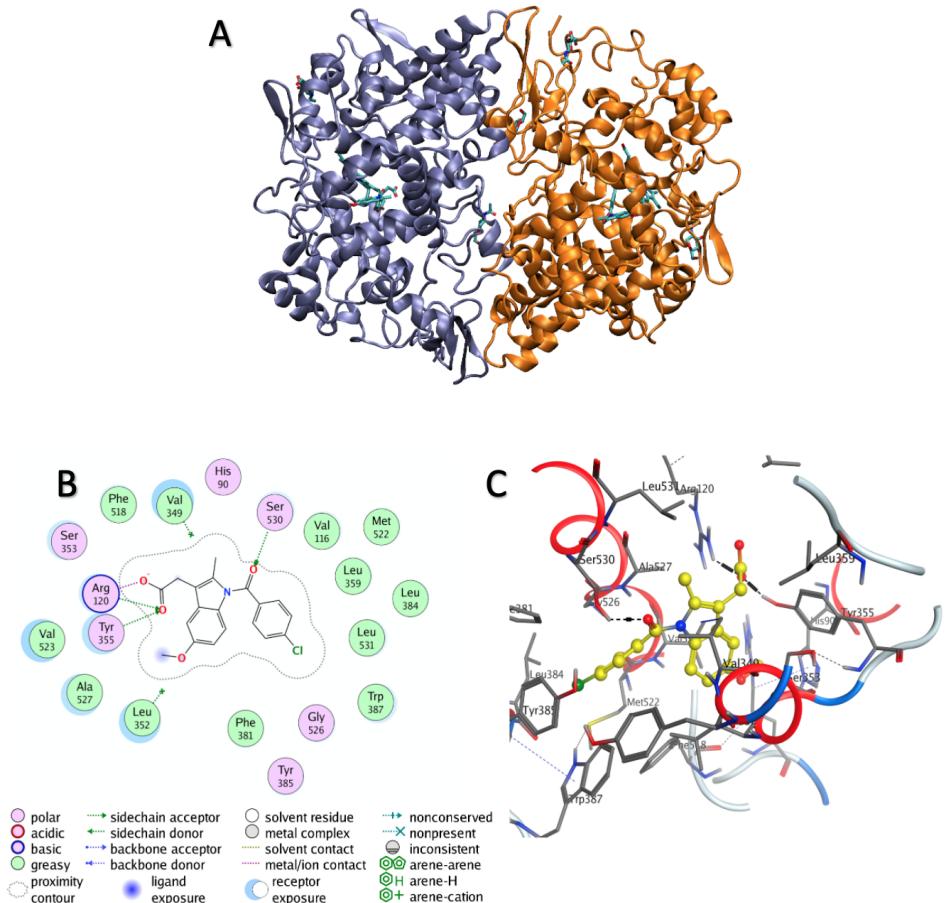


Figure 1.9: Indomethacin bound to COX-2. A) Crystal structure of the complex (PDB code 4COX), B) interactions of indomethacin in the active site of COX-2, C) 3D docking view of indomethacin binding site of COX-2. B and C modified from Bhat, et al.²¹.

Prostaglandins (family named from PGA to PGF) and other eicosanoids are made by the conversion of arachidonic acid. Nonactive arachidonic acid is changed into an active one by phospholipases. After that COXs change it to prostaglandin G2 and prostaglandin H2. From these, other prostaglandins (PGI, PGD, PGE, PGF), prostacyclins and thromboxanes, are obtained. The mechanism of the reaction converting arachidonic acid into prostaglandins G2/H2 is shown in Fig. 1.10.

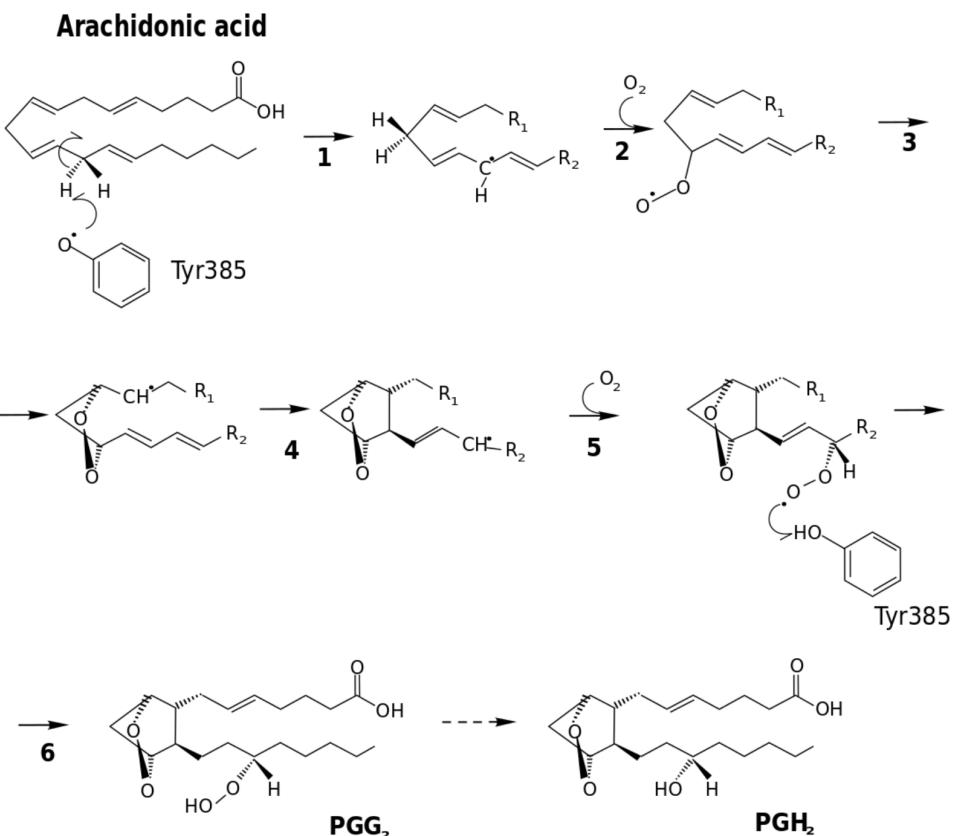


Figure 1.10: Mechanism of conversion arachidonic acid to prostaglandins with using the COX enzyme. PGG₂ denotes prostaglandin G₂, and PGH₂ denotes prostaglandin H₂. Source: Wikipedia²²

COX-1 gene is located in humans on chromosome 9 consisting of 11 exons (22 kb) code protein composed of 599 amino acids, with a molecular weight of 70 kDa^{18,23}. The gene can be only slightly regulated because of the lacking of binding sites for transcription factors. The expression of COX-1 protein is seen in most human organs and tissues. It is responsible for maintaining homeostasis in processes like cytoprotection of gastric mucosa or regulation of renal blood flow²⁴. Increased COX-1 expression has been detected in pathological conditions, i.e. in atherosclerotic plaques and rheumatoid disease (in synovial cells)^{18,23,25}.

Gene *COX-2* is located on chromosome 1 and consists of 10 exons. It has motifs (for example TATA box - non-coding sequence containing repeating T and A base pairs involved in pre-initiation complex formation²⁶) causing COX-2 transcription sensitive to stimulation by growth factors, such as epidermal growth factor, and factors involved in the inflammatory response, e.g. interleukin 1, bacterial liposaccharides, tumour necrosis factor-alpha, as well as transcription factors and oncogene proteins^{23,24,27-29}. COX-2 protein is made of 604 amino acids and has a molecular weight of 72 kDa^{23,30}. COX-2 is detected in the central nervous system, gastrointestinal tract, the blood system of the kidneys and vascular endothelium, in the placenta, as well as in the heart, cartilage and lungs²⁵.

It is also known that COX-1 is constitutive isoform, what is considered as enzyme which expression is wide in most tissues under normal conditions¹⁶, while COX-2 is an inducible enzyme, which expression can be up-regulated in pathological processes in response to pro-inflammatory compounds, such as cytokines, mitogens or tumour promoters.¹⁶ However there are studies suggesting upregulation of both enzymes in pathological conditions³¹. That can be seen especially in inflammation, pain, fever and in tumour cells.

The expression of COX-2 can be suppressed by anti-inflammatory factors - interleukins-10, interleukin-4, interleukin-13 and glucocorticoids^{30,32}.

The protein is easily found in cells of several different solid tumour: cancers of stomach, colon, pancreas, lung, breast, head and neck, endometrium, prostate and bladder^{15,27,33-38}. Moreover, COX-2 overexpression is found in all stages of cancer³⁹. Overexpression of COX-2 has been observed in conditions prior to lung cancer, such as atypical adenomatous hyperplasia, atypical alveolar epithelium, high-grade dysplasia, and carcinoma in situ^{36,40,41}. In case of non-small cell lung cancer, overexpression of COX-2 has been observed in 50-70% of adenocarcinoma cases^{36,42} and 20-30% of squamous cell carcinoma cases^{43,44}. Correlation between expression of the enzyme and features like age, sex or stage of the disease was not observed⁴⁴⁻⁴⁶.

Studies show that COX-2 protein is involved in many processes leading to the formation and development of cancer⁴⁷⁻⁵², by inducing phenomena of stopped apoptosis, intensification of neoangiogenesis, increase in the activity of matrix metalloproteinase and stimulation of the synthesis of adhesive proteins and weakening of the activity of the immune system. We describe here in brief three most significant cases in which COX-2 is overexpressed:

- 1) The programmed cell death, called apoptosis, is crucial for proper functioning of a multicellular organism, which is very organized and strongly regulates the number of its cells. Proper functioning of the apoptosis mechanism depends on the balance between pro-apoptotic factors (including Bax, Bak proteins) and anti-apoptotic factors such as Bcl-2 and Bcl-X1 proteins^{53,54}. It has been observed that increased COX-2 expression inhibits cell death by increasing the activity of the anti-apoptotic protein Bcl-2⁵⁵.
- 2) Neoangiogenesis is a process of the creation of blood vessels network. It is necessary for the growth and progression of neoplastic tumours with a diameter greater than 1-2 mm⁵³. There are many factors involved in the formation of new blood vessels within a growing tumour. They include, among others vascular epithelial growth factor (VEGF), transforming growth factor α and β (TGF α and β), basic fibroblast growth factor (bFGF) and the chemokines ENA-78 and IL-8. It has been found in both in vitro and in vivo that an increase in COX-2 expression in tumours leads to the formation and growth of new vessels^{56,57} at many stages.
- 3) The infiltration of the surrounding tissues by the neoplastic tumour and the formation of distant metastases is related to the interaction of tumour cells with the extracellular matrix and adhesive proteins. Adhesion to the extracellular substance is the most important stage that initiates metastasis formation. This process in many cancers depends on the CD44⁵⁸ protein. CD44 is a surface receptor for hyaluronan glycosaminoglycan, which is the main component of the extracellular matrix. Lung cancer cell lines overexpressing COX-2 also overexpress CD44. The addition of specific CD44 inhibitors significantly reduced the invasive capacity of the cells of the cell lines. Previous studies suggest that overexpression of COX-2 influences the increased invasiveness of non-small cell lung cancer by CD44^{59,60}.

Key to differentiate the tumour cells from normal inflammation is the presence of COX-2 in either its monomeric or dimeric form. In fact, it has been seen that COX-2 in cancer cells are overexpressed in the dimer form¹⁵(see Fig. 1.11).

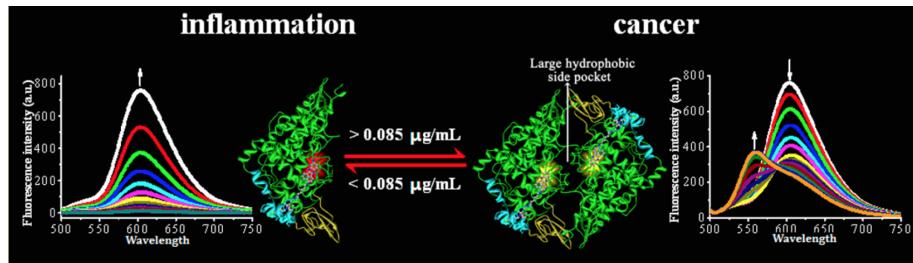


Figure 1.11: Diagram showing the difference between COX-2 in inflammation and cancer. Image from Zhang et al., 2013¹⁵

As a result of the different expression forms, COX-2 is one of the most interesting targets for the detection of cancer cells. Conventional methods to the detection of COX-2 in situ, such as immunohistochemical analysis, electrochemical analysis and tomography has poor spatial and temporal resolution. A better way to achieve this goal is imaging the presence of the protein using fluorescence probe-based spectroscopy. This allow for an accurate, non-invasive, high-resolution imaging of the COX-2⁶¹. The use of fluorescent probes is one of the best ways to assess the presence of Cox-2 in either its monomeric or dimeric form. In order to take full advantage of the fluorescent probes' ability, two different types of probes are commonly used. The first type is an active probe which fluorescence enhances after selectively binding to a protein. We call this "turn-on" mechanism. The second type of probe is named activatable probes (or "off-on" mechanism). Such a probe has no or very weak fluorescence before binding to the enzyme, because of the quenching processes. However, after attaching to the protein, it exhibits strong fluorescence⁶¹. The diagram showing the two types of fluorescent probe mechanisms is shown in Fig. 1.12.

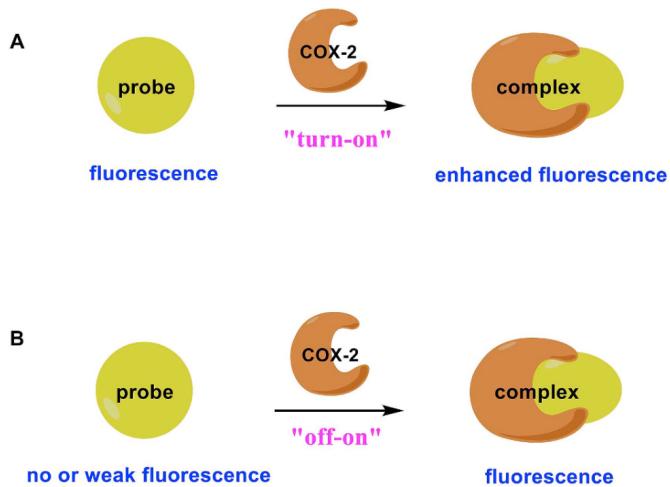


Figure 1.12: Two main types of fluorescent probes. Image from Chen et al., 2019⁶¹.

Weak fluorescence may be caused by quenching processes, such as Förster resonance energy transfer (FRET), photon-induced electron transfer (PET) or intramolecular charge transfer (ICT) .

1.4. Subject of the study

1.4.1. ANQ-ICM-6 as a fluorescent probe

The probe considered in this study is the ANQ-IMC-6 probe. It consists of the COX-2 inhibitor indomethacin (IMC) and the fluorophore acenaphtho[1,2-b]quinoxaline (ANQ). These two moieties are linked together by a six-atom aliphatic linker⁶². The structure of the probe is shown in Fig 1.13. While the inhibitor binds to COX-2, the ANQ carries the fluorescent properties.

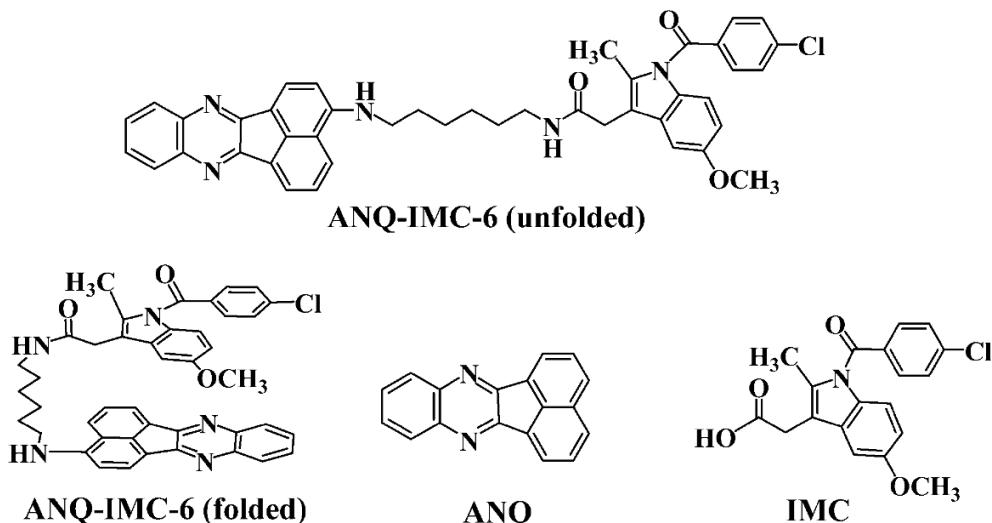


Figure 1.13: ANQ-IMC-6 probe structure. Image from Zhang et al., 2013¹⁵

The probe shows an 'off-on' mechanism: in its folded conformation, it shows weak fluorescence, while bound to COX-2 a very strong fluorescence is seen¹⁵ (see Fig. 1.14 and Fig. 1.15). ANQ-IMC-6 accumulates in the Golgi apparatus (where concentration of COX-2 has been observed⁶³) and uses COX-2 as target for imaging. In the work of Zhang et al., 2013¹⁵ it is suggested (by using molecular docking techniques) that the probe binds to COX-2 amino-acids Arg120, Tyr355 and Glu522. It interacts with the protein in the same manner as IMC to COX-2, known as reported in the literature^{64,65}. For the ANQ-IMC-6 probe, a large Stokes shift (90 nm) "due to the push-pull internal charge-transfer mechanism associated with the ANQ scaffold"¹⁵ is reported. The higher the concentration of the COX-2 molecules, the higher fluorescence intensity of the probe. Maximum of the fluorescence intensity is reported around 560 nm emission wavelength (see spectrum in Fig. 1.16)¹⁵.

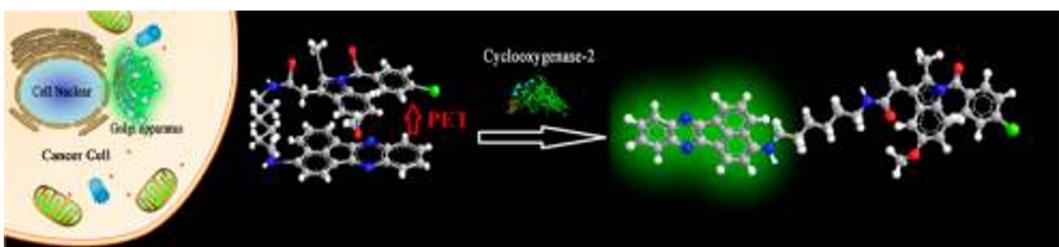


Figure 1.14: Fluorescence mechanism for identification presence of COX-2. Image from Zhang et al., 2013¹⁵

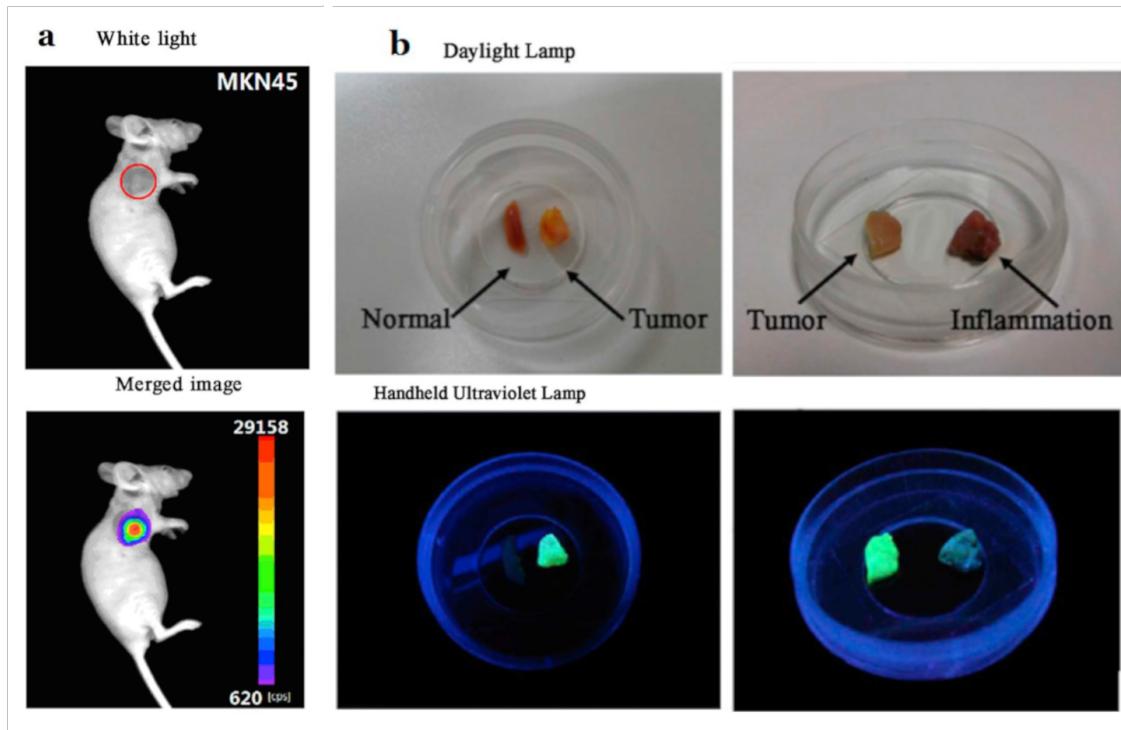


Figure 1.15: (a) In vivo imaging of carrageenan-induced COX-2 in mice with ANQ-IMC-6 probe ($30 \mu\text{m}$). (b) Visualization of tumour resection by the naked eye under ultraviolet illumination. Image modified from Chen et al., 2019⁶¹ based on work of Zhang et al., 2013¹⁵

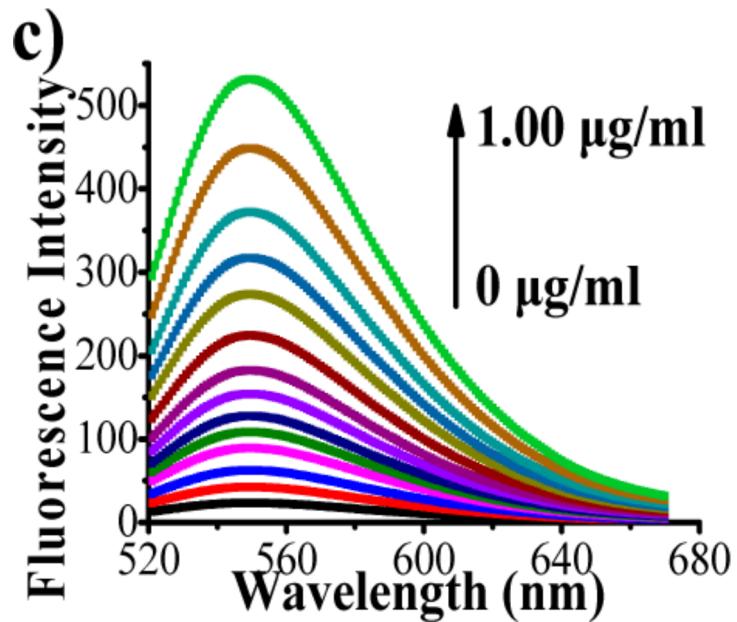
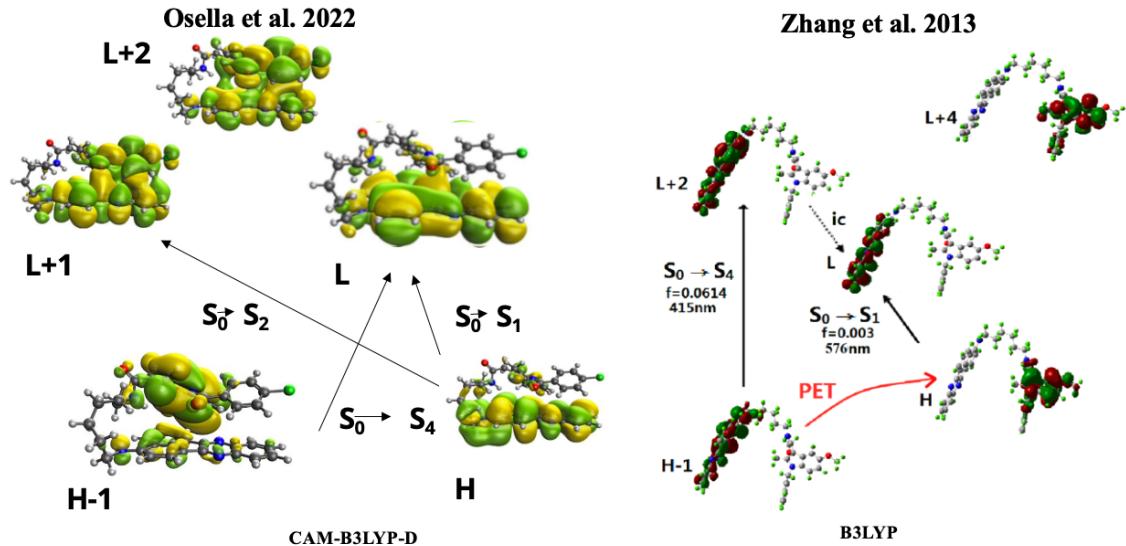


Figure 1.16: Fluorescence spectra for increasing amount of COX-2. Image modified from Zhang et al., 2013¹⁵.

In literature, 'off-on' mechanism of the probe is described by Zhang et al.¹⁵ as intense fluorescence is caused by restraining photo-induced electron transfer (PET) between ANQ and IMC after binding of the probe to the enzyme. The problem with this description is that the authors do not go into details of the PET process. Moreover, the orbital assignment (see Fig. 1.17) is completely different than in the work of our group⁶². Therefore the aim of this thesis is to make an in deep investigation of the mechanism of action of the probe when interacting with COX-2, via a novel multiscale computational approach developed in our group.

A. Folded conformation



B. Unfolded conformation

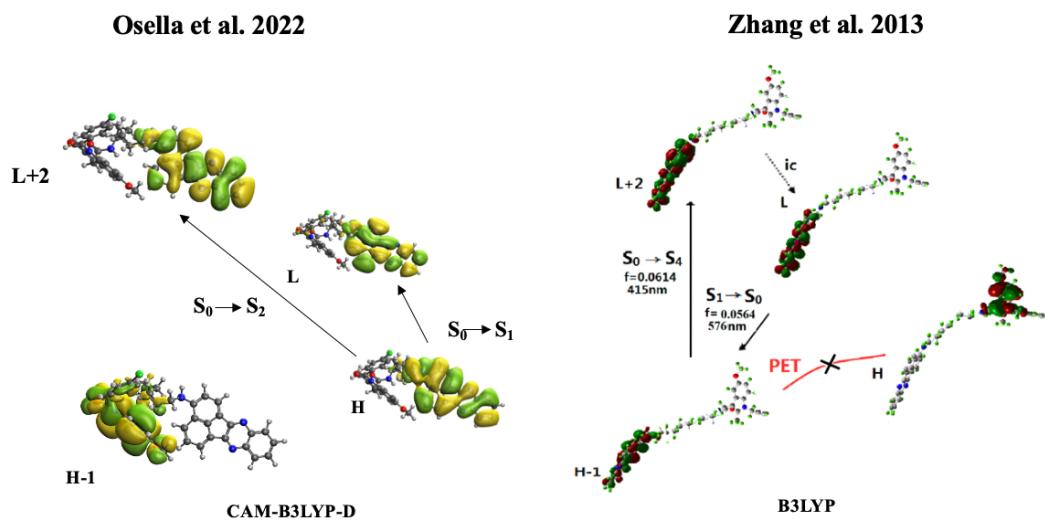


Figure 1.17: Comparison of the orbitals and electronic transitions between them. On the left side: work of our group (from supplemental information to Osella et al., 2022⁶²), on the right side: previous computation from the literature (Zhang et al., 2013¹⁵).

Chapter 2

Methods

2.1. Molecular dynamics

The study of the structure and dynamics of the systems consisting of biological macromolecules are crucial for the understanding of their molecular biology and the biophysics of their molecular interactions can be studied by a variety of methods. Molecular Dynamics (MD) among them is very powerful and can describe what is really difficult to get by experiments. MD can be used to get information about enzyme dynamics (mechanism and regulation), molecular transport (i.e. across bio-membranes), building viral capsids, DNA transcription control, DNA packaging in the nucleus, protein-protein interactions, macromolecular crowding, conformational changes, protein folding and more⁶⁶.

In this work MD simulation was used to achieve multiple different conformations of the probe-protein complex, which were later used as initial structures for ab initio hybrid quantum-classical dynamics, described in the next sections below.

2.1.1. Basic principles

MD simulations get information about atoms movement by integrating Newton's equations of motion for a system of N interacting atoms:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, i = 1, 2, \dots, N \quad (2.1)$$

while acting forces are defined as negative derivatives of potential $V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$

$$\mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i} \quad (2.2)$$

At each small time step, the simulation program solves equations. At each step, Cartesian coordinates are written to an output file, so we have access to the system geometry at regular intervals. Coordinates in function of time are called trajectory of the system. After a certain amount of time, the system reaches equilibrium, and from then, production MD run can be performed to collect and analyse the data. Much information about macroscopic behaviour of the system can be extracted by using the production trajectory⁶⁷.

2.1.2. Approximations

MD calculations are somehow limited by the use of different approximations.

At first, MD simulation are classical, which means that it use classical mechanics to

describe motion. This implies that we do not have any information about electrons in the system, what is sometimes needed. In addition, each system is described by a force field, in which empirical parameters are used to describe the chemical surrounding of each atom. Bonding parameters such as bond length, planar and dihedral angles, together with non-bonding interaction like Coulomb and van der Waals need to be fitted for each different chemical surrounding, if they are not present in the original force field.

2.1.3. Unit cell and periodic boundary conditions

When running MD simulations we also need to repair unnatural boundary condition, for example using periodic boundary conditions (PBC), meaning that we use a large number of unit cells (boxes) containing our system, so when an atom move through one border of the unit cell, it will re-appear on the opposite side with the same velocity value and direction⁶⁸. In the Fig. 2.1 PBC in 2 dimensions is shown. Added unit cells are images of our system. We can use different shapes of the box, but usually a cubic cell is considered.

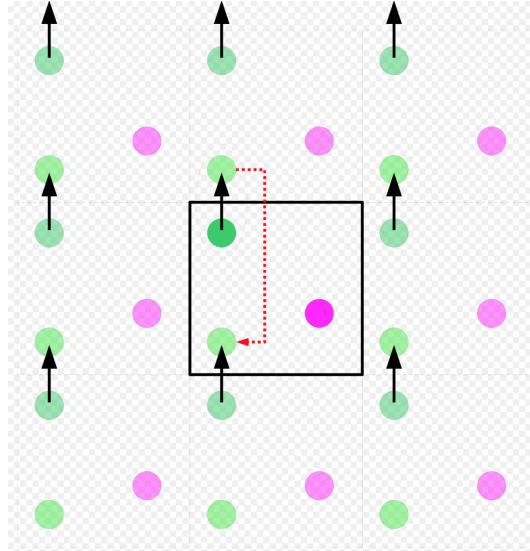


Figure 2.1: Periodic boundary conditions in 2D. Image from Wikipedia⁶⁸, made by I, Grimlock.

The box is described by image distance d , box volume V , box vectors \mathbf{a} , \mathbf{b} , \mathbf{c} and box vector angles $\angle ab$, $\angle ac$ and $\angle ab$. In the cubic box $V = d^3$. Box vectors equal to:

$$\begin{aligned}\mathbf{a} &= (d \quad 0 \quad 0) \\ \mathbf{b} &= (0 \quad d \quad 0) \\ \mathbf{c} &= (0 \quad 0 \quad d)\end{aligned}\tag{2.3}$$

and angles are all equal to 90° . When using PBC conditions, usually only the nearest image of each particle is considered for short-term non-bonded interactions.

2.1.4. Cut-off restrictions

Another approximation to consider is that long-range interactions (like Lennard-Jones or Coulomb interactions) are cut off by using an user defined radius. When PBC are applied, the distance of interactions considered has to be defined by half of the shortest box vector⁶⁷:

$$R_c < \frac{1}{2}\min(a, b, c)\tag{2.4}$$

so that the cut-off distance will not be more than one image.

The position of the atom is described by using position vector of the particle i:

$$\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i \quad (2.5)$$

The force F_{ij} is the force on i exerted by j

2.1.5. Global MD procedure

The algorithm for doing the MD simulation is presented below⁶⁷:

1. As an input initial conditions for MD simulation we need positions \mathbf{r} and velocities \mathbf{v} of all atoms in the system.
2. In each simulation step, we calculate forces \mathbf{F}_i by computing the force between non-bonded atom pairs:

$$\mathbf{F}_i = \sum_j \mathbf{F}_{ij} \quad (2.6)$$

and adding forces from bonded interactions and external forces.

3. Next step is to update the initial configuration by moving the atoms by Newton's equations of motion computed numerically.
4. After this, we can create an output file with atomic positions (geometry of the system), energy terms (i.e. potential, kinetic, Coulomb etc.), velocities, temperature and pressure and so on.

Steps 2,3 and 4 are then repeated for a defined number of steps, until the equilibrium is reached. After that, the procedure is repeated to obtain the production trajectory.

2.1.6. Computed variables

The kinetic energy of the system for N particles is given by:

$$E_k = \frac{1}{2} \sum_{i=1}^N m_i v_i^2 \quad (2.7)$$

The temperature of the system can be then calculated using equipartition theorem:

$$\frac{1}{2} N f k T = E_k \quad (2.8)$$

where f is the number of degrees of freedom:

$$f = 3N - \text{constraints} - 3 \quad (2.9)$$

We are subtracting 3 from degrees of freedom because of the 3 centre-of-mass velocities usually are set to zero⁶⁷.

The pressure tensor can be computed by subtracting the virial \mathcal{V} from kinetic energy E_k ⁶⁷:

$$\mathbf{P} = \left(\frac{2}{V} E_k - \mathcal{V} \right) \quad (2.10)$$

while the virial \mathcal{V} is given by:

$$\mathcal{V} = \frac{1}{2} \sum_{i < j} \mathbf{r}_{ij} \otimes \mathbf{F}_i \quad (2.11)$$

2.1.7. Integrating algorithms

For integration of the Newton's equations of motion we need another algorithm, which is called MD integrator. One of the most used integrators is the leap-frog algorithm⁶⁹. The leap-frog algorithm uses positions \mathbf{r} and time t and velocities \mathbf{v} at time $t - \frac{1}{2}\delta t$ and updates them via relations⁶⁷:

$$\mathbf{v} \left(t + \frac{1}{2}\delta t \right) = \mathbf{v} \left(t - \frac{1}{2}\delta t \right) + \frac{\delta t}{m} \mathbf{F}(t) \quad (2.12)$$

and

$$\mathbf{r}(t + \Delta) = \mathbf{r}(t) + \Delta t \mathbf{v} \left(t + \frac{1}{2}\Delta t \right) \quad (2.13)$$

Another possible integrator is the Verlet algorithm⁷⁰, which uses positions \mathbf{r} and velocities \mathbf{v} at time t and updates using relations⁶⁷:

$$\mathbf{v}(t + \frac{1}{2}\Delta t) = \mathbf{v}(t) + \frac{\Delta t}{2m} \mathbf{F}(t) \quad (2.14)$$

and

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t \mathbf{v}(t + \Delta t) \quad (2.15)$$

2.1.8. Conserving the temperature and pressure

Nosé-Hoover thermostat

For MD calculations as well as other simulations we may use thermostat to have constant temperature throughout the simulation. One of the examples of such thermostat is the Nosé-Hoover algorithm^{71?}. We then extend the Hamiltonian of the system with thermal reservoir and friction terms. The equations of motion then are replaced by⁶⁷:

$$\frac{d^2 \mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - \frac{p_\zeta d\mathbf{r}_i}{dt} \quad (2.16)$$

where p_ζ is the momentum of dynamic friction parameter ζ and own equation of motion:

$$\frac{dp_\zeta}{dt} = (T - T_0) \quad (2.17)$$

where T_0 is current instantaneous temperature and T is the target reference temperature.

Berendsen algorithm

We might also need constant pressure in the simulation (NPT ensemble). It can be done by using Berendsen algorithm⁷², which rescales the positions and box vectors in each step with a matrix μ , with the effect of a first-order kinetic relaxation of the pressure \mathbf{P} towards target reference pressure \mathbf{P}_0 :

$$\frac{d\mathbf{P}}{dt} = \frac{\mathbf{P}_0 - \mathbf{P}}{\tau_p} \quad (2.18)$$

with matrix μ defined by:

$$\mu_{ij} = \delta_{ij} - \frac{n_{PC}\delta t}{3\tau_p} \beta_{ij} [P_{0ij} - P_{ij}(t)] \quad (2.19)$$

where β is the isothermal compressibility of the system.

Parrinello-Rahman algorithm

Another procedure that can be used to conserve the pressure is Parrinello-Rahman^{73?} algorithm, which is more accurate and thus better for production runs.

In the Parrinello-Rahman barostat the matrix equation of motion is defined as:

$$\frac{d\mathbf{b}^2}{dt^2} = V\mathbf{W}^{-1}\mathbf{b}'^{-1}(\mathbf{P} - \mathbf{P}_{ref}) \quad (2.20)$$

where the matrix \mathbf{b} represents the box vectors, V is the volume and matrices \mathbf{P} and \mathbf{P}_{ref} denotes current and target temperature, respectively.

The Parrinello-Rahman modified equations of motion for the particles are described as:

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - \mathbf{M} \frac{d\mathbf{r}_i}{dt} \quad (2.21)$$

$$\mathbf{M} = \mathbf{b}^{-1} \left[\mathbf{b} \frac{d\mathbf{b}'}{dt} + \frac{d\mathbf{b}}{dt} \mathbf{b}' \right] \mathbf{b}'^{-1} \quad (2.22)$$

The strength of the coupling and possible deformations of the box are depicted by the inverse mass parameter matrix \mathbf{W}^{-1} :

$$(\mathbf{W}^{-1})_{ij} = \frac{4\pi^2\beta_{ij}}{3\tau_p^2 L} \quad (2.23)$$

where β_{ij} are isothermal compressibilities, L is the largest box matrix element and τ_p is the pressure time constant, presented in the input file.

2.1.9. Force field

A force field defines the functional of the potentials and a set of parameters needed to describe bonding and non-bonding interactions. Bonding interactions are chemical bonds, planar and dihedral angles, while non-bonding terms include intramolecular pair interactions like Coulomb, van der Waals and Lennard-Jones. These parameters must be defined for each different chemical environment of the atoms present in the investigated system. Several used force fields have been developed during the years, such as CHARMM⁷⁴, AMBER⁷⁵ or OPLS⁷⁶. Each force filed have a different optimization of bonded and non-bonded parameters and better describes different types of chemical systems. For protein simulations, the CHARMM force field is widely used. The functional form of the potential energy in CHARMM is described as⁷⁴:

$$\begin{aligned}
 V(r) = & \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \\
 & + \sum_{\text{dihedrals}} K_\chi(1 + \cos(n\chi - \sigma)) \\
 & + \sum_{\text{Urey-Bradley}} K_{UB}(S - S_0)^2 \\
 & + \sum_{\text{impropers}} K_\phi(\phi - \phi_0)^2 \\
 & + \sum_{\text{nonbonded atom pairs}} \left(\varepsilon_{ij} \left[\left(\frac{\mathbf{R}_{min,ij}}{\mathbf{r}_{ij}} \right)^{12} - \left(\frac{\mathbf{R}_{min,ij}}{\mathbf{r}_{ij}} \right)^6 \right] + \frac{q_i q_j}{\varepsilon_D \mathbf{r}_{ij}} \right)
 \end{aligned} \tag{2.24}$$

In above equation notation is the same as in MacKerell et al., 2001⁷⁴: "(...) K_b , K_θ , K_{UB} , K_χ , K_ϕ are the bond, valence angle, Urey–Bradley, dihedral angle, and improper dihedral angle force constants, respectively; b , θ , S , χ and ϕ are the bond length, bond angle, Urey–Bradley 1,3 distance, dihedral torsion angle, and improper dihedral angle, respectively. The subscript zero represents the equilibrium values for the individual terms. In the dihedral term n is the multiplicity and δ is the phase, which dictates the location of the minima and maxima. (...) ε is the LJ well-depth and R_{min} is the distance at the LJ minimum interaction energy; q_i is the partial atomic charge and ε_D is the dielectric constant; r_{ij} is the distance between atoms i and j , respectively. For explicit treatment of solvent ε_D is set to 1".

The above potential $V(r)$ includes all possible bonds, angles and dihedrals, as well as limited number of Urey-Bradley terms (cross-term accounting for angle bending using 1,3 nonbonded interactions) and improper dihedrals. As treating explicitly all nonbonded interactions have a big computational costs, a cut-off is considered. Directly truncating atom interactions can cause problems like change in energies and forces at the cut-off distance. To avoid that, CHARMM uses the Ewald method, which means that the interactions are splitted into direct and reciprocal spaces, which sums up to treat all the interactions present in the PBC systems⁷⁷. The newer implementation of this is paricle mesh Ewald (PME)⁷⁸, which lower the computational cost of the Ewald method and it is also included in the CHARMM force field.

2.2. Density Functional Theory

One of the most-used quantum-mechanical computational methods for the calculation of the electronic structure of atoms, molecules and solids in their ground state is the density functional theory (DFT), for which development Walter Kohn received the Nobel Prize award in 1998⁷⁹. An extension of it called time-dependent DFT (TDDFT) is used to predict electronic excitation energies.

The DFT method is widely used because of the good accuracy achievable with the proper choice of functional and basis set and its high scalability with the number of atoms in a system (N^3), while the so-called post-HF methods can be more accurate but with a higher computational cost (N^5 , N^6), making DFT the method of choice for computation

2.2.1. Schrödinger equation

By means of quantum mechanics, to get information about studied system, we need to solve the Schrödinger equation. For the non-relativistic case, when using Born-Oppenheimer approximation (so the motion of the electrons is independent of motion of the nuclei, because they have much less mass than the nuclei), time-independent Schrödinger equation for the wave function (of $3N$ coordinates when spin is omitted) Ψ is given by⁸⁰:

$$\hat{H}\Psi(r_1, r_2, \dots, r) = E\Psi(r_1, r_2, \dots, r) \quad (2.25)$$

where \hat{H} is the Hamiltonian operator of the system, r_i the coordinates of each electron and N is the number of electrons in the system.

The hamiltonian is then defined as sum of the kinetic energy operator \hat{T}_e , potential energy operator \hat{V}_e and external potential operator \hat{V}_{ex} which considers the interactions between electrons and nuclei:

$$\hat{H} = \hat{T}_e + \hat{V}_e + \hat{V}_{ex} \quad (2.26)$$

The external potential considers the interaction between electrons and nuclei.

Expected value of the Hamiltonian will get us the average of the total energy:

$$E[\Psi(r_1, \dots)] = \int \Psi^*(r_1, \dots) \hat{H} \Psi(r_1, \dots) dr_1 dr_2 \dots = \langle \Psi | \hat{H} | \Psi \rangle \quad (2.27)$$

The variational principle:

$$E[\Psi] \geq E_0 \quad (2.28)$$

allows us to find the ground state wave function by investigating all possible functions and getting the one that minimizes the total energy. However, this searching among all possible wave functions is very costly in terms of the computational performance. Moreover, the wave function has no physical meaning and cannot be measured by experiments. Better would be to use the electron density $\rho(r)m$ (function of only 3 coordinates!), since it allows to extract desired information from the density itself.

2.2.2. Hohenberg-Kohn Theorems

The connection between the electron density and the wave function of a given system is given by Hohenberg-Kohn theorems⁸⁰:

- a) The external potential and the Hamiltonian are determined by the electron density $\rho(r)$. The functional of $\rho(r)$ ³ is the ground state. It's because
- b) For any positive definite $\rho(r)$, for which $\int \rho(r)dr = N$, there is a variational principle, so:

$$E[\rho(r)] \geq E_0 \quad (2.29)$$

then the density which minimise the total energy is the ground state density^{81,82}.

The theorems introduce the notion of functional. By minimising it, the ground state density can be obtained. That is why the method is called Density Functional Theory. As we do not know the exact energy functional for our system, we can use the Kohn-Sham approach in which we consider known functional and make corrections. We can thus reformulate Eq. 2.26 in terms of density functionals:

$$E[\rho(r)] = T_e[\rho(r)] + V_e[\rho(r)] + V_{ex}[\rho(r)] \quad (2.30)$$

where the external potential functional is defined as:

$$V_{ext}[\rho(r)] = \int \hat{v}_{ext}\rho(r)dr \quad (2.31)$$

The idea introduced by Kohn and Sham is to consider an imaginary system of non-interacting electrons with the same ground state density as the real system with electrons which do interact and the same positions and atomic numbers of the nuclei. We can then simplify the Hamiltonian as consisting of the sum of one-electron operators and the energy functional is then reformulated as:

$$E[\rho(r)] = T_e[\rho(r)] + V_e[\rho(r)] + V_{ex}[\rho(r)] + E_{xc}[\rho(r)] \quad (2.32)$$

The kinetic energy consists now of only classical repulsion between electrons, while $E_{xc}[\rho(r)]$ is the new exchange-correlation functional, by which we correct all simplifications, such as the quantum mechanical exchange and correlation effects and non-classical corrections for repulsion effects. In terms of an orbital basis set it is now defined as⁸⁰:

$$E[\rho(r)] = \sum_i^N \left(\langle \chi_i | -\frac{1}{2} \nabla_i^2 | \chi_i \rangle - \langle \chi_i | \sum_k^{nucle} \frac{Z_k}{|r_i - r_k|} | \chi_i \rangle \right) + \sum_i^N \langle \chi_i | \frac{1}{2} \int \frac{\rho(r')}{|r_i - r'|} dr' | \chi_i \rangle + E_{xc}[\rho(r)] \quad (2.33)$$

where $\rho(r)$ is the exact density for Slater determinantal wave function:

$$\rho(r) = \sum_i^N |\chi_i|^2 \quad (2.34)$$

Finding the exact exchange-correlation term E_{xc} , for which the functional is not known, is now the only problem. This term will give us the exact density and energy. Since this is not possible, approximations in terms of the functionals have been introduced (see 2.2.5).

2.2.3. Time Dependent DFT

As conventional DFT solves only problems for ground-state, when examining excited states and electronic transitions, we need to employ the time-dependent version of DFT. Going back to the Schrödinger equation now in its time-dependent form:

$$i \frac{\partial}{\partial t} \Psi(r, t) = \hat{H}(r, t) \Psi(r, t) \quad (2.35)$$

gives us additional time-dependent field $\hat{V}(t)$ in our Hamiltonian:

$$\hat{H} = \hat{T}_e + \hat{V}_e + \hat{V}_{ex} + \hat{V}(t) \quad (2.36)$$

To be able to study the optical properties of the systems, we need to calculate the excitation energies and the corresponding oscillation strength (giving us the probability of a transition) and employing the DFT methods since it is very good in terms of performance cost. As TDDFT is very good compromise between accuracy and computational cost, it is widely used to calculate excite states of large molecules⁸³

2.2.4. Runge-Gross Theorem

Again, like for ground-state DFT, we need a theorem to connect the density with the wave function. For this we can use the theorem developed by Runge and Gross⁸⁴. This theorem states that the time-dependent external potential $v_{ext}(r, t)$ is determined by the exact time-dependent electron density $\rho(r, t)$ up to a constant and the wave function $\Psi(r, t)$ within a time-dependent phase factor. This implies that the time-dependent wave function is a functional of the time-dependent density⁸⁰:

$$\Psi(r, t) = \Psi[\rho(t)](t) e^{-\alpha(t)} \quad (2.37)$$

2.2.5. Functionals and Basis Sets

Equation 2.32 introduced the exchange-correlation energy (XC) functional term, which, in comparison to other exactly solvable terms in the equation needs to be approximated. On this approximation, results' quality depend, so the advances in DFT-based methods are primary oriented on developing XC functionals. There have been proposed many different functionals, which can be used. Choice of the functional depends on the system we study, as there are many different chemical systems in nature. It is important to understand that no existing functional is highly accurate for any system and even within a system not for all its properties⁸⁵.

Functional have been developed with increasing accuracy, following the 'Jacob's ladder' method, in which each next generation is an improvement of the previous one. The first functionals considered only local density approximations (LDA) – density approximated as for uniform electron gas density at a point; next, generalized gradient approximations (GGAs) – at each point use the density and its gradient; hybrid functionals – mixing some exact exchange with GGA; fully nonlocal functionals – include unoccupied orbitals too. The example of a popular hybrid functional is B3LYP⁸⁶ and in this work we used

a modification of it, called CAM-B3LYP⁸⁷, which combines B3LYP with the long-range correction, proposed by Tawada et al. in 2004⁸⁸.

For quantum chemistry calculations such as DFT, beside the functional, also the basis set must be specified. Generally, a basis set is a set of vectors defining a space (for example i, j,k defining a Cartesian 3D space). In quantum chemistry, a basis set is the collection of non-orthogonal one-particle functions, which are used to build molecular orbitals⁸⁹. In other words, the basis set is a set of functions by which we can turn the model partial differential equations into equations suitable to be calculated on the computer⁹⁰.

The orbitals are then approximated as a linear combination of the basis functions:

$$|\Psi\rangle \approx \sum_{\mu} c_{\mu i} |\mu\rangle \quad (2.38)$$

where expansion coefficients $c_{\mu i}$ are defined as:

$$c_{\mu i} = \sum_{nu} \langle \mu | \nu \rangle^{-1} \langle \nu | \psi_i \rangle \quad (2.39)$$

Basis set can describe atomic orbitals of Gaussian type, Slater type, or numerical atomic orbitals⁹¹. Bigger set will usually give more accurate results, but it has a higher computation cost and the computational time is higher.

Minimal basis sets are called STO-nG and consists n primitive Gaussian orbitals fitted to a single Slater-type orbital (STO). For example, for carbon atom to have 1s, 2s and three 2p orbitals, STO-3G basis set would be needed. Expansion upon the minimal STO-3G set is the so-called "split-valence basis set", in which more than one function per valence electron⁹² is considered. Example of this basis set type is 3-21G – small, not costly basis set, where single basis functions consisting of three Gaussians (**3-21G**) describe non-valence electrons and two basic functions describe a valence electron (**3-21G**). Similarly, medium size basis set can be 6-31G⁹³ and big 6-31G(d) (where d denotes the added Gaussian polarization function for heavy atoms (d))⁹⁴.

2.2.6. Configuration Interaction Singles (CIS)

A different approach to calculate the excitation energies and get information about transition processes than DFT-base methods is to use ab initio wave function based methods. For the interpretation of the electronic spectra, configuration interaction (CI) among all single determinants using a Hartree-Fock reference state, can be used⁹⁵. In CI, a ground-state reference Hartree-Fock single-determinantal wave function Ψ_{HF} is selected:

$$\Psi_{HF} = (n!)^{-1/2} \det[\chi_1 \chi_2 \dots \chi_i \chi_j \dots \chi_n] \quad (2.40)$$

where n is the number of electrons and χ are spin orbitals represented in a basis of N atomic basis functions Φ_{μ} :

$$\chi_p = \sum_{\mu} c_{\mu p} \Phi_{\mu} \quad (2.41)$$

The notation is the same as in Foresman, 1991 paper introducing "CI-singles" (CIS) method: "subscript notation (...) $\mu, \nu, \lambda, \sigma, \dots$, denote atomic basis functions; i, j, k, I, \dots , denote molecular orbitals which are occupied in the ground state; a, b, c, d, \dots , denote virtual

molecular orbitals unoccupied in the ground state; p, q, r, s, \dots , denote generic molecular spin orbitals"⁹⁵

Hartree-Fock equations can be solved by using the self-consistent field (SCF) procedure:

$$\sum_{\mu} (F_{\mu\nu} - \varepsilon_p S_{\mu\nu}) c_{\nu p} = 0 \quad (2.42)$$

where $F_{\mu\nu}$ is the Fock matrix:

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda\sigma} \sum_i c_{\mu i} c_{\nu i} (\mu\lambda||\nu\sigma) \quad (2.43)$$

where $H_{\mu\nu}$ is the one-electron core Hamiltonian and the antisymmetrized two-electron integrals $(\mu\lambda||\nu)$ are given by:

$$(\mu\lambda||\nu) = \int \int \Phi_{\mu}(1)\Phi_{\nu}(2)(1/r_{12})[\Phi_{\lambda}(1)\Phi_{\sigma}(2) - P\phi_{i\sigma}(1)\Phi_{\lambda}(2)]dr_1 dr_2 \quad (2.44)$$

while $S_{\mu\nu}$ is the overlap matrix:

$$S_{\mu\nu} = \int \Psi_{\mu} \Psi_{\nu} dr \quad (2.45)$$

and ε_p is the energy of orbital p for one electron.

Once the above equations are solved, the total energy of the ground state is given by a single determinant defined as:

$$E_{HF} = \sum_{\mu\nu} P_{\mu\nu}^{HF} H_{\mu\nu} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\mu\nu}^{HF} P_{\lambda\sigma}^{HF} (\mu\lambda||\nu) + V_{nuc} \quad (2.46)$$

while P^{HF} is the Hartree-Fock density defined as sum over the occupied orbitals:

$$P_{\mu\nu}^{HF} = \sum_{i=1}^n c_{\mu i} c_{\nu i} \quad (2.47)$$

and V_{nuc} is the energy of repulsion of the nuclei.

The above equations describe only one determinants for the electronic wave function.

When an occupied spin orbital is replaced with a virtual spin orbital, $n(N-n)$ possible determinants are made. The wave functions and its energies, can be described as:

$$\Psi_{ia} = (n!)^{-1/2} \det [\chi_1 \chi_2 \dots \chi_a \chi_j \dots \chi_n] \quad (2.48)$$

$$E_{i\sigma} = E_{HF} + \varepsilon_{\alpha} - \varepsilon_i - (ia||ia) \quad (2.49)$$

The "CI-singles" (CIS) approach is used to write the excited-state wave function as linear combination of all possible single excited determinants:

$$\Psi_{\text{CIS}} = \sum_{ia} \psi_{ia} S \quad (2.50)$$

Configuration interaction (CI) coefficients are then normalized eigenvectors of the Hamiltonian matrix:

$$\langle \psi_{ia} | H | \psi_{jb} \rangle = [E_{HF} + \varepsilon_a - \varepsilon(i)] \delta_{ij} \delta_{ab} - (ja||ib) \quad (2.51)$$

2.2.7. Equation-of-Motion Couple Cluster Singles and Doubles (EOM-CCSD)

Electronically excited states can be also described similarly to coupled-cluster ground-state theory by applying equation-of-motion methods (EOM). EOM is similar to CIS and the EOM states can be found by diagonalization of the similarity transformed Hamiltonian⁹⁶:

$$\bar{H} = e^{-T} H e^T \quad (2.52)$$

$$\bar{H}R = ER \quad (2.53)$$

with a general excitation operators T and R with respect to the reference determinant $|\Psi_0\rangle$. In EOM-CCSD methods however, T and R are cropped at single and double excitations. The amplitudes T have to satisfy the coupled-cluster equations:

$$\langle \Phi_i^a | \bar{H} | \Phi_0 \rangle = 0 \quad (2.54)$$

$$\langle \Phi^a b_{ij} | \bar{H} | \Phi_0 \rangle = 0 \quad (2.55)$$

The scaling of computation for EOM-CSSD is the same as for CIS(D).

The EOM-CCSD is good for geometry optimization of molecules, when high accuracy calculations for excited states are needed⁹⁷.

2.3. Ab initio molecular dynamics

These two approaches, classical MD simulation and quantum chemistry methods (DFT or CIS) can be mixed together as an hybrid approach in the following way - the dynamics of atoms (atomic nuclei) is governed by the Newton dynamics on the surface of the Born-Oppenheimer (BO) potential energy, assuming that for each instantaneous position of atoms, the electron energy takes the value determined by its eigenstate (ground or excited one). This method is called "ab-initio molecular dynamics" (AIMD). Technically, in AIMD classical MD trajectories propagate on the BO potential energy surface, which is generated in quantum mechanical model (e.g. CAM-B3LYP/6-31G) by computing the energy and its gradient at each time step of the MD simulation. As a result, each step has the same cost as the geometry optimization step.

Since the investigated system in this work consists of the protein, molecule and solvent (total atoms number: 2546) it would be too costly to run ab initio calculations for the whole system, while standard MD calculation would not allow us to get any information about the electronic structure and optical transitions. We used therefore AIMD as a hybrid method. The strategy is to treat both protein and water classically at the MD level of theory, and at the same time the molecular probe properties are calculated using TDDFT or CIS approach (see Evaluation section in the Results chapter). The AIMD method then allow us to connect the two parts (MD and TDDFT) in order to consider the environmental effect over the optical properties of the probe.

The followed workflow is the follow: a 300 ns long ground state MD simulation for the whole system (protein, water and probe) is performed, to equilibrate the probe in the biological environment. Then, a number of snapshots are extracted every 5 ns from the equilibrated MD and used as input for the AIMD calculations. From each of these calculations, the fluorescence properties are analyzed. The diagram of the used workflow is shown in Fig. 2.2.

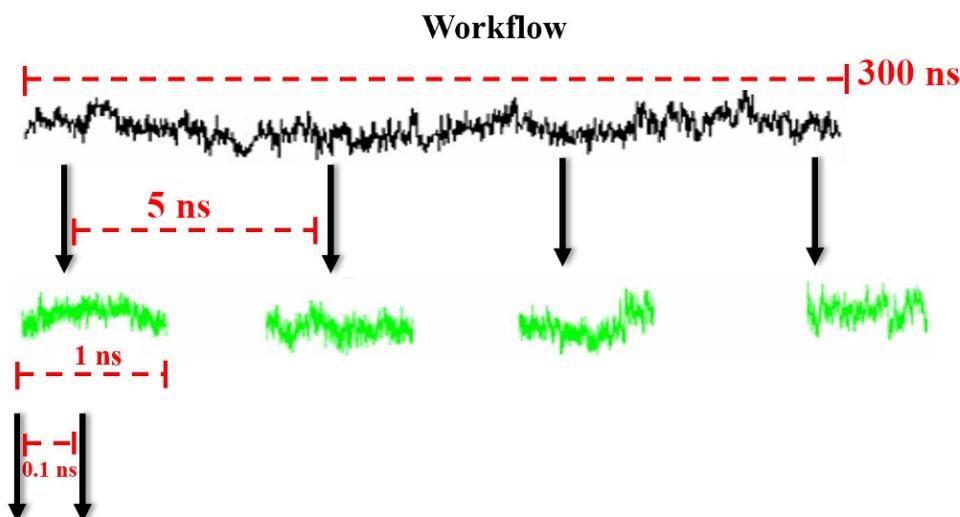


Figure 2.2: An example of a proposed workflow of the next steps in this study. MD in the ground state simulations noted in black, AIMD in the excited state noted in green.

Chapter 3

Results

3.1. Optimization in the excited state

Since there are no parameters for MD available in literature for S1 on the probe, we had to derive them from the optimization of the different parts of the molecule. ANQ and IMC fragments were therefore separately optimized in the first excited state (S1) in Gaussian 16⁹⁸ software using the Equation-of-Motion Coupled Cluster method with restriction to single and double excitations (EOM-CCSD) together with the 6-31G basis set. The linker was optimized in the S1 state using time-dependent density functional theory (TDDFT) with the CAM-B3LYP functional and the 6-31G(d) basis set. To find the energy minimum of the flexible linker, a relaxed scan, with a step size of 30 degrees on the dihedral angles in the linker, was performed in the S1 state employing TDDFT together with the CAM-B3LYP functional and the cc-pVDZ basis set. The scan was repeated for each dihedral present in the linker; for this reason, four sections of the linker (NH-CH₂-CH₂-CH₂, CH₂-CH₂-CH₂-CH₂, CH₂-CH₂-CH₂-CH₂ and CH₃-NH-CH₂-CH₂) were considered. A final single point EOM-CCSD was performed for each optimized structure to refine the energies. These calculations were needed to fit the force field parameters for the linker, since it is the most flexible part of the molecule and can lead to wrong results if not treated properly.

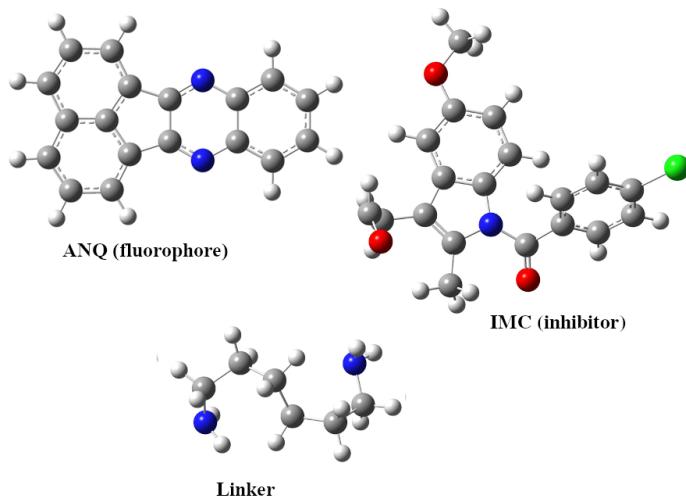


Figure 3.1: Optimized structures of the molecules constituting the ANQ-IMC-6 probe in the S1 excited state.

3.2. MD

The molecular dynamics simulations were computed using GROMACS 2018⁹⁹. The CHARMM36¹⁰⁰ force field was modified to introduce the parameters of the linker in the excited state, as well as the charges for the whole probe. In Fig. 3.2 a map of a the molecule with annotated CHARMM atom types is shown, while Table 3.1 contains non-existing CHARMM bonds, angles and dihedral types obtained as modifications of existing types with similar atom types that were used.

The system was built by adding the probe at the interface between the dimer of the protein and solvated it in water considered with the TIP3P model. A cubic box with sides equal to 19,8 nm was used with the periodic boundary conditions. We performed a 300 ns MD simulation with the whole system, using the Leap-Frog integrator with Coulomb and vdW cut-offs equal to 1.5 nm (Verlet cut-off scheme), time step equal to 2 fs and hydrogen bond constraints. Electrostatics was done with the use of Particle Mesh Ewald method. We used the NPT ensemble in which temperature was defined using Nosé-Hoover thermostat and pressure was defined using Parrinello-Rahman algorithm.

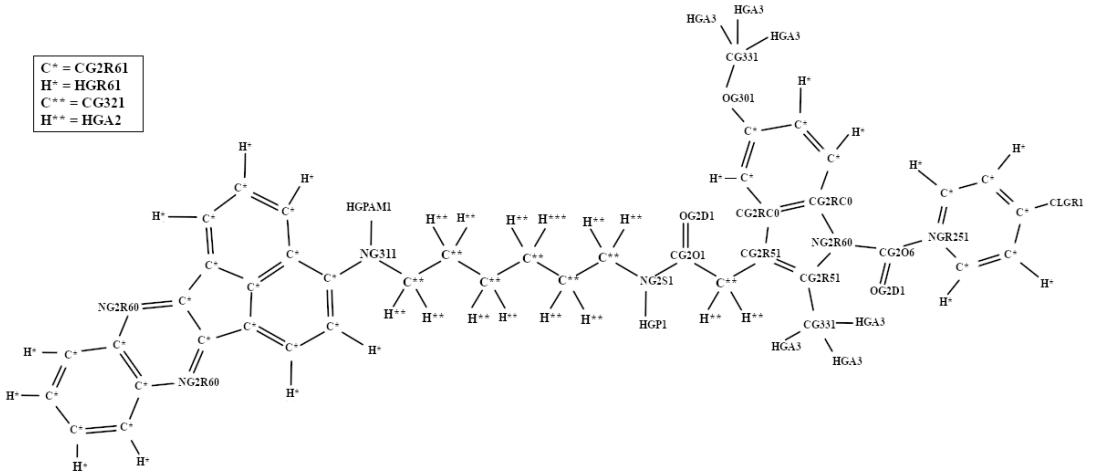


Figure 3.2: The molecule with annotated atom types used to modify the CHARMM36 force field.

Table 3.1: Force field modifications compared to the reference CHARMM36 one.

	our force field	modification from existing
bond	CG2O6 CG2R61	CG2O6 CG2R61
angles	CG2O1 CG321 CG2R51	CG2O1 CG321 CG2R61
	CG2O1 NG2R51 CG2RC0	CG2O6 NG2R51 CG2R51
	CG2R61 CG2O6 NG2R51	CG2R61 CG2O1 NG2S1
	CG2R61 CG2O6 OG2D1	CG2R61 CG2O1 OG2D1
	CG2O6 NG2R51 CG2RC0	CG2O6 NG2R51 CG2R51
dihedrals	CG321 CG321 NG311 CG2R61	CG321 CG321 NG311 CG2R64
	CG2R51 CG2O1 NG2S1 CG321	CG2R61 CG2O1 NG2S1 CG321
	CG2RC0 CG2R51 CG321 CG2O1	CG2R61 CG2R61 CG321 C2O1
	CG2R51 CG2R51 CG321 CG2O1	CG2R61 CG2R61 CG321 C2O1
	OG2D1 CG2O6 NG2R51 CG2RC0	OG2D1 CG2O6 NG2R51 CG2R51
	CG2RC0 CG2R61 CG2R61 OG301	CG2R61 CG2R61 CG2R61 OG301
	HGR61 CG2R61 CG2R61 CLGR1	HGR61 CG2R61 CG2R61 HGR61
	CG2R61 CG2RC0 NG2R51 CG2O6	CG2R51 CG2R51 NG2R51 CG2O6
	CG2RC0 CG2R51 CG2R51 CG331	CG2RC0 CG2R51 CG2R51 CG3C52
	OG2D1 CG2O6 CG2R61 CG2R61	OG2D1 CG2O1 CG2R61 CG2R61
	NG2R51 CG2O6 CG2R61 CG2R61	NG2S1 CG2O1 CG2R61 CG2R61
	CG2RC0 CG2RC0 NG2R51 CG2O6	CG2R51 CG2R51 NG2R51 CG2O6
	X NN1 CN1A X NN1	CG2RC0 NG2R51 CG2O6 CG2R61
	X NN1 CN1A X NN1	CG2R51 NG2R51 CG2O6 CG2R61
	CG2O6 CG2R61 CG2R61 HGR61	CG2O1 CG2R61 CG2R61 HGR61
	NG2R51 CG2R61 CG331 HGA3	NG2R51 CG2R51 CG331 HGA3
	CG2O6 CG2R61 CG2R61 CG2R61	CG2O1 CG2R61 CG2R61 CG2R61
	OG2D1 CG2O1 CG321 CG2R51	OG2D1 CG2O1 CG321 CG2R61
	CG321 CG2R51 CG2RC0 CG2R61	CG321 CG2R51 CG2RC0 CG2R61
	NG2S1 CG2O1 CG321 CG2R51	NG2S2 CG2O1 CG321 CG2R61
	CG2RC0 CG2R61 CG2R61 HGR61	CG2R61 CG2R61 CG2R61 HGR61
	CG331 CG2R51 NG2R51 CG2RC0	CG331 CG2R53 NG2R51 CG2RC0

It is worth noting here that we start our MD simulation from an already equilibrated system which was used previously to study the absorption properties of the probe embedded in the COX-2 dimer⁶². The results of the ground state MD simulation of the probe embedded in the protein are presented here. In Fig. 3.3 and 3.4 we show the root mean square displacement (RMSD) plots for the molecule and the protein. The RMSD for the protein is lower than 5 Å and for the probe it is below 2 Å, confirming that the system is fully equilibrated. Fig. 3.5 reports the plot of the total energy and its different contributions oscillating around constant values. The small drift in total energy, equal to -45 kJ/mol, is another confirmation that the system is equilibrated (we consider the system equilibrated when the total drift is less than 50 kJ/mol). Both analyses confirm that the probe inserted in the COX-2 dimer is equilibrated, and snapshots can be extracted for the next step of our study.

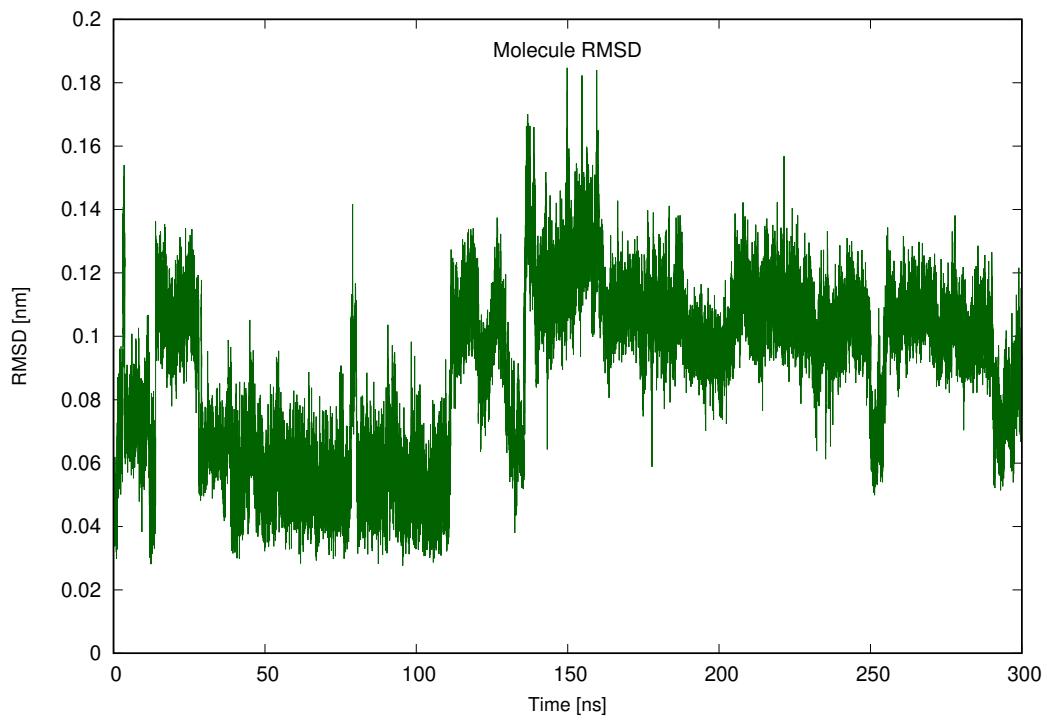


Figure 3.3: ANQ-IMC-6 RMSD plot.

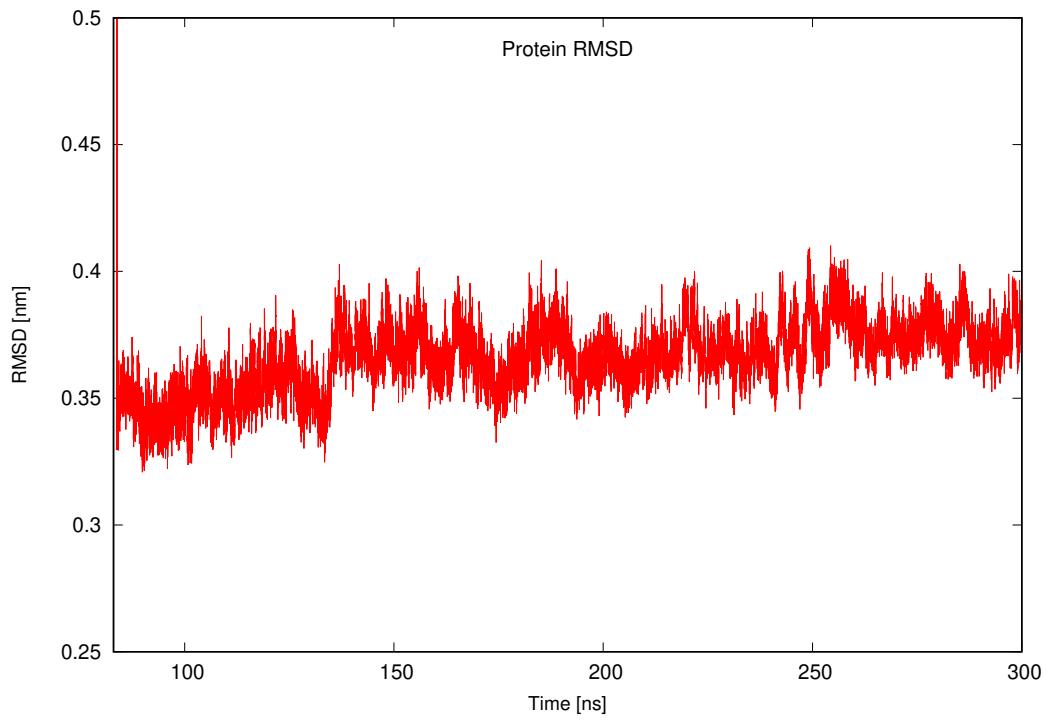


Figure 3.4: Protein RMSD plot

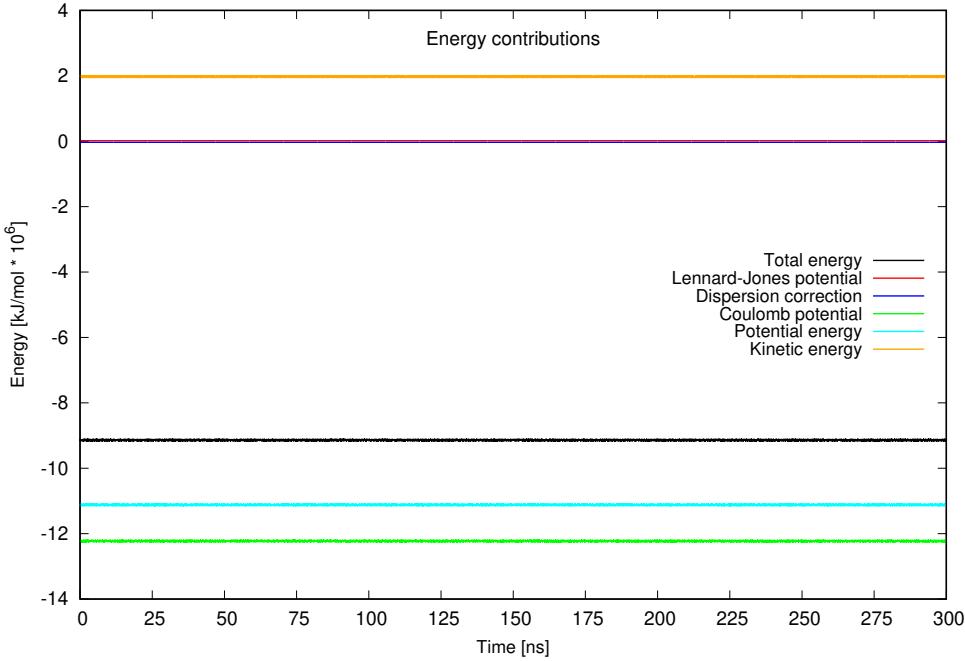


Figure 3.5: Evolution in time of the total energy (black) and its different contributions.

Fig. 3.6 shows a representative snapshot from the MD simulation showing the molecule bound to COX -2 in the active site on the protein at the interface between the two monomers.

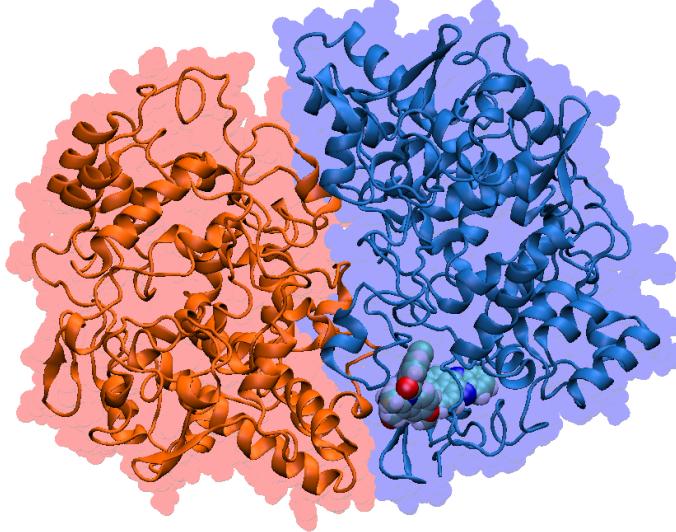


Figure 3.6: ANQ-IMC-6 bonded at the interface of the COX-2 dimer. The image is based on a snapshot from the ground state MD simulation.

From the trajectory file, frames were extracted every 5ns using GROMACS *trjconv* tool. Since thermal effects are present in the MD simulations, the different extracted frames not only differ in geometry due to the probe, but also due to the thermal energy. An overlap of all the extracted frames is presented in Fig. 3.8. From this figure it is clearly visible the contribution of thermal energy especially on the ANQ moiety of the probe, while the IMC is more affected by the change in geometry of the protein surrounding it. Interestingly, the probe keeps its unfolded conformation due to the presence of the protein, but instead of being fully ‘open’ as reported in literature, it assumes a ‘L-shaped’ conformation.

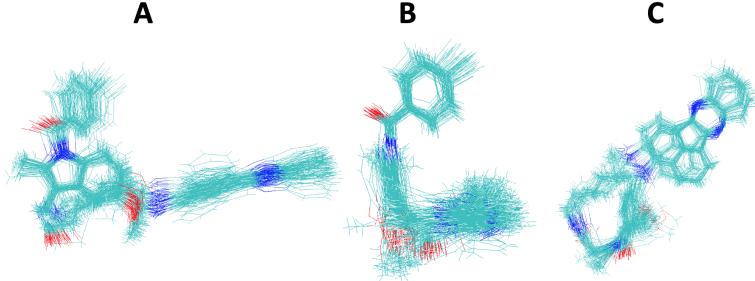


Figure 3.7: Overlap of the different frames extracted from MD. A, B, C denotes views from different directions. Stable unfolded conformation with thermal fluctuations influence can be seen.

From the MD simulation we do not only extract the probe geometry, but also part of the protein and water molecules surrounding the probe itself, since they might play an important role in affecting the overall fluorescence properties. In order to generate a reasonable model system as input for the AIMD calculations, only the protein within a 12 Å sphere radius from the probe was considered explicitly. For water molecules, the radius was decreased to 8 Å. The composition of the AIMD system considered is shown in Fig. 3.8, in which the probe (90 atoms) is treated at the ab-initio level of theory and the environment (protein and water molecules) as point charges. For the MD part of the AIMD we reparametrized the bonding terms of the force field, as reported in this section before, as well as the charges of the probe in the first excited state (optimization of the geometry and calculation of the charges with the ESP method).

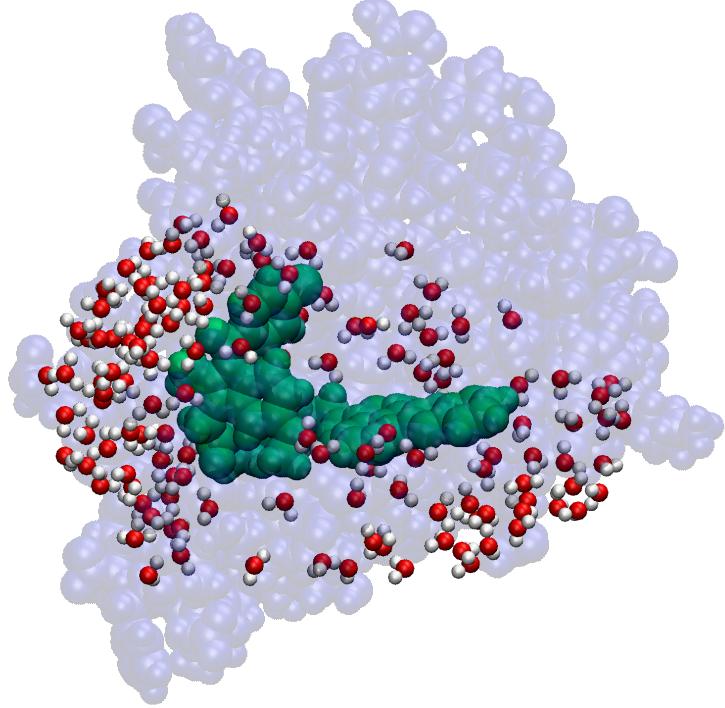


Figure 3.8: Structure of the system used for ab initio molecular dynamics consisting of the probe, protein aminoacids within 12 Å and water within 8 Å from that probe.

3.3. AIMD Validation

Since the used methodological protocol is novel, non-standard and is currently being developed in our group, validation of its accuracy is needed. For the AIMD procedure, we use the QChem 5.4 software¹⁰¹, which allows for the use of AIMD coupled with both DFT and post-HF methods, as well as for a strong control over different key parameters that can affect the computation. Three main parameters need to be validated: (i) AIMD timestep, (ii) method (TDDFT and CIS) and (iii) basis sets. First, we consider the timestep, which dictates how fast (or slow) the AIMD should proceed to get accurate results and reach the 1 ps time set as final target. Considering the CAM-B3LYP functional and the 3-21G basis set (to speed up the computation), we observed that among the different selected timesteps, the optimal one is 5 a.u., which correspond to 0.12 fs (Fig. 3.9 and 3.10). For this validation procedure, only one frame was considered. We used the 3-21G basis set to assess the validity of the proposed method (although we are aware than we have to increase it to get reliable results).

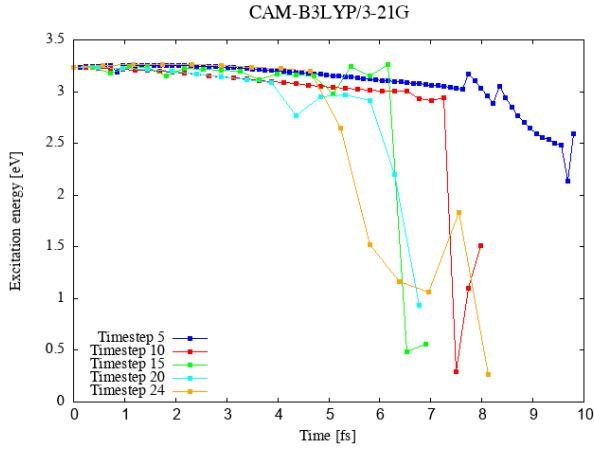


Figure 3.9: Excitation energy for different AIMD time step values using TDDFT method, CAM-B3LYP functional and 3-21G basis set.

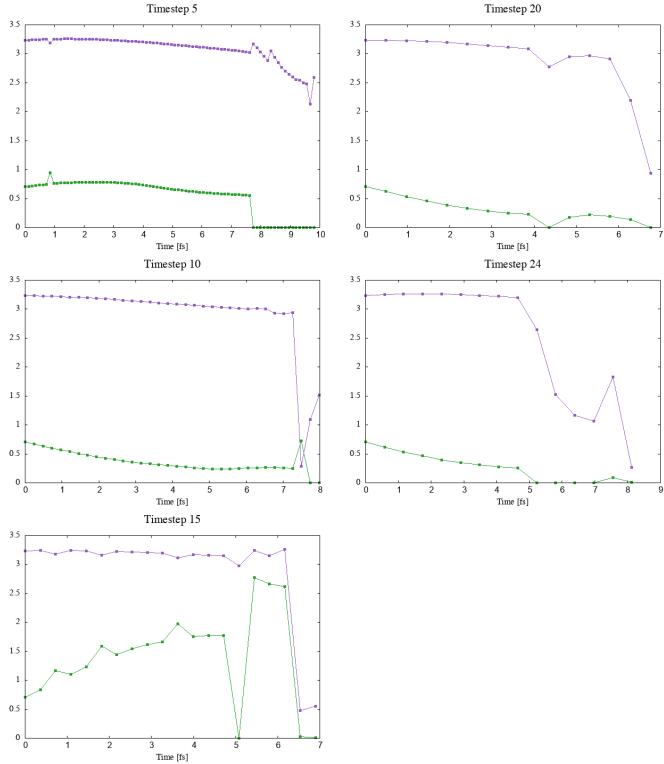


Figure 3.10: Excitation energy in eV (purple) and oscillator strength (green, multiplied by 10) for different AIMD time step values using TDDFT method, CAM-B3LYP functional and 3-21G basis set.

Interestingly, even when increasing the basis set to 6-31G (Fig. 3.11 and 3.12) and 6-31G(d) (Fig. 3.13 and 3.14), the 5 a.u. basis set is the best performing, which translates in a longer AIMD run among the different time steps considered. When comparing the different plots, it is visible that the energy of the first excited state does not strongly depends on the size of the basis set, but its overall trend in time is strongly affected by it and it can lead to instabilities on the long run (see Fig. 3.13). Moreover, the computational time strongly increases when larger basis sets are used.

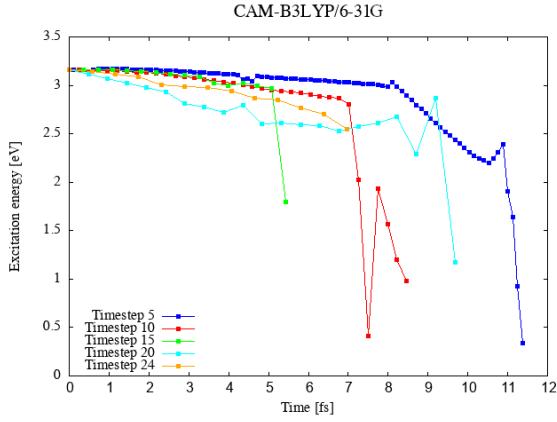


Figure 3.11: Excitation energy for different time step values using TDDFT method, CAM-B3LYP functional and 6-31G basis set.

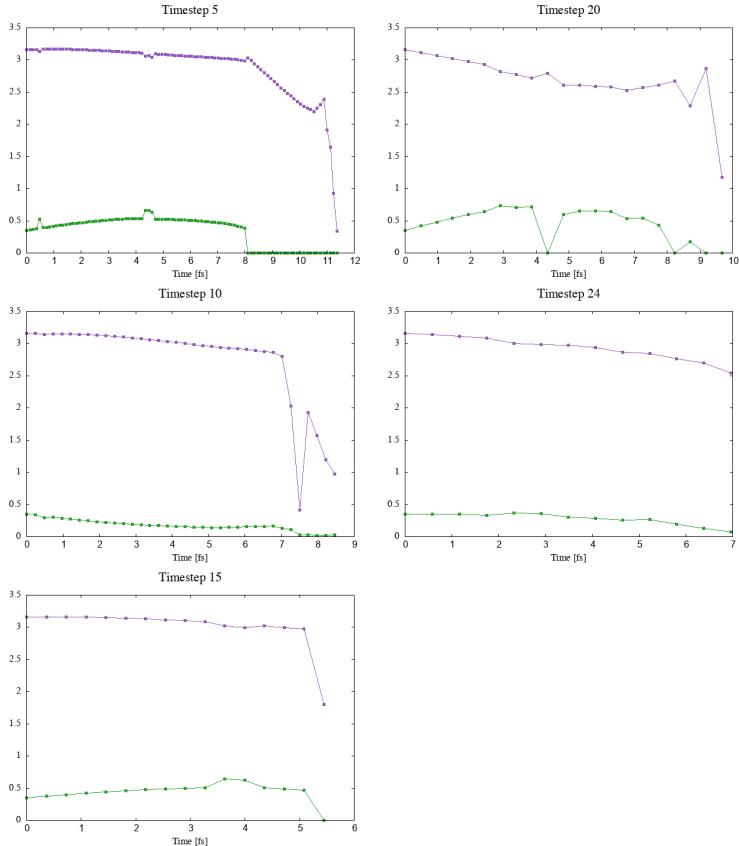


Figure 3.12: Excitation energy in eV (purple) and oscillator strength (green, multiplied by 10) for different time step values using TDDFT method, CAM-B3LYP functional and 6-31G basis set.

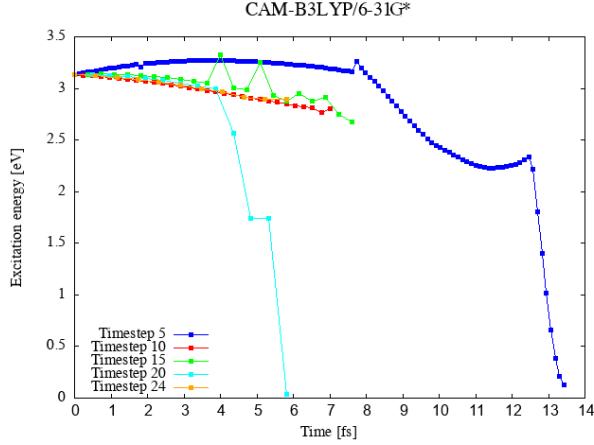


Figure 3.13: Excitation energy for different time step values using TDDFT method, CAM-B3LYP functional and 6-31G* basis set.

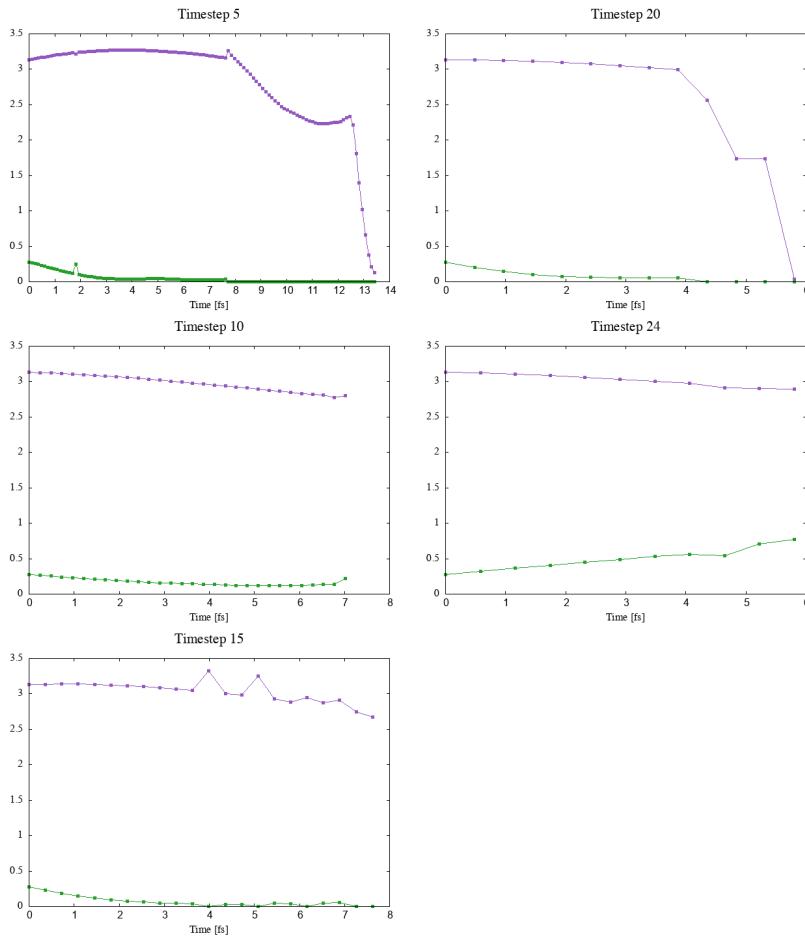


Figure 3.14: Excitation energy in eV (purple) and oscillator strength (green, multiplied by 10) for different time step values using TDDFT method, CAM-B3LYP functional and 6-31G* basis set.

As a next step, we consider the same timesteps as previously described and repeated the calculations considering this time the CIS method, and again the same basis sets as described before. The first striking difference is that for both 3-21G (Fig. 3.9 and 3.10) and 6-31G basis sets (Fig. 3.11 and 3.12), the 5 a.u. timestep is the most stable in time, while for the more accurate 6-31G(d) (Fig. 3.13 and 3.14) an instability arises just after 13 fs of AIMD run.

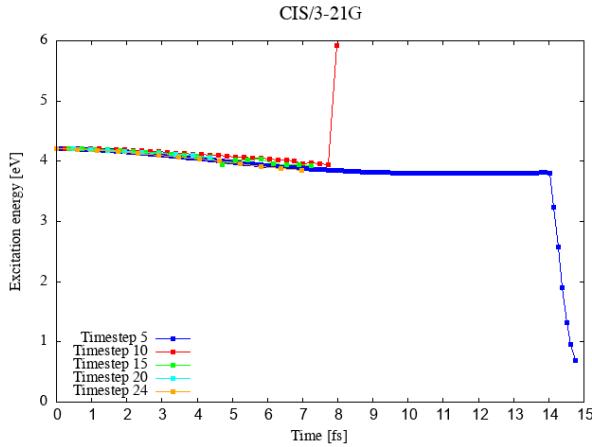


Figure 3.15: Excitation energy for different time step values using CIS method and 3-21G basis set.

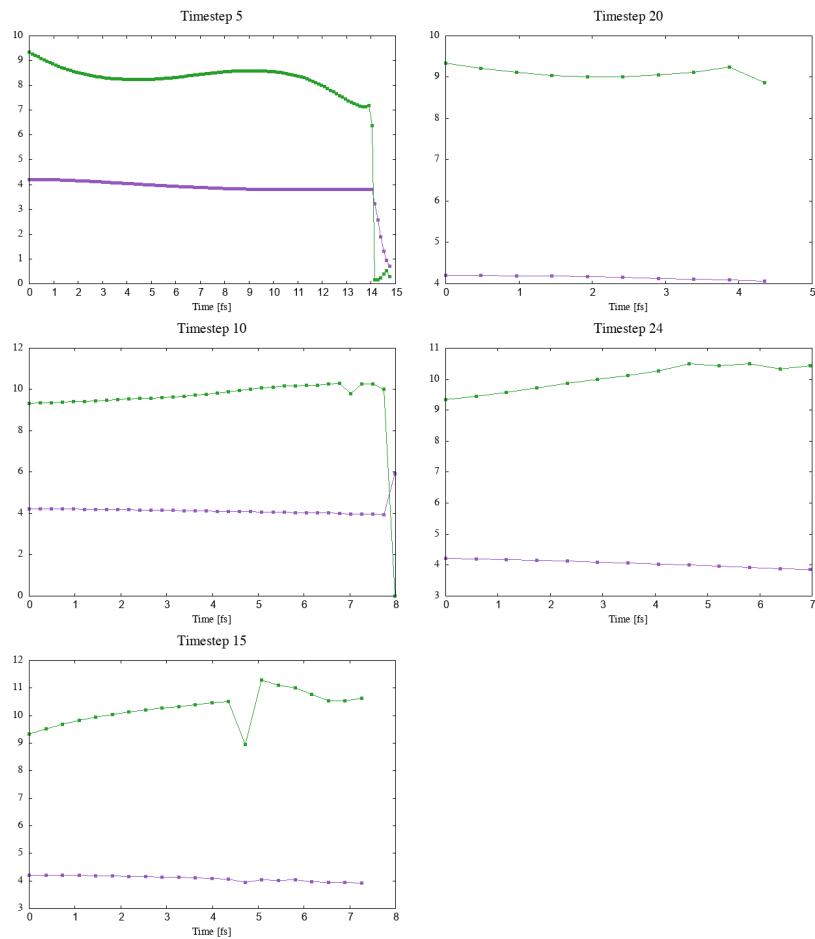


Figure 3.16: Excitation energy in eV (purple) and oscillator strength (green, multiplied by 10) for different AIMD time step values using CIS method and 3-21G basis set.

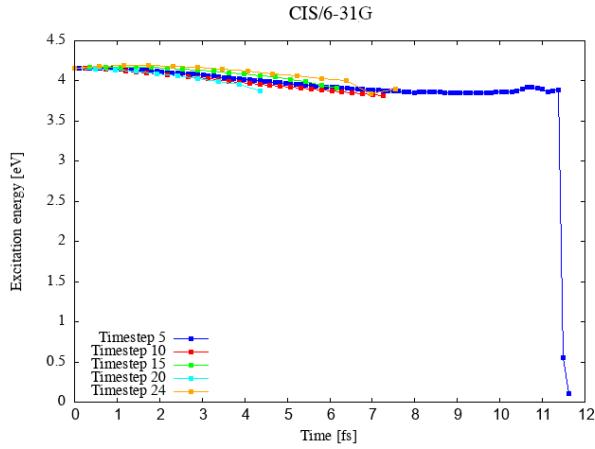


Figure 3.17: Excitation energy for different time step values using CIS method and 6-31G basis set.

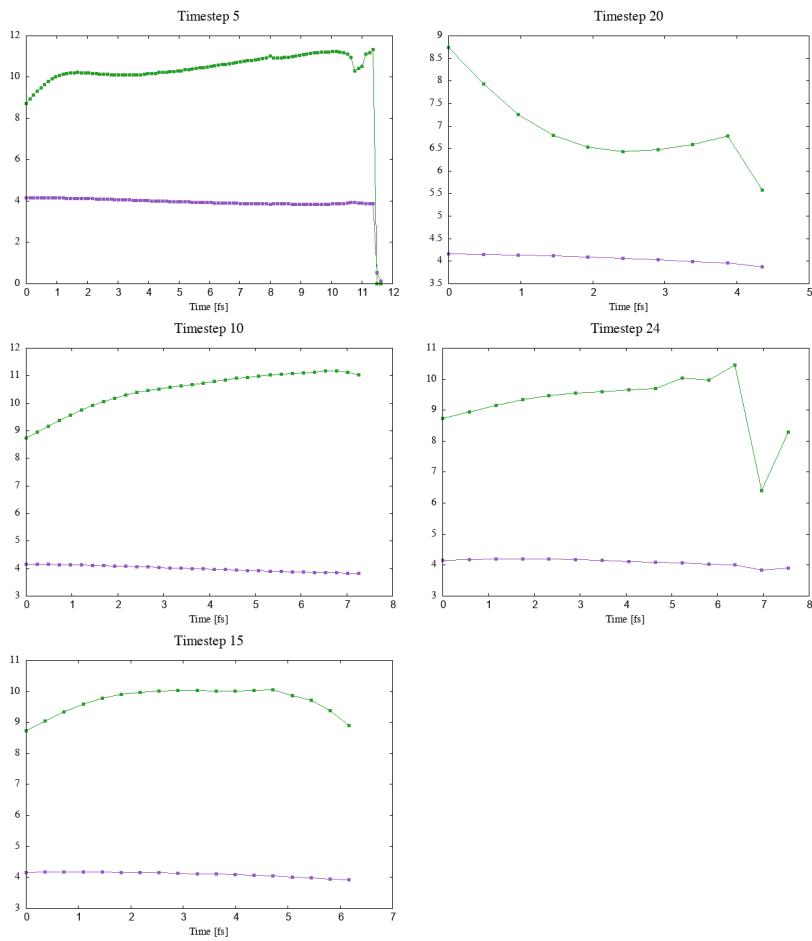


Figure 3.18: Excitation energy in eV (purple) and oscillator strength (green, multiplied by 10) for different AIMD time step values using CIS method, and 6-31G basis set.

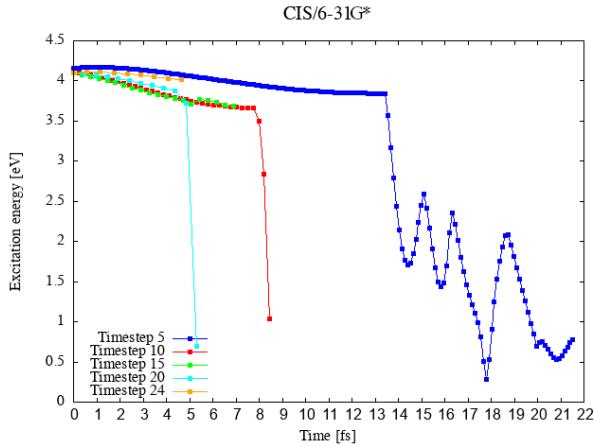


Figure 3.19: Excitation energy for different time step values using CIS method and 6-31G* basis set.

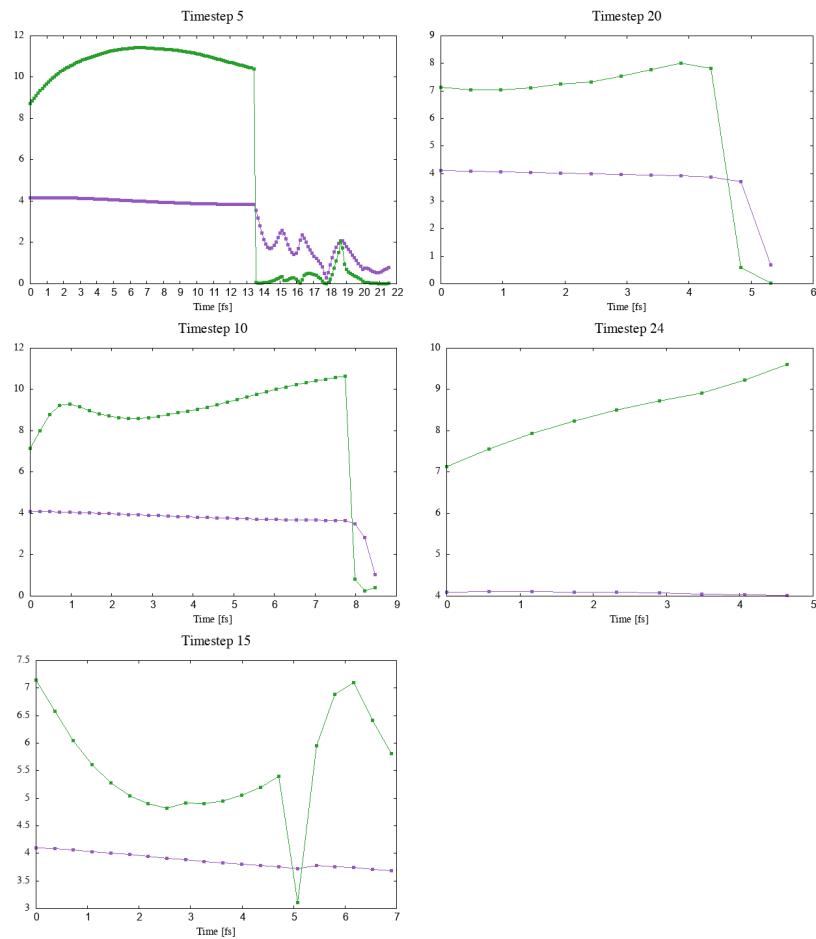


Figure 3.20: Excitation energy in eV (purple) and oscillator strength (green, multiplied by 10) for different AIMD time step values using CIS method and 6-31G* basis set.

At this point in the validation process, having set the timestep to 5 a.u., we need to decide what method (either DFT or CIS) to use, as well as which basis set. As a remark, we would like to point out that a shorter time step such as 1 a.u. (0.024 fs) would have been even better to have higher accuracy, but the computational cost associated to it is too high, requesting more than 10 times to reach the same AIMD time when compared to 5 a.u.

Fig. 3.21 show an overview of the performances of the different combinations of methods and basis sets with a timestep of 5 a.u.

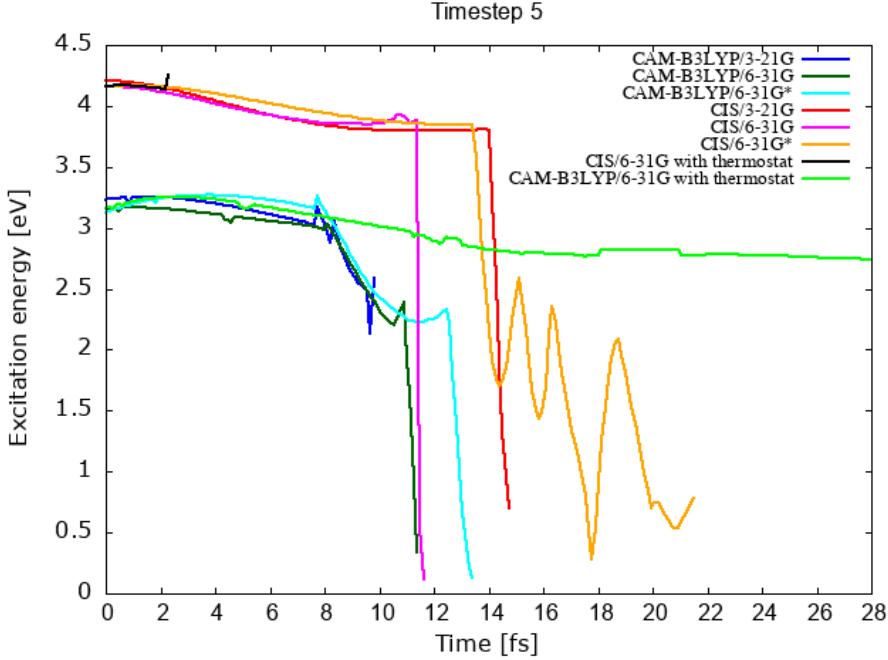


Figure 3.21: Excitation energy for different methods with the same time step 5.

From this plot we can draw our conclusions regarding the validation of the method:

- 1) Considering that the experimental fluorescence data present a peak at around 550 nm (ca. 2.3 eV), it is clear that the CIS method has a too high (less accurate) excitation energy compared to TDDFT with CAM-B3LYP. CIS also shows stability problems at longer time scales (after 14 fs). Hence, TDDFT with the CAM-B3LYP functional was considered as method of choice for this protocol.
- 2) Within the same method (either DFT or CIS) there is little or no effect of the size of the basis set on the accuracy of the energy description. We are aware that use of polarization functions can improve the results, however we observe very little difference between 6-31G and 6-31G*. Moreover, simulations with the bigger basis set are longer and have bigger computational cost. We thus compromise accuracy vs. computational cost, and chose the 6-31G basis set as the optimal one.

Being puzzled by the short lived simulation time (all the jobs failed after reaching a time window between 10 and 20 fs), we postulated that this was due to the presence of an early conical intersection to the ground state, which cannot be described by the DFT method (due to its multi-configurational character). Yet, after discussion and trials, we concluded that the observed instabilities were caused by the lack of a thermostat which equilibrates the system at a given temperature (in our case 298 K). As can be seen in Fig. 3.21, with the introduction of the Nosé-Hoover thermostat, the problem has been solved.

3.4. Ab initio MD

As result of the previous section (3.3. Validation), the TDDFT method along with the CAM-B3LYP functional and the 6-31G basis set has been chosen to perform the ab-initio MD calculations over the extracted frames. 34 frames extracted from the classical MD simulation and modified as described before (see MD section) as initial structures were used. At the time of the thesis submission, all calculations were still running. Results for first 70 fs of the calculations were however obtained, and conclusions over the short time non-equilibrium MD can be drawn. The evolution of the calculations in terms of the excitation energy in time is shown in Fig. 3.22.

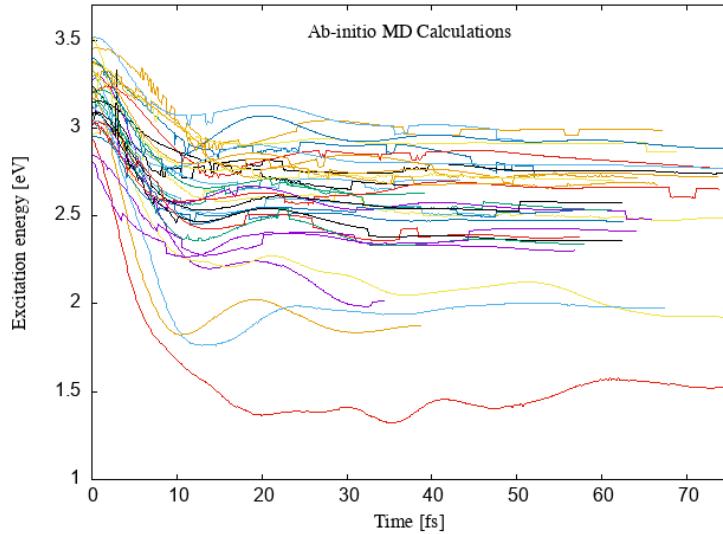


Figure 3.22: Evolution of the excitation energy in time for the 34 extracted frames.

Already for this short 70 fs evolution time, we can observe that the excitation energy for the vast majority of the frames relaxes from an initial value around 3.3 eV to 2.7 eV. For a few frames, this values is lower, ranging from around 1.5 to 2 eV. It has to be seen if this is due to a different environment surrounding the probe (i.e. different amount of water molecules, different protein conformation) or if is intrinsic to probe, due to changes in conformation.

From this time evolution, we slide the above plot every 10 fs and analyze the evolution of the excitation energy for the different frames (i.e. one time, all frames overlapped together and the data convoluted with a Gaussian function to obtain the fluorescence spectra).

When the calculation starts (at 0fs, see Fig. 3.23 there is only one peak with maximum around 359 nm, which corresponds to HOMO-2 \rightarrow LUMO electronic transition (46% of the frames). This transition between the orbitals (shown also in Fig. 3.23 describe the absorption for the probe.

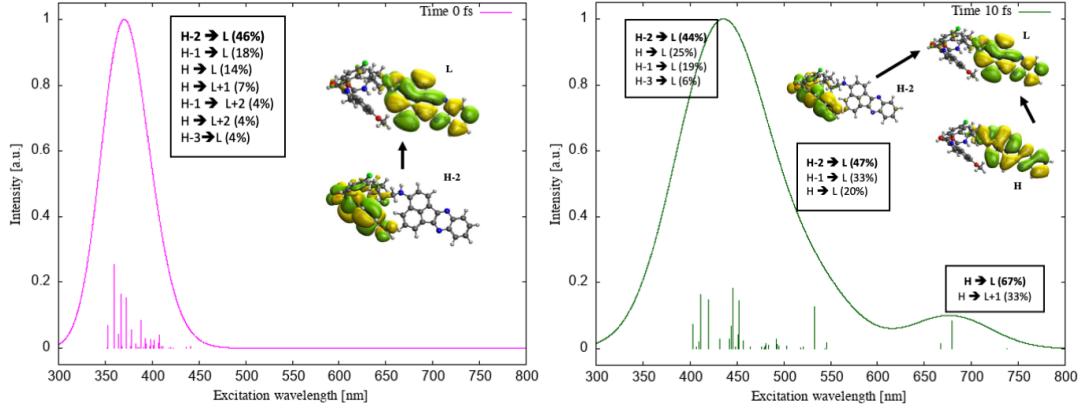


Figure 3.23: Spectrum for time 0 fs (left) and 10 fs (right). The peak is the result of a convolution of the data with a Gaussian function. Intensities were normalized by dividing the intensities by their maximum value. Types of transitions with their percentage corresponding for the peaks and the orbitals, are shown in the inset.

Already after 5 fs, we see changes in fluorescence peak wavelength and intensities, and new peaks emerge. At 10 fs (see Fig. 3.23) we see changes in the position of the main peak which is now redshifted around 446 nm. By subtracting this value to the absorption maximum, we obtain a 90 nm Stokes shift, which is in excellent agreement with the experimental 90 nm value. It is worth nothing in here that all the presented plots refer only to the S1 state. The main transition for this peak is still HOMO-2 \rightarrow LUMO transition, but we can also see that HOMO \rightarrow LUMO transition start to become important (25 % from 14 %, see Table 3.2), which is an indication that already after such a short timescale, the fluorescence becomes to be important. For the same S1 excited state, we also observe a shoulder and a new peak at longer wavelengths. The shoulder peaks around 533 nm and possess both HOMO-2 \rightarrow LUMO and HOMO \rightarrow LUMO transitions, 47% and 20%, respectively. The third peak around 679 nm has a majority of HOMO \rightarrow LUMO transition (67%), which confirm the absence of quenching due to PET already at shorter times.

Comparable situations are present for 20 fs and 30 fs (see Fig. 3.24) with a similar wavelength for the main peak, for which a HOMO-2 \rightarrow LUMO transition takes place. The shoulder and the other peak at longer wavelengths are also present, with the same character as described before. It is important to notice that the peaks at the longest wavelength are decreasing the HOMO \rightarrow LUMO character, while for the shoulder at around 500 nm this is steadily increasing.

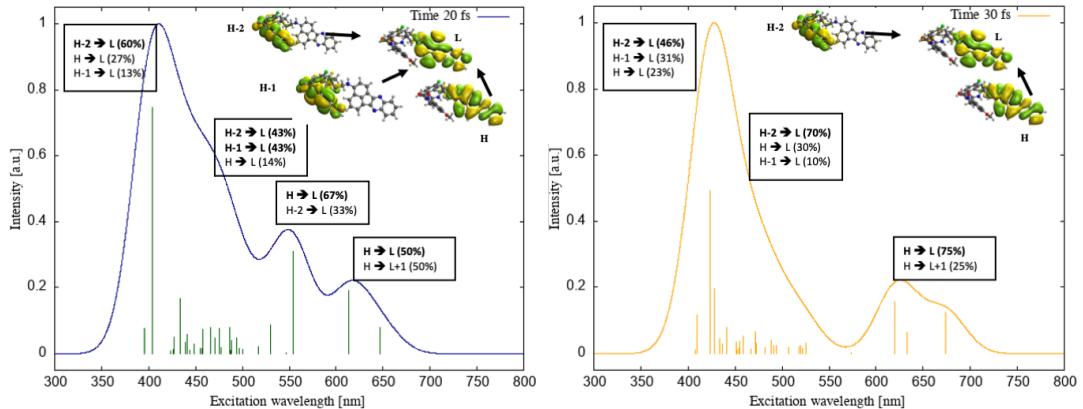


Figure 3.24: Spectrum for time 20 fs (left) and 30 fs (right). The peak is the result of a convolution of the data with a Gaussian function. Intensities were normalized by dividing the intensities by their maximum value. Types of transitions with their percentage corresponding for the peaks and the orbitals, are shown in the inset.

For 40 fs (see Fig. 3.25 and Table 3.3) the main peak is shifted towards shorter wavelength, at around 419 nm, and the transition is now a competition between HOMO-1 \rightarrow LUMO (43%) and HOMO \rightarrow LUMO (36%), while the shoulder still presents a predominant PET character and the peak around 640 nm still presents a strong fluorescent character due to the HOMO \rightarrow LUMO transition. After 50 fs from the absorption of a photon, the predominance of the HOMO \rightarrow LUMO transition (50%, 31% and 100% for the three peaks, respectively) is present in all the observed peaks. This situation is present up to 70 fs for the main peak at around 420-430 nm, while for the shoulder the competition between different transitions is still present. Interestingly, for the longest wavelength peak, after 50 fs the transition is a pure HOMO \rightarrow LUMO. Moreover, this peak seems to disappear at 70 fs.

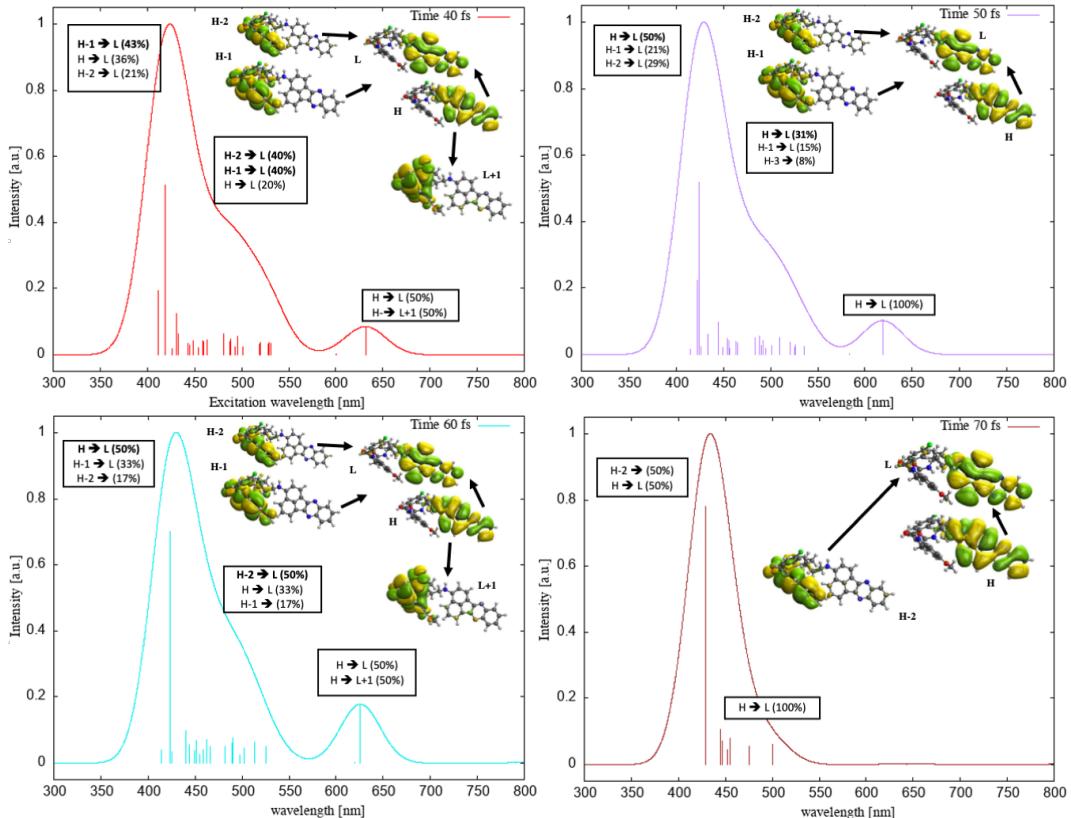


Figure 3.25: Spectrum for time 40, 50, 60 and 70 fs. The peak is the result of a convolution of the data with a Gaussian function. Intensities were normalized by dividing the intensities by their maximum value. Types of transitions with their percentage corresponding for the peaks and the orbitals, are shown in the inset.

Because calculations are still running and not all of them reached 60 fs, the last two plots must be taken with care. In fact, only a few frames for 60 and especially for 70 fs are present, and the discussed results might not be statistically significative yet, but still we observe a progression of the main peak from around 350 nm to 425 nm after 70 fs and an increasing in HOMO \rightarrow LUMO character of the transition over the HOMO-2 \rightarrow LUMO and HOMO-1 \rightarrow LUMO ones. As a result, we can already state that there is a strong competition between quenching due to PET and fluorescence at short time, and the

fluorescence become the main decay mechanism even just after short time. This is also represented by one frame for the 70 fs time (see Fig. 3.25 and corresponding structure), for which the biggest intensity in oscillator strength has been found. In this specific case, the peak is redshifted of 65 nm and represents a HOMO \rightarrow LUMO transition instead of the HOMO-2 \rightarrow LUMO one observed in the absorption spectra.

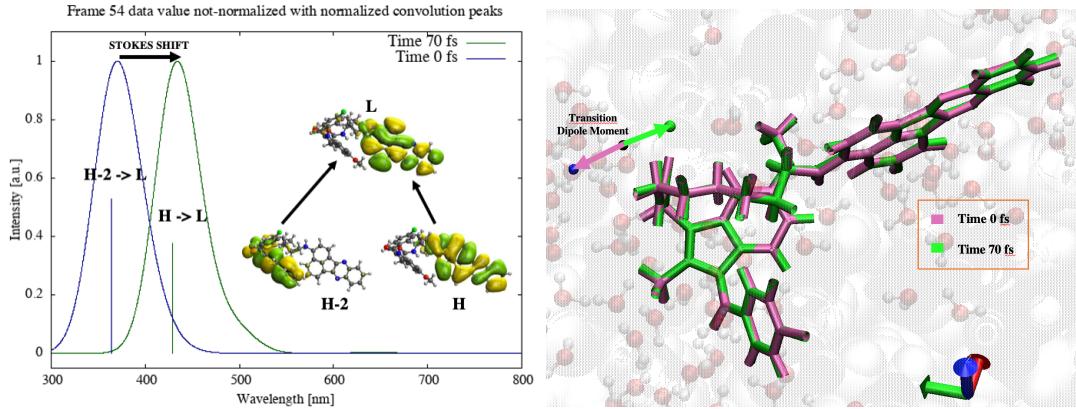


Figure 3.26: (up) Spectra for a selected frame for time 0 and 70 fs. The data (vertical lines) are not normalized to show comparable intensities. Peaks were, however, normalized by dividing intensities by their maximum value for a better display. (down) Structures of the molecule in the environment for the same frame. Absorption and emission transition dipole moments are presented in the vector representations (with an angle of 3.8° between each other).

Already after this short emission time, we can reveal the mechanism underneath the fluorescence emission of the probe in COX-2. After excitation, the S1 excited state is found at 360 nm in wavelength and is characterized by a HOMO-2 → LUMO transition which is representative of a PET process (since HOMO-2 and LUMO are localized over different moieties of the probe). At 70 fs after excitation, the fluorescence spectrum presents a major bright peak at around 420 nm, characterized by a HOMO → LUMO transition, for which the orbitals are localized over the same ANQ moiety, thus allowing for fluorescence to take place. This is also corroborated by the different orientation of the absorption and emission dipole moments, which both lie on the long molecular axis of the ANQ moiety but with an angle of 3.8 degrees between each other (see Fig. 3.26).

In Fig. 3.27 we summarize our results, reporting the overlap of all the fluorescence spectra but with two distinct normalizations, either by the intensity or by the number of frames. This allows us to describe different important observations. First, when considering the normalized intensity, we observe that after an initial oscillation in energy for the main fluorescence peak, already after 30 fs it does equilibrates in energy and it remains stable at around 430 nm. In addition, we see the appearance of the shoulder at longer wavelength (around 500 nm) and a second peak at longer wavelengths which is blue shifting in time, until its disappearance at 70 fs. This is even more evident when we consider the normalization by the number of frames. Yet, as mentioned above, this can also be due to the fact that the number of frames considered for the 70 fs is still not large enough to be statistically significant. Finally, we also observe the relative decrease and subsequent increase in the main fluorescence peak in time (Fig. 3.27, down), which might be attributed to the change in transition taking place for this peak in time. We recall here that for short time the main transition has a HOMO-2 → LUMO character, which should result in a non-radiative decay due to the presence of PET, while after 50 fs the HOMO → LUMO ones starts to become predominant with a strong, bright peak which indicates the rising of fluorescence character in the process.

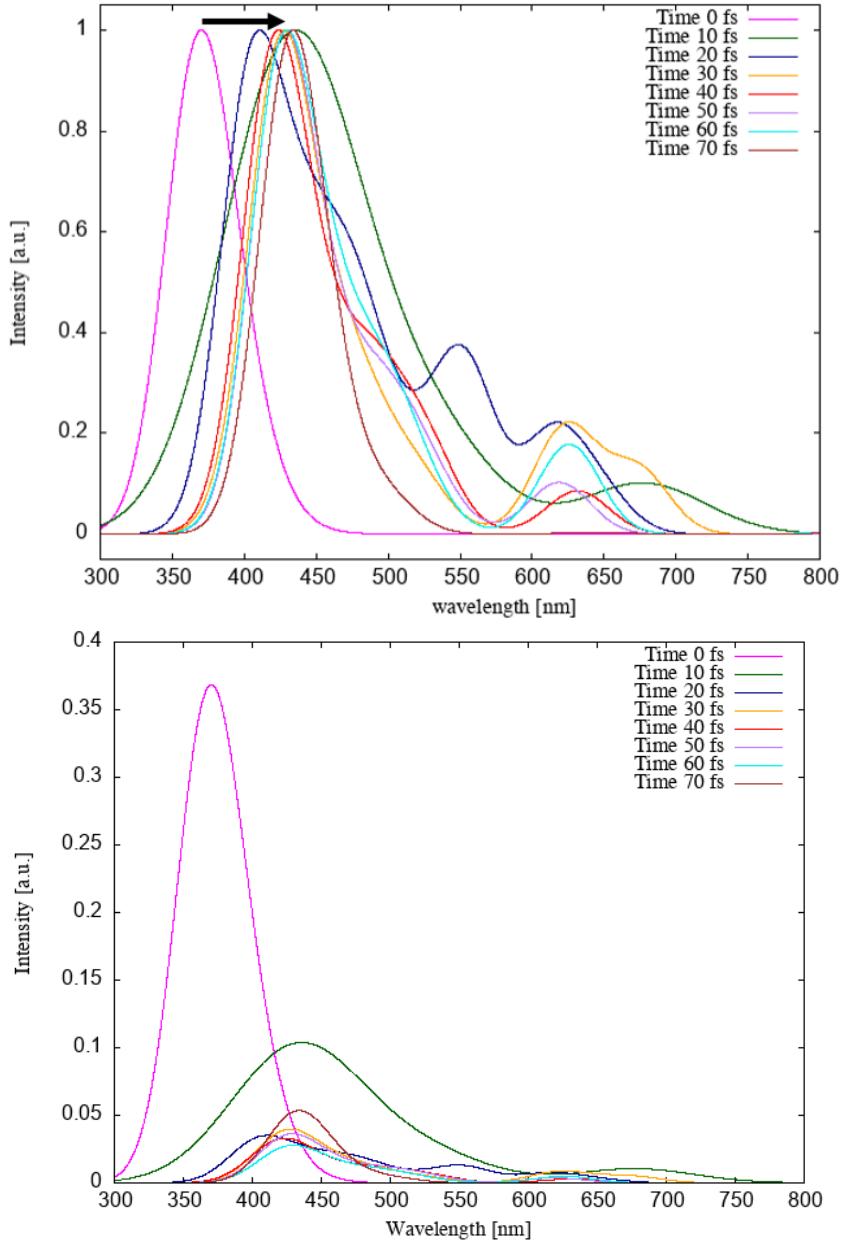


Figure 3.27: Resulted spectra from ab-initio MD as a convolution fits to data points obtained as extracted values from all frames for 0, 10, 20, 30, 40, 50, 60 and 70 fs time. Peaks intensity (corresponding oscillator strengths) were normalized by dividing the intensities by their maximum value (up) or by the number of frames (down).

The tables with the average and maximum emission wavelengths, intensity and transitions nature and probability (in percentage) for each peak and selected time are reported in the following tables.

Table 3.2: The results for 0, 10, 20 and 30 fs times. Maximum wavelength is the wavelength corresponding to maximum intensity (oscillator strength) for the peak. Transitions percentage describes how often a certain transition occurs.

Time (fs)	Average wavelength (nm)	Maximum wavelength (nm)	Maximum intensity	Transitions (%)
0	393	359	0.07	H-2 → L 46% H-1 → L 18% H → L 14% H → L+1 7% H → L+2 4% H-1 → L+2 4% H-3 → L 4%
10	433	446	0.63	H-2 → L 44% H-1 → L 19% H → L 25% H-3 → L 6%
	506	533		H-2 → L 47% H-1 → L 33% H → L 20%
	695	679		H → L 67% H → L+1 33%
20	438	434	0.19	H-2 → L 60% H-1 → L 13% H → L 27%
	489	466	0.09	H-2 → L 43% H-1 → L 43% H → L 14%
	544	530	0.10	H → L 67% H-2 → L 33%
	631	614	0.22	H → L 50% H → L+1 50%
30	438	423	0.65	H-2 → L 46% H-1 → L 31% H → L 23%
	493	472	0.09	H-2 → L 70% H → L 30% H-1 → L 10%
	625	620	0.21	H → L 75% H → L+1 25%

Table 3.3: The results for 40, 50, 60 and 70 fs times. Maximum wavelength is the wavelength corresponding to maximum intensity (oscillator strength) for the peak. Transitions percentage describes how often a certain transition occurs.

Time (fs)	Average wavelength (nm)	Maximum wavelength (nm)	Maximum intensity	Transitions (%)
40	440	419	0.49	H-1 → L 43% H → L 36% H-2 → L 21%
	500	481	0.06	H-1 → L 40% H-2 → L 40% H → L 20%
	617	632	0.08	H → L 50% H → L+1 50%
50	442	465	0.03	H → L 50% H-1 → L 21% H-2 → L 29%
	503	536	0.02	H-2 → L 46% H → L 31% H-1 → L 15% H-3 → L 8%
	602	619	0.10	H → L 100%
60	444	466	0.03	H → L 50% H-1 → L 33% H-2 → L 17%
	496	513	0.04	H-2 → L 50% H → L 33% H-1 → L 17%
	623	626	0.10	H → L 50% H → L+1 50%
70	451	475	0.03	H → L 50% H-2 → L 50%
	500	500	0.03	H → L 100%

Conclusions

In this thesis, we present a novel methodology to describe the fluorescence properties of the ANQ-IMC-6 probe when embedded in a biological environment such as COX-2, to evaluate its mechanism of action in cancer detection.

We have developed a method based on multi-scale computations, in which first the entire system is equilibrated with classical molecular dynamics simulations, and in a following step, ab initio MD simulations are performed. Such a combination of different methods allows obtaining information about the optical properties of the molecule without neglecting the influence of the surrounding environment (i.e. protein, water molecules), since it can strongly affect the intrinsic properties of the fluorophore. After the validation of the proposed computational protocol, we were able to perform ab initio MD (AIMD) simulations on a large number of the probe and protein conformations extracted from the MD simulations.

By performing AIMD simulations, it was possible to assess the early emission mechanism after the absorption of a photon by the probe. Already in the first 70 fs after absorption, interesting conclusions can be drawn on the competition between the PET quenching mechanism and the observed fluorescence. Yet, at this early stage of time evolution, we cannot exclude the presence of different quenching mechanisms that may compete with the fluorescence. Indeed, AIMD can be a very useful tool to compute excited state electronic properties and show how these properties change over time, since it takes time for the system to reach an equilibrium state. This methodology is efficient but also computationally expensive, as a large number of initial configurations is needed to obtain reliable and statistically significant results. Yet, even with the limited number of conformations discussed in this work, changes at very early stages of emission can already be observed, which is often overlooked, but extremely important for assessing the mechanism of action of fluorophores and their changes over time.

Our results show that already after 10 fs after absorption there is a competition between fluorescence and quenching, but the process equilibrates quite quickly and after 70 fs we observe that quenching ceases to be predominant and that the fluorescence decay channel increases in strength. As already stated before, these data are not final, as the calculations are still ongoing, but are already reliable enough to unravel the early stages of the fluorescence process. In the next step, we will focus on the evolution of fluorescence at longer times. Our work shows for the first time an effective way to reliably evaluate the fluorescence properties of a probe embedded in a complex biological environment.

Bibliography

- [1] Yan, K.-C., Steinbrueck, A., Sedgwick, A. C. & James, T. D. Fluorescent chemosensors for ion and molecule recognition: The next chapter. *Frontiers in Sensors* **2** (2021).
- [2] Terai, T. & Nagano, T. Fluorescent probes for bioimaging applications. *Current Opinion in Chemical Biology* **12**, 515–521 (2008).
- [3] Silva, A. P. D. *et al.* Signaling recognition events with fluorescent sensors and switches. *Chem Rev* (1997).
- [4] Miyawaki, A. Visualization of the spatial and temporal dynamics of intracellular signaling gfp-based fret for generating fluorescent indicators. *Developmental Cell* **4**, 295–305 (2003).
- [5] Dou, W. T. *et al.* Fluorescent probes for the detection of disease-associated biomarkers. *Science Bulletin* (2022).
- [6] Sliney, D. H. What is light? the visible spectrum and beyond. *Eye (Basingstoke)* **30**, 222–229 (2016).
- [7] Rennie, R. & Law, J. *Oxford Dictionary of Physics* (2019).
- [8] Lakowicz, J. R. *Principles of Fluorescence Spectroscopy Third Edition* (2006).
- [9] Hamzah, K. Singlet vs. triplet state - atomic structure (2021).
- [10] Jabłoński, A. Uber den mechanismus der photolumineszenz yon farbstoffphosphoren. *Zeitschrift fur Physik* **94**, 38–46 (1935).
- [11] Berardozzi, R. *et al.* Introduction to fluorescent proteins (2021).
- [12] Hussain, S. A. An introduction to fluorescence resonance energy transfer (fret) (2009). URL <https://arxiv.org/abs/0908.1815>.
- [13] Yuan, L., Lin, W., Zheng, K. & Zhu, S. Fret-based small-molecule fluorescent probes: Rational design and bioimaging applications. *Accounts of Chemical Research* **46**, 1462–1473 (2013). URL <https://doi.org/10.1021/ar300273v>. PMID: 23419062, <https://doi.org/10.1021/ar300273v>.
- [14] Tian, X., Murfin, L. C., Wu, L., Lewis, S. E. & James, T. D. Fluorescent small organic probes for biosensing. *Chem. Sci.* **12**, 3406–3426 (2021). URL <http://dx.doi.org/10.1039/DOSC06928K>.
- [15] Zhang, H. *et al.* Fluorescence discrimination of cancer from inflammation by molecular response to cox-2 enzymes. *Journal of the American Chemical Society* **135**, 17469–17475 (2013).

- [16] Simmons, D. L., Botting, R. M. & Hla, T. Cyclooxygenase isozymes: The biology of prostaglandin synthesis and inhibition. *Pharmacological Reviews* **56**, 387–437 (2004).
- [17] Hemler, M., Lands, W. E. & Smith, W. L. Purification of the cyclooxygenase that forms prostaglandins. demonstration of two forms of iron in the holoenzyme. *Journal of Biological Chemistry* **251**, 5575–5579 (1976).
- [18] Chandrasekharan, N. V. & Simmons, D. L. The cyclooxygenases. *Genome Biology* **5** (2004). URL <http://genomebiology.com/2004/5/9/241>.
- [19] Vane, J. R. Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. *Nature New Biology* **231** (1971).
- [20] Kato, M., Nishida, S., Kitasato, H., Sakata, N. & Kawai, S. Cyclooxygenase-1 and cyclooxygenase-2 selectivity of non-steroidal anti-inflammatory drugs: investigation using human peripheral monocytes. *Journal of Pharmacy and Pharmacology* **53**, 1679–1685 (2010).
- [21] Bhat, M. A. *et al.* Indole derivatives as cyclooxygenase inhibitors: Synthesis, biological evaluation and docking studies. *Molecules (Basel, Switzerland)* **23** (2018).
- [22] Wikipedia. Cyklooksygenaza (2022).
- [23] Sitarz, R., Kolasinska-Bzoma, M., Skórzewska, M., Polkowski, W. P. & Maciejewski, R. Cyklooksygenaza 2 i jej rola w karcenogenezie. *Zeszyty Naukowe Towarzystwa Doktorantów UJ Nauki Ścisłe*, **5** (2012).
- [24] Regulska, M. *et al.* Cox-2 inhibitors: A novel strategy in the management of breast cancer. *Drug Discovery Today* **21**, 598–615 (2016).
- [25] Kargman, S. *et al.* Characterization of prostaglandin g/h synthase 1 and 2 in rat, dog, monkey, and human gastrointestinal tracts. *GASTROENTEROLOGY* **111**, 445–454 (1996).
- [26] Watson, J. D. *et al.* *Molecular Biology of The Gene* (2004).
- [27] Soslow, R. A. *et al.* Cox-2 is expressed in human pulmonary, colonic, and mammary tumors. *Cancer* **89**, 2637–2645 (2000).
- [28] Rees, B. P. V. *et al.* Cyclooxygenase-2 expression during carcinogenesis in the human stomach. *Journal of Pathology* **196**, 171–179 (2002).
- [29] Atula, T. *et al.* Cyclooxygenase-2 expression in squamous cell carcinoma of the oral cavity and pharynx: Association to p53 and clinical outcome. *Oncology Reports* **16**, 485–490 (2006).
- [30] Appleby, S. B., Ristimaki, A., Neilson, K., Narko, K. & Hla, T. Structure of the human cyclo-oxygenase-2 gene. *Biochem. J* **302**, 723–727 (1994).
- [31] Zidar, N. *et al.* Cyclooxygenase in normal human tissues—is cox-1 really a constitutive isoform, and cox-2 an inducible isoform? *Journal of cellular and molecular medicine* **13**, 3753–63 (2009).
- [32] Hinz, B. & Brune, K. Cyclooxygenase-2—10 years later. *The Journal of pharmacology and experimental therapeutics* **300**, 367–75 (2002).

- [33] Chan, G. *et al.* Cyclooxygenase-2 expression is up-regulated in squamous cell carcinoma of the head and neck. *Cancer research* **59**, 991–4 (1999).
- [34] Hasegawa, K. *et al.* Expression of cyclooxygenase-2 (cox-2) mrna in human colorectal adenomas. *European journal of cancer (Oxford, England : 1990)* **37**, 1469–74 (2001).
- [35] Hasegawa, K. *et al.* Expression of cyclooxygenase-2 in uterine endometrial cancer and anti-tumor effects of a selective cox-2 inhibitor. *International journal of oncology* **26**, 1419–28 (2005).
- [36] Hida, T. *et al.* Increased expression of cyclooxygenase 2 occurs frequently in human lung cancers, specifically in adenocarcinomas. *Cancer research* **58**, 3761–4 (1998).
- [37] Tucker, O. N. *et al.* Cyclooxygenase-2 expression is up-regulated in human pancreatic cancer. *Cancer research* **59**, 987–90 (1999).
- [38] O’Kane, S. L., Cawkwell, L., Campbell, A. & Lind, M. J. Cyclooxygenase-2 expression predicts survival in malignant pleural mesothelioma. *European journal of cancer (Oxford, England : 1990)* **41**, 1645–8 (2005).
- [39] Subbaramaiah, K. & Dannenberg, A. J. Cyclooxygenase 2: A molecular target for cancer prevention and treatment. *Trends in Pharmacological Sciences* **24**, 96–102 (2003).
- [40] Mascaux, C., Martin, B., Verdebout, J.-M., Ninane, V. & Sculier, J.-P. Cox-2 expression during early lung squamous cell carcinoma oncogenesis. *The European respiratory journal* **26**, 198–203 (2005).
- [41] Wolff, H. *et al.* Expression of cyclooxygenase-2 in human lung carcinoma. *Cancer research* **58**, 4997–5001 (1998).
- [42] Achiwa, H. *et al.* Prognostic significance of elevated cyclooxygenase 2 expression in primary, resected lung adenocarcinomas. *Clinical cancer research : an official journal of the American Association for Cancer Research* **5**, 1001–5 (1999).
- [43] Fang, H. Y. *et al.* Cyclooxygenase-2 in human non-small cell lung cancer. *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* **29**, 171–7 (2003).
- [44] Yuan, A. *et al.* Total cyclooxygenase-2 mrna levels correlate with vascular endothelial growth factor mrna levels, tumor angiogenesis and prognosis in non-small cell lung cancer patients. *International journal of cancer* **115**, 545–55 (2005).
- [45] Hastürk, S. *et al.* Expression of cyclooxygenase-1 and cyclooxygenase-2 in bronchial epithelium and nonsmall cell lung carcinoma. *Cancer* **94**, 1023–31 (2002).
- [46] Petkova, D. K. *et al.* Overexpression of cyclooxygenase-2 in non-small cell lung cancer. *Respiratory medicine* **98**, 164–72 (2004).
- [47] Ashton, A. W. *et al.* Inhibition of endothelial cell migration, intercellular communication, and vascular tube formation by thromboxane a(2). *The Journal of biological chemistry* **274**, 35562–70 (1999).
- [48] Byun, J. H. *et al.* Association between cyclooxygenase-2 and matrix metalloproteinase-2 expression in non-small cell lung cancer. *Japanese journal of clinical oncology* **36**, 263–8 (2006).

- [49] Ermert, L., Dierkes, C. & Ermert, M. Immunohistochemical expression of cyclooxygenase isoenzymes and downstream enzymes in human lung tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research* **9**, 1604–10 (2003).
- [50] Gately, S. & Li, W. W. Multiple roles of cox-2 in tumor angiogenesis: a target for antiangiogenic therapy. *Seminars in oncology* **31**, 2–11 (2004).
- [51] Leahy, K. M., Koki, A. T. & Masferrer, J. L. Role of cyclooxygenases in angiogenesis. *Current medicinal chemistry* **7**, 1163–70 (2000).
- [52] Sánchez-Alcázar, J. A., Bradbury, D. A., Pang, L. & Knox, A. J. Cyclooxygenase (cox) inhibitors induce apoptosis in non-small cell lung cancer through cyclooxygenase independent pathways. *Lung cancer (Amsterdam, Netherlands)* **40**, 33–44 (2003).
- [53] Szostakiewicz, B. *Kliniczne znaczenie stopnia ekspresji cyklooksygenazy-2 (COX-2) w niedrobnokomórkowym raku płuca (NDRP)*. phdthesis, Gdańsk University Medyczny (2008).
- [54] Eric K, R. Targeted induction of apoptosis in cancer management: The emerging role of tumor necrosis factor–related apoptosis-inducing ligand receptor activating agents. *Journal of Clinical Oncology* (2005).
- [55] Tsujii, M. & DuBois, R. N. Alterations in cellular adhesion and apoptosis in epithelial cells overexpressing prostaglandin endoperoxide synthase 2. *Cell* **83**, 493– 501 (1995).
- [56] Masferrer, J. L. et al. Antiangiogenic and antitumor activities of cyclooxygenase-2 inhibitors. *Cancer research* **60**, 1306–11 (2000).
- [57] Tsujii, M. et al. Cyclooxygenase regulates angiogenesis induced by colon cancer cells. *Cell* **93**, 705–16 (1998).
- [58] Bartolazzi, A., Peach, R., Aruffo, A. & Stamenkovic, I. Interaction between cd44 and hyaluronate is directly implicated in the regulation of tumor development. *The Journal of experimental medicine* **180**, 53–66 (1994).
- [59] Dohadwala, M. et al. Non-small cell lung cancer cyclooxygenase-2-dependent invasion is mediated by cd44. *The Journal of biological chemistry* **276**, 20809–12 (2001).
- [60] Dohadwala, M. et al. Autocrine/paracrine prostaglandin e2 production by non-small cell lung cancer cells regulates matrix metalloproteinase-2 and cd44 in cyclooxygenase-2-dependent invasion. *The Journal of biological chemistry* **277**, 50828–33 (2002).
- [61] Chen, Y. Design and construction of cox-2 specific fluorescent probes. *Molecular and Cellular Probes* **48** (2019).
- [62] Osella, S., Marczak, M., Murugan, N. A. & Knippenberg, S. Exhibiting environment sensitive optical properties through multiscale modelling: A study of photoactivatable probes. *Journal of Photochemistry and Photobiology A: Chemistry* **425** (2022).

- [63] Wlodkowic, D., Skommer, J., McGuinness, D., Hillier, C. & Darzynkiewicz, Z. Er-golgi network—a future target for anti-cancer therapy. *Leukemia research* **33**, 1440–7 (2009).
- [64] Uddin, M. J., Crews, B. C., Ghebreselasie, K., Tantawy, M. N. & Marnett, L. J. [i]-celecoxib analogues as spect tracers of cyclooxygenase-2 in inflammation. *ACS medicinal chemistry letters* **2**, 160–164 (2011).
- [65] Uddin, M. J., Crews, B. C., Ghebreselasie, K. & Marnett, L. J. Design, synthesis, and structure–activity relationship studies of fluorescent inhibitors of cyclooxygenase-2 as targeted optical imaging agents. *Bioconjugate Chemistry* **24**, 712–723 (2013). URL <https://doi.org/10.1021/bc300693w>. PMID: 23488616, <https://doi.org/10.1021/bc300693w>.
- [66] Hospital, A., Goni, J., Orozco, M. & Gelpi, J. Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem* **8**, 37–47 (2015).
- [67] Abraham, M., Hess, B., van der Spoel, D., Lindahl, E. & the GROMACS development team. *GROMACS User Manual version 5.0.4*, www.gromacs.org access date: 26.08.2022 (2014).
- [68] Wikipedia. Periodic boundary conditions, access date: 26.08.22 (2022).
- [69] Hockney, R., Goel, S. & Eastwood, J. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics* **14**, 148–158 (1974).
- [70] Swope, W., Andersen, H., Berens, P. & Wilson, K. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics* **76**, 637–649 (1982).
- [71] Hoover, W. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **31**, 1695–1697 (1985).
- [72] Berendsen, H., Postma, J., Van Gunsteren, W., Dinola, A. & Haak, J. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684–3690 (1984).
- [73] Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
- [74] MacKerell, A. D., Banavali, J. N. & Foloppe, N. Development and current status of the charmm force field for nucleic acids. *Biopolymers (Nucleic Acid Sciences)* **56**, 257–265 (2001).
- [75] Cornell, W. *et al.* 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).
- [76] Kaminski, G., Friesner, R., Tirado-Rives, J. & Jorgensen, W. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B* **105**, 6474–6487 (2001).
- [77] Ewald, P. *Ann Phys* **64**, 253–287 (1921).

- [78] Darden, T., York, D. & Pedersen, L. Particle mesh ewald: An nlogn method for ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089–10092 (1993).
- [79] Cramer, C. Christopher j. cramer: Essentials of computational chemistry: Theories and models. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **108**, 367–368 (2013).
- [80] Thomitzni, B. Time-dependent density functional theory (tddft) - fundamentals. theory seminar of the drew group at the university of heidelberg. (2018).
- [81] Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Physical review* **136**, B864 (1964).
- [82] Levy, M. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the $\langle i \rangle v \langle /i \rangle$ -representability problem. *Proceedings of the National Academy of Sciences* **76**, 6062–6065 (1979).
- [83] Dreuw, A. Quantum chemical methods for the description of excited electronic states of medium-sized and large molecules: Theory and applications. habilitation, fakultat fur chemie, biochemie und pharmazie der johann wolfgang goethe-universitat frankfurt.
- [84] Runge, E. & Gross, E. Density-functional theory for time-dependent systems. *Physical Review Letters* **52**, 997–1000 (1984).
- [85] Rappoport, D., Crawford, N., Furche, F. & Burke, K. *Approximate Density Functionals: Which Should I Choose?* (John Wiley & Sons, Ltd, 2011).
- [86] Becke, A. D. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of Chemical Physics* **98**, 5648–5652 (1993). URL <https://doi.org/10.1063/1.464913>. <https://doi.org/10.1063/1.464913>.
- [87] Yanai, T., Tew, D. P. & Handy, N. C. A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp). *Chemical Physics Letters* **393**, 51–57 (2004). URL <https://www.sciencedirect.com/science/article/pii/S0009261404008620>.
- [88] Tawada, Y., Tsuneda, T., Yanagisawa, S., Yanai, T. & Hirao, K. A long-range-corrected time-dependent density functional theory. *The Journal of Chemical Physics* **120**, 8425–8433 (2004). URL <https://doi.org/10.1063/1.1688752>. <https://doi.org/10.1063/1.1688752>.
- [89] Sherrill, C. D. Basis sets in quantum chemistry. school of chemistry and biochemistry course in georgia institute of technology; <http://vergil.chemistry.gatech.edu/courses/chem6485/pdf/basis-sets.pdf>; access date: 27.08.22.,
- [90] Wikipedia. Basis set (chemistry), access date: 27.08.22.
- [91] Lehtola, S. A review on non-relativistic, fully numerical electronic structure calculations on atoms and diatomic molecules. *International Journal of Quantum Chemistry* **119**, e25968 (2019). arXiv:1902.01431.
- [92] Schulz, A. Basis sets used in molecular orbital calculations,. https://www.schulz.chemie.uni-rostock.de/storages/uni-rostock/Alle_MNF/Chemie_Schulz/Computerchemie_3/basis_sets.pdf, access date: 27.08.22.

- [93] Ditchfield, R., Hehre, W. J. & Pople, J. A. Self-consistent molecular-orbital methods. xii. further extensions of gaussian-type basis sets for use in molecular-orbital studies of organic molecules. *Journal of Chemical Physics* **56**, 2257 (1971).
- [94] Rassolov, V. A., Pople, J. A., Ratner, M. A. & Windus, T. L. 6-31g* basis set for atoms k through zn. *The Journal of Chemical Physics* **109**, 1223–1229 (1998). URL <https://doi.org/10.1063/1.476673>. <https://doi.org/10.1063/1.476673>.
- [95] Foresman, J. B., Head-Gordon, M. & Pople, J. A. Toward a systematic molecular orbital theory for excited states. *Journal of Physical Chemistry* (1991).
- [96] Berquist, E., Gilbert, A. & Herbert, J. *Q-Chem 5.4 User's Manual*. Q-Chem, Inc. (2021).
- [97] Gaussian website. eom-cssd optimization. <https://gaussian.com/g16eomcc-opt/>.
- [98] Frisch, M. J. *et al.* Gaussian 16 rev. c.01 (2016).
- [99] Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015).
- [100] Brooks, B. R. *et al.* Charmm: The biomolecular simulation program. *Journal of Computational Chemistry* **30**, 1545–1614 (2009).
- [101] Shao, Y. *et al.* Advances in molecular quantum chemistry contained in the q-chem 4 program package. *Mol. Phys* **113**, 184–215 (2015).