Universitat Autònoma de Barcelona

THEORETICAL ANALYSES OF PHENOMENA INVOLVING PHOTOACTIVE PROTEINS: A MULTISCALE APPROACH

MARC NADAL FERRET

Tesi doctoral

Doctorat en Química

Directors: Ricard Gelabert Peiri i Miquel Moreno Ferrer

Departament de Química

Facultat de Ciències

2014

DECLARATION

Memòria presentada per aspirar al Grau de Doctor per

Marc Nadal Ferret

Vist-i-plau

Ricard Gelabert Peiri Miquel Moreno Ferrer

Bellaterra (Cerdanyola del Vallès), September 25, 2014

Als meus pares, Montserrat i Artur; i al meu germà Alexandre

I a tots els altres familiars i amics que han format part del meu camí. Gràcies!

"Learn from yesterday, live for today,
look to tomorrow, rest this afternoon"

Charles M. Schulz, Charlie Brown's Little Book of Wisdom

ACKNOWLEDGEMENTS

Per a desenvolupar el treball de recerca que es plasma en aquestes pàgines he comptat amb l'ajuda de diferents persones i institucions.

Quant a la part humana he de dir que, a tall de *captatio benevolentiae*, se'm fa prou difícil abordar-la. Costa encertar les paraules que descriguin bé el paper que ha tingut cadascú en tot aquest periple, i em causa un cert esglai la possibilitat que la memòria em traeixi i certes persones importants quedin oblidades o poc reconegudes. Espero sortir-me'n tan bé com pugui.

Per començar, he d'agrair als meus directors de tesi, Ricard Gelabert i Miquel Moreno, així com al cap de grup Josep Maria Lluch, el suport valuós i constant que m'han ofert. Abordar un treball de recerca és complicat quan no se n'ha fet mai cap, i jo he tingut la sort d'estar a l'empara d'uns directors dels quals no només he après molt, sinó que també hi he xalat força amb les hores de converses, xerrades, discussions... que anaven des de dubtes de Fortran fins a la filosofia de la ciència. A part, he comptat amb la companyonia de la resta de membres –antics i actuals– del grup de *Dinàmica i mecanismes de les reaccions químiques i bioquímiques*. Al Juanma Ortiz i al Juanma Ramírez els aprecio l'ajuda i el tracte tan amable que em van donar quan jo tot just era un passerell en aquest ram i no sabia –com aquell qui diu– obrir una consola de Linux. A l'Edgar Mixcoha li agraeixo tot el temps que em va dedicar quan jo em trobava encallat en les beceroles del CHARMM. Amb ell i amb l'Ayax Pérez –que m'ha proporcionat inputs, suport i bons consells per a l'ús del ChemShell– hem tingut molts cafès que en què hem arreglat la ciència, el futbol, Catalunya, Mèxic i el món en general, com se sol dir. No m'he d'oblidar de Hansel Gómez, que ha estat un bon guia quan m'he hagut d'endinsar als "averns" del ChemShell. De Carlos Randino, a part d'haver-hi compartit interessants converses, dinars, excursions i inputs, cal dir que ha estat un fidel company en les penes i en les alegries que ens donaven les proteïnes fluorescents, i la recerca amb ell ha estat ben fructífera. A la Mireia Garcia li agraeixo

el seu coneixement en camps de forces i dinàmiques moleculars, així com també l'experiència d'haver-hi compartit tasques docents i d'haver-me motivat per la pedagogia.
I em deixo molta gent que ha passat pel grup, com la Lea Toledo, l'Edu Puig, l'Isaías Lans, el Reynier Suardíaz... i que en forma part, com la Laura Masgrau, l'Àngels González, el Pau Armengol, la Patricia Saura, la María Fernanda Mendoza... Així com també la gent d'altres grups de la Unitat de Química Física amb qui he conviscut aquest temps: el Jorge Alí, el Carles Acosta, la Irene Canales, el Javi Navarro, la Irene Reche... i molts i molts altres als quals ja em disculpo, d'antuvi, d'haver-me'n oblidat.
Pel que fa al suport institucional, he d'agrair a la Universitat Autònoma de Barcelona la beca de Personal Investigador en Formació (PIF) que vaig gaudir entre l'any 2009 i el 2011. Així mateix, a la Secretaria d'Universitats i Recerca (SUR) del Departament d'Economia i Coneixement (DEC) de la Generalitat de Catalunya i al Fons Social Europeu (ESF) per la beca d'investigació del programa FI-DGR, de la qual n'he estat beneficiari entre el 2011 i el 2014. També cal fer menció al Centre de Serveis Científics i Acadèmics de Catalunya (CESCA), que ha facilitat inestimables recursos computacionals. El projecte de recerca ha gaudit del finançament proporcionat pel Ministerio de Economía y Competitividad mitjançant el projecte CTQ2011-24292 i per la Generalitat de Catalunya a través del projecte 2009SGR409.

CONTENTS

ACRONYMS

ALDA Adiabatic Local Density Approximation

DFT Density Functional Theory

DVR Discrete Variable Representation

ECFP Cyan-Emitting Fluorescent Protein

ESPT Excited-State Proton Transfer

GFP Green Fluorescent Protein(s)

GGA Generalized Gradient Approximation

GIAO Gauge-Invariant Atomic Orbital

HB Hydrogen Bond

HOMO Highest Occupied Molecular Orbital

KIE Kinetic Isotope Effect

LBHB Low-Barrier Hydrogen Bond

LDA Local Density Approximation

LSS Large Stokes Shift

LUMO Lowest Unocuppied Molecular Orbital

MCTDH Multi-Configurational Time-Dependent Hartree

MD Molecular Dynamics

MM Molecular Mechanics

PDB Protein Data Bank

PES Potential Energy Surface

PYP Photoactive Yellow Protein

QM Quantum Mechanical

QM/MM Quantum Mechanical/Molecular Mechanical

RFP Red Fluorescent Protein(s)

RHF Restricted Hartree-Fock

TDDFT Time-Dependent Density Functional Theory

TDSE Time-Dependent Schrödinger Equation

TMS tetramethylsilane

ZPE Zero Point Energy

wt wild type

Part I

EXORDIUM

"A thinker sees his own actions as experiments and questions –as attempts to find out something. Success and failure are for him answers above all"

Friedrich Nietzsche

"Lasciate ogni speranza, voi che entrate"

Dante Aligheri

1

INTRODUCTION

Natural sciences have provided a huge amount of priceless information and knowledge throughout centuries. Mankind can definetely be grateful for the progress that physicists, chemists, biologists... have achieved, and it is obvious that any current research today would not be possible without the legacy of the history of sciences. It is beyond the scope of our work to explain in detail and from the very beginning the development of the milestones of science such as quantum mechanics, genetics or thermodynamics. But, assuming that we cannot reproduce the whole *movie* of what ancient Greeks used to call *natural philosophy*, we would like to highlight two recent dates that are very relevant and meaningful for the present thesis: 2008 and 2013. In 2008, the Royal Swedish Academy of Sciences awarded the Nobel prize in Chemistry to Professors Martin Chalfie, Osamu Shimomura and Roger Y. Tsien *for the discovery and development of the Green Fluorescent Protein (GFP)*.[1] Five years later, in 2013, the Nobel prize in Chemistry awarded Professors Martin Karplus, Michael Levitt and Arieh Warshel for the *development of multiscale models for complex chemical systems*[2] using both classical and quantum mechanical theory. The research presented in these pages has been carried out, approximately, within this period. It reflects that these are hot topics right now, and their importance has motivated many research groups to undertake this path, and so have we. These two subjects, namely the multiscale modeling and the biological systems –particularly proteins–, constitute the *leitmotiv* of the present thesis.

On the one hand, proteins are the chemical building blocks of life. To understand how cells and living beings work, an incredibly complex machinery of proteins has to be studied. Besides, apart from their natural function, some proteins –like GFP– can also act as biomarkers when connected to other proteins by DNA technology, which supposes an outstanding achievement in biomedicine by enabling the detection, tracing and imaging of living cells. Given their role in the *machinery* of life and their current and potential applications, proteins are the *target* of this thesis. On the other hand, since proteins are very complex entities, they require the implementation and even the development of new methods to be able to tackle their study.

This matches with the relevance of multiscale models, such that more accurate and demanding techniques can be combined with more feasible levels of calculation.

In a nutshell, this thesis is an itinerary through a collection of methodologies that we have designed and implemented, aimed to the study of biological systems. These methodologies have in common the need of a quantum mechanical analysis of different phenomena that involve proteins, but also a larger scale analysis, like the inclusion of the environment, the electrostatic effects or the dynamical effects.

1.1 MULTISCALE MODELS

During the last half century, chemistry and –particularly– biochemistry have witnessed amazing progresses. At the beginning of this period, most efforts were spent in determining the protein structure (by X-ray crystallography or by NMR-spectroscopy).[2] But the experimental procedures devoted to it had a theoretical and computational background too: the analysis of the diffraction pattern from an X-ray investigation or the spin-spin couplings obtained from a NMR experiment require a computer program that takes potentials obtained both empirically and theoretically into account. This is so because there is not enough experimental information to uniquely determine the structure of the studied system. It is a first and meaningful example of how theory and computation are crucial, also for experimental chemists.

Today, the aim of the research does not so much lie at the structure, but rather on the *function*. The experiments, helped by computing, can provide a reasonable knowledge of the structure of, say, a protein. However, questions regarding chemical mechanisms involved in that protein are much more difficult to answer empirically. To characterize transition states, to figure out if a system has enough Zero-Point Energy to populate the region of an energy maximum, to determine how a proton can be transferred and across what areas of the configuration space... are issues that are difficult –or impossible– to know experimentally. This makes theoretical modelling an essential tool as a complement to experiment. In this sense, the main achievement for which Karplus, Levitt and Warshel were awarded the 2013 Nobel Prize in Chemistry was the development of methods that used both classical and mechanical theory to manage such complex chemical systems and reactions.

At the beginning of the 70's Arieh Warshel visited Martin Karplus at Harvard so as to study molecules similar to retinal,

a chromophore responsible for animal vision that Karplus had studied.[3] Warshel could provide his knowledge of inter- and intra-molecular potentials and Karplus his quantum chemical experience. They constructed a computer program that calculated ground and excited state potential surfaces. Their method[4] consisted of treating the effects of the σ-electrons and the nuclei with a classical approach, and the π electrons were modeled using a PPP[5, 6] (Pariser-Parr-Pople) quantum chemical approach. For the first time, a hybrid method that successfully combined the advantages of classical and quantum approaches had been constructed. It had its limitations though: it was restricted to planar systems where symmetry makes a natural separation between π and σ electrons.

Arieh Warshel and Michael Levitt presented in 1976 a more general scheme[7] when studying an enzyme-substrate complex: they included the quantum mechanical energies associated with bond cleavage and charge redistribution of the substrate and the classical energies of steric and electrostatic interactions between the substrate and the enzyme. The surrounding water molecules were simulated by a microscopic dielectric model. Still some fundamental problems had to be solved, namely, to properly describe the interaction of the QM and MM parts, and the interaction of these with the dielectric surrounding.

All this work was the starting point for very valuable advances in biochemical modeling, like those developed by J. Gao,[8] F. Maseras and K. Morokuma,[9] U. C. Sing and P. Kollman,[10] and H. M. Senn and W. Thiel.[11] Taking into account these latter contributions, let us explain in Section 1.1.1 the Quantum Mechanical/Molecular Mechanics (QM/MM) method, nowadays one of the most important multiscale models.

1.1.1 *The QM/MM method*

The study of most chemical phenomena implies the calculation of the energy. For a dynamics simulation, for instance, the evaluation of the energy is essential to know the force, its gradient. If one wants to optimize the geometry of a given system, the minimum energy is sought. Or, if we want to know the absorption or emission wavelength of a given species, we will need to calculate the electronic ground and the corresponding excited states. For this reason, knowing the energy of a system is the key to understand the processes involving it.

As far as we know today, quantum mechanics is the only correct theory to describe the behavior of nature. From the Hamil-

tonian operator $\hat{H}$ (kinetic plus potential energy), the eigenfunctions ψ_n and their eigenvalues E_n furnish the stationary states and energies, according to the Time Independent Schrödinger Equation:

$$\hat{H}\psi_n = E_n\psi_n \tag{1}$$

As happens with the Newton equation, the time evolution of the system is highly related to its energy according to the Time Dependent Schödinger equation:

$$i\hbar\frac{\partial}{\partial t}\psi = \hat{H}\psi \tag{2}$$

If the system happens to be in a state ψ which is not an eigenstate of the Hamiltonian operator, but a linear combination of its eigenvalues, the mean (expected) value of the energy is given by the expression:

$$\langle E\rangle = \int \psi^*\hat{H}\psi d\tau \tag{3}$$

However, for most systems it cannot be solved exactly, which requires the use of numerical methods when integrating the Schrödinger equation. These methods are basically: variational, perturbative or a combination of both. However, even if these methods can be simplified, they are fairly demanding in terms of computational costs. When dealing with proteins, which might contain thousands of atoms, to include them all in terms of wave functions is absolutely impossible, no matter how coarse the method to integrate the Schrödinger equation is.

A cheaper alternative to the Schrödinger equation is provided by force fields. A force field is a set of functions depending on parameters that provide the energy of the system from a given nuclear configuration: distances, angles, torsions, Coulomb charges, Lennard-Jones van der Waals interactions... This approach is commonly known as MM (molecular-mechanics). A prototypical MM energy expression would be:

$$\begin{aligned} E_{MM} = &\sum_{\text{bonds}} k_d(d-d_0)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 \\ &+ \sum_{\text{dihedrals}} k_\phi[1+\cos(n\phi+\delta)] \\ &+ \sum_{A,B}\left\{\left[\epsilon_{AB}\left(\frac{\sigma_{AB}}{r_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}}\right)^{6}\right] + \frac{1}{4\pi\epsilon_0}\frac{q_A}{q_B}\right\} \end{aligned} \tag{4}$$

where d, θ and ϕ correspond to bond lengths, angles and torsions, respectively. As for n and d, they designate the torsional

multiplicity and phase. The bonded force constants are k_d, k_θ and k_ϕ. The sum over A and B is done only for those pairs which are not bonded; r_{AB} is the distance that separates them, and ϵ_{AB} and σ_{AB} are the Lennard-Jones parameters. Regarding q_A and q_B, they are the atomic partial charges of atoms A and B. These parameters are usually derived from quantum calculations or from experimental data that are fitted to the analytical functions that appear in Equation 4.

Calculating the energy by means of molecular mechanics allows for determining the stability of a given configuration of a system (say, a protein, for instance), by minimizing its energy (optimizing the geometry), or by checking its dynamics evolution by integrating the second Newton equation. However, no chemical reaction can be described with a force field as the function that describes the energy of two bonded atoms is harmonic and thus bond cleavage is impossible. Force fields are useful to understand the global behavior of a protein, but sometimes one can be interested in describing a particular region of the protein with especially good precision. Besides, parameters are usually adjusted only for ground state, so the treatment for excited states becomes trickier.

In these cases, there is always the possibility of "reducing" the system and including only a given number of atoms: the ones that are more directly involved in a chemical reaction or those that are going to be more relevant to describe particular features of the system, like calculating spectral wavelenghts, reactivity, transition states... However, although this *cluster* approach can provide reasonable results for some systems, there is plenty of accummulated experience showing that proteins are highly complex entities and this method fails to reproduce the effects of the environment. For instance, in Section 3.3, when calculating the absorption wavelenghts of the mKeima protein, it has been seen that the full quantum mechanical calculations that include a considerably large number of atoms do not always correctly describe the experimental absorption wavelenghts.

The hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) approach is a method to calculate electronic energies by means of combining quantum mechanics (that is, by solving the Schrödinger equation) with a classical potential energy (the force field).

It is a fact that a way to combine both kinds of calculation is clearly required for situations in which a group of atoms needs to be accurately described –that is, with a QM approach– but the rest of the environment (protein, solvent...) needs to be

also included, albeit in an less rigorous but affordable way. This division is the key of the QM/MM approach.

This way, the atoms that are most relevant for the studied phenomenon, are included in the QM region and the Schrödinger equation of their non-relativistic Born-Oppenheimer Hamiltonian is then solved (for example, throughout the research reflected in this thesis, we have not been able to include more than 100 atoms in the QM region). The rest of the system is included as a configuration of atoms that contribute to the total energy according to the force field.

Once the whole system S is divided into QM and MM regions, there are two possible schemes to calculate the QM/MM energy: substractive or additive. Substractive QM/MM schemes consist of calculating the MM energy of the entire system, calculating the QM energy of the QM region, and finally calculating the MM energy of the QM region. This last term is then substracted to the sum of the first two ones to avoid double counting. This yields

$$E^{sub}_{QM/MM}(S) = E_{MM}(S) + E_{QM}(QM) - E_{MM}(QM) \tag{5}$$

When following the additive scheme, the MM energy is calculated just in the MM region. This energy and the energy of the QM region are summed. To this sum, a QM/MM interaction term has to be added. This interaction term has to be thought of as a correction term. In all, it yields

$$\begin{aligned} E^{add}_{QM/MM}(S) = E_{MM}(MM) + E_{QM}(QM) \\ + E_{QM/MM}(QM, MM) \end{aligned} \tag{6}$$

The QM/MM allows for different approaches to solve the Schrödinger equation and for multiple force field. Furthermore, each particular way to calculate the interaction term defines a particular QM/MM method. Most QM/MM methods are additive. In Section 2.3, more specific details about how we have applied the QM/MM will be given.

1.2 BEYOND QM/MM

The combination of classical and quantum approaches is a solid asset that allows the theoretical study of biological systems that otherwise could not be done. However, not all the theoretical issues regarding biochemistry can be solved by only selecting the corresponding QM and MM regions, the QM and MM methods and a suitable way to assemble them. QM/MM is a very valuable tool from which further analyses have to be done.

Commonly, the majority of available QM/MM software calculates the potential energy of the system within the Born-Oppenheimer approximation in the electronic ground state. As a result, only the electronic Hamiltonian is solved for the QM region, being the nuclei-nuclei repulsion a constant term. This implies that the nuclei are treated like point masses. On the other hand, most force field are parameterized for the electronic ground state and, although some QM interfaces support the electronic excited-states calculations, its implementation is not easy and sometimes restricted to single point evaluations in which the MM region can polarize the QM region (but not the other way around). This practically circumscribes the QM/MM optimizations and dynamics simulations to the ground-state level. The aim of this thesis is to open Pandora's box by facing these issues. In other words, a QM/MM approach based on Born-Oppenheimer, static and ground-state calculations is not sufficient: we intend to go beyond it. QM/MM methods will help us tackle specific problems that will require a deep discussion and development of new methodologies. These problems mainly involve the determination of electronic excited states and the quantum treatment of the hydrogen atom, together with a dynamical approach.

One of the issues we are focusig on is the determination of the electronic absorption spectrum of a protein. This requires the calculation of electronic excited states that take into account the whole system. In order to have a realistic ensemble of configurations at a given temperature, a dynamics simulation is also required. Calculating electronic excited states is also needed when dealing with photochemical reactions, such as excited-state proton transfers in fluorescent proteins. This is another topic we intend to study in this thesis, for which QM/MM techniques will be followed, but further steps will have to be done. Furthermore, for the excited-state proton transfers, the proton might not behave just as a point mass, but as a quantum entity. The treatment of the hydrogen atom as a wavefunction is also adopted when determining its position in a short hydrogen bond in order to figure out if there is a Low-Barrier Hydrogen Bond (LBHB), which is another of the issues that we are studying. In the next sections, our challenges regarding the Low-Barrier Hydrogen Bonds and the excited-state proton transfers in fluorescent proteins are explained. In Section 1.5, an outline of the objectives is given.

1.3 BRIEF INTRODUCTION TO LOW-BARRIER HYDROGEN BONDS

Low-barrier hydrogen bonds are a particular kind of short hydrogen bonds (less than 2.5 Å between donor and acceptor) in which the hydrogen atom is not covalently bonded to either of the electronegative atoms, but shared between both of them. According to the explanation given by Cleland and Kreevoy,[12] when two electronegative atoms (say oxygens, for instance) have similar pK_a values, the hydrogen atom is covalently bonded to one of them, and the interaction with the other is mainly electrostatic. As the two oxygens become closer to each other, the barrier between the two possible hydrogen positions becomes lower, and eventually is low enough so that the zero point energy (ZPE) level is at or above the barrier, as shown in Figure 1.

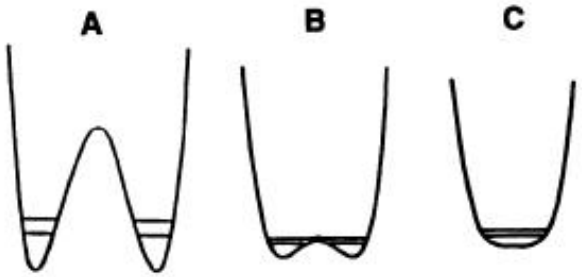


Figure 1: Potential functions for a double-well hydrogen bond (A), a low-barrier hydrogen bond (B) and a single-well hydrogen bond. The two first vibrational levels are given. This image has been taken from reference [12]

So as to generalize this statement to more dimensions, let us think of the potential energy hypersurface associated with a hydrogen bond as a multidimensional double well: the two minimum energy structures, each one corresponding to the proton attached to one or the other of the electronegative atoms, are separated by a classical potential energy barrier (by classical we mean that no zero-point energy is included). In a more generalized definition of LBHB, the classical energy barrier is low enough so that the nuclear wave function corresponding to the ground vibrational level of the double well reaches its maximum values at the region of that energy barrier. Thus, the proton can freely move in the region between the two electronegative atoms.[13–15]

Two of the most used physicochemical parameters that are symptomatic of the existence of an LBHB are a short distance between the hydrogen bond donor and acceptor atoms (< 2.55 Å for O–H–O and < 2.65 Å for O–H–N) and a far-downfield

proton ^{1}H NMR chemical shift (17–19 ppm). These values are meaningful features that are correlated with the existence of the LBHB, but they do not guarantee that the proton will actually be found in the middle of the two atoms. Let us bear in mind this latter remark, since it is a major point of the approach adopted by us in this work. When dealing with systems that are prone to host an LBHB, the quantum effects of the proton should be taken into account to determine its actual expected position.

Let us recall that a hydrogen nucleus is not a point mass, but a particle that accomplishes the Heisenberg uncertainty principle, and thus cannot be "located" at a given point. To be able to understand the underlying physics of a given problem, the Schrödinger equation must be solved and the wavefunctions obtained (ground and excited states). As it is well known in quantum mechanics, the wavefunction of the system (it is usually in the ground state, but any state in which the system is) contains all the information about it, that can be obtained by means of the operators (observables). The expected value of a certain observable as, for instance, a geometry $\boldsymbol{R}$ (an ensemble of positions), would be

$$\langle \boldsymbol{R} \rangle_E = \langle \Psi_E | \hat{\boldsymbol{R}} | \Psi_E \rangle \tag{7}$$

where Ψ_E is the eigenstate with energy E, that is to say, the vibrational state of the system. A different energy hypersurface (a different Hamiltonian) would lead to a different solution of the Schrödinger equation, that is, to a different wavefunction. And, consequently, to different expected values of the observables.

However, it is a very widespread approximation in quantum chemistry to treat the nuclei as point masses. By calculating the quantum electronic energy at a given nuclear geometry, a potential energy surface (PES) of the system can be found within the Born-Oppenheimer approximation. The minimum of this surface is thought to correspond to the stable coordinates of the nuclei. If the atom is a point mass, it will be found at that minimum in absence of kinetic energy. Even in cases in which the atom is treated as a wavefunction (for instance, when calculating the vibrational levels of energy), it is often assumed that its mean position (expectation value of the position) is the same as the minimum of the potential energy surface. What lies beneath this assumption is the harmonic approximation: if we consider that a minimum is a harmonic well, the minimum of the well coincides with the expectation value of the geometry (and, therefore, with where the nuclei will be most likely found). But when the system presents anharmonicities,

i.e., when the potential energy surface $V(x) = \sum_i k_i x_i^2$ (x_i being the corresponding elongation) does not properly describe the actual area around the minimum, the approximation is not valid anymore. This is the case of LBHBs: being the energy of the first vibrational state equal to or higher than the barrier, an anharmonical analysis of the nuclear vibrational wavefunctions is mandatory. Not only the coordinates of the nuclei need to be calculated with this approach, but a completely analogous discussion could be issued on the values of NMR-related quantities, such as chemical shifts and H–H coupling constants.[14–17] In fact, other apparently unrelated systems such as elongated dihydrogen (and compressed dihydride) complexes,[16–20] are examples that could only be theoretically studied by means of anharmonical vibrational analysis, thus revealing the limitations of the harmonic approximation. Besides, if the system has a certain kinetic energy (not only the ZPE), for it is at a certain temperature, the limitations of finding a minimum to predict the actual positions of the nuclei are still more evident, as seen in Chapter 3.

As for the LBHBs, it is widely accepted that hydrogen bonds can stabilize transition states or intermediates in enzymatic reactions.[21–23] Particularly, about 20 years ago LBHBs were proposed to play an important role in enzyme catalysis.[12, 13, 24–30]. This possibility triggered intense research that led to opposing viewpoints about the real existence of LBHBs in proteins and their chemical properties and functions in enzyme catalysis.[14, 15, 31–41] Nonetheless, the proof of their existence in enzymes had been elusive so far.

It is worth mentioning that many of the reported systems containing an LBHB were simple molecules studied in the gas phase or in crystal. Systems studied in solutions had properties more similar to a structure with the proton localized near the donor atom.[42, 43] According to Perrin,[43] this could be understood taking into account that even two completely symmetrical potential energy minima for the process in the gas phase would be solvated differently due to different instantaneous configurations of the solvent with the system. This statement has to be remembered very carefully because, as it will be shown all over Chapter 4, the differences between the system in crystal and in solution are crucial. The peremptory need to include the effects of the surrounding atoms and the motion of the whole system is one of the central concerns of this thesis (as it will be stressed for electronic absorption spectra calculations in Chapter 3). This way, we ought not to neglect the presence

of solvent and the constant movement of the system, for they may eventually break the symmetry of the double well.

In 2009, an LBHB was found in a protein for the first time. Using neutron diffraction techniques, Yamaguchi *et al.*[44] detected and reported an LBHB in the Photoactive Yellow Protein (PYP) crystal structure. Nevertheless, very recent (2012) theoretical studies[45] based on electronic structure calculations and Nuclear Magnetic Resonance (NMR) on PYP in solution claimed that there was not such LBHB.

By means of combining the calculation of QM/MM electronic energy surfaces with the obtention of the nuclear vibrational wavefunctions, some light will be shed on the PYP's LBHB issue. It will be proved how important it is, when dealing with such systems, to solve the nuclear Schrödinger equation and the relevance of considering under which conditions (crystal or solution) we are analyzing the problem, which will lead to a more reasonable and conciliating understanding of the abovementioned studies.

1.4 FLUORESCENT PROTEINS: THE GFP

1.4.1 *History*

To get to where we are today, the study of fluorescent proteins has undergone a very long and fascinating trajectory. In the sixties, Osamu Shimomura observed that green light was emitted from the jellyfish *Aequorea victoria* in the Pacific ocean (see Figure 2, left). At that time, it was thought that the light of all bioluminiscent organisms was produced by the reaction of luciferin and luciferase, but it was impossible to extract luciferin and luciferase from the rings of the jellyfish. Shimomura came up with an idea: the luminiscence reaction probably involved a protein, which meant that luminiscence could be reversibly inhibited at a certain pH. Experiments done by Shimomura himself proved him right. The overall mechanism, though, turned to be fairly more complex:[46] the jellyfish *Aequorea* emits green light whereas the photoprotein aequorin (extracted from the same organism) emits blue light in the presence of Ca^{2+} ions. The jellyfish contains another protein –the GFP– that, when exposed to that blue light, emits the green light, in what is known as an *in vivo* energy transfer process. The GFP presented an excitation maximum at 460 nm and an emission maximum at 508-515 nm.

In 1962 Shimomura isolated GFP[48] and in 1979 characterized the structure of its chromophore,[49] which was proposed

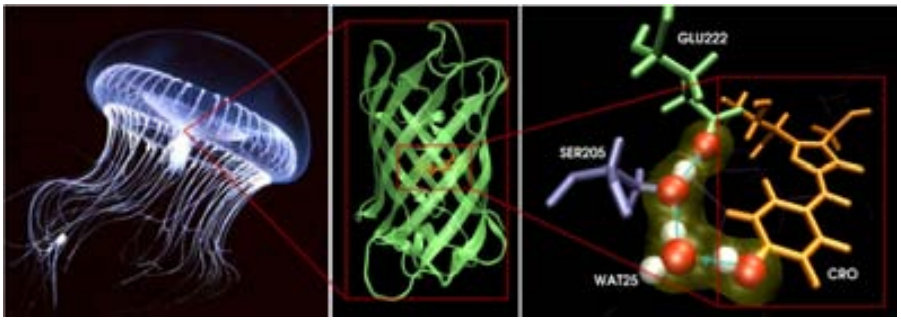


Figure 2: On the left, the *Aequorea Victoria* jellyfish. In the middle, the barrel-shaped structure of the GFP. On the right, the residues that are responsible for the photochemical reaction that causes the fluorescence.[47]

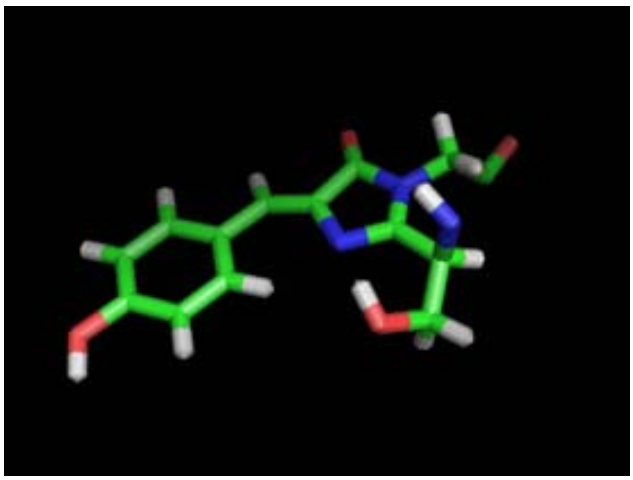


Figure 3: Representation of the GFP chromophore.

to be a substituted 4-(*p*-hydroxybenzylidene)-5-imidazolone moiety (see Fig 3). This chromophore is derived from the primary amino acid sequence through the cyclization of serine-dehydrotyrosine-glycine[50–53]. GFP consists of 238 amino acids and contains about 20,000 atoms. Once crystalized, its structure consists of eleven beta sheets set out in a barrel shape, with an alpha helix that goes through it (Figure 2, in the middle). According to Shimomura's own words at the Nobel lecture in 2008,[1] this protein remained useless for the next 30 years after the discovery. It was the fact that Douglas Prasher reported GFP's cloning and nucleotide sequence[54] what triggered the intense research for practical applications. GFP is particularly suitable as a biomarker for its brightness given its high quantum yield (that is, the ratio between number of photons emitted and number of photons absorbed), of 0.8 and its high photostability (that is to say, the chromophore emits a large number of photons before the fluorescence is irreversibly lost).[46] Martin Chalfie and

co-workers[50] discovered that the GFP is generally non-toxic, and its gene can be fused to the gene of a given target protein. After this engineered protein is expressed in the cellular medium it is obtained with a smallish tag (the GFP) attached. After this, if one wants to track this target protein, it is possible since it keeps doing its normal activity, but is now linked to the GFP, which means that can be followed by microscopic monitoring of the GFP fluorescence. That is what Chalfie and coworkers did: they expressed the GFP gene clone in *Escherichia coli* and in *Caenorhabditis elegans* (see, for instance, Fig 4). After that, it was discovered that GFP could also be expressed in other organisms, which made it a potential universal genetic tag. GFP could be used to monitor where a protein was, how fast it was moving, how it reacted (or responded to a chemical inhibitor), when genes were expressed... Since this achievement, tens of thousands of publications involving GFP (and similar proteins) have appeared: it has implied a revolution in the biological sciences.

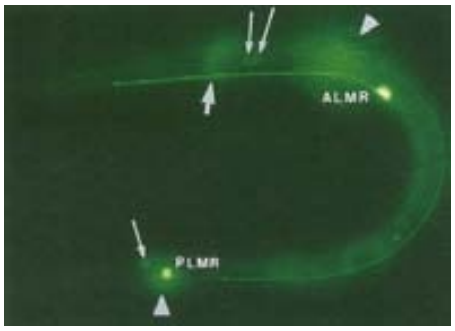


Figure 4: Expression of GFP in a small number of neurons (ALMR and PLMR) in a *C. elegans* larva. This image was published in the *Science* journal in 1994.[50]

1.4.2 *Broadening the palette*

The next step in the GFP research was to modify it to obtain more specific proteins. The mutagenesis of GFP allows for protein *engineering* that leads to improving its brightness and photostability, and can have profound effects on its spectral properties. Roger Y. Tsien[55] performed several mutations on the GFP that resulted in a palette of proteins that glowed cyan, blue, green and yellow. However, no red colors were obtained by direct mutation of GFP. Pity, as red light presents several advantages that make it particulary convenient for research to obtain red emitting fluorescent proteins. First, it penetrates biological tissues more easily since they are more transparent to

it. Second, the intensity of the Rayleigh-scattered light is lower for radiation with a longer wavelength, as it obeys the proportion $I \propto \lambda^{-4}$. Finally, each photon from red light is less energetic than those blue or green, which makes it less harmful for living tissues around. For all this, there was an interest in obtaining red or near-infrared fluorescent proteins.

Tsien went on doing important contributions in the design of fluorescent proteins apart from the GFP. Lukyanov and coworkers[56] published the discovery of drFP583, better known as DsRed, a fluorescent protein responsible for the red coloration around the oral disk of a coral of the *Discosoma* genus. The DsRed protein, which is an obligate tetramer, aroused a huge interest as complementary tracer. In the DsRed chromophore, the π-system of the GFP chromophore is extended by an additional N-acylimine moiety.[57] By means of extensive mutations the DsRed protein could be transformed into a monomer,[58, 59] which opened many doors to its potential use as a biological tag. Tsien an co-workers characterized the DsRed chromophore[60] and developed many fluorescent proteins from DsRed and from mRFP1[61] (another red fluorescent protein derived from that) with a greater photostability: mCherry, mStrawberry, mTangerine, mOrange, mBanana, mHoneydew...[62] (series commonly known as the "mFruits", m for monomeric) These mutated proteins were mutated again and again until a library of fluorescent proteins potentially suited for multicolor imaging had been created:[63] (mOrange R17H, G40A, T66M, A71V...)

The development of DsRed-like proteins led to the obtention of LSSmKate1 and LSSmKate2[64]. Besides, another protein containing a DsRed-like chromophore was obtained from the *Montipora* sp stony coral: mKeima[65]. LSSmKate1 and 2 and mKeima are characterized by its important Large Stokes Shift (LSS), that is, an outstanding difference between the absorption and emission wavelengths. For instance, mKeima is excited in the blue at about 440 nm, but it is red fluorescent (620 nm). The LSSs in the fluorescence emission make these proteins useful for multicolor two-photon *in vivo* imaging. As mentioned, the red-shifted FPs are most valuable for the deep-tissue and whole-body imaging since they present reduced autofluorescence, low light scattering and minimal absorbance at longer imaging wavelenghts.[66] For instance, a combined use of mKeima with a cyan-emitting fluorescent protein (ECFP) allowed for a two-color imaging, as seen in Fig 5.[67] In Chapter 3 we will come back to the DsRed like chromophores.

There is, understandably, an intense pressure to develop FPs that have convenient absorption/emission wavelengths to fit

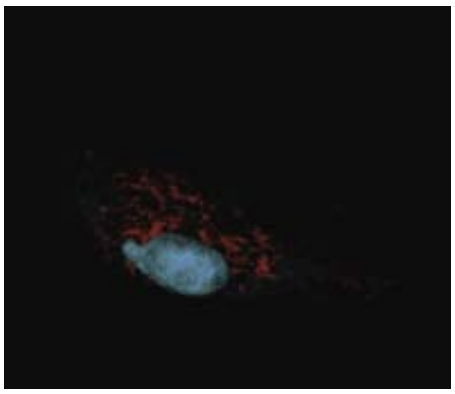


Figure 5: Visualization of the mitochondria and nucleous of a Vero cell by two-photon laser scanning microscopy using mKeima and ECFP.[67]

new uses in imaging. How do scientists in the field go about this demand? Random mutagenesis is a common procedure in molecular biology that can furnish many different proteins with widespread properties. It is commonly used, but it is after all a random process so many efforts are fruitless as unusable or less-than-optimal FPs are developed. However, if the chemical reaction that causes that given property is well understood and the involved residues are identified, the mutations can be rationally designed and chosen to obtain proteins with particular properties.[64] The need for engineering specific marker proteins pushes for a good understanding of the internal mechanism. That is why theoretical and computational approaches are essential for the progress of the biological sciences.

1.4.3 *The nature of the uorescent species: the proton wire in GFP*

As pointed out, the theoretical study of a protein mechanism is fundamental to rationally design mutants that can perform more efficiently when used in biomedical applications.

As for the GFP, there are many experimental data about it, its mutants and their emission. However, the process that takes place at the chromophore has not been fully dilucidated and shedding some light on it is part of the aim of this thesis. *What happens* within the GFP that causes the green fluorescence?

To begin with, Chattoraj *et al.*[68] observed that the GFP absorption spectrum has a maximum at 398 nm and another one at 478 nm (Figure 6). It was conjectured that the first one (more energetic) corresponded to the protonated chromophore, whereas the second one (less energy), to the anionic chromophore. One plausible explanation[52, 55] to this is the following: by increas-

ing the pH of the system, the less energetic absorption peak becomes higher and shifts towards higher wavelenghts (~ 525 nm). Therefore, it seems reasonable that the chromophore is more likely to be deprotonated with a higher OH^- concentration. Besides, the fluorescent green light emission has a wavelength of 509 nm, more similar to the one that is thought to correspond to the anionic chromophore. From this remarkable difference between both absorption bands, and considering that one of these bands is similar to the emission wavelength, one can infere that the photoexcitation produces a chemical reaction: both absorption bands arise from two different species: the proton transfer converts one into the other. From now on, we will call A the form of protonated chromophore and B the form with ionized chromophore. It is nowadays accepted[52, 69] that chromophore excitation causes the motion of a proton from the chromophore to the crystallographic water molecule Wat25. As a result, from this very water molecule a proton would then be transferred to Ser205, and finally another proton from Ser205 to Glu222, as seen in Fig 7.

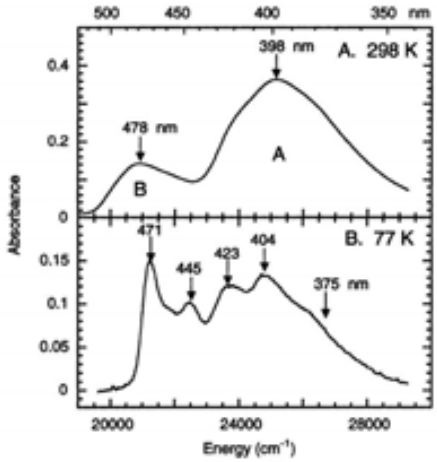


Figure 6: GFP absorption spectrum observed by Chattoraj *et al.* (obtained from the very same article)[68]

Basically, from the observations of Chattoraj *et al.* it is suggested that a chemical species A is photoexcited to A^* (see Figure 8). A^* would evolve to B^* (photoreaction), and from B^* to B, emitting radiation. Nonetheless, the state A^* vanishes very fast, whereas the formation of B^* is significantly slower. This led to think of the existence of one or two intermediate species that could correspond to non-relaxed forms of the species B. This way, A^* would evolve to I^*, and I^* would change to the ground state I emitting the well-known green light at 504 nm. However, part of this I^* would slowly change its conformation

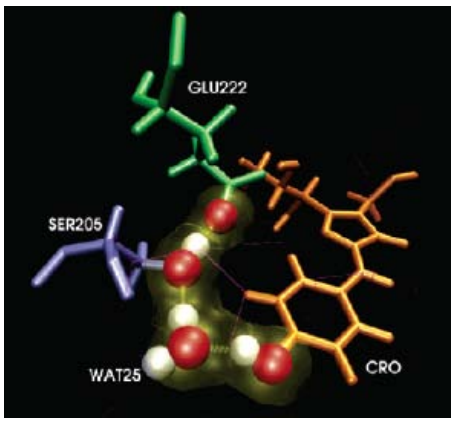


Figure 7: GFP proton wire[70]

until becoming B^* (see Figure 8 and Figure 35). Emission of radiation has also been observed at 482 nm, but more seldom: this is thought to correspond to the $B^* \rightarrow B$ transition.

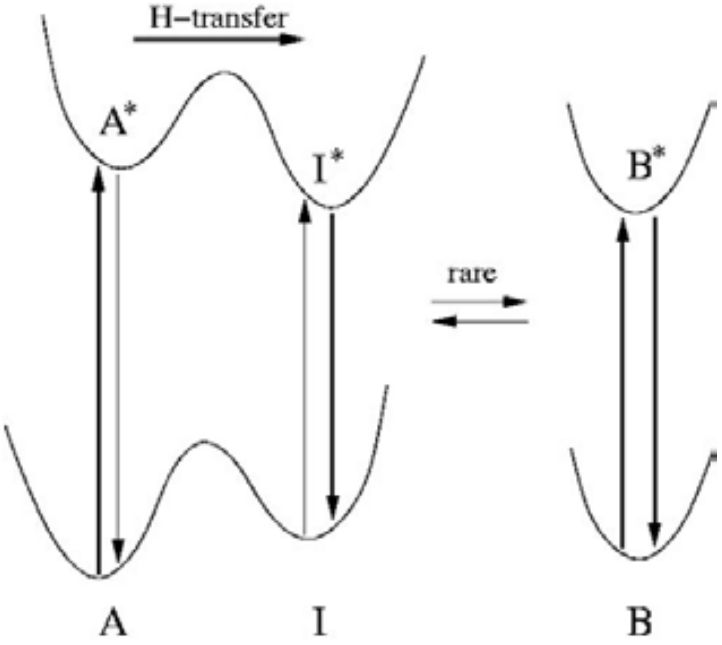


Figure 8: Schematic representation of the conjectured evolution of chemical species due to the absorption and emission of radiation. The proposed intermediates I_1 and I_2 are considered too[71]

Chattoraj *et al.*[68] detected that, in the GFP chemical reaction, when the labile hydrogen atoms had been replaced by deuterium atoms, the emission time-scale was 5 times greater. This implies, first of all, that the hydrogen is the *main character* of the reaction, because when its mass is increased the reaction is remarkably slower. To account for this, a proton transfer reaction is proposed. Secondly, the Kinetic Isotope Effect (KIE) is

pretty outstanding: the reaction with hydrogen atoms occurs five times faster than with deuterium, being this latter only twice as massive as the former, which leads to think that the system goes through a classically forbidden region. This phenomenon, that is, the fact that the potential energy of a system can be higher than its total energy is impossible in classical mechanics: the presence of wavefunction population in these regions is commonly known as *tunneling*. Tunneling is known to be crucial to understand the kinetics of certain chemical reactions, and this need prompted us to find a mathematical and computational procedure aimed to calculate its contribution.

The wealth of data on GFP formed a very suggestive picture, yet one that is difficult to clarify only with experimental approaches. Theoretical studies could do a lot to shed light on the mechanism operating in GFP, but this was a complicated task, as it involves a biological system, excited electronic states and proton transfer with significant tunneling (or suspicious thereof). In the past, our research group calculated the potential energy surface (in the ground and in the photoactive electronic excited states)[70] using the largest system that could be afforded then: no environment, and CASSCF/CASPT2-level energy calculations. With these approximations, the potential energy profiles suggested that the most favorable series of events would imply the Ser65-Glu22 proton to be transfered first. This was a counter-intuitive result, but no definite conclusions could be stated until a quantum dynamics simulation were performed. Experimental data relying on ultrafast time-resolved infrared spectroscopy managed to detect the formation of protonated Glu222.[72] Di Donato *et al.*,[73] by means of time-resolved mid-infrared spectroscopy, detected the formation of Glu222 even before the deprotonation of the chromophore. In any case, the calculated potential energy surface would allow for the integration of the time-dependent Schrödinger equation, that is, the evolution of the wave packet of the system.[74]. That work was actually the first quantum dynamics simulation of the operation of the proton wire in the GFP. The quantum dynamics simulation was performed both for perprotio and perdeutero cases. It was seen that, after photoexcitation, the hydrogen bond between the chromophore and Wat25 shortened suddenly for the first 20 fs. Two qualitatively different regimes were observed: firstly, during the first tens of picoseconds, when the system had excess energy, the dynamics was very fast and impulsive and a large fraction of the system developed an ionized chromophore. Later on, a slower pace of formation appeared. In all, it was concluded that the reaction was essentially concerted,

and faster than expected according to experimental data.[68] Part of this is due to the fact that, as mentioned, environment was missing from the system (or, in other words, the results were for a wire in gas phase). This can certainly be improved with QM/MM models.

Afterwards, a question was looming in the horizon: provided that a reliable (in chemical sense) quantum simulation were available, how can we obtain an estimation of the fraction of reaction that occurs via tunneling? This discussion will be presented in Chapter 5.

1.5 OBJECTIVES

In our path to a deeper analysis in multiscale modelling, the first issue that we shall present is the treatment of the absorption spectra in red fluorescent proteins, in Chapter 3. QM/MM calculations are required to account for the interactions of the whole protein with the chromophore, and their importance will be proved. But in this study, a combination of dynamics simulations and Excited-State Electronic Calculations (ESEC) has been proposed and both electrostatic and thermal effects have been revealed to be essential to correctly reproduce the experimental absorption spectra. Dealing with very common tools in QM/MM, a further step has been done designing a specific methodology that has confirmed that chromophores cannot be treated as rigid, isolated entities and, furthermore, has shed some light on the discussion of reactivity and proton wires in red fluorescent proteins mKeima and LSSmKate2.

The next step has consisted of going beyond the Born-Oppenheimer approximation. In Chapter 4, we will deal with the discussion of the existence of an LBHB in the PYP. An LBHB is essentially a quantum-mechanical phenomenon. As mentioned, it is a short hydrogen bond in which the energy barrier between reactants and products is low enough so that the vibrational function of the hydrogen atom may have its mean value at the position of the energy maximum. This implies that the atom must not be treated as a mass point, but as a delocalized wavefunction. To be able to figure out if there is or not such LBHB, and in which conditions, QM/MM energy surface calculations and dynamics simulations have had to be done, but anharmonical vibrational calculations have been needed as well.

Finally, more complicated methods will be exposed in Chapter 5. We will undertake the study of the chemical reaction that causes the fluorescence in the GFP. To tackle this study, a quantum mechanical approach will be needed to treat the nuclei,

and a new methodology to calculate probability flux will be presented. That is, we are going to find out the rate at which the wavepacket goes from reactants to products, and across which regions. Besides, a study of the excited-state reactivity that includes the electrostatic effects within a QM/MM scheme will be proposed.

The study of the chemical reaction in GFP in Chapter 5 will be presented as follows: firstly, we will thoroughly explain the method to compute the probability flux as a surface integral. With this, not only will it be possible to know the rate of the proton transfer, but also to know across which regions of space: this will allow for determining the tunneling and the recrossing, for example. This novel method is a powerful tool and can have many more applications. However, it has been applied to an already existing quantum dynamics simulation[74] perfomed by some members of our group. This simulation started from QM energy calculations of the chromophore and nearest residues, but the available resources had not allowed for considering the electrostatic effects of the whole protein. For this reason, a new potential energy surface will be proposed for electronic ground and excited states. This work is currently in progress.

From what has been exposed so far, we can list the main objectives of the present thesis:

- Find a way to correctly reproduce electronic absorption spectra in fluorescent proteins. To this end, QM/MM electronic embedding combined with TDDFT excited-state electronic calculations, optimizations and, especially, dynamics simulations, will be required. So as to find a proper description of absorption spectra of a fluorescent protein, it will be of utmost relevance to include the electrostatic effects of the protein environment on the chromophore. Also, thermal effects will be essential, and will be included by means of dynamics simulations that furnish a plethora of configurations that, statistically counted, yield a more reliable shape of the electronic absorption spectra. The calculations will be done on the RFP mKeima, with the objective of figuring out when its chromophore is protonated, even having different pH values or isomeries. The same methodology will be applied to the RFP LSSmKate2.

- Undertake an anharmonic analysis of the expected value of a hydrogen atom in short hydrogen bonds to understand whether or not a hydrogen bond can be considered a Low-Barrier Hydrogen Bond, and under which conditions. This will imply QM/MM calculations of the poten-

tial energy surface, and a further obtention of the nuclear wavefunctions of the hydrogen atom. Calculations of the chemical shift will also be needed. The aim of this analysis is to discern whether (and when!) there is an LBHB in the PYP. This will demand a separate analysis of the protein in crystal and in solution, and the results of the potential energy surfaces, the nuclear wavefunctions and the expected positions and chemical shift values will be analyzed for each case. Not only is our objective to shown the importance of an anharmonical analysis, but also to prove the relevance of taking into account the conditions in which the protein actually is, namely crystallized or solvated.

- Find a novel method to calculate the surface integral of the probability current (probability flux) in a generic system of coordinates, and a way to implement it (numerical methods and software development). This has to allow an analysis of the tunneling and recrossing contributions in a given chemical reaction.

- Study the photochemical reaction that produces the fluorescence in the GFP: a triple excited state proton transfer. By means of the mentioned method to calculate the probability flux, and taking the wavefunction from a previous quantum dynamics simulation by Vendrell *et al.*[74] on the GFP proton transfer, we want to compute the total flux according to: the regions it goes through –particularly if they are classically permitted or not– or to the sign of the flux –whether there is recrossing or not. Many more information can be pried out with this method, and the implementation by means of an adaptative Monte Carlo integration can furnish a detailed description of the total flux, tunneling, recrossing... according to the configuration of the system.

- Improve the Potential Energy Surface of the GFP in the ground and photoactive states, such that it takes into account the whole protein environment by means of QM/MM methods, and with a TDDFT excited-state electronic calculations. From a proper potential energy surface, a quantum dynamics simulation on the GFP proton wire can be done, with more reliable results. These results can be analyzed and the probability flux could be computed, thus having a still more accurate vision of *how* the GFP proton transfer occurs.

More generally, the whole objective of this thesis is to apply different multiscale methodologies to big systems such as proteins, and develop new strategies that shed some light on particular issues of their internal molecular procedures. These issues usually involve what lies beyond the already known QM/MM methodologies: calculation of electronic excited states, anharmonical vibrational analysis, tunneling...: very specific and demanding phenomena that have to be studied in such complex and important systems as proteins.

2 METHODS

2.1 DENSITY FUNCTIONAL THEORY

The Density Functional Theory provides a way to find the electronic structure of a system, particularly the energy of the ground state, by considering the energy as a functional of the density. That is, the energy is a function whose argument is another function: the electronic density.

The Hohenberg-Kohn theorems state the following:[75] first, that an external potential V_{ext} uniquely determines the ground-state density, thus the potential is a unique functional of the density. And second, that a universal functional for the energy $E[\rho]$ can be defined in terms of density, and that the exact ground state is the global minimum value of this functional. In other words, there is a variational principle that guarantees that $E(\rho \neq \rho_{exact}) > E_{exact}$.

To tackle the obtention of the ground-state energy, a ficticious system is proposed in which the electronic density is the same as the real system, but in which the electrons do not interact between them. From this system, and taking a Slater determinant as the initial guess with a set of orbitals χ_i (Kohn-Sham orbitals, from which the density can be obtained: $\rho(\vec{r}) = \sum_{i=1}^{N_{oc}} |\chi_i(\vec{r})|^2$), the Kohn-Sham equations are presented, which are self-consistent equations to be solved iteratively, analogous to the Roothan-Hall equations for the Hartree-Fock method:

$$\hat{h}_{KS}\chi_i = \epsilon_i\chi_i \tag{8}$$

where

$$\hat{h}_{KS} = -\frac{1}{2}\nabla^2 + v_{eff}(\vec{r}) \tag{9}$$

The term v_{eff} is the sum of a Coulombic term (that includes the electron-nucleus attraction v_n and the electron-electron repulsion) plus the exchange-correlation potential v_{XC}:

$$v_{eff}(\vec{r}) = v_n(\vec{r}) + \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} + v_{xc}(\vec{r}) \tag{10}$$

The exchange-correlation potential $v_{xc}(\vec{r})$ includes corrections to the kinetic energy of the ficticious system, and the exchange-correlation term. The trickiest issue of the Kohn-Sham equa-

tions is that $v_{xc}(\vec{r}, t)$ cannot be obtained exactly, thus some approximations are required.

2.1.1 *Time-Dependent Density Functional Theory*

As has been mentioned, the Kohn-Sham Density Functional Theory is limited to ground electronic states (or lowest states of a given symmetry/multiplicity). But a step further is needed when dealing with excited electronic states, and that is why Time-Dependent Density Functional Theory (TDDFT) is presented. In particular, TDDFT is very necessary for our purposes since it has been established as a suitable and proper way to study and analyze the ESPT processes like those from the fluorescent proteins,[76–78], provided a hybrid Coulomb-attenuated functional, such as CAM-B3LYP, is used, as explained in Section 2.1.2. TDDFT calculations for excited-states have been performed in Section 3.3 and Section 3.5 for the mKeima protein, in Section 3.6 when dealing with the LSSmKate2 protein, and in Section 5.8 when computing ground and excited electronic PESs for the GFP proton wire.

As for the TDDFT methodology, Runge and Gross formulated[79] an expression analogous to the Kohn-Sham equations, but with a time-dependent electronic density. Let us explain their mathematical development. To begin with, the Runge-Gross theorem is required, which is the time-dependent analogue of the first Hohenberg-Kohn theorem: it states that the time-dependent electron density, $\rho(r, t)$ determines the time-dependent external potential $V(r, t)$, up to a spatially constant, time-dependent function $C(t)$. This implies that the wavefunction is a functional of the electron density:

$$\Psi(\vec{r}, t) = \Psi[\rho(\vec{r}, t)]e^{-i\alpha(t)} \tag{11}$$

As happened with the second Hohenberg-Kohn theorem, one variational principle can be formulated. If the time-dependent wavefunction $\Psi(\vec{r}, t)$ is a solution of the time-dependent Schrödinger equation with the initial condition $\Psi(\vec{r}, t_0) = \Psi_0(t)$ then the wavefunction corresponds to a stationary point of the quantum mechanical action integral, defined as

$$A[\rho] = \int_{t_0}^{t_1} dt \, \langle \Psi(\vec{r}, t) | \, i\frac{\partial}{\partial t} - \hat{H}(\vec{r}, t) \, | \Psi(\vec{r}, t \rangle \tag{12}$$

This action is a functional of $\rho(\vec{r}, t)$ according to the Runge-Gross theorem. According to the least-action principle applied to quantum mechanics, the action must be stationary. Thus,

$$\frac{\partial A[\rho]}{\partial \rho(\vec{r}, t)} = 0 \tag{13}$$

Now, let us define the so-called universal functional $B[\rho]$:

$$B[\rho] = \int_{t_0}^{t_1} dt \, \langle \Psi[\rho](\vec{r}, t)| \, i\frac{\partial}{\partial t} - \hat{T}(\vec{r}) - \hat{V}_{el-el}(\vec{r}) \, |\Psi[\rho](\vec{r}, t)\rangle \tag{14}$$

This way, the action $A[\rho]$ can be thought of as the sum of two parts: $B[\rho]$, that depends only on the system of electrons, and the other depending on the applied potential $v(\vec{r}, t) = V_{el-nuc}(\vec{r}) + V(\vec{r}, t)$:

$$A[\rho] = B[\rho] - \int_{t0}^{t_1} dt \int d\vec{r} \rho(\vec{r}, t) v(\vec{r}, t) \tag{15}$$

In analogy to the time-independent Kohn-Sham equations, Runge and Gross start by assuming the existence of a time-dependent noninteracting reference system under an external one-particle potential $v_s(\vec{r}, t)$, whose electron density $\rho_s(\vec{r}, t)$ is equal to the electron density of the real system $\rho(\vec{r}, t)$. As was done with the Kohn-Sham equations, the noninteracting system is described by a Slater determinant $\Phi(r, t)$, consisting of single-electron orbitals $\phi(\vec{r}, t)$. Therefore, the density is

$$\rho(\vec{r}, t) = \rho_s(\vec{r}, t) = \sum_{i}^{N} |\phi_i(\vec{r}, t)|^2 \tag{16}$$

and the time-dependent Schrödinger equation yields:

$$i\frac{\partial}{\partial t}\phi_i(\vec{r}, t) = \left(-\frac{1}{2}\nabla_i^2 + v_s(\vec{r}, t) \right) \phi_i(\vec{r}, t) \tag{17}$$

It can be shown[80] that the relationship between the real potential $v(\vec{r}, t)$ and the noninteracting potential $v_s(\vec{r}, t)$ is

$$v_s(\vec{r}, t) = v(\vec{r}, t) + \int d\vec{r}' \frac{\rho(\vec{r}', t}{|\vec{r} - \vec{r}'|} + \frac{\partial A_{xc}[\rho]}{\partial \rho(\vec{r}, t)} \tag{18}$$

where A_{xc} is the exchange-correlation part of the action integral. As well as happened with the exchange-correlation potential v_{xc} in Equation 10, A_{xc} takes into account the exchange-correlation part of the potential and the correction between the real and the ficticious system's kinetic energy. If we include Equation 18 in Equation 17 we obtain:

$$i\frac{\partial}{\partial t}\phi(\vec{r}, t) = \big(-\frac{1}{2}\nabla_i^2 + v(\vec{r}, t) + \int d\vec{r}' \frac{\rho(\vec{r}', t)}{|\vec{r} - \vec{r}'|} + \frac{\partial A_{xc}[\rho]}{\partial \rho(\vec{r}, t)} \big) \phi_i(\vec{r}, t) \tag{19}$$

or, equivalently,

$$i\frac{\partial}{\partial t}\phi_i(\vec{r},t) = \hat{F}^{KS}\phi_i(\vec{r},t) \tag{20}$$

which is the time-dependent analogue to Equation 8. As happened with the Kohn-Sham equations, $\frac{1}{2}\nabla_i^2$ describes the kinetic energy of the electrons, $v(\vec{r},t)$ describes the external time-dependent potential and the Coulombic interactions of other electrons with the electron under consideration. As it has been said, A_{xc} takes into account the exchange-correlation part of the potential, but, as happened with v_{xc}, the exact expression of the functional is not known, therefore some approximations are required. It is worth mentioning that it is very common, as a first approximation, the so-called *adiabatic local density approximation* (ALDA). When ALDA is assumed, the originally time-dependent local exchange-correlation functional is replaced by the local time-independent one. This implies that the density varies slowly with time, and allows the use of local ground-state exchange-correlation functionals for TDDFT.

Gross and co-workers[81] exploited the fact that the frequency-dependent linear response of a finite interacting system has discrete poles at the excitation energies $\Omega_m = E_m - E_0$ of the unperturbed system. This way, the energy of the unperturbed (time-independent) system is the ground-state energy, and the excited states are obtained from the first-order time-dependent perturbation. Drew and Gordon[80] formulate it as follows: the time-dependent Kohn-Sham orbitals (Equation 20) can be expressed as a linear combination of M time-independent basis functions (with time-dependent coefficients):

$$\phi_p(\vec{r},t) = \sum_j^M c_{pj}(t)c_{pj}(t)\chi_j(\vec{r}) \tag{21}$$

Then, Equation 20 can be expressed as

$$i\frac{\partial}{\partial t}\mathbf{C} = \mathbf{F}^{KS}\mathbf{C} \tag{22}$$

where the ith column of matrix $\mathbf{C}$ contains the time-dependent expansion coefficients of $\phi_i(\vec{r},t)$, and $\mathbf{F}^{KS}$ is the matrix representation of the time-dependent Kohn-Sham operator in the given basis. When right multiplying Equation 22 by $\mathbf{C}^\dagger$ and substracting its Hermitian transpose, the following equation is obtained:

$$\sum_q \{F_{pq}P_{qr} - P_{pq}F_{qr}\} = i\frac{\partial}{\partial t}P_{pr} \tag{23}$$

Commonly, the linear-response approach is used. That is, the excitation energies are obtained from the linear time-dependent response of the time-independent ground state electron density to a time-dependent external electric field. In other words, the ground-state is obtained when the right term of Equation 23 is zero

$$\sum_{q}\{F_{pq}P_{qr} - P_{pq}F_{qr}\} = 0 \tag{24}$$

and, once the equation is solved, the time-dependent external field is applied, and the first-order (linear) response to this perturbation is analyzed.

$$P_{pq} = P_{pq}^{(0)} + P_{pq}^{(1)} \tag{25}$$

and

$$F_{pq} = F_{pq}^{(0)} + F_{pq}^{(1)} \tag{26}$$

From this, it can be shown[80] that a tensorial expression can be reached:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B^*} & \mathbf{A^*} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \tag{27}$$

where **X** and **Y** are, respectively, the vectors with the excitation and deexcitation coefficients, and the eigenvalues ω are the energies of the excited states. A and B are matrices derived from the time-dependent Kohn-Sham equations:

$$A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + (ia|jb) + (ia|f_{xc}|jb) \tag{28}$$

$$B_{ia,jb} = (ia|bj) + (ia|f_{xc}|bj) \tag{29}$$

In which the last term is defined as the response function of the chosen exchange-correlation potential.

A further simplification is the Tamm-Dancoff approximation,[82] which consists of neglecting the matrix **B**, thus simplifying Equation 27:

$$\mathbf{AX} = \omega\mathbf{X} \tag{30}$$

2.1.2 *B3LYP and CAM-B3LYP*

As previously mentioned, the exact exchange-correlation functional is not known, so some approximations are required. Some

well-known approximations are the Local Density Approximation (LDA), in which the exchange-correlation energy is assumed to be a function of the density. In other approximations such as the Generalized Gradient Approximations (GGA), the energy includes a non-local term that depends on the density gradient. The following step in approximations is formed by the Meta-GGA, in which the energy functional depends on the density, the density gradient, but also on the density Laplacian –that is, the kinetic energy.

The so-called hybrid functionals include a term calculated from the exchange energy obtained by the Hartree-Fock method, E_x^{HF}. One of the most well-known and used hybrid functional is the Becke three-parameter method (B3LYP)[83], in which the exchange-correlation term is:

$$E_{xc}^{B3LYP} = E_x^{LDA} + a_0(E_x^{HF} - E_x^{LDA}) + a_x(E_x^{GGA} - E_x^{LDA}) + E_c^{LDA} + a_c(E_c^{GGA} - E_c^{LDA}) \quad (31)$$

with $a_0 = 0.20$, $a_x = 0.72$ and $a_c = 0.81$. B3LYP is one of the most commonly used functionals in computational chemistry. In our case, though, in Chapter 3 we have had to deal with TDDFT calculations in fluorescent proteins in which charge-transfer exited states are involved, as some recent works regarding the LSSmKate2[76, 84] have disclosed. The point is that TDDFT remarkably underestimates the excitation energies when used with pure DFT funcionals or with B3LYP, for they describe the long-range exchange effects poorly.[76, 85, 86] This is why we have decided to work with the hybrid Coulomb-attenuated CAM-B3LYP functional[86], specifically designed to avoid such a spurious effect by including one term that depends on a flexible parameter that accounts for the proportion of the Hartree-Fock term in the calculations. This has given good results when charge-transfer states are involved.[87]

2.2 MOLECULAR-MECHANICS DYNAMICS SIMULATIONS

A classical dynamics simulation is the integration of the classical Newton equation of a system (that is to say, find its trajectory by means of classical dynamics). Such a procedure has been done for the mKeima protein in Section 3.4. For this case, the energy has been calculated by means of a force field.

Force fields, though more inexact than quantum wavefunctions, are a powerful tool to predict the energy of a system as a function of its nuclear configuration. Let us focus on the CHARMM22 force field[88–90]. This force field includes a har-

monic term for distances between bonded atoms, a harmonic term for the angles, a torsion (cosine) term for dihedrals, a harmonic term for coplanar atoms, and partial charges and van der Waals parameters for the non-bonding interactions. This force field, implemented in CHARMM-35 software[91, 92], had already been used in our group when dealing with molecular dynamics in the GFP protein[70]. In fact, it was particularly useful to the purpose of this mentioned paper since the CHARMM-22 parameters had previously been obtained by Thiel and coworkers[93] specifically for the GFP chromophore as, being a non-standard residue, did not have the corresponding parameters. This has also saved a considerable amount of work and has been useful, in the present thesis, for similar chromophores like mKeima or LSSmKate2, as explained in Section 3.4. For this reason, the MM approach that has been adopted in this thesis has been the CHARMM22. It is worth recalling though, that other force fields such as Amber[94] or Gromacs[95–98] consist of very similar kinds of functions to describe the interactions, so no great changes should be expected from using one or another. As for the MM dynamics simulations on the mKeima protein, the CHARMM-35 software[91, 92] has been used.

Having the energy as a function of the nuclear coordinates, one can obtain the gradient, that is to say, the force. In order to integrate the Newton equation (and to have the corresponding coordinates and velocities at different times) we have used stochastic boundary conditions (see Section 3.4). This means that the system is divided in three regions. There is a Newtonian zone in which Newton equations govern the dynamics. Then, a Langevin zone where Langevin (dissipative) dynamics governs the equations. Finally, there is the reservoir region in which atoms cannot displace at all. The dissipative terms included in the Langevin dynamics account for the omitted degrees of freedom from the reservoir zone. The equations of motion (Newton's or Langevin's, depending on the region) for the heating, equilibration and dynamics have been integrated with a leapfrog algorithm,[99] as implemented in the CHARMM program.

Nevertheless, the MM description of a system, although powerful and cheap in terms of computational effort, can sometimes fail to properly describe a chemical system. This will happen whenever an interaction outside those considered to fit the force field form and parameters appears in the system. Such a case, as will be described ahead, arises when dealing with LBHBs. At this point, it may be convenient to combine the MM description with the quantum mechanical approach. This

combination, known as QM/MM, has already been explained in Section 1.1.1.

The Newtonian (or Lagrangian) classic equations can also be solved with the QM/MM energy. This is called QM/MM dynamics simulation: it is not a proper quantum dynamics, as the Newton equation still rules, but the energy term is more accurate since it has been obtained –partially– with a QM approach. A QM/MM dynamics simulation has been done for the Photoactive Yellow Protein (PYP, see Section 4.2.2). More details about QM/MM dynamics simulations are given in Section 2.3.2.

2.3 THE CHEMSHELL QM/MM

As mentioned, the QM/MM approach allows for many different ways to assemble QM and MM regions. In order to combine the QM and MM calculations, we have used the ChemShell package[100]. This package is able to connect software for QM calculations (such as Gaussian, MNDO, Turbomole, GAMESS-UK...) with software for MM calculations (CHARMM, GROMOS..) from specific inputs. In our case, we have chosen to use DL-POLY[101], a module included in ChemShell that calculates the energy from a given force field. To treat the QM/MM coupling, we have used the *hybrid* option. As explained in the ChemShell manual,[100], the total energy of the system is essentially the sum of the QM and the MM energies obtained from the modules, but ChemShell calculates the coupling between both regions. For this purpose, link atoms are used. Link atoms are fictional hydrogen atoms that are placed between each pair of bonded QM and MM atoms. They are not taken into account by the QM or MM modules. Actually, they are not part of the molecule definition, but the fact that their positions are determined by QM and MM atom coordinates allows the analytic derivative of the QM/MM energy to be corrected for forces acting on link atoms. Angle terms that involve one MM atom and two QM are not considered, since a change in the position of the link atom already changes the QM energy. The electrostatic coupling includes the MM centers as point charges in the QM code with the so-called *shift* correction, in which the charge of MM atoms bonded to QM atoms is shifted away, and a point dipole is placed on the recipient atom of this charge shift. The polarization of the MM region due to QM atoms has not been taken into account.

2.3.1 *Electronic embedding*

One particular case of the QM/MM calculation is the electronic embedding. We have used this method for single-point calculations in which the calculated energy is that corresponding to the QM region (no MM energy is calculated). However, the effect of the MM part on the QM portion of the system is included: the point charges of the MM region polarize the QM part, modifying its orbital description. This idea was developed by Hall and Smith.[102, 103] For our purposes, this kind of electronic calculation requires that the QM software is capable of considering the MM charges when calculating the orbitals, and implies that point charges of the MM region are known. When optimizing or running dynamics simulations, the MM energy terms need to be calculated as well, so this approach would not be useful any more. However, for single point calculations, it has been a powerful tool to us.

In our research, QM/MM calculations have been really important, both for ground and excited electronic states. It is possible to use ChemShell to assemble the MM and QM regions (considering the corresponding force field) and then run the calculation with Gaussian 09,[104] thus being able to use functionals that are not supported by ChemShell, like the CAM-B3LYP.[86] Gaussian 09 allows the electronic embedding (by means of the *charge* or the *massage* keywords) and includes the CAM-B3LYP functional. The point charges have been taken according to CHARMM22 force field, as it has already been stressed. In particular, in Section 4.2, a grid of single point QM/MM energy calculations is obtained by means of this procedure. In Section 3.3, excited-state electronic calculations have been performed following this very scheme. So as to simulate the absorption spectra, the electronic embedding has been used to calculate excited states on different snapshots taken from the corresponding molecular dynamics simulation. This will be explained in detail in Section 3.5 (mKeima) and Section 3.6 (LSSmKate2).

2.3.2 *QM/MM optimizations and dynamics*

ChemShell QM/MM energy calculations do not only provide single-point calculations, but this software includes devices to perform energy minimizations and dynamics simulations. For instance, in Section 3.3, when studying the absorption spectra of mKeima, we have wanted to analyze the effect of QM/MM optimizing the geometries so that they reach energy minima.

Also, in Section 4.2.2, before starting the QM/MM molecular dynamics simulation on the solvated Photoactive Yellow Protein, a geometrical optimization has been performed. For these, ChemShell hybrid internal coordinate optimizer (HDLC-opt[105]) has been used (with the default parameters). HDLC-opt, when performing geometry optimizations, uses the low memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm, which is a quasi-Newton optimization algorithm using BFGS for Hessian update. For both cases of QM/MM optimization, the QM energy has been calculated with Gaussian 09,[104] using DFT[106] theory with the B3LYP functional[107] and the 6-31+G(d,p) basis set. Even though CAM-B3LYP has been reported to perform sometimes better than B3LYP also in the ground state[108, 109] (see Section 4.2.1), the ChemShell interfaces for QM/MM optimizations and dynamics do not support it.

It needs to be pointed out that, when optimizing a system, sometimes not all of the atoms are included either in the QM or in the MM region: some of them –usually, the more external ones– are not even taken into account for the optimization. The QM and MM atoms whose coordinates are optimized are called the *active region*. The ChemShell software takes into account the restraints of the atoms that are fixed.

As for the QM/MM dynamics simulations, the DL-POLY package supports a range of direct dynamics integration algorithms which can be coupled with any other ChemShell module (as long as it provides energy and gradients). In Section 4.2.2, a QM/MM molecular dynamics simulation is carried out on the Photoactive Yellow Protein (PYP), with the MNDO software[110] and the AM1 semiemipirical method[111] for the QM treatment. In that section, another advantage of the QM/MM is shown: in some particular cases, such as when Low-Barrier Hydrogen Bonds (LBHBs) are suspected, no parameters are available in the force fields (at least in those known to us), and they would not make a lot of sense anyway. It is not only a matter of accuracy what requires here the QM treatment, but the fact that an LBHB is a qualitatively different phenomenon, that can only be explained by means of quantum mechanics.

Part II

RESULTS AND DISCUSSION

"If you try and take a cat apart to see how it works, the first thing you have on your hands is a non-working cat"

Douglas Adams

3

ABSORPTION SPECTRA: ELECTROSTATIC AND THERMAL EFFECTS

In this chapter, a methodology in the study of fluorescent proteins is presented: the analysis of the electrostatic effects caused by the residues around the chromophore and the thermal agitation of the protein, and their influence on the modulation of the absorption spectra of fluorescent proteins. We are going to prove that this methodology is very necessary when dealing with such systems. So as to prove this need, we have taken the mKeima protein as a target to perform the different calculations. A succint introduction to absorption spectra and its importance is given in Section 3.1. In Section 3.2 the mKeima and its main features will be described. In Section 3.3, the effect of the surrounding atoms and the geometric minimizations will be evaluated and discussed. In Section 3.4, the molecular dynamics (MD) simulations on the mKeima will be explained, and combining the results of the MD with the electrostatic environment effects, a method to simulate the spectra will be presented in Section 3.5. This method will be applied to a similiar fluorescent protein, the LSSmKate2: the corresponding results will be shown in Section 3.6. Finally, some concluding remarks will be given in Section 3.7.

3.1 INTRODUCTION

Absorption and emission spectra are one of the most valuable sources of information of chemical species, and, particularly, of fluorescent proteins, since they allow the characterization of, respectively, reactants and products in a photochemical reaction. In this chapter, we are focusing on the simulation of the absorption spectra of Red Fluorescent Proteins (RFP).

A chemical reaction is often described as the transformation of a chemical species into another. However, the definition of "chemical species" becomes trickier as more atoms are involved in the process. For instance, a protein such as the Green Fluorescent Protein (GFP) contains more than $2,000$ atoms. If it is solvated, the entire system may require more than $10,000$ atoms to be properly described. Still, it is possible to distinguish reactants from products: in the case of the GFP, where an Excited-State Proton Transfer (ESPT) is thought to occur,

the protein with protonated chromophore is considered a reactant, while considered a product when the chromophore is anionic: both species have different spectroscopic properties.[53] But even though reactants and products can be classified, their features are not unique any more given the huge amount of possible configurations that the system can adopt. For this reason, it would be more precise to think of the reactants as a collection of configurations that are likely to become products by means of a photochemical reaction (and products should be thought of as an ensemble of configurations as well). As a result, if there is an ensemble of multiple reactants, there is also a corresponding ensemble of spectroscopic wavelenghts that correspond to the individual excitation of each of these structures. We should bear in mind that a single protein at a given temperature can show different wavelengths, as it displays a bunch of configurations in which charges, hydrogen bonding networks or long-range interactions are constantly changing. In other words, the experimental data available concerning radiation absorption is, in general, the absorption spectrum, which describes how much radiation is absorbed at a given wavelength. It is usually a band encompassing a wide range of wavelengths, more or less symmetric and showing at least a maximum. The band can be thought of as consisting of individual absorption signals (specific spectral lines) originated in different radiation-molecule interaction events, each one corresponding to a particular structure of the system, and what is registered is usually the envelope.

The fact of having such a plethora of available configurations reveals the highly complex and changing environment of the chromophore. Presently, there is solid agreement that it is the way in which a chromophore has been matured what determines the fluorescent properties of a protein,[57] more than the similarity of the amino acids that compose it. For instance, the common tripeptide Met-Tyr-Gly is able to form chromophores spanning a 175-nm range for the maxima of emission spectra.[112] Despite this, the atoms that surround the chromophore need to be taken into account as well. It has been reported that local environmental features such as the position of charged amino acids residues, hydrogen bonds or hydrophobic interactions can yield shifts of up to 40 nm in absorption and emission maxima.[112]. Let us focus on the kind of RFP represented by drFP583[113] (commercial name DsRed). In this kind of chromophores, the π-system of the GFP chromophore is extended by an additional N-acylimine moiety, altough the mechanism of the chromophore maturation is

not fully understood yet.[57] Recently developed proteins like mKeima,[65] LSSmKate1[114] and LSSmKate2[114] (440 nm - 463 nm absorption, 605 nm – 624 nm emission), or proteins like mNeptune[64], eqFP670[115] and TagRFP675[116] (592 nm – 611 nm absorption, 646 nm – 670 nm emission), all of them containing a DsRed-like chromophore, have exhibited novel photochemical properties by means of rationally modifying the chromophore environment.[64] Such a phenomenon suggests that not only does the chromophore determine the absorption and fluorescence spectra, but also the surrounding residues and the solvent play an important role.

We are not trying to underrate the fact that the nature of the chromophore and its maturation are crucial to account for the fluorescence wavelength, but we definetely claim that a correct description of the absorption properties of a protein must consider these two factors, namely, the large number of atoms that surround the chromophore, and the great amount of configurations that change over time. In other words, the complexity within the complexity.

3.2 MKEIMA

The monomeric Keima (mKeima) is a red fluorescent variant of a protein from the *Montipora* sp. stony coral.[65] Its backbone structure consists of eleven beta sheets displaying a barrel-like shape with one alpha helix going through it (see Figure 9), very reminiscent of wt-GFP. Its chromophore is a DsRed-like chromophore, formed by a tripeptide that involves glutamine, tyrosine and glycine (see Figure 10). mKeima absorbs and emits light maximally at 440 nm and 620 nm respectively. This remarkable difference between the absorption and emission wavelengths has been referred to as Large Stokes Shift (LSS)[64, 65] although, strictly speaking, the absorbing and emitting species are not the same. The LSS makes mKeima a suitable target for dual-color single-laser fluorescence cross-correlation spectroscopy.[65] Besides its tantalizing usefulness, this protein is particularly suited to theoretical studies since several crystallographic structures have been recently published,[117, 118] showing significant differences between them.

Violot *et al.*[117] have observed two absorption maxima at the wavelengths of 440 nm and 590 nm, and one emission maximum at 620 nm. It has been assumed[117] that the 440 nm absorption maximum corresponds to the neutral chromophore, whereas the anionic one yielded the 590 nm maximum. From their work, three crystallographic structures are available at the

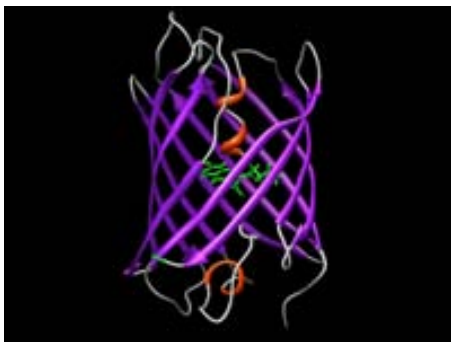


Figure 9: Backbone structure of the mKeima protein, with the chromophore in its center

Protein Data Bank (PDB): one obtained at pH 3.8 (ID 2WHS), which presents a *cis* chromophore; one obtained at pH 8.0 (ID 2WHU), which shows a *trans* chromophore; and a third one presenting both chromophores at a time, obtained at pH 5.6. On the other hand, Remington and co-workers[118] do not mention the 590 nm absorption maximum and the crystallographic structure supplied by them has been obtained at pH 7.0 (ID 3IR8) possessing a *cis* chromophore. They do not find evidence of a *trans* chromophore though its presence with less than a 10% proportion is not excluded. In Figure 10, both configurations of DsRed-like chromophore are displayed. In both isomers, an ESPT is thought to be responsible for the observed fluorescence but the proton wire responsible of the ESPT process is foreseen to be different from that originally found in wild type (wt) GFP involving the motion of three protons.[71] From the crystallographic data, a proton wire of two protons is predicted for the *cis* chromophores whereas just one proton is expected to transfer in the *trans* isomer. But neither is there agreement with the isomerization of the chromophore nor with its protonation state, for Blanchoin and co-workers[117] think that there is a counter-intuitive reverse relation with the pH of the medium: when analyzing the absorption spectra, they observe that the signal of the species considered anionic vanishes as the pH rises. They suggest that the chromophore is protonated at pH 8.0 because Asp157 should be deprotonated (see Figure 11), which forces Ser142 to H-bond to it, whereas at pH 3.8 Asp157 is protonated, so Ser142 is able to stabilize an anionic chromophore. As for Remington and co-workers, they propose a situation in which the *cis* chromophore is protonated at pH 7.0 and the proton is transferred from the chromophore to Ser142, and from Ser142 to Asp157 (see Figure 12).

This lack of agreement makes mKeima particularly suitable to undertake a theoretical study of the modulation of the spec-

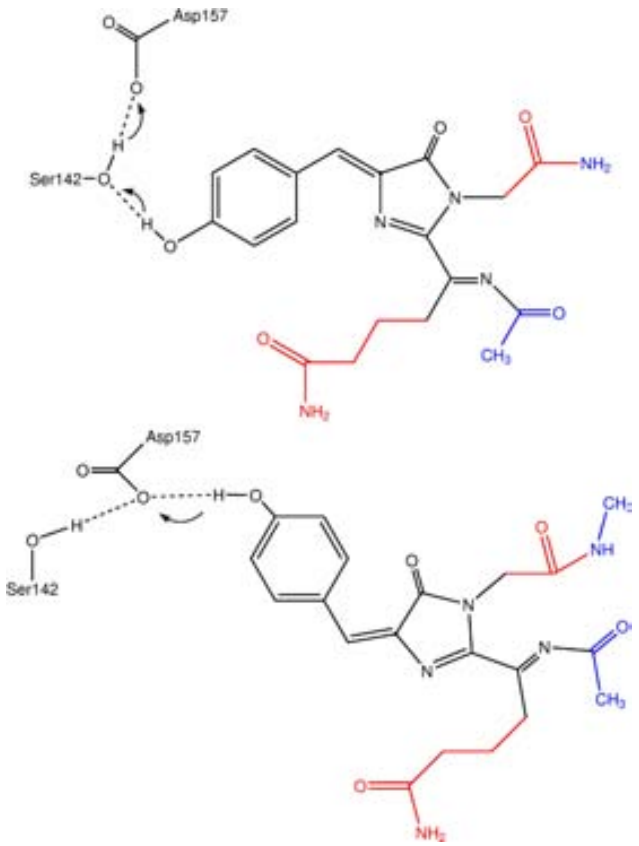


Figure 10: The protonated form of the mKeima *cis* (up) and *trans* (down) chromophore and atoms of its immediate surroundings. Arrows indicate the putative proton transfer(s) that lead(s) to the deprotonated (anionic) chromophore structure, as some authors have suggested.[117, 118] Wave lines indicate covalent bonds that have been cut and saturated with hydrogen atoms in some reduced models of the protein considered here (see text for the details). In black: atoms considered in the *optimized chromophore* model. In red: atoms that have been included in the *only chromophore* (unoptimized) model. In blue: atoms of the Ser142 and Asp157 residues additionally considered in the *optimized wire* model

trum mediated by the environment that could shed some light on the particular issues of its isomerization and protonation state. It is a matter of fact that the different X-ray structures were found following different procedures, but no more information on mKeima's structure or plausible reaction pathways has been published.

To deal with the mentioned topic, several theoretical procedures have been considered here. To begin with, a series of increasingly sophisticated static models have been undertaken. Ranging from the chromophore to the whole protein, we

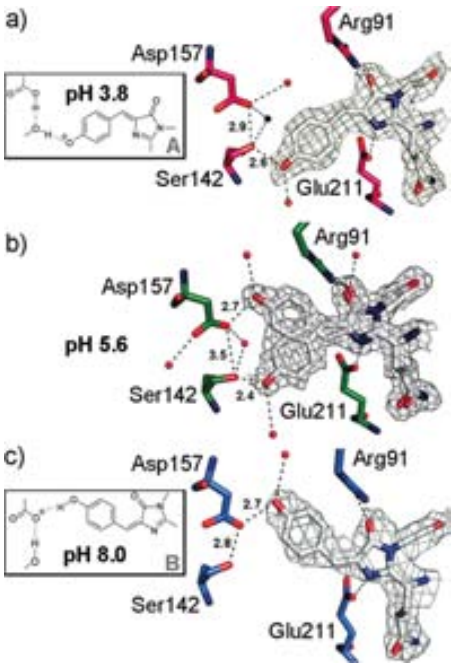


Figure 11: Representations of the mKeima chromophore structures at different pH obtained by Violot *et al*. This figure has been taken from reference [117]

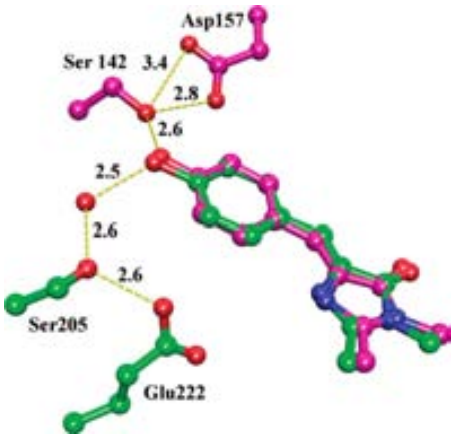


Figure 12: Ball and stick figure of the proton relay of mKeima (purple), showing the chromophore, Ser142 and Asp157. Hydrogen bonds are shown as dashed lines with lengths given in Å. The corresponding proton relay of GFP is superimposed in green. This figure has been taken from reference [118]

have evaluated the absorption wavelenghts changes by including more and more atoms in the models, and by optimizing some of them.

From the results of these static models it is concluded that the effect of the environment is highly remarkable, but not enough to satisfactorily account for the absorption spectra. At

this point, an extensive MD simulation has been carried out, with a double goal: to ensure to what extent the crystallographic structures represent also stable dynamical structures during a MD simulation and to provide an ensemble of structures from which the excited states' properties could be calculated, thus allowing for thermal fluctuations that yield a largely improved description of the observed physical and chemical properties of mKeima.

3.3 EXCITED-STATE ELECTRONIC CALCULATIONS ON STATIC MKEIMA STRUCTURES

The calculations of excited states on static structures have been done by means of Gaussian 09.[104] All these calculations have been performed for each of the three mentioned crystallographic structures, and for each protonation state of the chromophore. For the ground state, Density Functional Theory (DFT)[106] calculations have been used, with the CAM-B3LYP[86] functional and the 6-31+G(d,p) basis set. A total of 15 excited electronic states have been evaluated using Time-Dependent Density Functional Theory (TDDFT) with the same functional and basis

The mKeima chromophore (see Figure 10) comes from the tripeptide sequence Gln-Tyr-Gly. At an earliest stage, our calculations have involved only the chromophore (with some atoms belonging to the contiguous residues, shown in blue in Figure 10), directly taken from the corresponding crystallographic structure. Then the chromophore has been geometrically optimized. To prevent the glutamine contained in the chromophore to bend too close to the two coplanar rings, it has been partly removed, preserving the π-system of the chromophore, and so have been some atoms from the glycine (atoms shown in red in Figure 10). All dangling bonds have been properly saturated with hydrogen atoms.

Next, in order to consider the influence of the surrounding atoms and residues, we have calculated the excited states on a model (referred by us as *wire*) formed by the crystallographic chromophore and the residues Ser142 and Asp157 (putative proton donors or acceptors) and two or three (depending on the case) crystallographic water molecules prone to H-bond them. Again, an analogous optimization of the former has been done to assess the steric effects, removing water molecules and some atoms from the chromophore glutamine and glycine and some atoms from the residues Ser142 and Asp157. Finally, we have taken a larger model (which we name *shell*) by adding

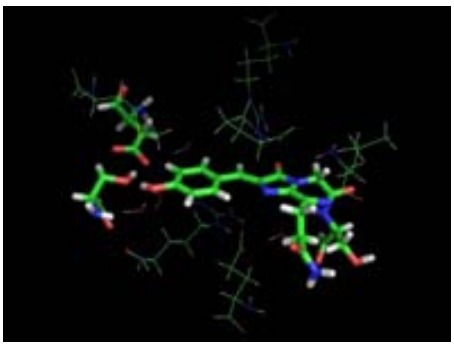
most residues containing at least one atom within 3.5 Å of the chromophore in the crystallographic structure. In all, residues Ser142, Asp157, Arg91, Glu211, Met159 and Arg193 are included for all *shell* models. Then, for 3IR8 and 2WHS structures (*cis* chromophore) residues Pro59, Leu61 and Ser65 have been included, whereas for the 2WHU structure (*trans* chromophore) Glu144 and Phe173 have been included. 3IR8 *shell* contains 4 crystallographic water molecules, 2WHS *shell* contains 3 crystallographic water molecules and 2WHU *shell* contains 2 crystallographic water molecules. *Shell* models for structures 3IR8, 2WHS and 2WHU are shown, respectively, in Figure 13a, Figure 13b and Figure 13c.

Our results (see Table 1) show that when the chromophore alone is geometrically optimized (*optimized chromophore*) it exhibits absorption wavelengths noticeably shorter than the ones obtained with the geometry directly taken from the X-ray structure. This is so because the optimization of the ground state stabilizes it more than the other states. On the other hand, by adding more atoms (*wire* and *shell* models), a variety of results is obtained. In the *wire* model, the wavelengths tend to be longer than in the crystallographic structure. However, in the even bigger *shell* model, the wavelengths are either longer or shorter depending on the crystallographic structure. For this reason, we cannot draw conclusions about which is the most suitable model. The *wire* and *shell* models describe fairly well the 590 nm maximum associated to the anionic chromophores, but yield poor results for the protonated ones. Conversely, the *optimized wire* models furnish good results for the protonated chromophore, but bad for the anionic one, so these small models fail to properly describe the whole absorption spectra of mKeima. A similar result has been obtained for the closely related LSS-Kate models,[76] so we conclude that these reduced models, incidentally quite habitual in theoretical studies of photochemistry, are not suited to deal with large systems such as fluorescent proteins.

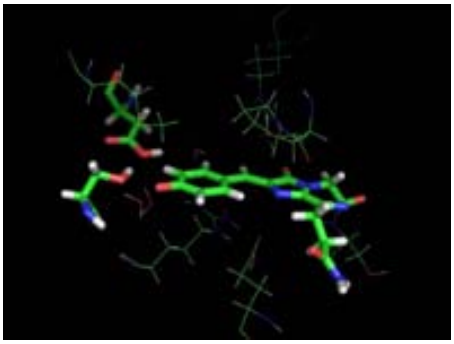
For this reason, the whole protein has been taken into account. To this purpose, the crystallographic structures have been solvated and partly MM-optimized by means of CHARMM22 force field[88–90] just to eliminate bad contacts. From the resulting structures, the electronic excited states have been calculated within the electronic embedding scheme, as explained in Section 2.3.1. The QM region includes the *wire* structure (with some of the Ser142 and Asp157 atoms not included) and the rest of the protein are taken as point charges. It is nonetheless true that the polarization of the protein due to the chromophore

Table 1: Calculated absorption wavelenghts (nm) according to the model and structure used. In parentheses, the oscillator strength is given. Letters N and A refer to Neutral or Anionic chromophore, respectively. The structure 3IR8[118] and 2WHS[117] have a *cis* chromophore, whereas the structure 2WHU[117] has a *trans* chromophore.

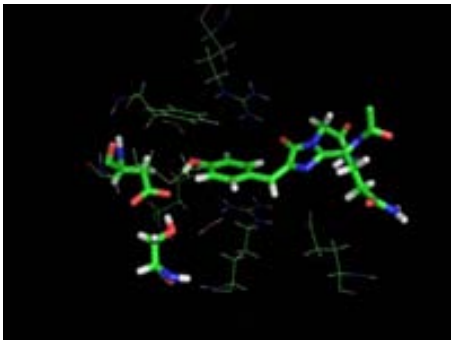
structure	only chromophore	optimized chromophore	wire	optimized wire	shell	whole system	ChemShell optimized whole system
3IR8 N	551 (0.67)	369 (0.77)	561 (0.63)	430 (0.99)	557 (0.72)	422 (0.97)	391 (1.09)
A	575 (0.89)	486 (0.1.07)	593 (0.92)	472 (1.11)	588 (0.75)	492 (1.17)	448 (1.22)
2WHS N	519 (0.59)	369 (0.77)	568 (0.56)	430 (0.99)	489 (0.69)	457 (0.95)	426 (1.13)
A	561 (0.94)	486 (1.07)	571 (0.97)	472 (1.11)	517 (0.82)	504 (1.19)	460 (1.20)
2WHU N	494 (0.66)	363 (0.83)	505 (0.83)	420 (1.30)	490 (0.46)	444 (1.25)	427 (1.28)
A	560 (1.00)	458 (1.43)	550 (1.03)	439 (1.42)	530 (0.67)	491 (1.37)	453 (1.40)
QM atoms	46-50	35-36	80-83	49	203-213	59-62	59-62
MM atoms	-	-	-	-	-	11685-12795	11685-12795



(a)



(b)



(c)

Figure 13: Representation of the so-called *Shell* model for structure 3IR8 (up, shown neutral), 2WHS (middle, shown anionic) and 2WHU (down, shown neutral)

is not taken into account (as the values of the point charges remain constant), but we think that this effect will not be very significant in the spectrum, see for instance the work by Murugan et al.[119] This model will be referred to as *Whole System*. The results obtained from these models (see Table 1) happen to be more reliable and significantly closer to the 440 nm maximum, specifically for the protonated chromophore. Nevertheless, although this 440 nm corresponding to the protonated chromophore is very well described within the *Whole System* model, the anionic chromophore maximum excitation

seems to shift too much towards the blue, when it has been suggested[117] that it could account for the 590 nm absorption maximum.

The next step in the exploration of the effects of the medium to the spectra has consisted of repeating the latter calculations but on a QM/MM optimized structure (*ChemShell-optimized Whole System* from now on). The initial structures in this optimization are the same as the *Whole System* ones, and the QM region is again the *wire* structure (again, with some of the Ser142 and Asp157 atoms not included). The structure has been optimized geometrically using the ChemShell[100] software within the QM/MM scheme, as explained in Section 2.3.2. The optimization's active region includes all residues at 20 Å from the chromophore. From these QM/MM optimized structures, we have used the same procedure as with *Whole System* model to calculate 15 excited electronic states.

A further improvement of the results could be expected after these structures have been optimized. Nonetheless, the wavelengths of the resulting excited states describe almost systematically worse the experimental absorption spectra (Table 1), particularly for the anionic *cis* chromophore models, in which the minimum structures yield partially disrupted proton wires (see Table 7 and Table 8), in Section 3.4). It is reasonable to think that the range of wavelengths experimentally observed could be as well a result of the fluctuations in the positions of the chromophore. Not necessarily does a minimum provide more realistic wavelengths if the minimum of potential energy finds itself in a region of configurational space seldom visited by the system in the conditions of temperature of the experiment. A minimization would produce, if anything, a structure close to what a low temperature sample would render, likely a solid or crystalline sample. For this reason, classical dynamics simulations have been run for all the structures. That way, not only is it possible to check the stability of the supplied crystallographic structures, but also to acquire a spread of structures from which the same *Whole System* calculations can be done and somehow reproduce the absorption spectra. In Section 3.4 these classical dynamics simulations are analyzed, and, following it, in the Section 3.5 we present a simple, yet reasonable way to simulate the absorption spectra taking into account the full dynamics of the protein.

Even though further calculations will be needed, we still can reach some conclusions from this sub-section. All in all, from data in Table 1 we can perform a very interesting kind of analysis that is in essence only possible from the theoretical point

of view: it shows that there are two effects that, originating in the environment of a given chromophore, contribute to the final absorption wavelength of a particular structure. First, the environment could play a mere geometrical effect by constraining the chromophore's structure to be different than in vacuo. This can be seen when comparing any two columns referring to the same kind of environment but differing solely on whether optimization has been done, like for instance columns 2 & 3 (where environment is not present), columns 4 & 5 (the *wire* model), and so on. This usually blue-shifts the absorption, because the optimization is done in the ground state and, usually, makes the excitation energy larger. Second, the effect can be of polarization type, where the environment can affect the charge distribution of the chromophore and, in this way, alter the energy difference between ground and photoactive state. This is seen comparing columns where the chromophore has the same geometry but differs by the presence of the environment like columns 2, 4, 6 and 7. In actuality both effects can, and do, act together.

3.4 PARAMETERIZATION AND CLASSICAL DYNAMICS

Once we have got to this point, one might pose the question of why we decided to run a classical dynamics simulation instead of a combined QM/MM simulation. Being the case that we did not have all the parameters of the chromophore and we had to parameterize them, a QM/MM approach would have been able to circumvect this problem, because the chromophore (and nearest residues) would have been treated quantum-mechanically. Besides, a quantum-mechanical treatment is in general more reliable than a force field one. However, we already started some QM/MM dynamics simulations on the different models and we saw that the time that would have to be spent was unaffordable. The CHARMM35 software[91, 92], which is parallellized for MM simulations, does not allow parallellization when there is a QM region. For instance, equilibrating the system for 100 ps could take up to two weeks: simulating a few nanoseconds would have definetely taken too long. With the ChemShell software[100, 120], something similar happened: when using a semiempirical Hamiltonian with MNDO[110], only one processor could be used at one time. Some programs like Gaussian or Turbomole could be used with the ChemShell interface, and they could be parallellized, but they scaled very badly (with 8 processors, the calculations would be twice as fast).

We were only too aware of the limitations of the MM dynamics simulations: no chemical bonds can be broken, so in a certain sense no "chemistry" can be simulated. But for our purpose, we did not intend to simulate a chemical reaction (which, probably, in the ground state would not have happened anyway), but to check the dynamical stability of the protein structure –the involved hydrogen-bond lengths in particular– and to obtain a statistically relevant ensemble of configurations, which required that the snapshots had to be the less correlated the better, so quite a long dynamics (several nanoseconds) would be required.

3.4.1 *Parameterization*

To run the classical dynamics simulations (for the three X-ray solved structures, 2HWS, 2WHU and 3IR8, each one with protonated and anionic chromophore, that is to say, six simulations) some previous steps have had to be done as the PDB structure is obtained at very low temperatures and may be significantly different at physiological temperature. Moreover, crystallographic structures do not include hydrogen atoms, so they have been placed by means of the PDB2PQR server.[121–123] The protonation of the chromophore, if needed, is quite straightforward. The only tricky decision has been excluding a hydrogen atom from the peptide bond between the chromophore and leucine 61, thus assuming that there is a double bond between the nitrogen (named N in Figure 14) and what would be the alpha carbon (CA1 in Figure 14) of the glutamine (both part of the chromophore). But in the PDB structure these atoms are coplanar with the two other atoms linked to the carbon, so a double bond can be expected, and some published works[64, 124] support this hypothesis. For all the following processes the CHARMM22[89] force field and the CHARMM-35b1[91, 92] software have been used.

Another delicate issue to deal with is that the chromophore is not a standard residue, so new parameters are needed to run MM dynamics simulations. Since mKeima chromophore is very similar to the GFP one, most of the necessary parameters have been taken from the parameterization done for GFP by Thiel and co-workers some years ago.[93] Compared to GFP, the mKeima chromophore has a glutamine instead of a serine, so for these very atoms, the CHARMM22 force field parameters for a glutamine have been taken. However, there are two atoms that differ from any other amino acid: the abovementioned nitrogen and carbon atoms that form a double bond

(Figure 10 and Figure 14). All bonds, angles, dihedrals and improper angles, partial charges and non-bonding parameters involving these two atoms have had to be found. This parameterization can be useful for other DsRed-like chromophores, like the LSSmKate2 chromophore (see Section 3.6).

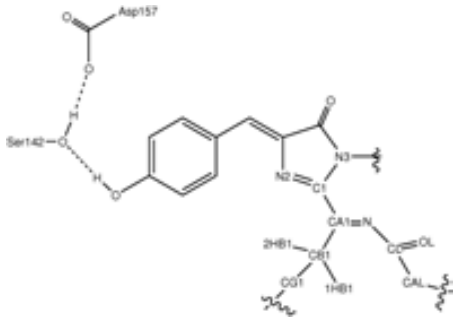


Figure 14: Structure of the *cis* chromophore. The atoms CL, OL and CAL belong to the residue Leu61

For the sake of consistency, we have tried to follow Thiel's method[93] as thoroughly as possible, although we have calculated each parameter separately from the rest (except for the nonbonding and partial charges). Besides, Thiel and co-workers have found the GFP chromophore parameters starting from already existing structures in amino acids, but no amino acid in the CHARMM topology file has this kind of double bond. Given this lack of structures to compare with, we have taken the whole mKeima *cis* chromophore for our calculations. Nevertheless, we have always followed the same level of calculation they use, so that our set of obtained parameters is consistent with theirs and with the CHARMM22 force field. For bonds, angles and improper angles, energy profiles have been scanned around plausible minima with B3LYP/6-31G(d). This is the level of calculation at which Thiel and co-workers have adjusted the parameters for the GFP chromophore. In this way a consistent set of parameters can be obtained. All these potentials are treated harmonically by CHARMM force field, so the force constants and equilibrium angles and distances have been found by adjusting the profile to a parabola. Dihedral angles have not been calculated as the steric effects do not allow for the exploration of the profile and anyway the values for dihedral parameters tend to be much lower than the other sort of parameters, so we have directly taken the parameters calculated by Thiel or, otherwise, the already existing ones in CHARMM22 force field.

Regarding the nonbonding Van der Waals interaction parameters and partial charges, they have been obtained with the

same algorithm. Two energy profiles for a water molecule H-bonded to one atom (either the involved carbon or nitrogen atom) have been calculated (sometimes removing some neighboring atom to avoid unwanted contacts) with an ab initio Restricted Hartree Fock (RHF)/6-31G(d) calculation (again, following the way Thiel and co-workers have proceeded). Then the CHARMM22 energy has been calculated with suitable initial parameters for partial charges and Van der Waals parameters. The squared differences between the electronic RHF and CHARMM22 energies for each point are summed, also over the two profiles so that both profiles have been consistently adjusted at a time. This sum is a function of the parameters to be found, so through minimization it is possible to obtain the desired values of them. Nonetheless, CHARMM topology files force the total charge of a group to be an integer. To avoid program failure, slight modifications upon some atom charges have been done. Parameters regarding the bonds (force constants and equilibrium distances) are shown in Table 2. As for the angles, shown in Table 3, it must be borne in mind that the calculations have been performed on the whole chromophore structure. It can be suspected that the presence of nearby atoms causing steric effects might increase the required energy for bending, resulting in fairly higher force constants for the angle parameters. In order to make sure that this is not the case, we have tried to calculate the energy profiles of the angles C1-CA1-CB1, N2-C1-CA1 and N3-C1-CA1 (see Figure 14) getting closer and closer to the minimum, in order to take into account only the small displacements in which the steric effects are much less noticeable. From these profiles we have obtained similar results (K_Θ = 121.178 kcal mol^{-1} rad^{-2} and Θe=115.104 degrees for C1-CA1-CB1; K_Θ = 326.029 kcal mol^{-1} rad^{-2} and Θe=125.117 degrees for N2-C1-CA1; and K_Θ = 455.677 kcal mol^{-1} rad^{-2} and Θe=122.478 degrees for N3-C1-CA1), so the obtained force constants in Table 3, slightly higher than common force constants for similar angles, might be due to the existence of a double bond between CA1 and N that prevents these from bending. In order to have the same criterion for each angle, we consider these latter fittings as a verification, but for our purposes of simulating a classical dynamics, we have taken into account the parameters shown in Table 3. Another detail to be mentioned is that for the angle CA1-N-CL we have first set the Leu61 Carbon (name CL in Figure 14) coplanar with C1, CA1 and N (as such coplanarity is observed at all X-ray structures). After that, the same procedure has been followed to obtain the parameters.

Table 2: Force constants and equilibrium distances for the referred bond parameters obtained after fitting. The bond potential in the CHARMM-22 force field is calculated as $V(r) = K(r - R_e)^2$. The atoms involved in the parameterization are depicted in Figure 14.

Bond	K (kcal mol^{-1} Å^{-2})	R_e(Å)
CA1-C1	376.10	1.42152
CA1-CB1	250.54	1.56093
CA1-N	539.04	1.35381
N-CL	414.93	1.39071

Table 3: Force constants and equilibrium angles for the referred angle parameters obtained after fitting. The angle potential in the CHARMM-22 force field is calculated as $V(\Theta) = K_\Theta(\Theta - \Theta_e)^2$. The atoms involved in the parameterization are depicted in Figure 14.

Angle	K_Θ (kcal mol^{-1} rad $^{-2}$)	Θ_e(degrees)
C1-CA1-CB1	126.02	114.822
C1-CA1-N	124.88	120.32
CA1-CB1-1HB1	73.56	106.87
CA1-CB1-2HB1	68.79	112.114
CA1-CB1-CG1	68.79	111.24
N2-C1-CA1	304.06	121.33
N3-C1-CA1	456.96	122.67
N-CL-CAL	126.67	117.69
N-CL-OL	124.29	122.011
CA1-N-CL	47.0967	146.309
CB1-CA1-N	112.31	122.005

Table 4: Force constants and equilibrium angles for the referred improper angle parameters obtained after fitting. The improper angle potential in the CHARMM-22 force field is calculated as $V(\Phi) = K_\Phi(\Phi - \Phi_0)^2$. The atoms involved in the parameterization are depicted in Figure 14.

Improper	K_Φ (kcal mol^{-1} rad $^{-2}$)	Φ_0(degrees)
C1-CA1-N2-N3	60.37	-2.97
CA1-N-CB1-C1	103.33	-2.41
CL-CAL-N-OL	56.28	0.48

Table 5: Electrostatic atomic parameters. The electrostatic potential in the CHARMM22 force field is calculated as the interaction between two atoms i and j by using the formula $V = \varepsilon[(R_{min,ij}/r_{ij})^{12} - 2(R_{min,ij}/r_{ij})^{6}]$, where ε_{ij} is the geometric mean of ε_i and ε_j, and $R_{min,ij}$ is the arithmetic mean of $R_{min,i}$ and $R_{min,j}$. The atoms involved in the parameterization are depicted in Figure 14.

Atom type	Partial charge (au)	ε (kcal mol^{-1})	$R_{min}/2$(Å)
N	-0.8651	-0.09649	1.9352
CA1	0.1339	-0.00043	2.0835

The improper parameters are shown in Table 4 and the electrostatic parameters for atoms N and CA1 are shown in Table 5. About these, it is important to recall that they constitute a "group" in the topology and parameters of the force field, which means that the total charge has to be zero, otherwise the CHARMM software does not allow to proceed with the calculations. That is why we have slightly modified their charges, but in order not to make that change so abrupt, we have included these two atoms in a bigger group to which multiple minor modifications can be done (that way, the N charge becomes -0.70 au, and the CA1 charge becomes 0.30 au, which at least can describe the existing dipole).

Note that, for the anionic chromophore, as the main changes are expected for the tyrosine moiety, we have not deemed it necessary to fit the parameters again.

3.4.2 *Classical Dynamics Simulations*

Once we have the missing parameters, we can already start the classical dynamics simulations (more details are given in Section 2.2). To begin with, a CHARMM automatic algorithm has been used to set different solvation water molecules all over a sphere centered at the origin (which lies very close to the chromophore) within a 30 Å radius. These water molecules first, and later the whole protein, have been roughly optimized (500 steps of *Adopted Basis Newton-Raphson method*) so as to avoid bad contacts. The structures resulting from these optimizations have been used for the *Whole System* excited states electronic calculations, as explained in Section 3.3.

Prior to running the MD simulation, the minimized structures have been heated and equilibrated. The heating has been done as follows: the minimized structure is heated up to 100 K. Random velocities are originally assigned following a Gaussian

distribution. The integration step is 1 fs. Every 5 fs, the system rises 5 K up to 100 K by rescaling of velocities. After 50 ps, the coordinates are saved and, starting from those, again random velocities are assigned and the same procedure is repeated for 50 ps to heat the system up to 200 K. Same thing is done till the system reaches the 300K. Afterwards, the system has been equilibrated for 150 ps, following the described procedure, but this time the velocities are not updated and the temperature remains constant. Once the system is fully equilibrated, a 20 ns dynamics simulation is run. This procedure has been carried out 6 times: each model (2WHS, 2WHU, 3IR8) has been simulated with the chromophore neutral and ionized. This corresponds to a total simulation for the production phases of 120 ns. The dynamics simulation has been treated with stochastic boundary conditions. Being the chromophore centered at the origin (more or less coincides with the center of the protein), the Newtonian zone includes all the residues containing at least one atom at 20 Å or less from the origin. The Langevin zone contains all residues not belonging to the Newtonian zone that have at least one atom at 24 Å or less from the origin. The rest of the system is part of the reservoir zone. This separation is justified by the fact that no great conformational changes are expected for the mKeima protein, as in terms of structure it is very similar to GFP, which has a fairly rigid structure.[68, 70] The reduction of degrees of freedom expected to be not so relevant has made calculations much more feasible. Besides, this method allows for a temperature control, so we have ensured that the dynamics simulations' production has been carried out at 300 K. More details are given in Section 2.2.

A first noticeable result of the MD simulations is that they prove that all the considered structures are fairly stable over time. In this sense, it is of uttermost relevance to pay attention to the distances between the residues that can host the proton transfer wire as no covalent bonds exist to bind them together. For protonated *cis* configurations, the proton is expected to be transferred from the chromophore to the Ser142 and from the Ser142 to the Asp157 (see Figure 10), whereas for protonated trans configurations the proton would be transferred directly from the chromophore to the Asp157, being in this case the Ser142, H-bonded to the Asp157, a mere spectator of the proton transfer as depicted in Figure 10.

In Table 6, Table 7 and Table 8 distances from the oxygen atoms implied in the proton wire are shown for the crystallographic structures and their corresponding geometrical optimizations. We have considered the structures obtained by Blan-

Table 6: Distances (in Å) between the donor and acceptor oxygen atoms of the cited residues for 3IR8 structure[118] (*cis* chromophore). The uncertainty range in the last column corresponds to the standard deviation.

	Crystal structure	Chromophore protonation state	Optimized wire	ChemShell optimized Whole System	Dynamics average
Cro-Ser	2.66	neutral	2.49	2.61	2.81 ± 0.14
		anionic	2.56	3.85	2.86 ± 0.15
Ser-Asp	2.67-2.74	neutral	2.49	2.79	2.91 ± 0.31
		anionic	2.54	4.35	2.91 ± 0.27
Cro-Arg91	2.83-2.86	neutral		2.72	2.86 ± 0.15
		anionic		2.70	2.80 ± 0.12
Cro-Glu211	2.78-2.84	neutral		3.14	3.24 ± 0.23
		anionic		3.25	4.19 ± 0.36

Table 7: Distances (in Å) between the donor and acceptor oxygen atoms of the cited residues for 2WHS structure[117] (*cis* chromophore). The uncertainty range in the last column corresponds to the standard deviation.

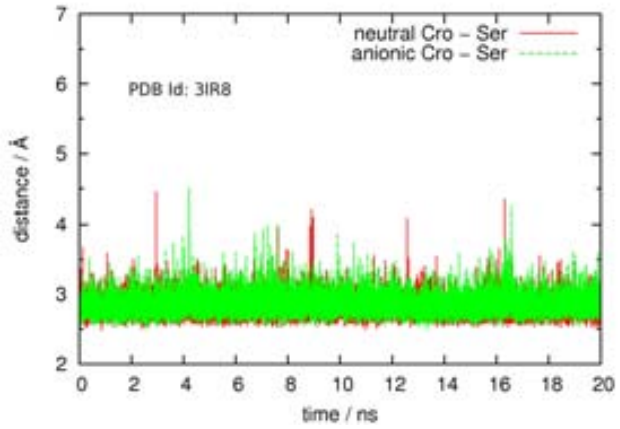
	Crystal structure	Chromophore protonation state	Optimized wire	ChemShell optimized Whole System	Dynamics average
Cro-Ser	2.70-2.77	neutral	2.49	2.57	2.89 ± 0.19
		anionic	2.56	2.65	2.95 ± 0.34
Ser-Asp	2.93-3.22	neutral	2.49	2.80	3.14 ± 0.43
		anionic	2.54	3.41	4.08 ± 0.62
Cro-Arg91	2.80-2.94	neutral		2.77	2.84 ± 0.13
		anionic		2.74	2.82 ± 0.13
Cro-Glu211	2.90-3.0	neutral		3.43	3.54 ± 0.28
		anionic	2.56	2.65	2.95 ± 0.34
		anionic		3.39	4.12 ± 0.41

Table 8: Distances (in Å) between the donor and acceptor oxygen atoms of the cited residues for 2WHU structure[117] (*trans* chromophore). The uncertainty range in the last column corresponds to the standard deviation.

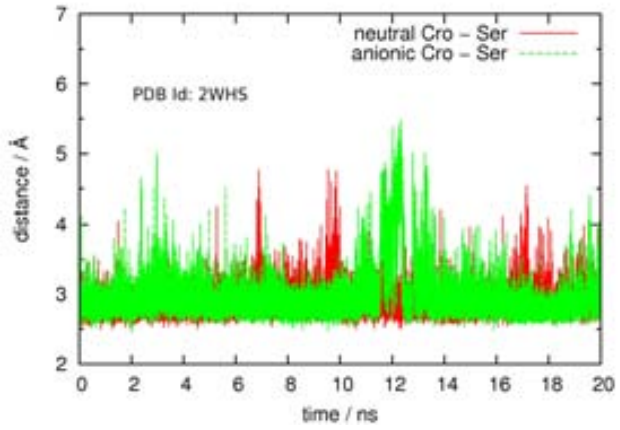
	Crystal structure	Chromophore protonation state	Optimized wire	ChemShell optimized Whole System	Dynamics average
Cro-Asp	2.92-3.16	neutral	2.45	2.58	2.70 ± 0.17
		anionic	2.44	2.52	2.73 ± 0.10
Asp-Ser	2.56-3.61	neutral	2.72	2.84	2.94 ± 0.41
		anionic	2.83	2.93	4.27 ± 0.41
Cro-Arg91	2.78-3.35	neutral		2.78	2.94 ± 0.18
		anionic		2.76	2.91 ± 0.22
Cro-Glu211	2.69-3.20	neutral		3.29	3.38 ± 0.26
		anionic		3.53	3.99 ± 0.38

choin and co-workers[117] and Remington's.[118] Although the protein is monomeric, the crystallographic structures contained two[118] or four[117] monomers, so the maximum and minimum distance values are shown for each case. Also, it must be considered that the Gaussian *wire* optimization has been the same for both *cis* chromophore models, so the distances are the same. Some of the distances displayed in Table 6, Table 7 and Table 8 that have been obtained with the ChemShell optimization are far greater than the corresponding ones from the X-ray structure, from the *wire* optimization or from the MD simulation. This is so because along the optimization a water molecule has ended up H-bonding the chromophore, thus pushing away the Ser142 and somehow disrupting the proton wire. The last column contains the average of the values all over the dynamics simulation, with the corresponding standard deviation.

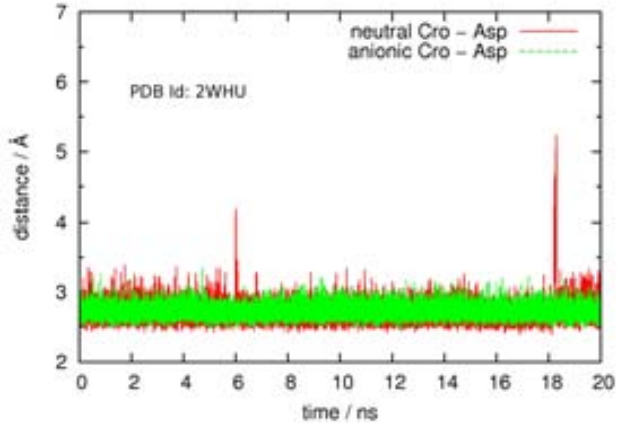
The results of the 20 ns-long MD simulations are presented in Figure 15 and Figure 16. It is clearly seen that the first link of the proton wire is well formed and stable over time for all the structures as the O-O distances betweeen the chromophore and the corresponding acceptor residue oscillate most of the time between 2.5 and 3 Å (Figure 15). The same can be said for the distances between the oxygen atoms of the serine and the aspartate residues (Figure 16) as far as the neutral chromophore is considered. However, for the structures obtained by Blanchoin and co-workers, this distance noticeably increases when the chromophores are anionic.



(a)



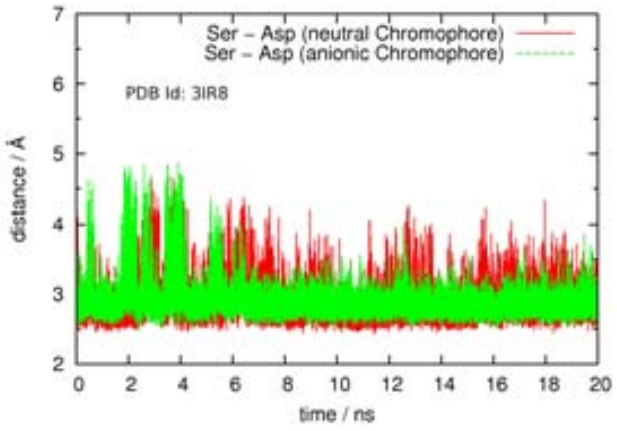
(b)



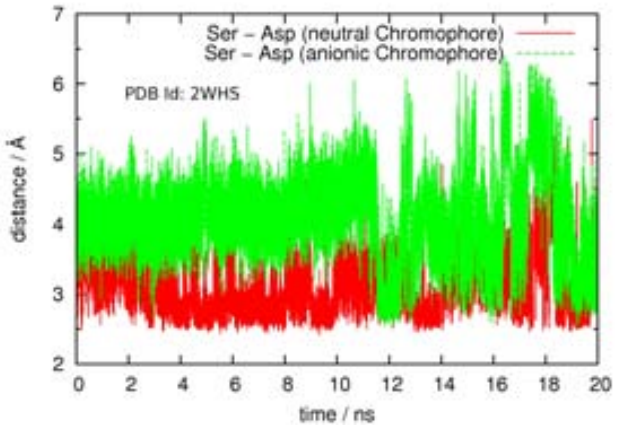
(c)

Figure 15: Distances between the oxygen atoms of chromophore and the corresponding residue throughout the molecular dynamics simulations.

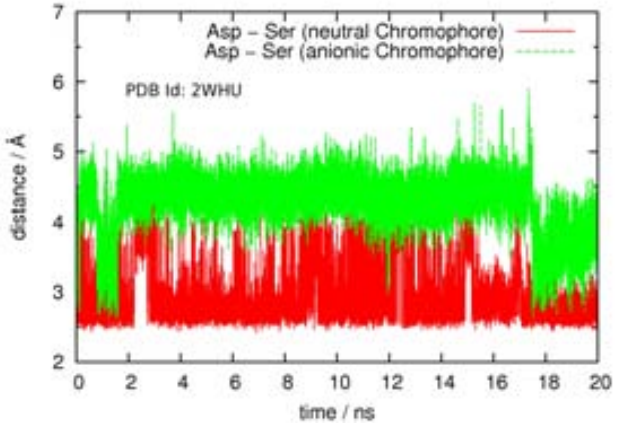
The last column of Table 6, Table 7 and Table 8 gives the average of these distances for each case so that we can compare



(a)



(b)



(c)

Figure 16: Distances between the oxygen atoms of the serine and the aspartic residues throughout the molecular dynamics simulations.

them with the X-ray data and the values previously obtained through energy minimizations. Qualitatively the averaged val-

ues along the dynamics are in good accordance with X-ray data though the former values are slightly larger. This is to be expected in part given the fact that the MD were devised to simulate the protein in physiological conditions where the averaged distances between different residues are expected to be larger than the values measured when the structure is frozen. At any rate it is worth noting that the X-ray values usually fall within the statistical range of variation of these distances along the whole simulation, measured by the corresponding standard deviation, a value which is also given in Table 6, Table 7 and Table 8.

As for the comparison with the previous optimizations, it is clearly seen in Table 6, Table 7 and Table 8 that optimizing just the fragment of the protein implied in the proton wire provides O-O distances clearly shorter than X-ray values (and so also much shorter than the MD-averaged values). This result clearly emphasizes the importance of taking into consideration the protein environment when analyzing the structure of the active center (in this case the proton wire). It is tempting to verify whether more accurate results should be obtained when optimizing the whole system. Results in Table 6, Table 7 and Table 8 show that, as it is the case when comparing the values of the absorption wavelengths, they lead to quite disparate results. This confirms that optimizing a large molecular system might provide a geometry that is not necessarily representative of the actual structure of the molecule as its dynamics can span a large region of the potential energy surface in a relatively short period of time. This latter assertion is not secondary: it is an important conclusion we get from our research, and deserves a deep reflection. For some purposes, it is necessary to properly locate the minima: for example, for some chemical reactions, a good geometrical description of the reactants and products will furnish reliable values for exothermicities, will allow for determining the spontaneity or will be a necessary starting point to find the transition state. But it is misleading to find energy minima expecting to obtain *the correct* structures from which the most realistic information of the system can be pried out. Firstly, the system has lots of degrees of freedom. We may find a local minimum that is not the one with the least energy. Not only this: the absolute energy minimum might not be either the most representative one for our reaction. But secondly, even if we had a representantive optimized structure, the system is not still: at a given temperature possesses some kinetic energy and will detach from the minimum. A starting minimization might be required to reduce an unrealistic excess of potential energy

(for instance, resulting from an inaccurate position of solvent molecules), but *not* to provide a more exact description of the geometry of the system.

Following with the analysis of the MD simulations, qualitative analysis of Figure 11 shows that the position of the chromophore and its planarity could also be stabilized because of the charged amino acids Arg91 and Glu211, that would have a nonbonding interaction with, respectively, one oxygen and one nitrogen atoms situated in one of the coplanar rings of the chromophore. It can be seen in Table 6, Table 7 and Table 8 that, when the chromophore is neutral, both the optimization and the MD lead to distances that account for a nonbonding stabilizing interaction. When the chromophore is anionic, the Cro-Arg91 distance remains small. As for the Cro-Glu211 distance, it becomes larger than found in the crystal structure after optimizing or during the MD, but not enough to make the nonbonding stabilizing interaction irrelevant. In all cases, we have considered the shortest interaction distance. Results in Table 6, Table 7 and Table 8 suggest that these two residues, especially Arg91, play an important role in the dynamics of the chromophore and, consequently, in modulating its photochemical properties.

3.5 THEORETICAL SIMULATION OF THE ABSORPTION SPECTRA

Each of the 20 ns-long simulations has been used to extract a snapshot every 0.5 ns (40 snapshots per simulation, 240 snapshots total). The snapshots derived from each molecular dynamics run form an ensemble of structures that is expected to be more representative of the state of the protein in solution in physiological conditions than the optimized minimum would be.

That way, we have repeated the *Whole System* calculations on the considered snapshots of our dynamics simulations. Then, all the different excited states of a single dynamics simulation with oscillator strength (f) greater than 0.1 are sorted by their frequencies and a wavelength histogram is done on the accumulated value of f in the bin, being 10 nm the size of each bin, and normalizing the integral (Figure 17). These histograms represent an average oscillator strength for each 10 nm wavelength window, weighted by the number of snapshots which have an excited state accessible ($f \geqslant 0.1$) through excitation at the wavelength range from the ground state. The sum of the counted oscillator strengths around a given wavelength value gives an

approximate idea of how likely it is to detect absorption around it regardless of the nature of the excited state involved. It is not a thorough reproduction of the spectra (since, for instance, polarization of environment is not accounted for), but it is a feasible and plausible way to simulate it. A similar procedure has already been used by other authors such as Thiel and co-workers,[125], Imhof[126], and Martínez and co-workers[127], to characterize the absorption spectra of other fluorescent proteins.

Apart from the fact that excited states with higher energies have been found, all protonated chromophore (both *cis* and *trans*) models exhibit an absorption band with a maximum between 400 nm and 450 nm, which matches pretty well with the 440 nm absorption maximum experimentally detected. As for the anionic chromophore models, both *cis* and *trans* show a band between 500 nm and 600 nm. This band is slightly blue-shifted with respect to the detected absorption maximum at 590 nm, yet it is not necessarily a bad description for it must be considered that the calculations involve energies, which do not depend linearly of wavelengths, so the same energy difference errors, for lower energies, yield greater errors in wavelengths.

In all the calculated absorption spectra shown in Figure 17 there appears a higher energy (lower wavelength) band between 300 nm and 400 nm. This band comes from transitions to excited states of higher energy that have a non-negligible oscillator strength ($f \geqslant 0.1$, as explained above). For the neutral chromophore cases (spectra on the left of Figure 17) this band is clearly less intense than the "main" absorption band centered at wavelengths slightly above 400 nm. On the contrary, for the anionic chromophores (spectra on the right of Figure 17), this higher energy band competes on equal grounds with the experimentally measured band at wavelengths above 500 nm and, in the case of the *trans* chromophore (Figure 17f), is clearly the most intense band. This band cannot be compared with the experiments as it lies outside the published spectral range considered in previous works.[114, 117]

We have deemed it interesting to analyse separately the effect on the spectra caused by thermal agitation (given the multiplicity of molecular configurations) and the electrostatic effect of the environment. We have tackled this by taking the different geometries corresponding to the *only chromophore* and *wire* models from the ensemble of configurations with which the spectra have been simulated. With these geometries, the same TDDFT calculations have been done to each snapshot and similar simulated spectra have been done. For the sake of clarity, the results

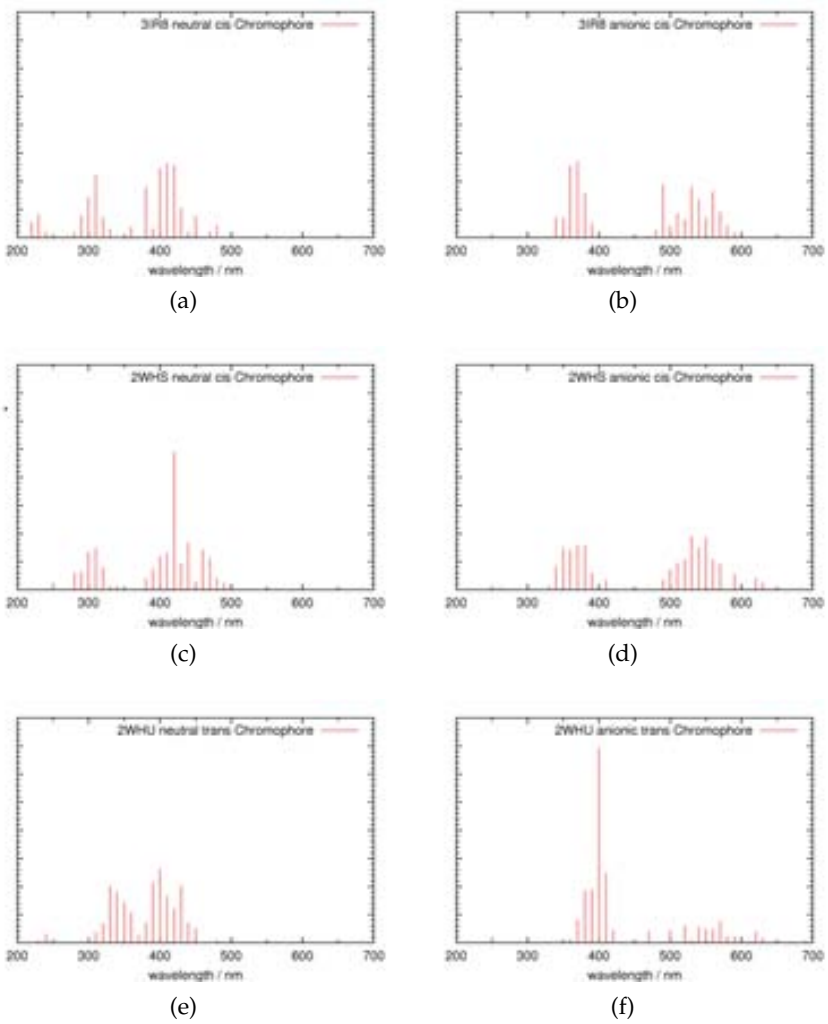


Figure 17: Representation of the absorption spectra calculated throughout the classical dynamics simulations (see text for the details). The ordinate axis indicates the scaled cumulative oscillator strength. The graphics correspond to the models 3IR8[118] (*a* and *b*), 2WHS[117] (*c* and *d*), and 2WHU[117] (*e* and *f*), the chromophore being neutral (left) or anionic (right).

are presented in Figure 18 as convoluted spectra with a convoluting Gaussian function of σ= 10 nm. A similar idea to assess the electrostatic effect of the environment in a protein has already been reported by Lopez and co-workers.[128]

Comparing Table 1 with Figure 18 we can easily notice the importance of taking an ensemble of structures instead of only one. For example, for the structure 3IR8, the neutral *only chromophore* presents an absorption wavelength at 551 nm (Table 1). However, Figure 18a shows that the blue line (which corre-

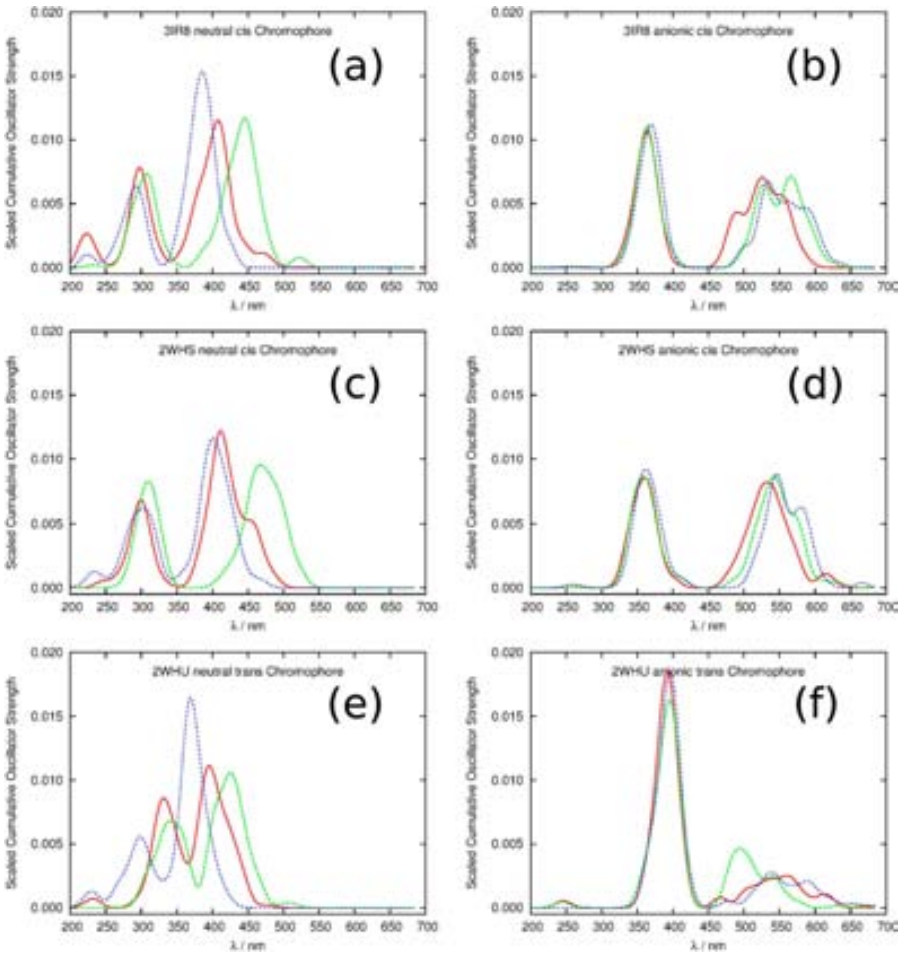


Figure 18: Representation of the absorption spectra considering the *Whole System* (red), the *wire* (green) and the *only chromophore* (blue) models. These correspond to the PDB structures 3IR8 with a protonated (*a*) and anionic (*b*) chromophore, 2WHS with a protonated (*c*) and anionic (*d*) chromophore, and 2WHU with a protonated (*e*) and anionic (*f*) chromophore. In all cases, a convoluting Gaussian function of σ= 10 nm has been used to smooth the statistical noise. Note that the red curves in spectra a-f above correspond to the convolution of data shown in Figure 17

sponds to the isolated chromophore) presents no signal at this wavelength. The optimized chromophore absorbs at 369 nm (Table 1) and Figure 18a shows some signal at this wavelength, but it is not the maximum. A similar analysis for the *wire* model (green line in Figure 18a) could be done with similar conclusions. For the anionic *only chromophore*, it happens the other way around. The crystallographic structure absorbs at 575 nm (Table 1), which in Figure 18b corresponds to a populated region (although not the maximum), but the *optimized chromophore* ab-

sorbs at 486 nm (Table 1), which corresponds to a completely flat area of the blue line in Figure 18b. Again a completely analogous analysis could be done for the *wire* model.

It can be seen that introducing electrostatic effects affects the position and width of the band. However, the separation between thermal and electrostatic effects is not thorough because all the structures have been generated with a molecular dynamics performed using a force field that includes all the electrostatic interactions from the environment with the chromophore, instead of making a simulation where these interactions had been taken out: these simulations would likely tear these models apart.

Starting from different crystallographic structures of the protein, we have asserted that all protonated chromophore (both *cis* and *trans*) models exhibit an absorption band with a maximum between 400 nm and 450 nm, which matches pretty well with the 440 nm absorption maximum experimentally detected. As for the anionic chromophore models, both *cis* and *trans* show a band between 500 nm and 600 nm.

3.6 THE CASE OF LSSMKATE2

The methodology that we have developed and implemented for the study of the mKeima absorption spectra has also been tested, in collaboration with our colleague Carlos Randino, for a similar RFP: the LSSmKate2, developed in 2010 by Piatkevich *et al.*[64, 114]. As well as mKeima, LSSmKate2 has a DsRed-like chromophore and shows a LSS. This protein allows the possibility of obtaining a multicolor image useful to select different parts of biological systems. It presents an absorption maximum at 460 nm and an emission maximum at 605 nm.

As happens with mKeima and with many other photochemical reactions in fluorescent proteins, a proton transfer from the chromophore to some acceptor atom is thought to cause the LSS on the fluorescence. For LSSmKate2 the mechanism involves two species: the protonated and deprotonated chromophore. The protonated chromophore absorbs light and starts a proton transfer reaction. Based on its X-ray structures, a two-link ESPT for LSSmKate2 is proposed: the first proton moves from the hydroxyl group of the chromophore to the hydroxyl group of Ser158 and the second one from Ser158 to the carboxylic group of Asp160 (Figure 19). Once the different proton transfers occur, the fluorescence in the deprotonated chromophore takes place.

To be able to simulate the absorption spectrum of LSSmKate2, a classical MD simulation[129] had to be run with a dual aim:

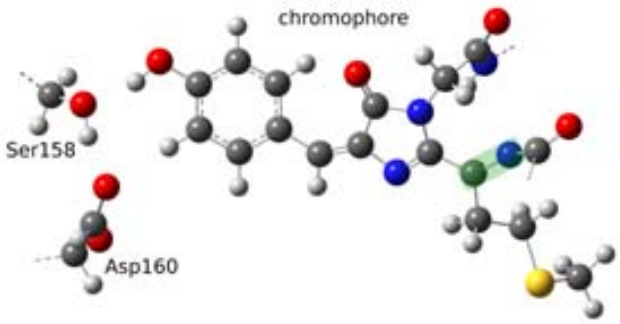


Figure 19: Representation of the chromophore and the acceptor amino acids (Ser158 and Asp160) involved in the ESPT in LSSmKate2. Color code: dark grey, carbon; white, hydrogen; red, oxygen; blue, nitrogen; yellow, sulphur.

on the one hand, to check the dynamical stability of the crystallographic structure[64] supplied at the Protein Data Bank, and on the other hand, to furnish an ensemble of structures from which further excited state electronic calculations can be done, thus providing a tool to introduce thermal effects in the spectrum, as explained in Section 3.6.1.

This MD simulation[129] had yielded interesting results to figure out possible paths available for the excited state proton transfer. To begin with, it can be seen that the protein is stable along the dynamics simulation, that is to say, there are no relevant conformational changes. As for the chromophore and the residues that are proposed to host the proton transfer, namely, Ser158 and Asp160, several hydrogen bonds are observed. By analyzing the distances between the oxygen atoms of Ser158 and Asp160 (Figure 20, *a* left) and the distances between the hydrogen atom of Ser158 and the oxygen atom of Asp160 (Figure 20, *a* right), several features can be noticed.

First of all, leaving apart some pronounced, short-lived and random fluctuations that appear occasionally, the range of the oscillations is small (about 0.5 Å, slightly more for the hydrogen-oxygen distance). Besides, the distance between the oxygen atoms oscillates between 2.5 Å and 3 Å, and the distance between the oxygen and the hydrogen atoms oscillates between 1.5 Å and 2 Å. This constant difference of 1 Å coincides with the average hydrogen chemical bond distance, which implies that the three atoms are almost collinear and the hydrogen bond is well formed and is maintained along the MD simulation. Following the same criteria, we can also indentify in Figure 20

b and *c* when the chromophore is hydrogen bonded either to Ser158 or Asp160. On the contrary, the corresponding hydrogen bond does not exist when the range of oscillation of the oxygen-oxygen distance is about 1 Å, while the oxygen-oxygen distance is beyond 3 Å (Ser158) or 4 Å (Asp160). Then, except for short temporary hydrogen bond disruptions, the chromophore is always hydrogen bonded to either to Ser158 or Asp160, but never to both simultaneously.

Considering all that, two kinds of hydrogen bond networks can be identified, which supposes two qualitatively different situations. Firstly, one in which, as it appears in the crystallographic structure, the chromophore is hydrogen bonded to Ser158 and Ser158 is hydrogen bonded directly to Asp160, thus enabling a putative two-link proton relay. This situation, as can be observed in Figure 20 *a* and *b*, is not the most probable in terms of residence time as it only occurs between 0 and 1.5 ns, and between 9.5 and 11.5 ns approximately. Secondly, another situation can be observed, in which the chromophore is hydrogen bonded to Asp160, which is also hydrogen bonded to Ser158 (Figure 20, *a* and *c*). This is the most likely situation in terms of residence time, as it happens between 1.5 ns to 9.5 ns, and from 11.5 ns to the end of the simulation. Nonetheless, from about 14 ns to 17 ns, about 22-23 ns and about 31-32 ns, all the distances shown in Figure 20 increase. This corresponds to temporary disruptions of the hydrogen bonds, and in these periods the above mentioned residues are hydrogen bonded to other residues (mainly water molecules) but they always return to the latter situation, in which the chromophore is hydrogen bonded to Asp160 despite some spurious peaks. This latter configuration would enable a single link proton relay, in which the proton could be transferred directly from the chromophore to Asp160, analogously to what is suggested to occur for LSSmKate1,[64] where the phenoxy moiety proton in the chromophore is thought to be directly transferred to Glu160 (as directly seen in the X-ray structure).

As has been checked by the analyses of the excited-state electronic structure calculations presented in Section 3.6.1, there are no appreciable differences between those two different configurations in terms of the nature of excited states found for each of them. Furthermore, the photochemistry of both configurations would eventually lead to the same chemical species: an anionic chromophore and an aspartic residue.

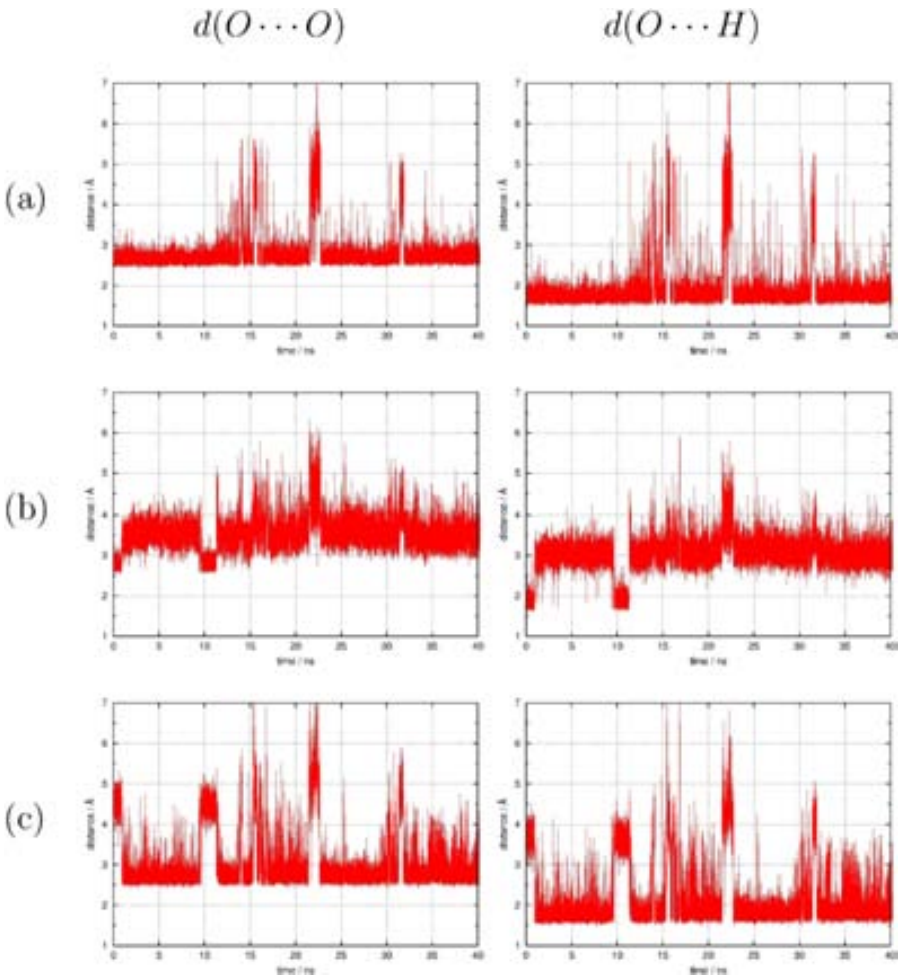


Figure 20: Distances between the residues involving the hydrogen bonds in the putative proton-wire along the MD simulation, taking into account the oxygen-oxygen distances (left) or the oxygen-hydrogen distances (right) for Ser158-Asp160 (a), Cro-Ser158 (b) and Cro-Asp160 (c) hydrogen bonds.

3.6.1 *Theoretical simulation of the LSSmKate2 absorption spectrum*

Excited state electronic structure calculations have been performed on 160 snapshots from the ensemble of structures obtained from the molecular dynamics simulation (one every 0.25 ns). As it has been explained in Section 3.5, the calculated absorption wavelengths constitute a fairly good approximation to the experimental absorption spectrum as it takes into account the thermal effects.

For this system, the QM region includes the chromophore in its entirety with the carbonyl group of the previous amino acid (Phe60) and the amino group of the next amino acid (Ser66). The side chain of Ser158 and a deprotonated Glu160, the two responsible amino acids participating in the ESPT, are also in-

cluded (in total, 53 QM atoms + 4 link atoms). All the QM atoms of the model are represented as spheres in Figure 19.

15 excited states have been calculated for each of the 160 frames. Then, again, all the different excited states with oscillator strength (f) greater than 0.1 have been sorted by their frequencies and a wavelength relative histogram has been done with a bin size of 10 nm. The histogram is shown in Figure 21.

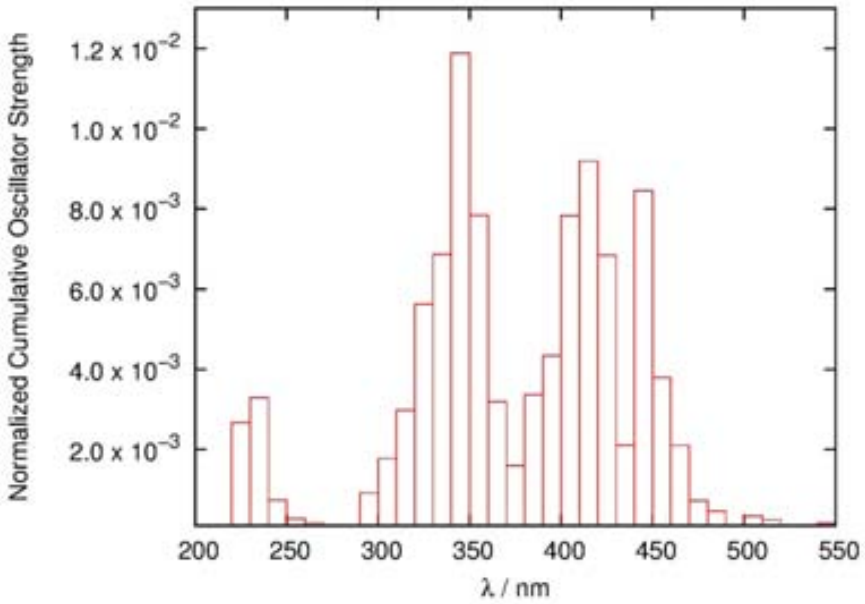


Figure 21: Simulated electronic absorption spectrum of LSSmKate2. The red columns represent the sum of oscillator strength for each set of 10 nm (properly scaled).

The simulated electronic absorption spectrum of LSSmKate2 (Figure 21) built as we have explained takes into account several configurations of the protein. Among them, both proton relay configurations (that is to say, the Cro-Ser158-Asp160 and Cro-Asp160 ones) have been considered for the simulated absorption spectrum. To figure out whether both proton-relay configurations present different absorption frequencies or not, the absorption frequencies corresponding to the 15 excited states calculated for each of the 160 snapshots (including only those transitions with oscillator strength equal to or greater than 0.1) have been represented versus $d_{Cro-Ser} - d_{Cro-Asp}$, where $d_{Cro-Ser}$ is the oxygen-oxygen distance between the chromophore and Ser158 and $d_{Cro-Asp}$ is the oxygen-oxygen distance between the chromophore and Asp160. When $d_{Cro-Ser} - d_{Cro-Asp}$ is positive, the chromophore is closer to Asp160 than to Ser158 (typical of the Cro-Asp160 proton-relay configuration). Conversely, when the value is negative, the chromophore is closer to Ser158

than to Asp160 (which would be typical of the Cro-Ser-Asp proton-relay). The results are shown in Figure 22. Most points are concentrated on the positive abscissa axis, this being a consequence of the already commented predominance of the Cro-Asp proton-relay configuration along the whole dynamics. Notwithstanding this, no obvious correlation can be observed between the abscissa value and the absorption wavelength, so both configurations appear equally spread along the spectrum.

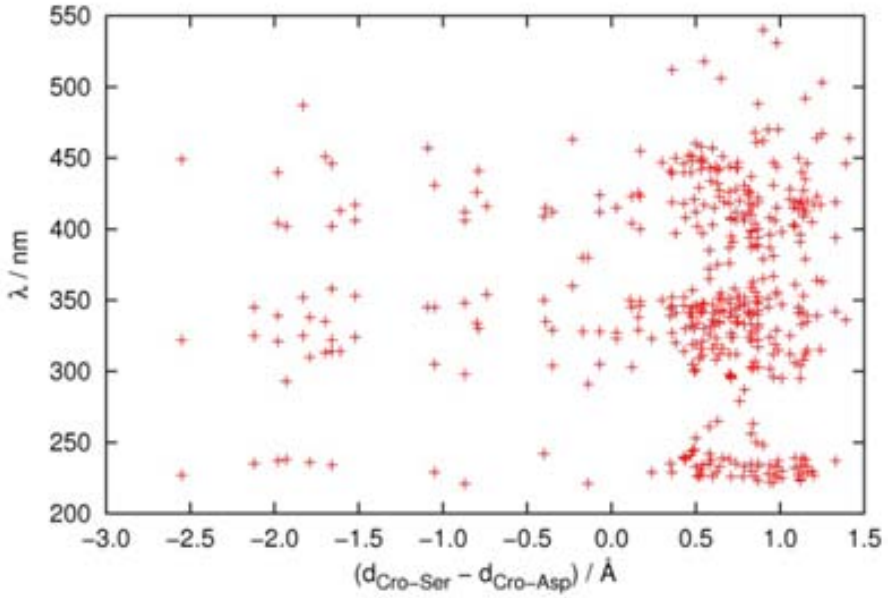


Figure 22: Representation of the wavelengths obtained while calculating the spectrum versus the difference of O-O distances between Cro-Ser158 and Cro-Asp160.

The mentioned spectrum (shown in Figure 21) presents three bands, two of which (the most intense) are in agreement with the available experimental data. The most intense band has a maximum at about 345 nm and the second one shows two peaks at 415 nm and 445 nm. The latter absorption band, the one experimentally analyzed and known to produce red fluorescence at 605 nm, is well reproduced when compared with the experimental absorption maximum (460 nm). Nevertheless, the higher energy band between 300 and 375 nm, is red-shifted by about 65 nm with respect to its corresponding experimental absorption band (graphically estimated at 280 nm).[64] Shifts of similar magnitude have been found in related studies.[119] Both bands present narrower bandwidths than the corresponding experimental values. As for the third band (~ 200-250 nm), it lies out of the range of experimental absorption wavelengths published[64]. From a statistical point of view a possible con-

cern could be the degree of convergence of this spectrum. To assess this point, we have proceeded as follows: we have computed different spectra based on data with increasingly large length of the simulation in steps of 10 ns and computed the spectra using exactly the same procedure, and convoluted with a Gaussian function of σ= 10 nm to attenuate statistical noise. The results are shown in Figure 23, and it can be seen that the overall shape of the spectra (number, intensity, width and position of the bands) is already stable from the beginning. As for the precise positions of the peaks, they undergo some small shifts and could change slightly if more samples were included.

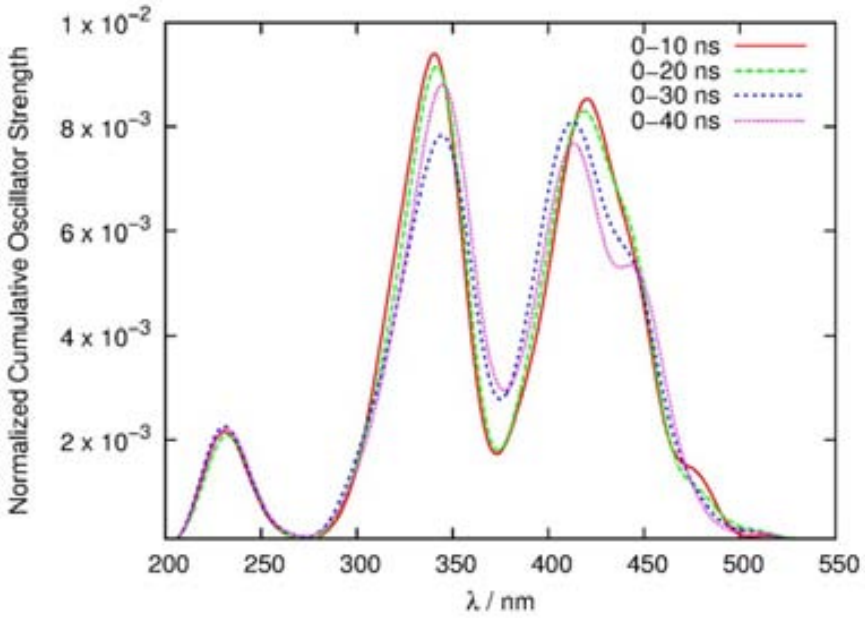


Figure 23: Convergence of spectrum along the simulation. Each of the spectra has been computed on an increasingly longer part of the simulation. A convoluting Gaussian function of σ= 10 nm has been used to smooth the statistical noise.

Despite having followed essentially the same method for mKeima in Section 3.5, one might have noticed some variations in the way that the spectra have been simulated for each protein. The simulations on mKeima showed a very stable regime: with the chromophore H-bonded to either a donor or an acceptor atom that remained the same during all the simulation. Nonetheless, for LSSmKate2 the proton wire presented two qualitatively different situations, namely, a two-link and a one-link proton wire that alternate along the simulation. Although for this particular case the hydrogen bond network did not visibly affect the absorption wavelenghts (as can be seen in Figure 22) it seemed reasonable to run a longer simulation (twice as long, 40 ns in-

stead of 20 ns) and to calculate the spectrum taking more snapshots (one every 0.25 ns instead of one every 0.5 ns, twice as frequent). All this implied more computational resources, but they were affordable since for LSSmKate2 we were dealing with only one molecular model (instead of the 6 models for mKeima). Besides, having come up with a significant change in the hydrogen bond network in the MD simulation prompted us be cautious and to verify the convergence of the spectrum along the simulation, as shown in Figure 23. Finally, for mKeima, the electrostatic effect was evaluated by calculating the excited states (Table 1) and by reproducing the spectra (Figure 17) on different models. We concluded from this that the surrounding atoms (the protein environment) definetely account for the absorption properties, and so do the thermal effects, therefore here we have taken it for granted and we directly simulated the spectrum for LSSmKate2.

Analyzing the description of the electronic states accessed upon photoexcitation in terms of the orbital description, we can observe three molecular orbitals that participate in the $\pi\pi*$ excitations leading to the UV-vis spectrum shown in Figure 21: HOMO (Highest Occupied Molecular Orbital), LUMO (Lowest Unoccupied Molecular Orbital) and LUMO+1. All of them are located over the chromophore, preserving the approximate molecular plane as a node. In Figure 24 we have depicted these three molecular orbitals for Cro-Asp160 (taken at 4 ns of the MD simulation) and Cro-Ser158-Asp160 (taken at 10.5 ns of the MD simulation) hydrogen bonded structures. It is clear that these molecular orbitals are identical for the two structures, so justifying that both are compatible with the experimental electronic absorption spectrum.

The right band (low energy) of the spectrum could then be assigned to a HOMO $\rightarrow$ LUMO transition. As can be seen in Figure 24, the HOMO $\rightarrow$ LUMO excitation represents an electron shift from the phenoxy moiety to the pyrimidine moiety of the chromophore. This represents effectively a "charge transfer" of intramolecular type and justifies the methodological choice made (CAM-B3LYP, see Section 2.1.2). Based on the nature of the molecular orbitals involved in this excited state, the excitation produces depletion of charge of the phenoxy moiety which likely brings about an increase of the acidity of the chromophore. Thus, this charge shift provides support for the ESPT hypothesis as a basis of the mechanism of generation of the fluorescent species. On the other hand, the higher energy band corresponds mostly to a HOMO $\rightarrow$ LUMO+1 excitation. In this case, LUMO+1 is a orbital evenly spread over the full

chromophore and the main difference with HOMO lies in the increased number of nodal planes perpendicular to molecular plane on the phenoxy moiety of the chromophore (from HOMO with 2 to LUMO+1 with 3).

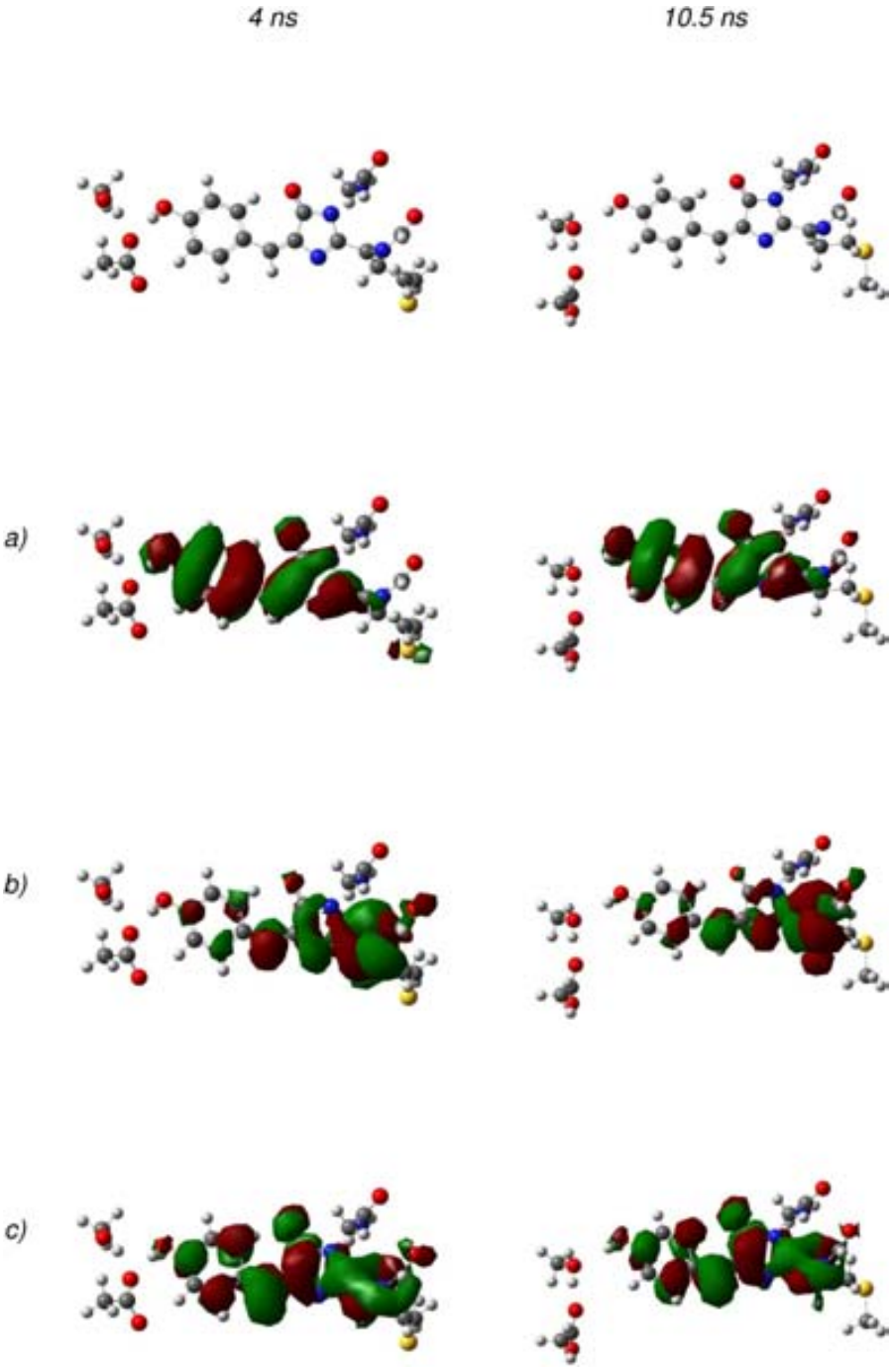


Figure 24: Orbital representations of two selected frames at 4 ns and 10.5 ns of the MD simulation. The HOMO (*a*), LUMO (*b*) and LUMO+1 (*c*) involved in the active $\pi\pi*$ transitions are shown.

Regardless of the fact that two different configurations allow the formation of the anionic chromophore along, respectively a one-link or two-link proton transfer, the molecular orbitals involved in the electronic excitation o the chromophore are not appreciably modified, which justifies that both configurations are compatible with the experimental electronic absorption spectrum. The two bands come from a $\pi\pi*$ orbital excitation. A transition between the HOMO to the LUMO accounts for the absorption band responsible for the red fluorescence The migration of the electronic density towards the pyrimidine ring facilitates the shift of the proton of the chromophore in the excited state, making a more favorable reaction. Conversely, the higher energy band can be assigned to the HOMO $\rightarrow$ LUMO+1 transition, implying a very minor displacement of the electronic density over the chromophore.

3.7 CLOSURE

To summarize this very section about the simulation of the RFPs mKeima and LSSmKate2 absorption spectra, let us bear in mind that ranging from a few atoms to the whole protein, and considering even QM/MM optimizations of the structure, it has been impossible to reproduce simultaneously the transitions corresponding to the neutral and anionic chromophores from individual structures. One plausible explanation of this is that the energy minimum is not necessarily populated in terms of residence time. The protein undergoes dynamical shifts that, at least, belong to different configurations. To avoid these problems, we have done an extensive sampling of the regions that are populated at 300 K, so that thermal effects are introduced by combining classical dynamics simulations and TDDFT electronic calculations. We have seen that the protein environment that surrounds the chromophore (both the identity of the residues and their thermal motion) has a noticeable effect on the absorption spectra (position and width of the bands).

4

LOW-BARRIER HYDROGEN BONDS IN PROTEINS

In this chapter, the different points of view concerning the existence of a low-barrier hydrogen bond in the PYP, will be shown. In Section 4.1, we present the state-of-the-art of the available experimental data and the discussion held about the existence of an LBHB in the PYP before we set up our studies. Next, in Section 4.2 the results of our calculations are shown, for crystal (Section 4.2.1) and solvated (Section 4.2.2) PYP. Finally, in Section 4.3, the whole issue is concluded and summarized.

4.1 THE PHOTOACTIVE YELLOW PROTEIN

The Photoactive Yellow Protein (PYP) is a blue light photoreceptor found in the halophilic photosynthetic bacterium *Ectothiorhodospira halophila*, and serves as a sensor for negative phototaxis to blue light –which means that induces the bacterium to move away from the blue light–.[130, 131] The chromophore responsible for perception of light and phototactic response is a deprotonated *trans*-4-hydroxycinnamic acid (also known as *p*-coumaric acid) covalently linked to a cysteine residue.[132] Its structure consists of two parallel planes of β sheets secondary structure, known as β clam (see Fig 25). The chromophore is located between these sheets. This protein was known to contain two short hydrogen bonds adjacent to the reaction center:[133, 134] one formed between the phenolic oxygen of the chromophore of PYP, *p*-coumaric acid (*p*CA), and the carboxylic oxygen of Glu46 (with a distance O···O of 2.58 Å) and the other between the phenolic oxygen of *p*CA and the phenolic oxygen of Tyr42 (with a distance O···O of 2.51 Å). Taking into account the O···O distance, both hydrogen bonds can be considered as possible LBHBs, although the hydrogen atoms involved in them had never been observed.[133–135] But quite recently, in 2009, Yamaguchi *et al.*[44] have provided the first direct demonstration of the formation of an LBHB, in the PYP. Yamaguchi *et al.*[44] prepared crystals with crystallization buffers made with 99.99% heavy water and identified 87% of the H/D positions in PYP by performing high-resolution (1.5 Å) neutron crystallographic analysis combined with high-resolution (1.25 Å) X-ray crystallography at room temperature. Their main results are

Table 9: Main measured interatomic distances (Å) corresponding to the *p*CA···Glu46 and *p*CA···Tyr42 hydrogen bonds in PYP. These data were obtained from a neutron diffraction experiment performed by Yamaguchi *et al.*[44]

*p*CA···Glu46 HB		*p*CA···Tyr42 HB	
O(*p*CA)···O(Glu46)	2.56	O(*p*CA)···O(Tyr42)	2.52
O(*p*CA)···D	1.37	O(*p*CA)···D	1.65
O(Glu46)···D	1.21	O(Tyr42)···D	0.96

summarized in Table 9. In Figure 25, the crystal structure of the PYP is shown. The mentioned short hydrogen bonds and the atoms that surround the *p*CA are shown in Figure 26. The average distance O–H (D) over all the measured hydrogen bonds was 0.95 Å.

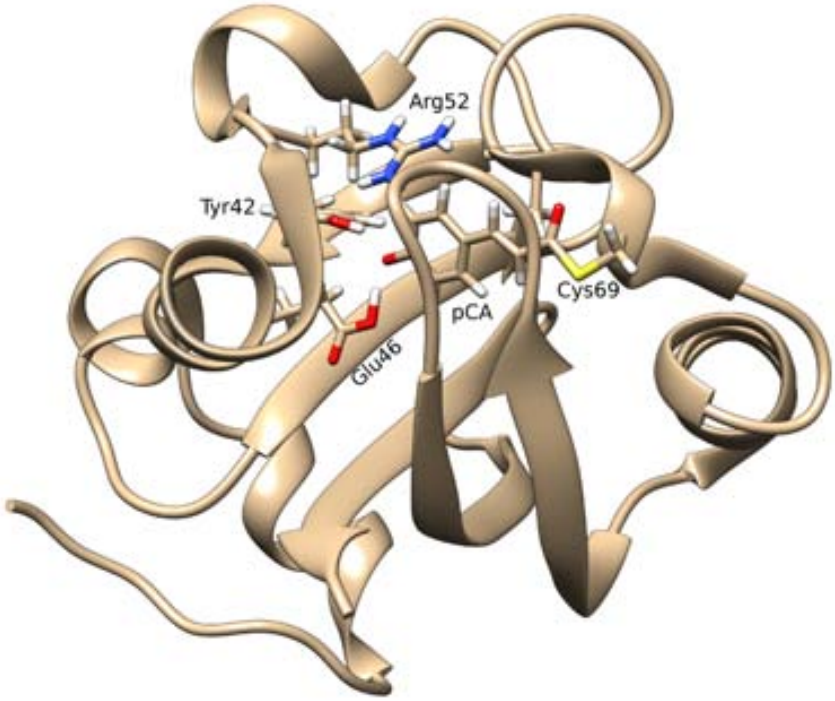


Figure 25: Structure of crystallized PYP as determined by neutron diffraction (PDB ID 2ZOI).

It is clear that the deuterium nucleus appears to be spread in the central region of the *p*CA···Glu46 hydrogen bond, thus being "shared" by the two oxygen atoms. Conversely, in the *p*CA···Tyr42, the deuterium atom is covalently attached to the phenolic oxygen of Tyr42. If we recall the definition of LBHB given in Section 1.3, the measurements of Yamaguchi *et al.*[44] furnish empirical evidence that *p*CA···Glu46 is actually an LBHB

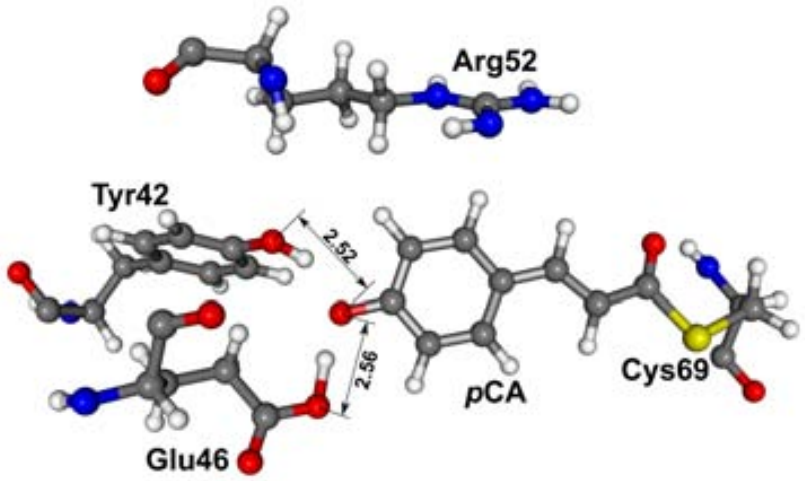


Figure 26: Detailed view of *p*CA and surrounding residues with distances to the closest proton acceptors (Å). Note that Arg52 is deprotonated

since the hydrogen atom has been detected to be in between the two electronegative atoms (oxygens). Analogously, *p*CA···Tyr42 is not an LBHB (see Table 9 and Figure 26). Another important counterintuitive finding was that Arg52, a residue at a distance of 6.34 Å from the phenolic oxygen of *p*CA, is deprotonated (neutral) in crystal, which prevents it from stabilizing the negative charge in the vicinity of the chromophore, and therefore contributing to the formation of the LBHB.

However, in very recent publications, Saito and Ishikita[45, 131, 136] have claimed that Arg52 should be protonated on the PYP protein surface and that the chemical properties of the *p*CA···Glu46 can be simply explained as a conventional -albeit short- hydrogen bond, without invoking the LBHB concept. To support such statment, they performed quantum mechanical/molecular mechanical (QM/MM) optimizations in order to reproduce the two short hydrogen bond distances of the crystal structure, obtaining 2.57 and 2.50 Å for *p*CA···Glu46 and *p*CA···Tyr42, respectively.[45] For both H-bonds, they say that there are no potential energy minima in the middle of the energy profile, but they are at 1 Å and 1.01 Å from, respectively, Glu46 and Tyr42. Therefore, the two hydrogen atoms should be covalently bonded to these residues, as far as they are concerned. Besides, they argue that the experimental ^{1}H NMR chemical shift (15.2 ppm) in solution[137] was assigned

to protonated Glu46, which is a value smaller than that for typical LBHBs. Moreover, QM/MM calculations by Saito and Ishikita[136] suggest that the experimental chemical shift for the *p*CA···Glu46 hydrogen bond should correspond to a geometry with the proton attached to the Glu46 moiety.

At this point the controversy is raised, and it is no longer clear whether an LBHB has already been definitely determined in a protein (in the PYP in particular). In the following sections we will describe a theoretical study of the two short hydrogen bonds found in the PYP aimed at discussing the existence of such LBHB and the source of discrepancy among the different experimental and theoretical results available so far. Let us remember that, so as to tackle the issue of LBHBs in PYP, a specific theoretical approach is adopted, as mentioned in Section 1.3. On the one hand, the fingerprint of an LBHB is the delocalization of the proton in the central region of the hydrogen bond, being this property determined by the nuclear probability density of the system, not just by the potential energy surface. This way, the fact that Saito and Ishikita do not find a minimum at the middle of the hydrogen bonds does not prove that there is not an LBHB, so the Schrödinger equation for nuclei will have to be solved. On the other hand, the chemical shift they obtain is 15.2 ppm,[131] certainly less than the common values for LBHBs (about 18-19 ppm), but this value was measured in solution, whereas the LBHB found by Yamaguchi was in a crystallized protein. To fully analyze these contradictory results, the protein will have to be studied for both cases: crystal and solvated, thus taking into account its conditions (environment, temperature...). This distinction will be shown to be determinant to understand whether (and when!) there is or there is not an LBHB in the photoactive yellow protein.

4.2 CALCULATIONS AND DISCUSSION

To this purpose, as we have already stated, we need to address this issue distinguishing when the protein is crystallized (Section 4.2.1) or solvated (Section 4.2.2). For both cases, the starting point was the neutron diffraction structure determined by Yamaguchi *et al.*[44] (PDB ID: 2ZOI), which contains about 87% of the hydrogen or deuterium atoms. Wherever an atom appeared in two positions with different occupation numbers, the one with the highest value was selected. Missing atoms were introduced at pD 9 (the conditions in which the structure was crystallized) using the PROPKA server.[121–123]

This analysis has implied, basically, the scanning of potential energy profiles or surfaces of the hydrogen bonds that we are studying and the integration of the nuclear Schrödinger equation. All these calculations have been performed on the PYP crystal structure. However, for the solvated PYP, the same calculations have been done, but previously a QM/MM molecular dynamics simulation has been carried out, so that relevant snapshots have taken as starting points.

Before tackling the rest of the problem, let us bear in mind that the neutron diffraction experiment of Yamaguchi *et al.* pinpointed in an unambiguous way the position of deuterium in *p*CA$\cdots$Glu46 very close to the midpoint of the hydrogen bond and that Arg52 was deprotonated. Although it may seem odd, it is nonetheless true that this statment comes from a *direct* measurement of the positions of protons and deuterons by means of the neutron diffraction, and as an experimental evidence must be considered a valuable piece of information. For this reason, for our calculations on crystallized PYP, Arg52 will be neutral. Nonetheless, on the other hand, we deemed it convenient to protonate Arg52 when the protein is solvated given its pK_a and the availability of protons in solution. As we mentioned in Section 1.3, the difference between the crystallized and solvated protein is thought to be remarkable: in particular, the (justified) decisions we have taken about protonation state of Arg52 will be shown to be relevant in the following subsections.

4.2.1 *Crystallized PYP*

Let us start showing the results on the O(*p*CA)$\cdots$D$\cdots$O(Glu46) hydrogen bond in crystallized PYP. As it has been already mentioned, we think that the possibility of an LBHB in the crystal form of PYP needs to be reexamined using proper theoretical methods that go beyond simple topographical studies of the potential energy surfaces. Nevertheless, so as to solve the corresponding nuclear Schrödinger equations, the two initial Born-Oppenheimer potential energy surfaces are required: the ones that involve *p*CA$\cdots$Glu46 and *p*CA$\cdots$Tyr42 hydrogen bonds. To this purpose, the potential energy surfaces (or profiles) have been computed within a QM/MM electronic embedding scheme, as explained in Section 2.3.1. In this study, the QM region includes the residues *p*CA, Glu46, Tyr42, Arg52, Thr50 and Ile31, and the QM energy has been calculated with Density Functional Theory (DFT)[106] using the CAM-B3LYP functional[86] and the split-valence 6-31+G(d,p) basis set. Although CAM-B3LYP was developed especially for studies of

charge-transfer excited states, for ground states it has been reported to perform better than B3LYP when computing classical potential energy barriers and NMR shielding constants.[108] More recent studies have confirmed a similar performance of both functionals for ground-state proton-transfer reaction barriers.[109] For this reason, CAM-B3LYP has been used in ground state DFT calculations.

Despite the fact that this choice is justified, we have wanted to make sure that the functional we have been using cannot affect our conclusions dramatically, so we have tested different methods of calculation on a 1-dimensional profile corresponding to the proton transfer in the Glu46$\cdots$*p*CA hydrogen bond. The profiles have been obtained by displacing the hydrogen atom involved in a short hydrogen bond along the line that links the two corresponding oxygen atoms. The results can be seen in Figure 27, and they indicate that all methodologies describe a very anharmonic potential energy profile, with two minima. The differences given by the different methodologies are small and concern the depth of the second minima describing the proton bound to *p*CA, which are summarized in Table 10.

All methods predict the existence of a marked shoulder with a secondary minimum with the proton bound to *p*CA, with a stability between 2.8–3.4 kcal mol^{-1}. The classical barrier to proton transfer is 3.1–3.7 kcal mol^{-1} from the Glu46–H(D)$\cdots$*p*CA minimum, and of 0.1–0.6 kcal mol^{-1} from the Glu46$\cdots$H(D)–*p*CA minimum.

As it has been already said, Arg52 is included in its neutral form. Neutron diffraction data have been obtained at room temperature,[44] so looking for a mininum (a suitable structure to describe the structure at best at 0 K) would be a debatable strategy, as seen in Chapter 3. For this reason, no geometry optimizations have been considered to describe the crystalline form of PYP.

When necessary, NMR isotropic shieldings have been obtained through use of the gauge-invariant atomic orbital (GIAO) method for each concerned structure (using the Gaussian 09 as well). Chemical shifts have been determined as the difference of isotropic shieldings of tetramethylsilane (TMS) protons and that of the concerned proton.

Given the fact that the O(*p*CA)$\cdots$D$\cdots$O(Glu46) hydrogen bond is not completely linear, we have chosen to study the motion of the H (D) on the plane defined by the positions of the three atoms involved. This has induced the definition of two orthogonal coordinates, x_1 and x_2 (see Figure 28), useful to analyze the potential energy surface and to perform the vibrational study,

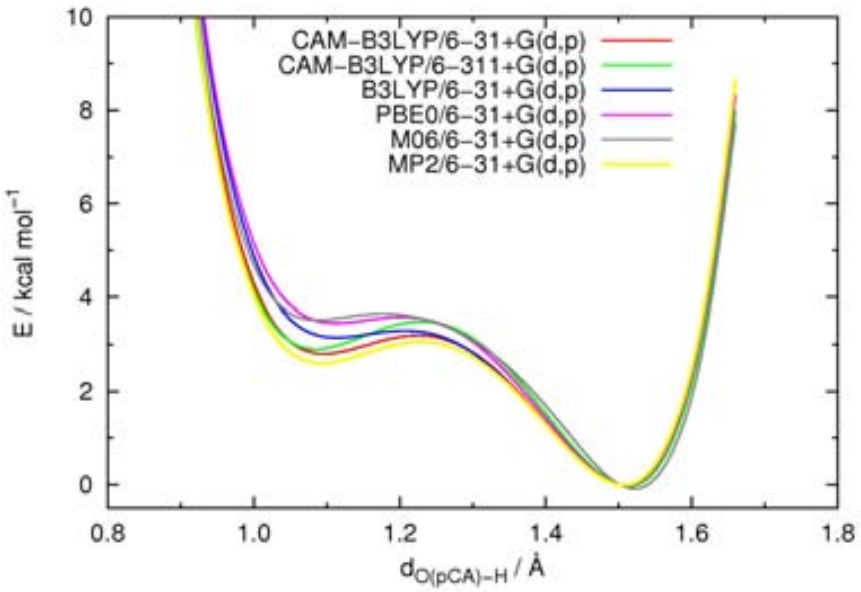


Figure 27: Computed 1-dimensional potential energy profiles with several methodologies (combinations of basis sets, functionals, etc.). The methodology used in this work to derive most of the results is shown in red (CAM-B3LYP/6-31+G(d,p)).

that are defined graphically in Figure 28. The origin for these coordinates has been conventionally assigned to the deepest minimum in the potential energy explored. For x_1 in particular, which denotes the position of H(D) along the line linking donor and acceptor atoms, positive values correspond to H(D) being closer to Glu46, and negative values to H(D) being closer to *p*CA than the minimum, respectively.

Using these coordinates, the PES has been computed and is represented in Figure 29. In agreement with the results of Saito and Ishikita the deepest minimum is found to be in a configu-

Table 10: Stabilities and barriers obtained with different methodologies

Methodology	Barrier pCA←Glu46 kcal mol^{-1}	Barrier pCA→Glu46 kcal mol^{-1}	ΔE kcal mol^{-1}
CAM-B3LYP/6-31+G(d,p)	3.19	0.39	2.81
CAM-B3LYP/6-311+G(d,p)	3.48	0.56	2.92
B3LYP/6-31+G(d,p)	3.29	0.13	3.16
PBE0/6-31+G(d,p)	3.57	0.11	3.46
M06/6-31+G(d,p)	3.66	0.13	3.29
MP2/6-31+G(d,p)	3.05	0.44	2.61

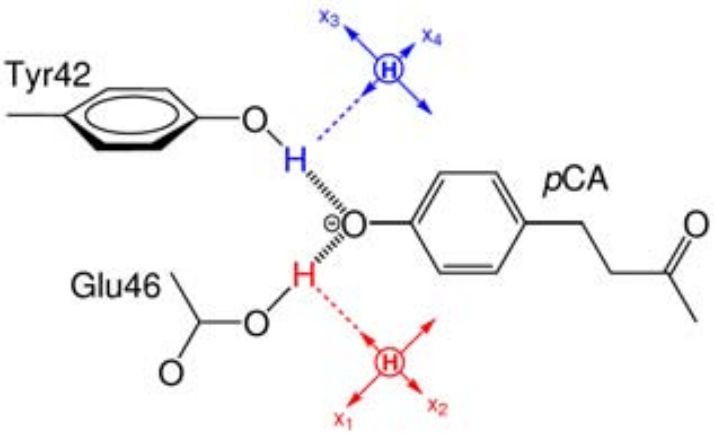


Figure 28: Description of the dynamical coordinate systems used to compute vibrational eigenstates and expected geometries of the hydrogen bonds. Coordinates for the Glu46–*p*CA hydrogen bond (x_1,x_2) are shown in red, and those for the Tyr42–*p*CA hydrogen bond (x_3,x_4) are shown in blue. The values for coordinates x_1 and x_3, which denote the main component of the position of the proton in the respective hydrogen bond, increase in going away from *p*CA.

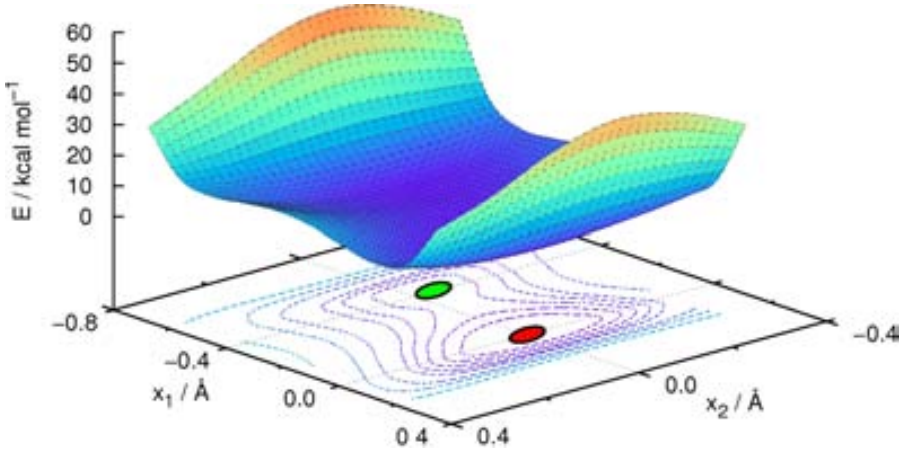


Figure 29: Potential energy surface built using the coordinates described in Figure 28. The red circle indicates the position of the absolute minimum –corresponding to the proton bound to Glu46– and the green circle that of the shallow minimum with the proton bound to *p*CA.

ration where the proton is bound to Glu46 (distances: Glu46—H(D)=1.05 Å, *p*CA—H(D)=1.51 Å). A remarkable feature we have found is a secondary minimum 2.31 kcal mol^{-1} above the former and corresponding to H(D) being bound to *p*CA (distances: Glu46—H(D)=1.43 Å, *p*CA—H(D)=1.13 Å). A similar minimum was also reported by Saito and Ishikita, but for a model system devoid of protein environment, that is, consist-

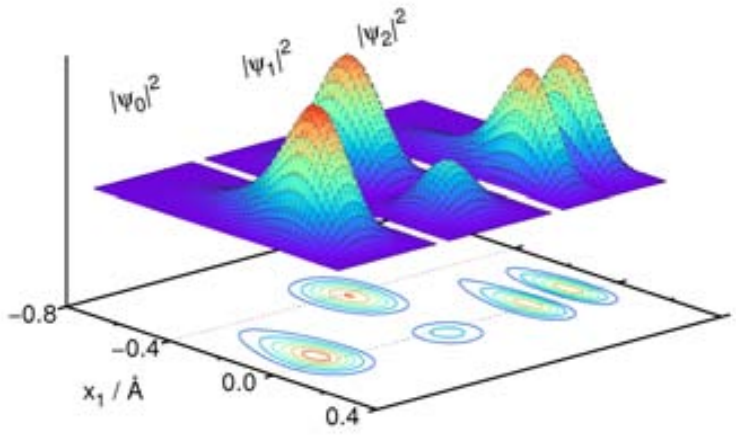


Figure 30: Side-by-side representation of the probability density functions of the ground (ψ_0), first excited (ψ_1) and second excited (ψ_2) vibrational states of the H-PYP system.

ing only of Glu46 and pCA [45]. The existence of this secondary minimum and its energetic properties indicate a degree of anharmonicity that is likely to have severe effects on the ZPE and, as a consequence, on the expected geometry of the hydrogen bond.

In concordance with the reasons given above, it is not the value of the coordinates of the minimum what ought to be compared to the experimental data on geometry, but rather the expectation values of geometrical parameters derived from the anharmonic vibrational wave function. To this purpose, from the potential energy surface, the nuclear wavefunctions have been obtained as follows: a total of 10 points in the x_1 and 6 points in the x_2 directions have been computed, for a total of 60 points (see Figure 28 for a description of the coordinates). The resulting potential energy surface has been fitted into cubic spline functional forms.[138] In all cases care has been exercised to ensure that the border of the potential energy profiles or surfaces has been "closed", that is to say, that all values on the border of the potential energy profiles or surfaces are sufficiently high with respect to the value of the energy in the minimum. This is relevant to avoid unphysical reflection effects on the vibrational wave functions computed.

To solve the nuclear Schrödinger equation, the generic discrete variable representation (DVR) proposed by Colbert and

Miller has been used.[139] For the DVR representation, a regular grid of 25 × 25 points spanning the domain of x_1 and x_2 (see Figure 28) has been used. After the DVR matrix representation of the Hamiltonian has been constructed, its diagonalization yields the eigenvalues E (vibrational energy levels) and eigenvectors Ψ (vibrational wavefunctions). The determined energy levels and wave functions have been tested for convergence by checking for stability upon enlargement of the DVR grid. The wave functions obtained in this way have been used to compute expected positions.

We want to emphasize that the 2-dimensional ZPE levels derived from this anharmonic vibrational study are high (4.07 kcal mol^{-1} for H-PYP, 2.74 kcal mol^{-1} for D-PYP) and are near or above the potential energy barrier for the proton-transfer. This will have important effects on the "spread" of the nuclear wave functions with consequences bearing on the expected geometry. Figure 30 shows the probability densities $|\psi_s|^2$ for the first few vibrational states. Taking the ground vibrational state as an illustrative example, the probability density peaks at *slightly negative* x_1 values (which means closer to *p*CA than the minimum of potential energy would indicate) with a large "tail" towards even more negative values. As a consequence, we would anticipate the expectation value of the distance H(D)–Glu46 to be *longer*, and H(D)–*p*CA to be *shorter*, than the position of the minimum would indicate. We have computed the expectation values for x_1 and x_2 and turned them into distances to donor and acceptor atom. The results are presented in Table 11.

The value of the D–Glu46 distance derived from the ground vibrational state is 1.10 Å, larger than for the potential energy minimum in our reduced PES (1.05 Å) but still far from neutron diffraction experiment (1.21 Å). For the D–*p*CA a complementary behavior is observed (1.46 Å from vibrational analysis, 1.51 Å from the PES' minimum, and 1.37 Å from neutron diffraction). The values coming from the ground vibrational state would correspond to what would be determined at 0 K, when only the ground state would be populated. Instead, experimental data was determined at 300 K. Then, what should be computed is the thermal average (according to an equilibrium Boltzmann distribution) of the expectation values for the distances for all vibrational levels that are appreciably populated. In this, it is useful to remark that the first excited vibrational state ψ_1, being completely above the barrier, has a markedly larger expectation value for the D–Glu46 distance, which will translate into a thermal increase of the measured length and a decrease of the complementary D–*p*CA distance. Indeed, at 300

Table 11: Vibrational energy levels, expected values for the distances between X=(H,D) and donor and acceptor atoms in the *p*CA$\cdots$Glu46 hydrogen bond, and chemical shifts for the 2-dimensional potential energy surface model. Population of chemically significant regions P_A is given for each isotopologue and for each vibrational state ψ_s, with A=Glu46 for "H(D) bound to Glu46", A=*p*CA for "H(D) bound to *p*CA", and $A = X$ for the intermediate case. Only levels within 5 kcal mol^{-1} of the ground state are shown. d_1 and d_2 are the distances between the expected value of the hydrogen (deuterium) atom and, respectively, O (Glu46) and O (*p*CA).

	State	E (kcal mol^{-1})	d_1 (Å)	d_2 (Å)	δ (ppm)	P_{Glu46}	P_X	P_{pCA}
H-PYP	ψ_0	4.07	1.14	1.43	18.7	0.64	0.30	0.06
	ψ_1	6.61	1.34	1.22	18.2	0.22	0.17	0.61
	ψ_2	8.77	1.13	1.44	18.3	0.70	0.25	0.05
D-PYP	ψ_0	2.74	1.10	1.46	18.5	0.78	0.20	0.02
	ψ_1	4.84	1.34	1.22	19.0	0.15	0.30	0.55
	ψ_2	5.73	1.09	1.47	18.1	0.82	0.16	0.02
	ψ_3	6.61	1.27	1.29	17.9	0.31	0.30	0.39

K the Boltzmann averaged distances resulting from this considerations yield 1.11 Å for D–Glu46 and 1.45 Å for D–*p*CA. The same procedure renders the values 1.14 Å for H–Glu46 and 1.42 Å for H–*p*CA.

To obtain a quantitative measurement of the degree of delocalization of the vibrational wave functions, a different procedure has been considered, based on deriving the population of certain chemically significant regions from the wave function itself, where the regions can be in this case "H(D) bound to Glu46", "H(D) bound to *p*CA", and "H(D) in-between". This can be done through this simple formula

$$P_A^s = \int_A |\psi_s|^2 d\tau \tag{32}$$

where P_A^s is the probability of finding H(D) in the region A when in vibrational state s, given by $|\psi_s|^2$. The regions are defined solely on the ranges of x_1 as follows: "H(D) bound to Glu46" corresponds to $x_1 > -0.1$, "H(D) in-between" to $-0.1 > x_1 > -0.3$, and "H(D) bound to pCA" to $-0.3 > x_1$. Table 11 also displays this information. As can be seen, already in the ground vibrational state, the D-PYP system suffers substantial delocalization as 22% of the density is found outside of the region corresponding to the deuterium bound to Glu46, this value increasing to 36% in the case of H-PYP. First excited states exaggerate this behavior notably (78% for H-PYP, 85% for D-PYP). The spread of the ground state wave function outside of the neighboring area to the minimum just described is responsible from the difference we have found in the expected geometry of this hydrogen bond and that denoted by the coordinates of the potential energy minimum. The same property for the first excited state sets the trend as temperature increases, which in both isotopologues is to shorten the H(D)–*p*CA distance.

Our results indicate that the reported crystal structure, taken as is, describes a substantially anharmonic vibrational profile for the H(D) in the Glu46$\cdots$*p*CA hydrogen bond, and that this enlarges notably the expected D–Glu46 distance in about 0.16 Å with respect to the average O$\cdots$H(D) distance over all hydrogen bonds determined in the neutron diffraction resolved structure. This enlargement reaches 0.19 Å for the expected H–Glu46 distance.

While it is not enough for a quantitative agreement with experiment, the trend shown is in the right direction and is not negligible. It must be borne in mind that effects like crystal packing are absent from the model used, and also that just one structure (in the sense of position of the heavy atoms) is considered: since experimental data comes from diffraction of a

crystal at room temperature, it would be necessary to include a sample of the configurations of the protein in the crystal as befits the thermal distribution at room temperature, which would make this study unattainable. Finally, it is known that even small imprecisions in the energies of the two minima along a double-well potential have noticeable effects on the degree of (de-)localization of the corresponding vibrational wave functions [140], which will manifest themselves when computing expectation values. Thus, small errors attributable to the electronic structure methodology used contribute to explain the discrepancies between experimentally measured distances and the values derived from the vibrational analysis.

Another parameter that is often invoked to discriminate the LBHB nature of a bond is the chemical shift. Saito and Ishikita computed the value of the isotropic shielding at the structure of the minimum of the PES and derived a value of 14.6 ppm [136]. Using the same rationale as above for expectation values of the geometrical parameters, the deuterium atom can be found in a plethora of different arrangements with respect to the donor and acceptor atoms, governed by the vibrational wave function. A more adequate form of computing an estimate of the chemical shift of this atom is to average the value of the chemical shift obtained at each of these arrangements using $|\psi_s|^2$ as weight, to produce vibrational level specific values of chemical shift, and then do a Boltzmann average to determine the expected value to be measured. This procedure has been applied before satisfactorily [14, 15, 17, 20]. The values of chemical shift for the first few vibrational levels are also presented in Table 11. The values for the ground state –adequate for very low temperature experiments– are 18.7 ppm for H-PYP and 18.5 ppm for D-PYP. These values fall well within the range commonly accepted to indicate the existence of an LBHB. This is true even for the position of the minimum in our PES (17.6 ppm).

Considering all the above, the ZPE level in the case of D-PYP lies close to the barrier of potential energy, and the wave function of this first vibrational state is spread over the full range of D coordinates. This is indicative that, in the crystal form, this hydrogen bond qualifies as an LBHB. This statement is strengthened by the value of 18.5 ppm for the chemical shift linked to the ground vibrational state of D-PYP. The LBHB character for H-PYP would be reinforced, as the ZPE is larger and thus, the vibrational wave function for the ground state is spread out more evenly still.

A point of controversy in the literature concerned the protonation state of Arg52, as the presence of a neutral Arg52 might

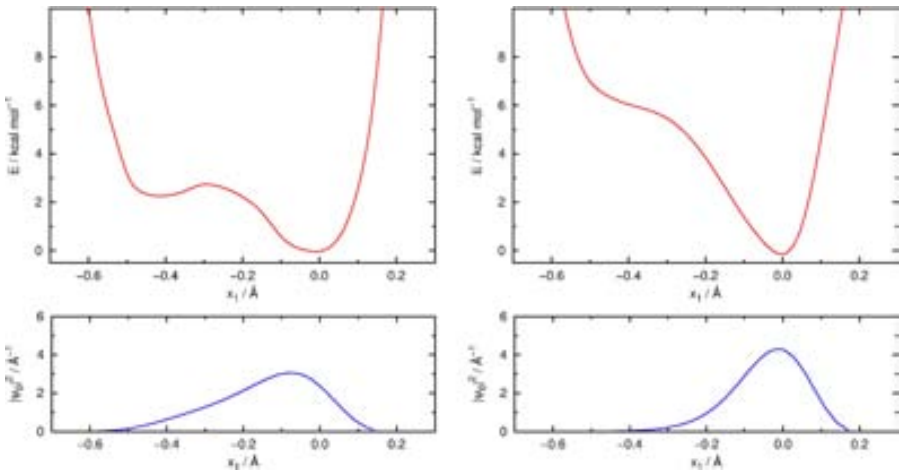


Figure 31: Top panels (in red): Comparison of 1-dimensional potential energy profiles of H(D) transfer in the Glu46$\cdots$*p*CA hydrogen bond for crystalline PYP when Arg52 is deprotonated (left) and protonated (right). Bottom panels (in blue): Probability densities of the ground vibrational states ($|\psi_0|^2$) of H-PYP when Arg52 is deprotonated (left) and protonated (right). For the definition of x_1 see Figure 28.

be seen as instrumental in stabilizing an LBHB between Glu46 and *p*CA. It is possible to assess what the role of the protonation state of this residue is by computing a 1-dimensional potential energy profile for the transfer of H(D) from Glu46 to *p*CA in the crystal structure for the two protonation states of Arg52. These profiles have been done evaluating 10 geometries in the x_1 direction, and the nuclear wavefunctions have been obtained following the same method as explained for the 2-dimensional surfaces. In this case, DVR representation has been taken to be 50 evenly spaced points spanning the domain of calculated geometries.

The results are displayed in Figure 31. A very striking difference is seen in both profiles, namely that the secondary shallow minimum describing the situation where H(D) is bonded to *p*CA, vanishes when Arg52 is protonated. Thus, changing the protonation state of Arg52 has an effect of about 4 kcal mol^{-1} on the stabilization of this secondary minimum. In fact, solving the 1-dimensional nuclear Schrödinger equation of the vibrational Hamiltonian on both profiles shows quite clearly that, if the protein is held at the structure determined in the crystal, the nuclear wave function of the ground state is essentially localized on the right well (corresponding to Glu46–H(D)$\cdots$*p*CA) when Arg52 is protonated. On the contrary, if Arg52 is deprotonated, the wave function is delocalized with measurable amplitude over both wells, including the central region over the bar-

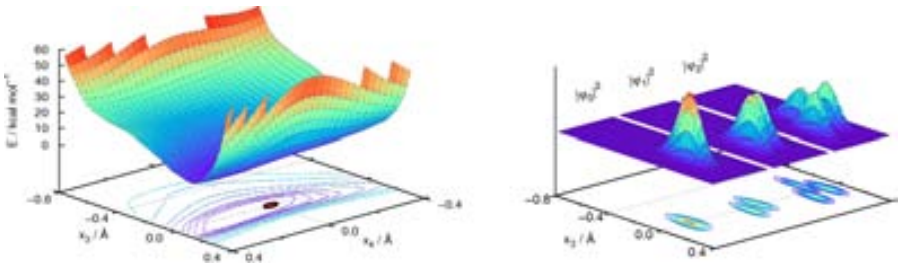


Figure 32: Results of the 2-dimensional anharmonic vibrational study of the Tyr42-pCA hydrogen bond. Left: Potential energy surface using the dynamical coordinates described in Figure 28. The red circle denotes the position of the absolute minimum, with the proton bound to Tyr42. Right: side-by-side representation of the probability density functions of the ground (ψ_0), first excited (ψ_1) and second excited (ψ_2) vibrational states of the H-PYP system.

rier. Thus, a situation akin to a LBHB arises only when Arg52 is deprotonated.

Let us now turn to the second short hydrogen bond, *p*CA$\cdots$Tyr42. In the crystal structure donor and acceptor atoms are 2.52 Å apart, but the shared D is found clearly in the immediate neighbourhood of Tyr42's oxygen. An analogous procedure to determine the 2-dimensional vibrational eigenstates has been carried out using the x_3 and x_4 dynamical coordinates defined in Figure 28. Using these coordinates, the potential energy surface corresponding to the motion of H(D) has been computed and is shown in Figure 32. The profile shows a deep minimum at short Tyr42–H distances, and a shoulder can be seen at short pCA distances that fails to become a secondary minimum, in contrast to what was found for the Glu46–pCA hydrogen bond. The vibrational states have been computed and a summary of the lowest lying states and expectation values of the geometry are shown in Table 12, while plots of the probability densities can be seen in Figure 32. The contrast with the Glu46–pCA hydrogen bond case is striking: the wavefunction for the ground state is confined closer to the minimum at $(x_3, x_4) = (0, 0)$, and the lowest-lying excited states spread actually along x_4 (which is a sideways motion of the proton) rather than along x_3, and hence the expected geometries within each vibrational state are very similar to each other and to that of the potential energy minimum ($d_1 = 1.01$Å, $d_2 = 1.51$Å). There is little doubt, in view of the above, that H(D) is to be found always bound to Tyr42. Analysis of the population of the chemically relevant areas of the Tyr42$\cdots$*p*CA hydrogen bond (Table 12) reveals that there is no effective population of the "H(D) bound to pCA" region, while at least 85% population is to be found in the "H(D) bound to Tyr42" region. A varying, but small portion of at most 15% of

Table 12: Vibrational energy levels, expected values for the distances between X=(H,D) and donor and acceptor atoms in the *p*CA$\cdots$Tyr42 hydrogen bond, and chemical shifts for the 2-dimensional potential energy surface model. Population of chemically significant regions P_A is given for each isotopologue and for each vibrational state ψ_s, with A=Tyr42 for "H(D) bound to Tyr42", A=pCA for "H(D) bound to pCA", and A=X for the intermediate case. Only levels within 5 kcal mol^{-1} of the ground state are shown. d_1 and d_2 are the distances between the expected value of the hydrogen (deuterium) atom and, respectively, O (Tyr52) and O (*p*CA).

	State	E (kcal mol^{-1})	d_1 (Å)	d_2 (Å)	δ (ppm)	P_{Tyr42}	P_X	P_{pCA}
H-PYP	ψ_0	3.68	1.03	1.50	14.0	0.85	0.15	0.00
	ψ_1	6.66	1.01	1.52	13.4	0.88	0.12	0.00
D-PYP	ψ_0	2.16	1.02	1.51	13.8	0.90	0.10	0.00
	ψ_1	4.23	1.01	1.52	13.3	0.92	0.08	0.00
	ψ_2	6.23	1.00	1.53	13.1	0.90	0.10	0.00

population is in the central area. This is in sharp contrast to the Glu46–pCA hydrogen bond (Table 11), where the central region was populated up to 30% in H-PYP. To summarize H(D) are much more localized in this hydrogen bond. In this analysis, we have used the definition that "H(D) bound to Tyr42" corresponds to $x_3 > -0.1$, "H(D) in-between" to $-0.1 > x_3 > -0.3$, and "H(D) bound to pCA" to $-0.3 > x_3$.

As could be expected, computation of expected chemical shifts for the vibrational ground states yield values of 14.0 ppm (H) and 13.8 ppm (D), well below the range where and LBHB can be suspected. Excited vibrational states yield values of chemical shift even smaller, which causes predicted values of chemical shift at temperatures higher than absolute zero to be smaller than those for the ground states.

Therefore, for all practical purposes the Tyr42$\cdots$*p*CA hydrogen bond is not an LBHB, but a short hydrogen bond with the H(D) clearly bound to Tyr42.

4.2.2 *Solvated PYP*

When analyzing the results for crystal PYP (see Section 4.2.1) we have shown that the neutron diffraction structure (including a deprotonated Arg52) brings about a situation with large anhamonicities for the H(D) motion along the proton-transfer coordinate between *p*CA and Glu46. This in turn causes the ground vibrational state to lie at or above the potential energy barrier and results in the largest probability of finding H(D) lying at the center of the hydrogen bond. In this sense, it can be said that the claim of Yamaguchi *et al.* that keeping Arg52 neutral would make the LBHB necessary to stabilize the negative charge on *p*CA is vindicated by our solid-state results.

Our calculations for crystallized PYP also predict a very low-field signal for the transferring proton, which further supports the claim that this is indeed an LBHB. Nevertheless Sigala et al. reported that the proton shared by *p*CA and Glu46 appears with a chemical shift of 15.2 ppm in solution NMR experiments [137]. This value lies well outside the range accepted for an LBHB (17–19 ppm), and this was one of the key points raised by Saito and Ishikita [45] to refute the existence of such an LBHB. In fact, the strongest criticism raised by Saito and Ishikita to the conclusions of Yamaguchi concerns the protonation state of Arg52. The rationale of this criticism was that Arg52 has a very high pK_a value and ought to be protonated, which should weaken the hydrogen bond between *p*CA and Glu46, such that under this conditions it will likely not be an LBHB.

We think that the disparate results on the existence or not of an LBHB in PYP can be explained by accepting that both experiments refer to *different* systems, or rather, the same system in two different states. It is not unheard of that a given biological macromolecule shows substantial differences in structure between crystalline state and solution [141]. The structure of PYP is such that *p*CA is not deeply buried in the structure and Arg52 is actually placed as if it were a "lid" over *p*CA and in direct contact to the solvent bulk (See Figure 25 and Figure 33). Now, if solution (as opposed to crystal) conditions are considered, it is reasonable to expect Arg52 to be protonated, as it is in contact with the solvent bulk. This will indeed affect the energetics of the *p*CA$\cdots$Glu46 hydrogen bond, weakening it from its situation when Arg52 is deprotonated and causing the chemical shift to drop out of the LBHB range. To test this end we have taken the crystal structure and protonated the Arg52 residue as proposed by Saito and Ishikita. We have determined the average chemical shift value derived from the 1-dimensional proba-

bility density, which produces a value of 18.5 ppm for the transferring proton in the ground vibrational state and 17.9 ppm for deuterium (see Figure 31), which is only slightly smaller than the value obtained for deprotonated Arg52 and would still be thought to correspond to an LBHB (although we have already shown above that the nuclear wave function in the crystal with protonated Arg52 is localized on the Glu46–H(D)$\cdots$*p*CA well). We conclude from this result that changing the protonation state of Arg52 does not seem to be the only reason for the low chemical shift of the transferring proton determined by Sigala *et al.* [137]

Now, if the system is studied in solution at room temperature, it would be subject to thermal jitter, which could affect the dynamical stability of the hydrogen bond network over time. It is reasonable to expect that this would cause the average donor-acceptor distance of any hydrogen bond to enlarge noticeably.

For these reasons we think that conditions in solution differ drastically from those in crystal state and need to be addressed in an adequate way if the trends introduced are to be assessed reliably. The proper procedure involves tracking the evolution of the system in solution along a dynamical trajectory at room temperature. The possibility that an LBHB be present in the relevant part of the system rules out the possibility of standard MM simulations as none of these has been parameterized to describe this kind of interaction, and a much costlier QM/MM trajectory has to be computed instead.

The starting point for the QM/MM molecular dynamics simulation is the crystal PDB, with Arg52 protonated (because the protein is in solution). A water sphere of 35 Å has been used to solvate the protein, placing its center at the center of mass of the protein. This 35 Å radius droplet reproduces the environment of the protein in solution. Figure 33 depicts the solvated system. While it is not possible to fully reproduce bulk solvent conditions in this way, this picture shows that a rather thick layer of solvent molecules surrounds the protein, and that it is legitimate to assume that the main effects of the interaction with the solvent will be reproduced in this setup.

Once the protein has been solvated, the geometry has been optimized within a QM/MM scheme (more details are given in Section 2.3.2). The quantum mechanical region includes the residues *p*CA, Glu46, Tyr42, Arg52, Thr50, and Ile31 and three crystallographic water molecules that interact with them, 88 atoms in total. The MM region includes 18270 atoms. The active region includes all residues with atoms within 19 Å of the chromophore.

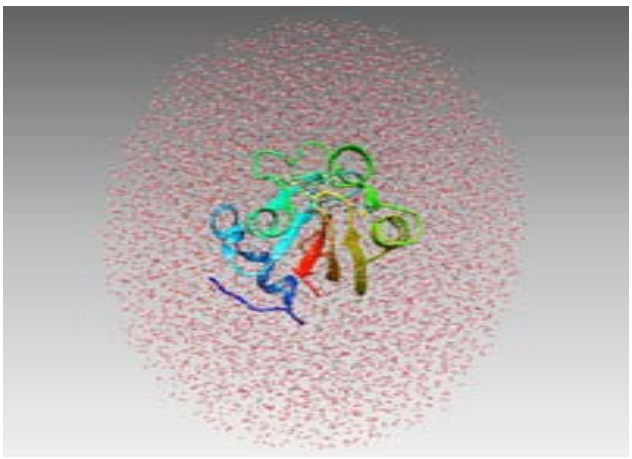


Figure 33: Depiction of the system used to study PYP in water solution. The diameter of the water droplet is 70 Å, and a spherical concentric layer of ~ 20 Å thickness surrounds the protein.

Starting from the optimized structure, a molecular dynamics simulation has had to be computed. Because LBHB interactions are absent from all force fields known to us, a QM/MM molecular dynamics simulation has had to be performed. The QM region includes the same residues as the optimization, while the MM region includes the rest of the protein and solvent molecules. The active region has consisted of all the residues containing at least one atom with 15 Å of any *p*CA atom. Because the computational costs are extremely onerous with respect to the MM molecular dynamics, the AM1[111] semiempirical Hamiltonian has been selected to deal with the QM part of the system. Even in these conditions, the simulation is very costly and only relatively short simulation times are affordable, even though the simulation suffices to assess the dynamical stability of the structure. The system has been heated to 300 K in steps of 10 K. After that, the system has been equilibrated for 10 ps. Finally, a 100 ps dynamics simulation has been run (in 1 fs time steps) in the NVT ensemble. More details about the QM/MM optimization and dynamical simulation are given in Section 2.3.2.

The dynamics shows that the *p*CA···Glu46 hydrogen bond is still well formed over time with mean elongation of ~2.9 Å (see Figure 34). This reveals a substantially stretched hydrogen bond compared to the situation in the crystal structure (2.56 Å). Even though the quantitative magnitude of the change has to be taken with care due to the approximations incurred in the simulation, it is still very reasonable that thermal agitation increases the average O–O distance from its value in the crystalline structure.

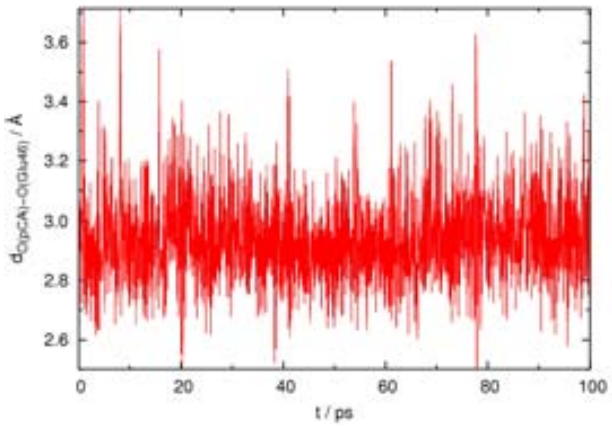


Figure 34: Time evolution of the *p*CA–Glu46 hydrogen bond distance along the QM/MM molecular dynamics simulation.

Table 13: Zero-point energies, expected values of the of the O(Glu46)···H distance and average chemical shift values for selected structures along the QM/MM molecular dynamics in solution at 300 K.

$d_{O(Glu46)--pCA}$ (Å)	E (kcal mol^{-1})	$\langle d_{O(Glu46)-H} \rangle$ (Å)	$\langle \delta \rangle$ (ppm)
2.61	3.62	1.06	16.4
2.66	3.52	1.07	15.1
2.71	4.27	1.02	13.7

The substantial enlargement of the *p*CA–Glu46 distance, together with Arg52 being protonated whilst in solution, represent a meaningful change in the energetics and dynamics of the putative LBHB. Does it still qualify as one? To answer this question we need access to the energy landscape of the proton-transfer in structures visited by the trajectory. We have picked at random some structures sampled by the trajectory that have selected *p*CA···Glu46 O–O distances. For each structure a 1-dimensional potential energy profile has been computed with a total of 10 distinct geometries, obtained by shifting the position of the H from donor to acceptor. The QM/MM methodology for the energy profiles and the obtention of the nuclear wavefunctions has been the same as explained for the crystal structure calculations in Section 4.2.1 (now the DVR representation has been taken to be 50 evenly spaced points spanning the domain of calculated geometries). Again, the nuclear wavefunctions have been used to compute the expected value of the chemical shift. A few of the values computed are shown in Table 13.

The range of *p*CA···Glu46 O–O distances explored covers the shortest range explored by the QM/MM MD simulation, namely between 2.61 Å and 2.76 Å (which represents about 5% of the structures). On the grounds of the chemical shift computed, we observe a very clear trend to lower chemical shifts upon increasing O–O distance, indicative, within current precision, of a trend towards a "normal" hydrogen bond. For the majority of other structures (those with *p*CA···Glu46 O–O distances beyond 2.76 Å), it is expected that the chemical shift will remain basically constant at this value.

Summarizing, inclusion of thermal jitter *and* protonation of Arg52 in solution causes the structure of the *p*CA···Glu46 hydrogen bond to weaken and become a conventional hydrogen bond.

The general concept that LBHBs determined in solid or gas phase vanish when the system is studied in solution is not new. Perrin *et al*. studied some years ago several systems that showed LBHB character in gas phase or in solid state and found that these systems existed as two tautomers in solution [42, 142–147]. They explained this behavior by invoking the concept of "solvatomers" [41, 43], meaning by this that the tautomers are affected by the solvent disposition, and in the understanding that even a molecule whose H-transfer could be described by a completely symmetric double well would see this symmetry lifted because of interactions with instantaneous solvent configurations.

4.3 CLOSURE

As we anticipated in Section 1.3, PYP behaves very differently in these two physically different situations, namely, in crystal and in solution.

Taking the neutron diffraction structure (that is, the crystal structure), with neutral Arg52, we have shown that the energy profile is very anharmonic. As for the O(*p*CA)···D···O(Glu46) hydrogen bond, the nuclear vibrational analysis has led to expected values of the H (D) position that are fairly different from the energy minimum. Even though the profile is not symmetric and the proton (deuterium) is not situated exactly at the midpoint between the two electronegative atoms, it can be asserted that it is not covalently bonded to any of them, so we actually found an LBHB in the crystal PYP. The value of the calculated chemical shift for this hydrogen bond (18.7 ppm in the ground vibrational state) corroborates this statement. All this proves the importance of the approach that we followed: being the

O(*p*CA)···D···O(Tyr42) hydrogen bond even shorter, both the neutral diffraction experiment and our simulations concluded that it is not an LBHB. It is convenient then to emphasyze that the only way to figure out whether a hydrogen bond is an LBHB or not is finding out the expected value of the proton (deuterium), either measuring it empirically or calculating it. And when it is intended to go on with the calculations, only by treating the proton (deuterium) as a wavefunction can this issue be tackled, as we have made clear.

On the other hand, when the protein is in solution, a different point of view has been taken. The only experimental datum indicative of the fact that O(*p*CA)···D···O(Glu46) is not an LBHB came from the 15.2 ppm value of the chemical shift in solution[45]. Because Arg52 sits between *p*CA and bulk solvent, it is reasonable that in solution this residue will be protonated. Moreover, the thermal jitter that would be present in solution at room temperature helps in determining an increased O–O distance for this hydrogen bond. In these conditions, the same procedure as in the crystal phase carried on different snapshots of a QM/MM molecular dynamics simulation has revealed that the proton is, in all cases, very localized next to Glu46. The computed chemical shift in these circumstances are in line with those for a conventional hydrogen bond. Summarizing, we have found out that there is no LBHB when PYP is in solution. Again, this proves the importance of the method we followed and the relevance of considering the solvent and the thermal effects.

In all, we have conciliated two opposite viewpoints about the existence of an LBHB in PYP: it exists when the protein is crystallized, but in solution it becomes a common hydrogen bond, much in line with what was found for many simpler systems by Perrin *et al.*[42, 43]. We have shown the importance of the adopted methodology so as to properly face the problem, and, more specifically, our results support the first direct experimental demonstration of the formation of an LBHB in a protein –albeit in crystal phase.

5

THE REACTION IN GFP AND HOW IT HAPPENS

As it has been mentioned in the Introduction, this chapter is devoted to the study of the Green Fluorescent Protein's reactivity. Some of our group's members had calculated the potential energy surface[70] and performed a classical dynamics simulation of the GFP proton wire[74]. This work, published between 2006 and 2008, is a valuable legacy that provided valuable insights on the proton tranfer that accounts for the GFP fluorescence. These insights will be explained in more detail in Section 5.1.

From the potential energy surface and from the quantum dynamics simulation, a method is proposed to compute the tunneling contribution and, more generally, to analyze the probability flux and to know its values in different regions of the hyperspace. The definition of tunneling is not unambigous: in Section 5.2, we present some considerations about the reasons why the tunneling contribution should be calculated from a dynamical point of view, that is, from the probability flux. A first explanation of the probability flux and its importance is given in Section 5.3. The method we propose to analyze the probability flux is carefully presented in Section 5.4, and its implementation is explained in Section 5.5. In Section 5.6, the results of this analysis are shown, and general conclusions of its importance and applications are given in Section 5.7.

However, the potential energy surface calculated by Vendrell *et al.* does not take into account the protein environment: it consisted of a QM calculation on the chromophore and nearest residues. A decade ago, the computational resources were not as powerful as today, and the software devices for QM/MM calculations (such as ChemShell) were not available yet. Throughout this thesis, the importance of the electrostatic effects of the environment have been stressed and proved. For this reason, we have started to find a new potential energy surface in the ground and excited electronic state such that takes into account the electrostatic interactions with the whole protein. It is aimed to perfom a more reliable quantum dynamics simulation and analyze the probability flux with the developed method. This work is still in progress, and it will be presented in Section 5.8.

5.1 INTRODUCTION TO THE GFP PROTON WIRE

In this section we are presenting the previous work by Vendrell *et al.*. The work published between 2006 and 2008 was a son of the technological level of the time.[1] Some preliminary work[148] had revealed that the photoactive $\pi\pi^*$ state could be lying close to a dark $\pi\sigma^*$ state and the possibility of a crossing could not be disregarded. This made calculations of the excited states complicated as three excited states (ground, $^1\pi\pi^*$ and $^1\pi\sigma^*$) had to be computed. Furthermore, the choice of QM/MM methods available at the time was narrow. This, together with the need to compute a sizeable extent of the PESs, made the authors opt for a cluster approach, devoid of environment, and computed at a multi-configurational level (CASSCF/CASPT2).

So as to figure out whether the geometry of the system was stable in physiological temperature, and thus to have a plausible structure to which further calculations can be done, a Molecular Dynamics (MD) simulation was done.[70] The MD simulation revealed that the hydrogen bonds that form the proton wire (Figure 35) show donor-acceptor distances of 2.6-2.8 Å and the chromophore (formed by two rings) remains essentially planar, as no more than 10° inclination of one ring respect to the other are reached. This planarity is not irrelevant: isolated from the protein, the chromophore practically has no quantum yield.[52, 71] The reason that accounts for this is that the chromophore, when it presents a certain torsion between its rings, can access a conical intersection that allows a nonradiative transition to the ground state. When surrounded by other amino acids and side chains, the chromophore cannot move freely. Crystallographic studies have disclosed that the protein has a beta barrel shape, with an alpha-helix going through it. The chromophore is placed in the middle of this central helix, protected from the solvent by the beta sheets that surround it.[69] This packaging ensures the planarity of the chromophore throughout the dynamics and, as a result, it is potentially emissive if properly excited.

As for the potential energy, Vendrell *et al.* studied the different electronic configurations[70] of the different protonation states. The model used for these electronic calculations is shown in Figure 35, in which the residues involved in the proton wire are included, plus some close residues that can be relevant. This study revealed the existence of an accessible $\pi\pi*$ state. The adjusted PES of the excited state[149] showed a very small energy

1 The declared goal was to perform a QD simulation to evaluate the mechanism, measure the velocity of the process and assess the fraction of tunneling in it

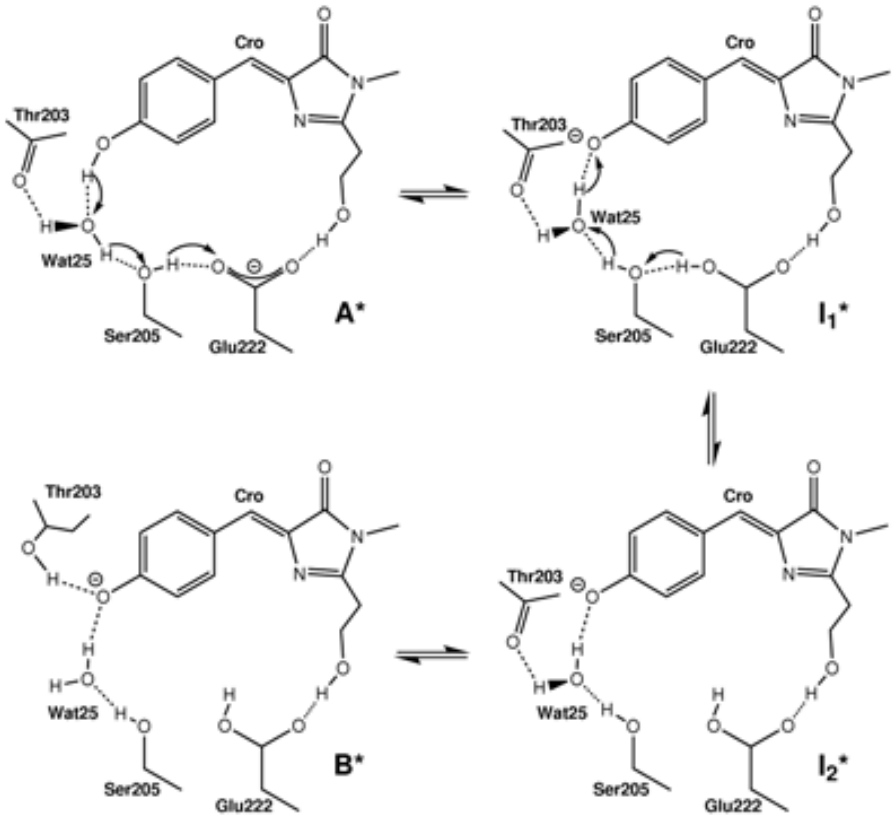


Figure 35: Schematic representation of the chemical species experimentally observed. These models are *mutatis mutandis* the ones used for electronic calculations, except for Thr203, which is not included in the calculations.[74]

barrier (1.9 kcal/mole, 0.8 kcal/mole including Franck-Condon energy). The energy profiles[70] suggested that, upon photoexcitation, the most favorable path was the following: the proton would first be transfered from Ser205 to Glu222, then from Wat25 to Ser205 and finally from the chromophore to Wat25, that is to say, the opposite from what could have been expected at first. This configuration would imply the existence of intermediate species with the least charge separation. Lill and Helms[71] postulated that the reaction happened in a stepwise scheme, but their simulation started from a deprotonated chromophore, and no quantum effects were considered.

At this point, Vendrell *et al.* performed a quantum dynamics simulation of the proton transfer.

A quantum dynamics simulation consists of integrating the time-dependent Schrödinger equation (Equation 2). For this, a detailed knowledge of the PES is required. From the potential energy calculated for different positions[70], a Potential Energy

Surface as a function of several parameters was interpolated[149], both for the ground (V^{S_0}) and for the photoactive (V^{S_1}) states.

The quantum dynamics simulation was solved on a reduced model, represented in Figure 36, that only included the atoms *directly* involved in the proton transfer. It was assumed that proton motion would occur only along the line that links each donor-acceptor atom pair. As depicted, the internal coordinates system is described with seven different coordinates: three r_i coordinates that indicate the signed distance between trasnferring proton *i* and the midpoint between its donor and acceptor atoms, three R_i coordinates that denote the distance between each donor-acceptor atom pair, and the center-of-mass. The TDSE was integrated using the MCTDH package and the time-evolved wave function analyzed.[74, 150] Even though this model is limited in its predictive capabilities (the atoms that are not involved in the proton wire are not taken into account and the environment is missing altogether), it is still one of the most detailed models available which includes quantum dynamical effects in the motion of the nuclei.

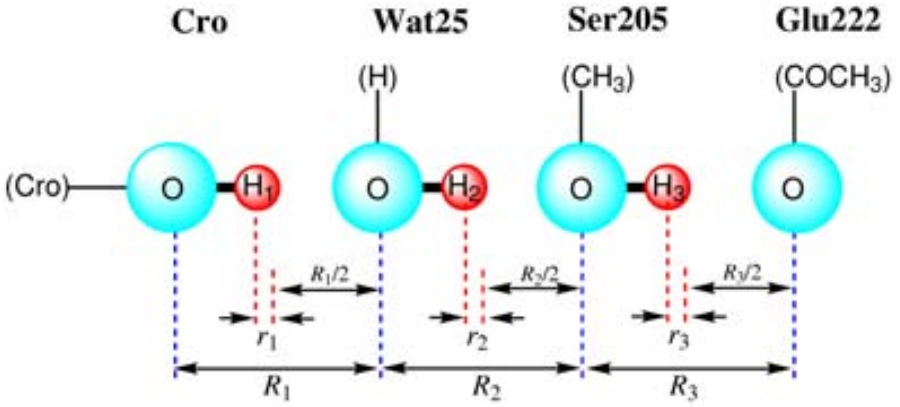


Figure 36: Dynamical model used to follow the dynamics of the proton relay in GFP.

Let us consider x_i as the Cartesian coordinate of each of the seven atoms ordered from left to right as shown in Figure 36, being x_1 the Chromophore oxygen. The internal coordinates can be expressed as functions of the Cartesian coordinates:

$$
\begin{aligned}
r_1 &= x_2 - \frac{1}{2}(x_3 + x_2) \\
R_1 &= x_3 - x_1 \\
r_2 &= x_4 - \frac{1}{2}(x_5 + x_3) \\
R_2 &= x_5 - x_3 \\
r_3 &= x_6 - \frac{1}{2}(x_7 + x_5) \\
R_3 &= x_7 - x_5 \\
R_{cm} &= \frac{\sum_i x_i m_i}{\sum_i m_i}
\end{aligned}
\tag{33}
$$

With these coordinates, the wavefunction was separated into two products: one that depended on the center-of-mass, and another one depending on the rest of the coordinates. This way, the quantum dynamics ran on six degrees of freedom instead of seven, which simplified the problem. From the Hamiltonian operator $\hat{H} = \hat{T} + \hat{V}$, the potential energy had already been obtained by interpolation.[149] The kinetic energy operator was adapted to the specific coordinates:

$$
\begin{aligned}
\hat{T} = &-\left(\frac{1}{2m_{H_1}} + \frac{1}{4M}\right)\frac{\partial^2}{\partial r_1^2} - \left(\frac{1}{2m_{H_2}} + \frac{1}{4M}\right)\frac{\partial^2}{\partial r_2^2} \\
&-\left(\frac{1}{2m_{H_3}} + \frac{1}{4M}\right)\frac{\partial^2}{\partial r_3^2} - \frac{1}{M}\left(\frac{\partial^2}{\partial R_1^2} + \frac{\partial^2}{\partial R_2^2} + \frac{\partial}{\partial R_3^2}\right) \\
&-\frac{1}{2M}\left(\frac{\partial}{\partial r_1}\frac{\partial}{R_2} + \frac{\partial}{\partial r_2}\frac{\partial}{\partial R_3} - \frac{\partial}{\partial r_2}\frac{\partial}{\partial R_1} - \frac{\partial}{\partial r_3}\frac{\partial}{\partial R_2}\right) \\
&-\frac{1}{4M}\left(\frac{\partial}{\partial r_1}\frac{\partial}{\partial r_2} + \frac{\partial}{\partial r_2}\frac{\partial}{\partial r_3}\right) + \frac{1}{M}\left(\frac{\partial}{\partial R_1}\frac{\partial}{\partial R_2} + \frac{\partial}{\partial R_2}\frac{\partial}{\partial R_3}\right)
\end{aligned}
\tag{34}
$$

Firstly, the wavefunction corresponding to the ground state Hamiltonian was obtained. This function was taken as the initial condition of the time-dependent Schrödinger equation. Then, the equation was solved for the electronic excited-state Hamiltonian. This was done to mimick the Franck-Condon principle of vertical excitation. The ground state wavefunction, eigenfunction of the V^{S_0} Hamiltonian, in the V^{S_1} Hamiltonian is not stationary any more, so it evolves with time. The function was propagated for 15 picoseconds.

If we analyze the perprotio dynamics immediately after the photoexcitation (initial stages of the simulation) it can be seen that the distance between chromophore and Wat25 (R_1) decreases

suddenly from 2.55 to 2.40 Å after only 20 fs (Figure 37, right). It can be expected that when the first proton transfer starts, the donor (chromophore) and acceptor (Wat25) oxygen atoms get closer, since the chromophore becomes more acid upon photoexcitation and facilitates the hydrogen leap. It can also be obseved (Figure 37) that R_2 and R_3 increase a little bit at the very beginning, but after a short while they decrease (this happens when the other two hydrogen atoms are transfered (after about 15 fs of simulation). When R_2 and R_3 start diminishing, R_1 has not been stabilized yet, which suggests a concerted reaction. On the left of Figure 37, it is shown that for the first 20 fs of simulation there is a fast proton transfer: all three protons are displacing almost concomitantly, although H_1 is a bit slower than the other two.

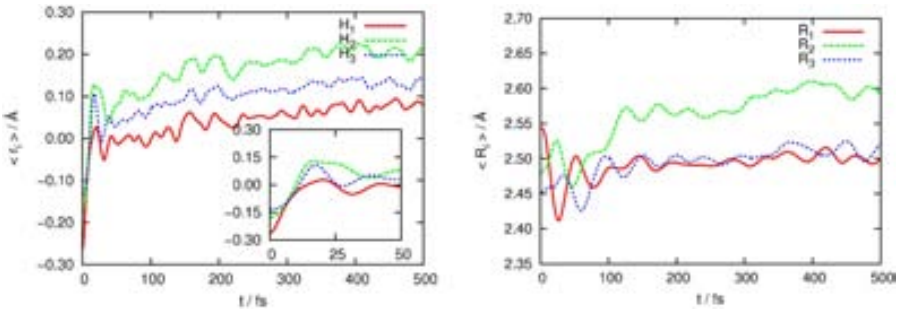


Figure 37: Expected values of the positions of the protons (left) and the distances of the consecutive oxygen atom pairs (right) for the first 500 fs of the perprotio dynamics simulation.[74]

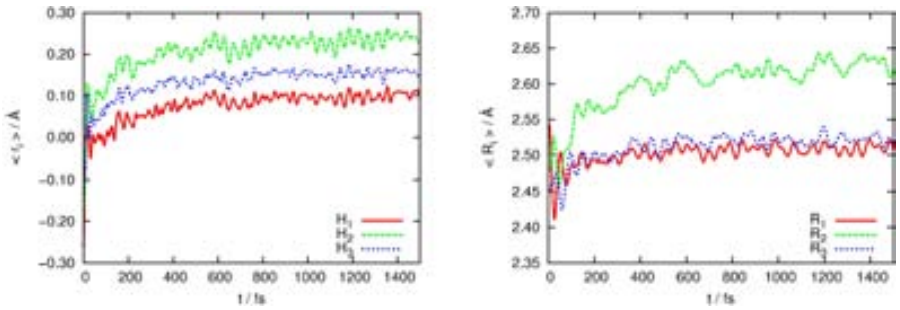


Figure 38: Expected values of the positions of the protons (left) and the distances of the consecutive oxygen atom pairs (right) for the first 1500 fs of the perprotio dynamics simulation.[74]

If we pay attention to longer simulation times (Figure 38) it can be seen that, despite the stabilization, the values of the coordinates still oscillate. These oscillations are probably a result of a closed dynamical model that does not take into account the chromophore nor the protein environment. As a result, the excess of energy coming from the photoexcitation cannot be re-

distributed into other coupled degrees-of-freedom (altough in latter works by Vendrell *et al.*[150] this conjecture could not be proved) and cannot disrupt the proton wire either. The main consequence of it is that the wavepacket bounces after reaching the product's region, so there is a reflux, which accounts for the final oscillations. This hypothesis is reinforced when protons are substituted by deuterons. The reaction occurs slower with deuterons than with protons (see insets of Figure 37 and Figure 39), which is in agreement with the detected KIE.

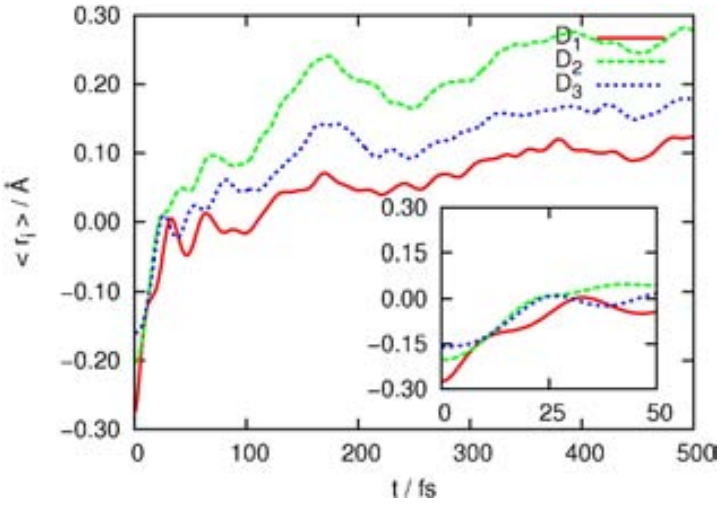


Figure 39: Expected values of the r_i coordinates for the first 50 fs of perdeutero simulation.[74]

The strong green fluorescence at 508-510 nm in GFP is originated at the ionized chromophore. Thus, the reactants and the products of the reaction that generates the fluorescent species in the excited state differ mainly in the position of the first proton in the wire. For this reason, the hypersurface that separates reactant from products has been defined as the ensemble of configurations in which the proton is at the midpoint between the chromophore oxygen and Wat25 oxygen, which in these internal coordinates is defined as $S \equiv r_1 = 0$. It must be borne in mind that, even though $r_1 = 0$ is not a closed surface, the dynamics simulation[74] was performed in a limited hypercube, that is, the internal coordinates had boundaries. This way, the wavepacket could not trespass the limits of this hypercube. The surface $r_1 = 0$ cuts the hypercube into two hypercubes: reactants and products, defined by closed surfaces. This will be important to make sure that the divergence theorem can be applied, as explained in Section 5.4.1.

Having univocally defined reactants and products, the population of each of them can be computed. From the Heaviside

step function,[2] a Population operator $\hat{P}$ can be defined such that $\hat{P} = 1 - \Theta(r_1)$. Applying this operator to the wavefunction leads $CXY = \langle\psi| 1 - \Theta(r_1) |\psi\rangle$,[3] which is the population of the wavepacket in the regions where $r_1 < 0$, that is, the reactants (the first hydrogen atom has not been transfered yet).

In Figure 40, the time evolution of the reactant's population CXY is shown. For short times (20 fs) the reaction is faster with protons than with deuterions, as mentioned. However, for longer times, the population of neutral chromophores is higher with protons than with deuterions. In other words, for longer times (from 300 fs) deuterions populate the product's region more than protons do. This behavior is counterintuitive since it could suggest that the chemical reaction is faster with deuterons than with protons. However, this is not so: the system is closed and the wavepacket bounces. It is due to its greater tunneling and greater ZPE in the products that the proton comes back to the reactants's zone more easily than deuterons do. The final oscillations in the coordinates and the refluxes are spurious consequences of the closed model. To correct this model and explore new strategies to let the system release the excess of energy is a challenge that lies beyond the scope of this thesis.

The results show that the time scale of the process's rate is the hundreds of femtoseconds. Experimental data suggest rates of several picoseconds. Having a barrierless path connecting reactants and products in the photoactive state is a plausible explanation to account for this discrepancy. In all, it is clear that the potential energy surface calculated by Vendrell *et al.* can be improved with the existing tools available now, such that the environment effects are included.

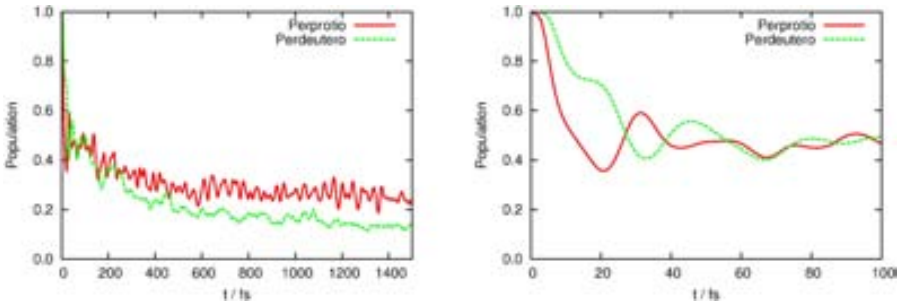


Figure 40: Protonated chromophore population (reactants) for perprotio (red) and perdeutero (green) simulations.[74]

2 The Heaviside step function Θ is defined as $\Theta(x) = 0$ when $x < 0$ and $\Theta(x) = 1$ when $x > 1$

3 CXY is the nomenclature adopted by Vendrell *et al.*[74]

5.2 APPROACHES TO COMPUTING THE TUNNELING CONTRIBUTION

A different, and rather complicated problem, lies at the question of *how much* tunneling is there, even if a full-blown QD simulation is available.

Stricto sensu, to define tunneling does not make much sense in quantum mechanics. Tunneling is usually thought of a correction to classical calculations: in classical mechanics, no system can have more potential energy than its total energy, but it is actually possible according to quantum mechanics. This implies that the physical meaning of tunneling is unclear without classical mechanics. Somehow, tunneling gives the idea of: how different the real behavior of a system is from the behavior predicted by a "wrong" theory. Nevertheless, the question of quantifying (how much) and qualifying (how and when) the tunneling is a tantalizing one that needed answer.

Even though tunneling is not a clearly defined observable, the total energy of the system and its potential energy are well defined in quantum mechanics. This way, one possible strategy to evaluate the tunneling contribution could consist of defining a Tunneling operator $\widehat{tun}$ from the Heaviside step function:

$$\widehat{tun} = \Theta(V - \langle \hat{H} \rangle) \tag{35}$$

This operator would furnish the probability (population) of finding the system in the classically non-permitted region, so this would yield the tunneling contribution in the reaction. We calculated the population in classically forbidden regions at different times. Pretty low values were obtained (see Figure 41): about 6% for hydrogen and 3.5% for deuterion (these data have not been published yet).

However, in terms of practical utility, this definition might not be the best one. The calculation of the population in a classically forbidden region gives a *static* perspective of the problem. That is to say, we can know at each time what percentage of population lies at given region, but these percentages do not indicate what is the *dynamical* variation of population at that region. Here comes the importance of the *probability* ux. Instead of identifying the fraction of the *population* at the classically forbidden region, we need to know the fraction of ux that goes through its surface. This might seem a subtle difference, but it is essential. To make this point clearer, let us think of a river. We can measure which regions are deeper than others, so the amount of water (wavefunction population) at each region can be known. But if one wants to know how much water flows at a

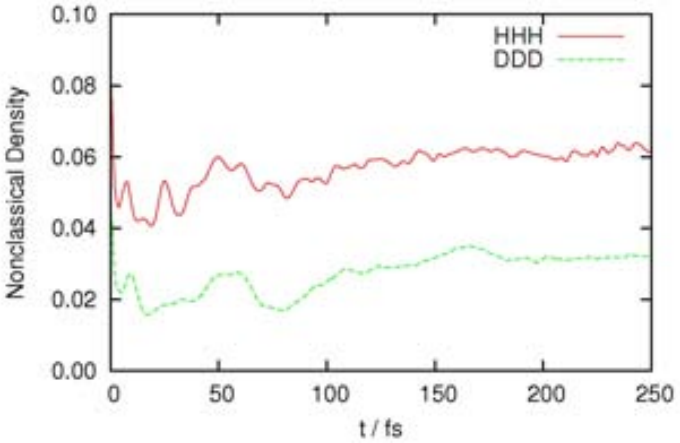


Figure 41: Population of the system in the classically forbidden region along the perprotio (red) and perdeutero (green) dynamics simulation.

given time and how, that is, through which areas there is more water flowing, it will be necessary to analyze the water current (probability current).

The probability flux can then be calculated as the **time variation of the population**. This time variation can be known from the wavefunction, but, calculated as a time-derivative, does not allow to discern the classical and non-classical flows. The probability flux, thus, has to be calculated as a surface integral of the *probability current* through the hypersurface that separates reactants from products ($r_1 = 0$, as mentioned).

This calculation, explained in the following sections, will provide an insight on the tunneling and will furnish important tools for the study of the reaction kinetics of GFP and, more generally, of any chemical reaction.

5.3 THE PROBABILITY FLUX

Quantum dynamics simulations are a very powerful tool to simulate chemically complex systems of small and medium size in which quantum effects are expected. The integration of the time-dependent Schrödinger equation (TDSE) for systems with several degrees of freedom and for times up to several picoseconds has become feasible by means of highly efficient algorithms like the Multi-Configurational Time-Dependent Hartree (MCTDH).[151–156]

Once the TDSE is solved, and the wavefunction is obtained, it might still be tricky to analize the results. Although it is well

known in quantum mechanics that the wavefunction *contains all the information about the system*, it is nonetheless true that if the wavefunction obtained from the TDSE depends on certain variables, to get the wanted information is not straightforward and can require a mathematical developement.

For the sake of generality, let us focus on a kind of problem that involves the heart of the chemistry, namely, the conversion of a chemical identity into another via a reaction. Assuming reaction like $A \rightleftharpoons B$, it may be interesting to compute the "amount" or "fraction" of each species at each time. In general, this can be computed easily with the expression

$$\frac{[A]}{[A]+[B]} = \langle \Psi | \, \Theta_A \, | \Psi \rangle \tag{36}$$

where Θ_A is a projection operator that returns the value 1 when in the areas of configurational space corresponding to species A, and 0 elsewhere. This expression can be generalized to more complex schemes, provided each chemical species is defined by a **region** of the configurational space. The *rate* at which a species appears or disappears is directly the time derivative of the right-hand side expression in Equation 36. It is possible, though, that the interest does not so much lie with the rate of formation of B but rather in knowing specifically over *which* regions of the division (hyper)surface between A and B the transformation occurs, and to what extent.

In this section, a thorough and novel mathematical developement is presented with the aim of calculating the quantum probability flux from reactants to products for generic coordinates, such that it is possible to select which part of the flux goes through which area. This procedure allows for discerning the flux through classically permitted areas and the so-called *tunneling* (see Section 5.1), or for distinguishing the regions across which the flux is positive or negative at a given time. Note that in the present work we distinguish probability current, which is defined in Equation 37,[157] from probability flux, being this latter the surface integral of the former one (in some references, both words are used as synonyms).

$$\vec{j} = \frac{-i\hbar}{2m}(\psi^* \vec{\nabla} \psi - \psi \vec{\nabla} \psi^*) \tag{37}$$

It must be borne in mind that the method to calculate the probability current that we are presenting is not a simple tool to get some meaningful information from a specific quantum dynamics simulation. So as to reach a mathematically consistent expression a long, complex and non-trivial procedure has been

followed. Besides, this development has been done without loss of generality, thus allowing for the probability current calculations from different quantum dynamics simulations in different sets of coordinates, which makes this work more valuable. Finally, so as to solve the flux integral, an adaptative Monte Carlo method has been proposed, with favorable scaling properties. This numerical approach has required the use of probability densities (in what is known as *importance sampling*), which has turned to be not only a tool to make the integration more efficient, but also a magnitude with a very useful physical and chemical meaning. For this reason, this development *per se* is a fundamental part of the research.

5.4 MATHEMATICAL DEVELOPEMENT

As it has already been stressed, we are interested in calculating the flux through a hypersurface. The mentioned electronic calculations (ground and excited states) carried out by Vendrell *et al.*[70] can be interpolated, so the potential energy surface can be obtained. Besides, knowing the wavefunction, it is possible to find the total energy at a given point. This way, it is possible to distinguish classically permitted and non-permitted regions. What will be shown in this section is the mathematical procedure that we have developed to get a useful expression of the probability flux.

5.4.1 *Physical background*

To begin with, let us remind, or rediscover, what the probability current is, and how we can transform and relate it to other known magnitudes that might be of interest for our problem.

Let us start from the Schrödinger equation for a system with a particle of mass m:

$$i\hbar\frac{\partial\psi(\vec{r},t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi(\vec{r},t) + \hat{V}\psi(\vec{r},t) \tag{38}$$

From now on, let us refer to $\psi(\vec{r},t)$ as ψ. If we left multiply Equation 38 by ψ^* we obtain

$$i\hbar\psi^*\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\psi^*\nabla^2\psi + \psi^*\hat{V}\psi \tag{39}$$

Analogously, we can take the complex conjugate of Equation 38 and left multiply by ψ:

$$i\hbar\psi\frac{\partial\psi^*}{\partial t} = -\frac{\hbar^2}{2m}\psi\nabla^2\psi^* + \psi\hat{V}\psi^* \tag{40}$$

By substracting Equation 40 from Equation 39 the following expression is obtained:

$$i\hbar(\psi^*\frac{\partial\psi}{\partial t}+\psi\frac{\partial\psi^*}{\partial t})=\frac{-\hbar^2}{2m}(\psi^*\nabla^2\psi-\psi\nabla^2\psi^*) \tag{41}$$

It is well know that $|\psi|^2=\psi\psi^*$. This way,

$$\frac{\partial}{\partial t}(\psi\psi^*)=\psi^*\frac{\partial\psi}{\partial t}+\psi\frac{\partial\psi^*}{\partial t} \tag{42}$$

The gradient operator must be applied as a derivative. It can be easily shown that

$$\vec{\nabla}(\psi^*\vec{\nabla}\psi-\psi\vec{\nabla}\psi^*)=\psi^*\nabla^2\psi-\psi\nabla^2\psi^* \tag{43}$$

Introducing Equation 42 and Equation 43 into Equation 41, and dividing by $i\hbar$ yields

$$\frac{\partial|\psi|^2}{\partial t}=-\vec{\nabla}\left(\frac{-i\hbar}{2m}(\psi^*\vec{\nabla}\psi-\psi\vec{\nabla}\psi^*)\right) \tag{44}$$

If the probability current (see Equation 37) is taken into account, the following expression can be obtained

$$\frac{\partial|\psi|^2}{\partial t}+\vec{\nabla}\vec{j}=0 \tag{45}$$

Equation 45 is a kind of continuity equation: at a given spot, the time variation of the probability density equals de probability current exiting from it: in other words, the part of the wavepacket that leaves a region becomes a current that escapes from that region. When, instead of probability density ($|\psi|^2$) we deal with charge density or mass density, analogous continuity equations are obtained: the subjacent idea of these continuity equations is that some physical properties, like mass or charge, cannot disappear, so when they decrease in a region, there must be a flux outwards, and viceversa.

When dealing with a whole region of space, we can integrate for that volume

$$\int|\psi|^2dV=-\int\vec{\nabla}\vec{j}dV=\Phi \tag{46}$$

Where Φ is the *probability ux*. The Gauss-Ostrogradsky[158] or divergence theorem guarantees that, in orthogonal coordinates, the volume integral of a divergence of a vector field $\vec{j}$ ($\vec{\nabla}\vec{j}$) equals the integral of this field across the surfaces that encloses that volume, as written in Equation 47

$$\int\vec{\nabla}\vec{j}dV=\oint\vec{j}d\vec{S} \tag{47}$$

Equation 46 shows how we will tackle the issue of the flux integral. From the results of the quantum dynamics simulation[74] we can know the probability density at different regions of hyperspace and different times, so the total flux (left part of Equation 46) can be easily calculated. But what we want to know is how much of the flux Φ (that is, the wavepacket from reactants to products) goes through classically permitted areas, and how much does not. If the flux is calculated as a surface integral, we can know, spot by spot, if the potential energy is higher or lower than the total energy, that is, if the flux is crossing a classically permitted area or not, as shown in Equation 48. To find a way to obtain this is the main purpose of this chapter, and it will be seen that the achievemnt of such a procedure leads to a much more solid understanding of the isotopic effects of chemical reactions.

$$\Phi = \Phi_{classic} + \Phi_{tunnel} = -\oint \Theta(E-V)\vec{j}d\vec{S} - \oint \Theta(V-E)\vec{j}d\vec{S} \quad (48)$$

5.4.2 *Obtaining a convenient expression for the flux*

In this subsection, the well-known equations shown in Section 5.4.1 are the starting point of a complex mathematical transformation that will provide a suitable expression for the calculation of the flux. Through all this, a deep and thorough mathematical discussion is required.

To begin with, there are some problems that need to be considered. Firstly, if we laid out the problem in Cartesian coordinates, the divergence theorem would be valid since they are orthogonal: each coordinate defines the position of only one particle, so these can vary independently. Nevertheless, the wavefunction is known in internal coordinates, which are not orthogonal. Secondly, the probability current is defined in Equation 37 for an ideal case of a one-mass system: the model adopted to describe the GFP simulation by Vendrell *et al.* has seven masses (particles). Let us explain how these issues have been circumvected.

It should be recalled, first of all, that the time variation of the population is the same regardless of the coordinate system used:

$$-\frac{\partial}{\partial t}\int |\psi|^2 dV = \int \vec{\nabla}^\alpha \vec{j}^\alpha dV^\alpha \quad (49)$$

Being α any coordinate system. Whatever α is, the surface is the same (and, in our case, it is closed). Let us express Equation 49 in *mass-weighted* Cartesian coordinates, that are defined as

$$x_i^c = m_i^{-1/2} x_i^w \tag{50}$$

Where c stands for Cartesian, and w stands for *mass-weighted* Cartesian. Since mass-weighted Cartesian coordiantes are orthogonal

$$\int \vec{\nabla}^w \vec{j}^w dV^w = \oint \vec{j}^w d\vec{S}^w \tag{51}$$

This way, the problem of dealing with Equation 37, in which the current depends on the mass, is solved: now the masses are implicit in the coordinates. There will be no need, then, to define a reduced mass. Mass-weighted Cartesian coordinates allow for a definition of probability current and obey the Gauss-Ostrogradsky theorem. Now, instead of trying to figure out under which circumstances can the divergence theorem be applied, our strategy consists of performing the appropriate transformations on Equation 51.

The probability current, in mass-weighted Cartesian coordinates, is expressed as

$$\vec{j}^w = \frac{1}{2}(\psi^{*w}\vec{\nabla}\psi^w - \psi^w\vec{\nabla}\psi^{*w}) \tag{52}$$

Now, it is necessary to transform the wavefunction, its gradient and the volume element from internal coordinates to mass-weighted Cartesian. To carry out such a transformation, the Jacobian matrix will be required. The Jacobian matrix $\mathbf{J}$ is defined as $\mathbf{J} = \frac{\partial \vec{x}^{int}}{\partial \vec{x}^w}$, such that $\vec{x}^{int} = \mathbf{J}\vec{x}^w$. The internal coordinates are ordered such that

$$(r_1, R_1, r_2, R_2, r_3, R_3, R_{cm}) = (x_1^{int}, x_2^{int}, x_3^{int}, x_4^{int}, x_5^{int}, x_6^{int}, x_7^{int}) \tag{53}$$

The Jacobian matrix is defined as

$$\mathbf{J} = \frac{\partial \vec{x}^{int}}{\partial \vec{x}^w} = \begin{pmatrix} \frac{\partial x_1^{int}}{\partial x_1^w} & \frac{\partial x_2^{int}}{\partial x_1^w} & \cdots & \frac{\partial x_7^{int}}{\partial x_1^w} \\ \frac{\partial x_1^{int}}{\partial x_2^w} & \frac{\partial x_2^{int}}{\partial x_2^w} & \cdots & \frac{\partial x_7^{int}}{\partial x_2^w} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1^{int}}{\partial x_7^w} & \frac{\partial x_2^{int}}{\partial x_7^w} & \cdots & \frac{\partial x_7^{int}}{\partial x_7^w} \end{pmatrix} \tag{54}$$

That is:

$$\mathbf{J} = \begin{pmatrix} -\frac{1}{2}m_O^{-1/2} & -m_O^{-1/2} & 0 & 0 & 0 & 0 & \frac{m_O^{1/2}}{M} \\ m_{H_1}^{-1/2} & 0 & 0 & 0 & 0 & 0 & \frac{m_{H_1}^{1/2}}{M} \\ -\frac{1}{2}m_O^{-1/2} & m_O^{-1/2} & -\frac{1}{2}m_O^{-1/2} & -m_O^{-1/2} & 0 & 0 & \frac{m_O^{1/2}}{M} \\ 0 & 0 & m_{H_2}^{-1/2} & 0 & 0 & 0 & \frac{m_{H_2}^{1/2}}{M} \\ 0 & 0 & -\frac{1}{2}m_O^{-1/2} & m_O^{-1/2} & -\frac{1}{2}m_O^{-1/2} & -m_O^{-1/2} & \frac{m_O^{1/2}}{M} \\ 0 & 0 & 0 & 0 & m_{H_3}^{1/2} & 0 & \frac{m_{H_3}^{1/2}}{M} \\ 0 & 0 & 0 & 0 & -\frac{1}{2}m_O^{-1/2} & m_O^{-1/2} & \frac{m_O^{1/2}}{M} \end{pmatrix} \tag{55}$$

Let us continue changing the wavefunctions in Equation 52. The value of the wavefunction depends on the coordinate system used (since its square value is a density): it is such that its integral value over all space is normalized. Besides, the volume elements are related through the determinant of the Jacobian matrix[158]. In this case, the determinant of the Jacobian matrix defined in Equation 55 is *a*) a finite number that does not depend on the integration limits and *b*) different from zero. Now, the normalization condition yields

$$1 = \int |\psi^{int}|^2 dV^{int} = \int |\psi^{int}|^2 \det \mathbf{J} dV^w = \int |\psi^w|^2 dV^w \tag{56}$$

Which means that the expressions of the wavefunction, being either in one or the other coordinate systems, will be

$$|\psi^w|^2 = |\psi^{int}|^2 \det \mathbf{J} \rightarrow \psi^w = e^{i\alpha} (\det \mathbf{J})^{1/2} \psi^{int} \tag{57}$$

Where the phase α is a real number, that will be taken 0 for the sake of simplicity. This decision does not affect the physical meaning of the magnitudes since the properties of the wavefunction belonging to the same unitary ray are identical. Now, after having transformed the wavefunction, let us transform the gradients that appear in Equation 52

$$\vec{\nabla}\psi^{w} = \frac{\partial\psi^{w}}{\partial\vec{x}^{w}} = (\det\mathbf{J})^{1/2}\frac{\partial\psi^{int}}{\partial\vec{x}^{int}} \tag{58}$$

Looking at the Jacobian matrix $\mathbf{J}$ defined in Equation 55, it can be shown that

$$\left(\frac{\partial\psi^{int}}{\partial\vec{x}^{w}}\right) = \mathbf{J}\left(\frac{\partial\psi^{int}}{\partial\vec{x}^{int}}\right) \tag{59}$$

Equation 58 and Equation 59 yield

$$\vec{\nabla}\psi^{w} = (\det\mathbf{J})^{1/2}\mathbf{J}\left(\frac{\partial\psi^{int}}{\partial\vec{x}^{int}}\right) \tag{60}$$

Introducing Equation 57 and Equation 60 into Equation 52 we obtain

$$\vec{j}^{w} = \det\mathbf{J}\frac{\hbar}{2i}\mathbf{J}(\psi^{*int}\widetilde{\nabla}\psi^{int} - \psi^{int}\widetilde{\nabla}\psi^{*int}) \tag{61}$$

Where $\widetilde{\nabla}\psi^{int} = \left(\frac{\partial\psi^{int}}{\partial\vec{x}^{int}}\right)$ is not *stricto sensu* a gradient since the internal coordinates are neither orthogonal nor rectangular, so $\widetilde{\nabla}\psi^{int}$ does not point the direction of maximum growth of the function.

Well, with these transformations, we can know the probability current in mass-weighted Cartesian coordinates: we know the value of the wave function and its gradient in internal coordinates and we know the Jacobian matrix. As for the surface differential, it can be thought of as a volume differential multiplied by a Dirac delta function of the subspace that defines our surface, in this case, $r_1 = 0$. The flux can now be written as

$$\Phi = -\int(\vec{j}^{w}\cdot\vec{n}^{w})dV^{w}\delta(r_1)dV^{w} = \int(\vec{j}^{w}\cdot\vec{n}^{w})\det\left(\frac{\partial\vec{x}^{w}}{\partial\vec{x}^{int}}\right)dV^{int} \tag{62}$$

Considering that

$$\det\left(\frac{\partial\vec{x}^{w}}{\partial\vec{x}^{int}}\right) = \left[\det\left(\frac{\partial\vec{x}^{w}}{\partial\vec{x}^{int}}\right)\right]^{-1} = [\det\mathbf{J}]^{1} \tag{63}$$

and introducing Equation 61 into Equation 62 the following expression is obtained:

$$\Phi = -\int dV^{int}\delta(r_1)\frac{\hbar}{2i}\vec{n}^{w}\mathbf{J}(\psi^{*int}\widetilde{\nabla}\psi^{int} - \psi^{int}\widetilde{\nabla}\psi^{*int}) \tag{64}$$

The limits of the integral in Equation 64 are those of the quantum dynamics simulation. The determinants of the Jacobian matrix have been simplified.

As for the value of the normal vector in mass-weighted Cartesian coordinates, $\vec{n}^w$, it is easy to find. If the surface to be parameterized is $r_1 = 0$:

$$S \equiv r_1 = 0 \rightarrow r_1 = m_{H_1}^{-1/2} x_2^w - \frac{1}{2} m_O^{-1/2} (x_3^w + x_1^w) = 0 \quad (65)$$

Since $\vec{n}^w = \vec{\nabla} S$

$$\vec{n}^w = (-\frac{1}{2} m_O^{-1/2}, m_{H_1}^{-1/2}, -\frac{1}{2} m_O^{-1/2}, 0, 0, 0, 0) \quad (66)$$

This vector will have to be normalized in mass-weighted Cartesian metrics.

5.4.3 *How to deal with the center-of-mass and the surface normal vector*

Once we have reached this point, we have to deal with another issue: the treatment of the center-of-mass. This is not trivial, since the quantum dynamics simulation was solved in a six-dimensional space, explicitly excluding the center-of-mass. But, when integrating over this coordinate, we do not know the integral limits.

First of all, it will be assumed that the wavefunction in internal coordinates depends on the center-of-mass as a planar wave (see Equation 67), which is an eigenfunction of the linear momentum operator P. There is no energy exchange between the center-of-mass and the other degrees of freedom, so the wavefunction is separable (see Equation 68).

$$\chi(R_{cm}) = e^{\frac{iP_{cm}R_{cm}}{\hbar}} \quad (67)$$

$$\psi^{int}(\mathbf{x}^{int}) = \Psi(r_1, R_1, r_2, R_2, r_3, R_3)\chi(R_{cm}) \quad (68)$$

Ψ is the wavefunction that we know numerically, obtained from the previous quantum dynamics simulation. This way, Equation 64 can be written as

$$\Phi = -\int \frac{\hbar}{2i} \delta(r_1) dV^{int} \vec{n}^w \mathbf{J} \left((\Psi\chi)^* \widetilde{\nabla}(\Psi\chi) - (\Psi\chi) \widetilde{\nabla}(\Psi\chi)^* \right) \quad (69)$$

Now, let us split the pseudo gradient vector $\widetilde{\nabla}$ into two parts: $\widetilde{\nabla} = (\widetilde{\nabla}^{rel} | \widetilde{\nabla}^{cm})$, where $\widetilde{\nabla}^{rel}$ is the derivative respect to the relative coordinates, that is, the internal coordinates without the center-of-mass (it has 6 dimensions). $\widetilde{\nabla}^{cm}$ is the derivative respect to the center-of-mass, and has got one dimension. Considering this distinction, the Jacobian matrix can be written as follows:

$$\mathbf{J} = \begin{pmatrix} \frac{\partial x_1^{int}}{\partial x_1^w} & \frac{\partial x_2^{int}}{\partial x_1^w} & \cdots | & \frac{\partial x_7^{int}}{\partial x_1^w} \\ \frac{\partial x_1^{int}}{\partial x_2^w} & \frac{\partial x_2^{int}}{\partial x_2^w} & \cdots | & \frac{\partial x_7^{int}}{\partial x_2^w} \\ \vdots & \vdots & \ddots | & \vdots \\ \frac{\partial x_1^{int}}{\partial x_7^w} & \frac{\partial x_2^{int}}{\partial x_7^w} & \cdots | & \frac{\partial x_7^{int}}{\partial x_7^w} \end{pmatrix} \equiv \left(\mathbf{J}^{rel} | \mathbf{J}^{cm}\right) \tag{70}$$

Where $\mathbf{J}^{rel}$ is a 7×6 matrix: it derives the relative coordinates (6 in total) with respect to the mass-weighted Cartesian coordinates (7 in total). On the other hand, $\mathbf{J}^{cm}$ is a 7×1 matrix: it derives the center-of-mass respect to each mass-weighted Cartesian coordinate. This way, $\vec{j}^w = \mathbf{J} \cdot \vec{j}^{int}$ can be expressed as

$$\vec{j}^w = \left(\mathbf{J}^{rel} | \mathbf{J}^{int}\right) \begin{pmatrix} \chi\chi^* \left(\psi^* \widetilde{\nabla}^{rel} \psi - \psi \widetilde{\nabla}^{rel} \psi^*\right) \\ \hline \psi\psi^* \left(\chi^* \widetilde{\nabla}^{cm} \chi - \chi \widetilde{\nabla}^{cm} \chi^*\right) \end{pmatrix} \tag{71}$$

$\vec{j}^w$ is a column vector of 7 rows. This vector corresponds, in Equation 64, to the expression $\frac{\hbar}{2i}\mathbf{J}\left(\psi^{*int}\widetilde{\nabla}\psi^{int} - \psi^{int}\widetilde{\nabla}\psi^{*int}\right)$, except for the determinant of the Jacobian matrix, which eventually vanishes with the volume differential. Each k component of this vector $\vec{j}^w$ can be written as:

$$\begin{aligned} j_k^w = \left(\frac{\hbar}{2i}\right) \Bigg(\chi\chi^* \sum_{i=1}^{6} \frac{\partial x_i^{int}}{\partial x_k^w} \left(\psi^* \frac{\partial \psi}{\partial x_i^{int}} - \psi \frac{\partial \psi^*}{\partial x_i^{int}}\right) \\ + \psi\psi^* \frac{\partial R_{cm}}{\partial x_k^w} \left(\chi^* \frac{\partial \chi}{\partial R_{cm}} - \chi \frac{\partial \chi^*}{\partial R_{cm}}\right) \Bigg) \\ = j_k^{rel} + j_k^{cm} \end{aligned} \tag{72}$$

Each k component from vector $\vec{j}^w$ is the sum of seven contributions: the first six correspond to the relative coordinates, and the last one corresponds to the center-of-mass. Therefore, $\vec{j}^w$ can be thought of as the sum of two vectors: one with contributions in the first six coordinates ($\vec{j}^{rel}$) and another one with only one contribution in the last coordinate ($\vec{j}^{cm}$). Obviously, both vectors have seven components.

Now, Equation 69 can be written:

$$\Phi = -\int \frac{\hbar}{2i} \delta(r_1) dV^{int} \left(\vec{n}^w \cdot (\vec{j}^{rel} + \vec{j}^{cm})\right) = \Phi^{rel} + \Phi^{cm} \tag{73}$$

where Φ^{cm} is irrelevant. To prove such a statement is beyond the scope of the present thesis, but we can assert that it would

not make sense if the total flux across a internal hypersurface were affected by the degree of freedom of the center-of-mass. Otherwise, the separability of the center-of-mass would be violated. For this reason, we assume that $\Phi^{cm} = 0$. Now, the the integral that we want to calculate is

$$\Phi = -\int \delta(r_1) dV^{int} \left(\vec{n}^w \cdot \vec{j}^{rel} \right) = \int \chi\chi^* dR_{cm} \int \delta(r_1) dr_1 dR_1 dr_2 dR_2 dr_3 dR_3 \left(\vec{n}^w \cdot \vec{j}' \right) \tag{74}$$

Where $\vec{j}'$ is

$$\vec{j}' = \frac{\vec{j}^{rel}}{\chi\chi^*} = \frac{\hbar}{2i} \sum_{i=1}^{6} \frac{\partial \vec{x}^{int}}{\partial \vec{x}^w} \left(\Psi^* \frac{\partial \Psi}{\partial x_i^{int}} - \Psi \frac{\partial \Psi^*}{\partial x_i^{int}} \right) \tag{75}$$

So as to deal with Equation 74 some remarks must be done. The Dirac delta function has the property that

$$\int f(x_1, x_2 ... x_n) \delta(x_1) dx_1 dx_2 ... dx_n = \int f(0, x_2 ... x_n) dx_2 dx_n \tag{76}$$

To include the Dirac delta in the integrand has allowed to treat the surface integral as a volume integral, with which useful variable changes have been done. The Dirac delta function in Equation 74 implies that, from all the hypervolume, only those points in which $r_1 = 0$ will contribute to the integral. In other words, it converts a volume integral into a surface integral. In fact, *stricto sensu*, the Dirac delta is not a function, but a distribution. We can think of the Dirac delta as a distribution that consists of the limit of the Gaussian function in which the variance tends to zero: all the probability is concentrated in $r_1 = 0$. Dirac delta is normalized. As for R_{cm}, we do not care about it since $\int \chi(R_{cm}) \chi^*(R_{cm}) dR_{cm} = 1$.

There is another issue to be considered in Equation 74. The dot product $\vec{n}^w \cdot \vec{j}'$ is, as all dot products, a bilinear form. In orthonormal coordinates, the matrix associated to the dot product is the identity: this is assumed to such an extent that it is often taken for granted. The mass-weighted Cartesian coordinates are orthogonal, but not orthonormal: not all the directions are equivalent, since they are weighted by different masses. Moving along a given direction is not equivalent as moving along another. The matrix of the dot product in this metric is therefore orthogonal (diagonal) but different from the identity. First of all, the normal vector has to be normalized so that $\vec{n}^w \cdot \vec{n}^w = 1$, and afterwards, we will have to proceed as follows:

$$\vec{n}^w \cdot \vec{j}' = \begin{pmatrix} n_1^w & \cdots & n_7^w \end{pmatrix} \begin{pmatrix} m_O^{-1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & m_{H_1}^{-1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & m_O^{-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & m_{H_2}^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & m_O^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & m_{H_3}^{-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & m_O^{-1} \end{pmatrix} \begin{pmatrix} j_1' \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ j_7' \end{pmatrix} \tag{77}$$

Finally, the expression for the tunneling flux is

$$\Phi_{tunnel} = -\int dR_1 dr_2 dR_2 dr_3 dR_3 \left(\vec{n}^w \cdot \vec{j}'\right) \Theta(V - \langle \hat{H} \rangle) \tag{78}$$

Once we have an appropriate expression for the flux, a remark needs to be made. As it has already been stressed, the value of $\Phi = \frac{\partial}{\partial t} \int |\psi|^2 dV$ can be known –albeit approximately, since the derivative is not analytical, but obtained with finite differences. With this, it has been possible to verify whether the calculation of Φ is correct and to corroborate the adequacy of the mathematical developement hereby explained.

5.5 NUMERICAL PROCEDURES

All the analytical expressions obtained in Section 5.4 would be a simple curiosity if we were not able to solve the problem numerically. This implies a proper choice of numerical methods to calculate the integral, and implement them in an efficient way.

We have to bear in mind that the evaluation of the function $\vec{n}^w \cdot \vec{j}'$ is computationally demanding. Each point required a computational time of approximately 1.6 seconds.[4] A rectangle-like integration method would take too long. If, for instance, each degree of freedom were divided into 15 rectangles, this would imply the evaluation of $15^5 = 759735$ points, which would take 14 days of calculation. This time would be affordable, but we would not be sure if the integration step is correct or not. We can suspect that at the beginning of the dynamics there is not much flux and the function is not likely to have sharp changes, but in longer times of the simulation the wavepacket will bounce and therefore become much more structured. In these cases, a thinner step would be advisable.

4 With a Xeon processor of 8 cores/node. Intel, 64 bits (em64t)

However, with only 30 divisions per degree of freedom, the calculation time would be 280 days, and still we would probably not be sure if the integral has converged or not. In a similar way, a numerical quadrature method would not be useful either since a very structured function can require a great number of points, and it would take even longer to make sure that the calculation has converged.

Considering all this, a Monte Carlo method presents several advantages. By generating random points, after a certain time most part of the hypersurface is covered uniformly. After a given number of iterations, it is possible to estimate whether it has converged or not, without having to decide an integration step from the beginning. Besides, it is possible to follow statistical criteria to establish error estimations from the mean results and variances, so the confidence level of the results can be known. Another important advantatge of the Monte Carlo method is that, as each evaluation of the integrand does not require information of the previous step, it is possible to parallelize, thus leading to faster calculations.

We made the first attempts to calculate the integral by means of a random sampling in which all the points were equally probable (uniform distribution). It could be seen that in the majority of points the integrand was almost zero, and only in a certain region the contributions were actually relevant. To make the calculations more efficient, an approach based on importance sampling has been chosen in order to maximize the number of samples in which the integrand is in highly contributing regions of variable space. To this end, a version of Peter Lepage's VEGAS algorithm[138] has been implemented (adaptative Monte-Carlo).

We assume that the probability density function can be expressed as a product of one-dimension density functions, such that

$$\rho(x_1, x_2 ... x_n) = \rho(x_1)\rho(x_2)...\rho(x_n) \tag{79}$$

This implies assuming that the different degrees of freedom are not correlated. Initially, we start with a uniform probability density. The intervals that form the integration limits are divided into bins. The value of the probability density at each bin is a single number. When sampling and calculating the integral, not only the integral value is calculated, but also the accumulative absolute value of the function at each interval.

The calculation of the integral is done in steps, or generations: at each generation, the integral is (re)calculated, but also a density probability is found. With this method, not only the value

of the integral is more accurate at every generation, but also the density probability is improved. This way, if the integrand is concentrated at a given region, the probability density ρ will be bigger at that region and will decrease at those regions with neglegible contributions. Once a generation is completed, the density used for that generation and the calculated density are combined such that $\rho_{i+1}^{new} = (1-\alpha)\rho_i^{old} + \alpha\rho_i^{new}$. After several tests, it was seen that $\alpha = 0.1$ was most convenient for our purpose. A higher value would imply a risk of reducing too quickly the samples at areas in which there could be important contributions still not visited. A lower value would provide more reliable results but would make the convergence much slower. In the limit, $\alpha = 0$ would imply no importance sampling at all.

At every i generation a value of the integral I_i is calculated, as well as the standard deviation σ_i. After m generations, the final value of the integral will be[138]

$$I_{best} = \frac{\sum_{i=1}^{m} \frac{I_i}{\sigma_i^2}}{\sum_{i=1}^{m} \frac{1}{\sigma_i^2}} \tag{80}$$

To proceed with all these calculations, a software has been programmed. The software code is given in Appendix B. In Figure 42 its flux diagram is shown. For every point sampled (that is, for every ensemble of 6 numbers: the 0 for r_1 plus 5 more for the other coordinates) the integrand is evaluated. From the different evaluations a value of the integral is obtained, and so is the standard deviation. It was decided to set the number of samples per generation to 7500. Each degree of freedom was divided into 150 intervals. This means that each individual probability density function has 150 bins. When the generation is finished, the different contributions (in absolute value) of the integral according to the region and the degree of freedom are stored in a matrix. Once these values are normalized, this density and the former one are combined to obtain a new density for the next generation. After an infinite number of generations, the probability density would coincide with the absolute value of the integrand.

The generations with the best probability density (reasonably the latter ones) will have smaller standard deviations, so their weight in the final result will be greater. The confidence level of Equation 80 is calculated as[138]

$$\sigma_{best} = \left(\sum_{i=1}^{m} \frac{1}{\sigma_i^2} \right)^{-1/2} \tag{81}$$

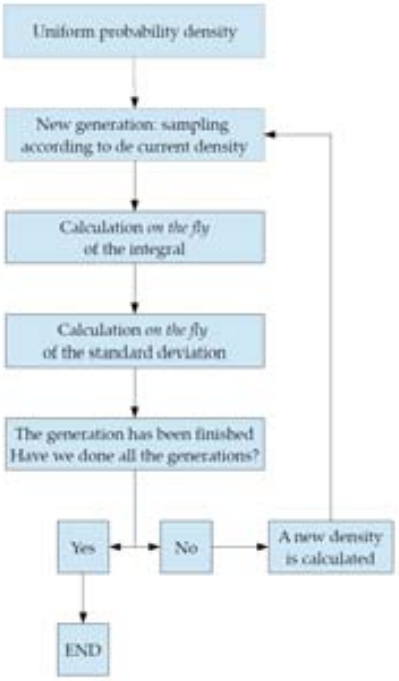


Figure 42: Flux diagram of the program implemented for the calculation of the flux

5.6 RESULTS OF THE ANALYSIS

Let us begin the discussion of the results by comparing to reference results. The total flux can also be calculated as the time derivative of the survival density (see Equation 46). From the dynamics simulation with MCTDH, the density at the region $r_1 = 0$ has been obtained and numerically derived respect to the time. Figure 43 compares both results: the flux calculated as the time derivative of a population, and as a surface integral.

In general terms there is good agreement between both results, which in practice validates the mathematics behind our approach and the implementation. A certain note of caution needs to be issued here, which is that the results coming from the numerical calculations obtained from the integration of the probability current integration are slightly smaller in absolute value than their reference values. To consider this effect a certain number of factors need to be taken into account: 1)reference values come from numerical derivatives of data output in this case every 1 fs, and 2) the MCTDH wavefunction in use (Ψ^{int}) is a numerical interpolation since in the MCTDH algorithm the wavefunction is represented on a finite primitive grid

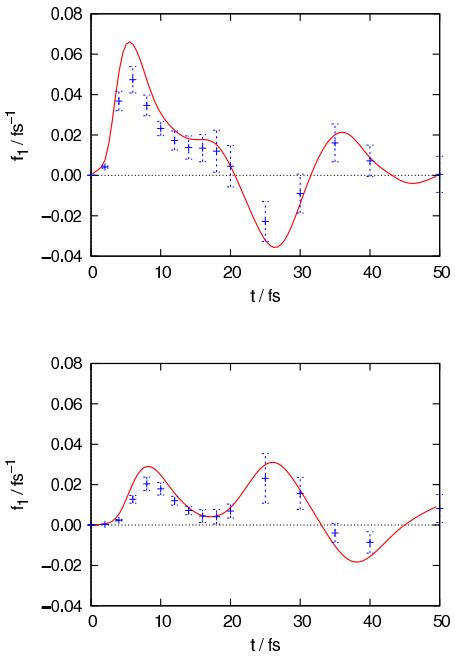


Figure 43: Total flux across the surface defined as $r_1 = 0$, for the perprotio (top) and perdeutero(bottom) isotopologues of the GFP simulation. In red, reference values computed as the time derivative. In blue, calculated fluxes as surface integrals. Vertical blue lines denote statistical uncertainty equivalent to one standard deviation.

of points. Nevertheless, we believe that a likely cause of discrepancy in the absolute values of the reference and computed flux arises from the actual execution of the adaptative Monte Carlo scheme used. Firstly, if the value of α were too large or the number of samples per generation were too scarce, the sampling densities could be slightly biased. Secondly, it is also possible that the variables are correlated. But if they are so to some extent, in theory the adaptative method used should not be more efficient than a straightforward Monte Carlo scheme.[138] In any case, the discrepancy is small and the achieved result suffices to illustrate that the method works. Therefore, a meaningful and detailed study of the tunneling and recrossing contributions is possible.

By means of "filtering" the flux with the step function $\Theta\left(V - \langle\hat{H}\rangle\right)$, the tunneling flux through $r_1 = 0$ has been calculated, as shown in Figure 44. Hydrogen and deuterium tunneling flux have a maximum at about 6 and 8 fs, respectively, which coincide with a maximum value of total flux. At these times, the tunneling contribution constitutes a 15.9% of the total flux for the perprotio and 29.4% for the perdeutero simulations. It could seem odd that that the perdeutero flux has a bigger percentage of

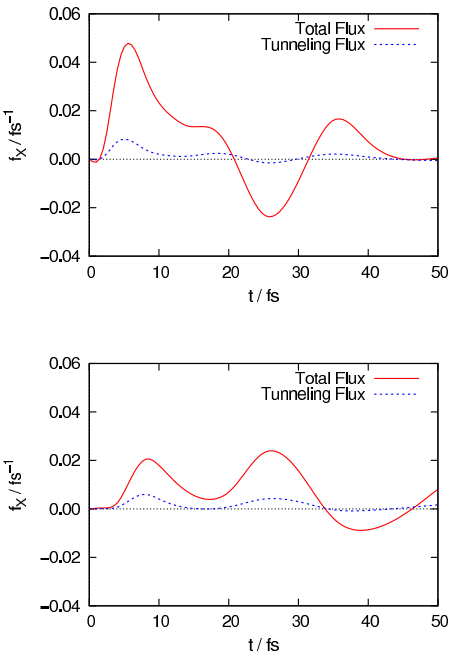


Figure 44: Total (red) and nonclassical (tunneling, blue) fluxes across the surface defined as $r_1 = 0$, for the perprotio (top) and perdeutero (bottom) of the GFP simulation.

tunneling contribution on the total flux, but in absolute terms the reactive tunneling flux in the perprotio simulation is $7.53 \cdot 10^{-3}$ fs^{-1}, larger than the perdeutero value, $5.99 \cdot 10^{-3}$ fs^{-1}. The fact that the ZPE is higher in the perprotio simulation accounts for its higher values of the flux: the perdeutero system runs into more difficulties in the transit to products, that is why in the perdeutero system tunneling plays a more outstanding role in relative terms.

Since the calculation of the surface integral has required the determination of the probability density for sampling, it is possible to take advantage of this information to understand more deeply what happpens in the chemical reaction. After all, not only is ρ a tool to achieve a more efficient sampling, but it is a *direct* indicator of what geometries are more responsible for certain flux contributions.

For instance, at the very beginning of the perprotio simulation, the main contributions are expected to come from heavy atom arrangements resembling those of the expected values of the different coordinates in the electronic ground state ($\langle R_1 \rangle$=2.55 Å, $\langle R_2 \rangle$=2.48 Å, $\langle R_3 \rangle$=2.45 Å),[74] The density probability of coordinates R_1, R_2 and R_3, shown in Figure 45, properly reproduces these values. After such a short time after initiation one would expect that the other hydrogen atoms would still be in the reactants' area (that is, negative), and this is seen in Fig-

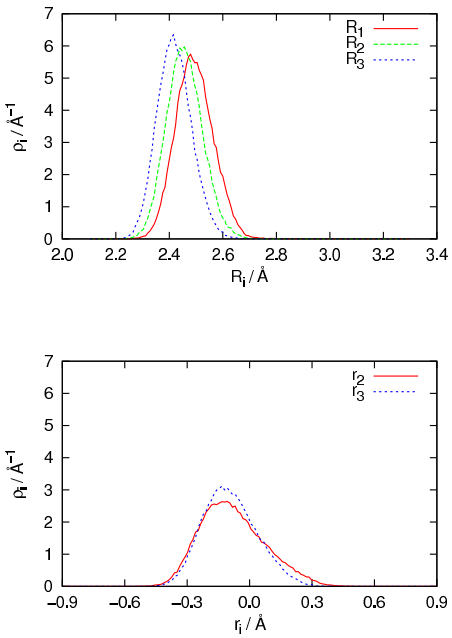


Figure 45: One-dimensional sampling densities (Equation 79 for the perprotio GFP simulation 6 fs after photoexcitation. Top: sampling densities for the heavy degrees of freedom (red: R_1, green: R_2, blue: R_3). Bottom: sampling densities for the light degrees of freedom (red: r_2, blue: r_3). All sampling densities are normalized to unit area.

ure 45 bottom, which shows a spreading tail toward positive values.

As for the perprotio simulation at 30 fs, when the flux is negative (see Figure 43), the order of values has changed: now R_1 is much shorter than R_3, being R_2 the largest. Light atoms have shifted towards positive values: r_3 is almost evenly centered at 0 and r_2 mostly displaced toward positive coordinate values with some contribution at negative coordinate values. This can be seen in Figure 46. At this point, one can pose the the question of how can the system contribute to reactive flux simultaneously from configurations where the second proton is already well on the product side or on the reactant side. This can be tackled by separating ρ_i into ρ_i^+ and ρ_i^-, which are computed as described for the total density, but discerning whether the evaluation of the sample has led to a positive or a negative value of the flux. These densities are not used for any sampling ends. Figure 47 shows these for r_2, $\rho_{r_2}^+$ and $\rho_{r_2}^-$ at 30 fs for the perprotio simulation and is complementary to Figure 46 (note that the densities are all normalized). It seems quite clear that the global reactive flux can be thought of as coming from two contributions: one centered at ~ 0.25 Å which yields negative

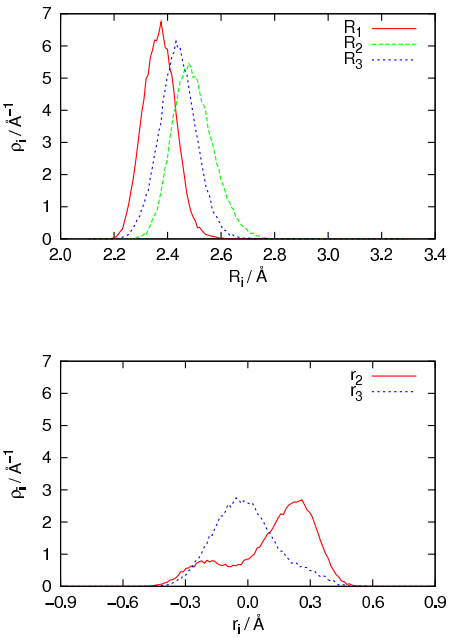


Figure 46: One-dimensional sampling densities (Equation 79) for the perprotio GFP simulation 30 fs after photoexcitation. Top: sampling densities for the heavy degrees of freedom (red: R_1, green: R_2, blue: R_3). Bottom: sampling densities for the light degrees of freedom (red: r_2, blue: r_3). All sampling densities are normalized to unit area.

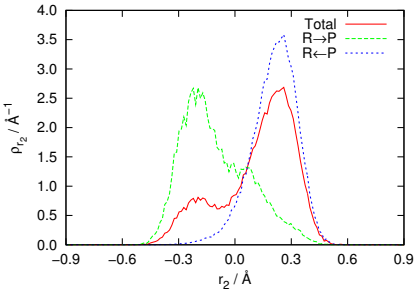


Figure 47: One-dimensional sampling function for r_2 in the perprotio GFP simulation 30 fs after photoexcitation (red: ρ_{r_2}) and its decomposition into total positive (green, indicative of transformation of reactants into products) and negative (blue, indicative of transformation of products into reactants) components. All densities are normalized to unit area.

fluxes and a smaller one at ~ -0.2 Å which contributes to the positive fluxes. In other words, Figure 47 is a snapshot of a recrossing situation.

Not only is it possible to store the probability densities into positive and negative contributions, but also the nonclassical and classical densities can be obtained. Let ρ_i^{tun} be the one-

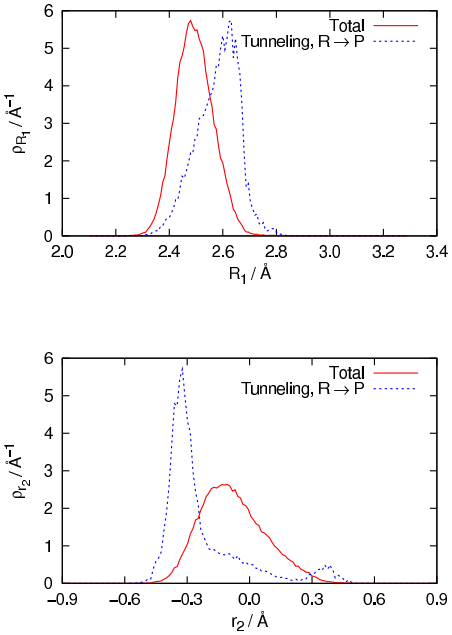


Figure 48: One-dimensional sampling function for R_1 (top) and r_2 (bottom) for the snapshot at 6 fs after photoexcitation of the perprotio GFP simulation. In both panels, the total density is shown in red, whereas tunneling density is shown in blue.

dimensional probability density in the i-th degree of freedom that reflects the amount of nonclassical flux. Figure 48 illustrates this in a practical way for degrees of freedom R_1 and r_2 after 6 fs for the perprotio simulation. Looking at the density functions for R_1, the tunneling contributions come from configurations with substantially larger R_1 values (the peak values appear at $\rho_{R_1} \sim 2.5$ Å, $\rho_{R_1}^{tun} \sim 2.65$ Å). This makes sense, as at longer O–O distances proton transfer is obstructed by larger potential energy. Thus, if density flows from reactants to products at this very region, it is likely to occur mainly nonclassically. Focusing now on $\rho_{r_2}^{tun}$, tunneling happens quite remarkably at very short O–H distances. Again, a highly compressed O–H bond implies a highly increased potential energy, so at this region, almost all the reactive flux must be nonclassic. Despite this, most reactive flux goes across classically permitted regions (see Figure 44) and occurs at shorter R_1 and longer r_2 values. Similar graphs could be produced for all variables and times, which would help to visualize what is actually going on (from a chemical point of view!) along the dynamics.

Another interesting aspect that can be analyzed is the positive (from reactants to products) and negative (from products to reactant) contributions of the flux that goes across $r_1 = 0$, which can be easily computed using the same methodology

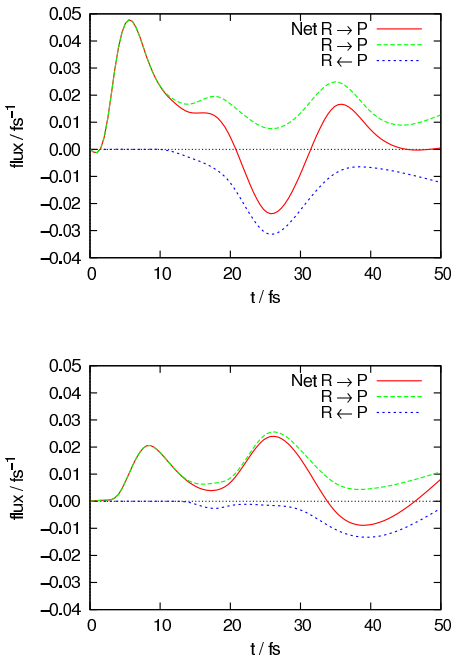


Figure 49: Overall net reactive flux for the perprotio (top) and perdeutero (bottom) simulations of the proton-wire. In red, the total flux; in green it is shown the net positive component of this flux (reactants yielding products), and in blue the net negative component of this flux (products yielding reactants).

(now, with a different Heavyside step function in Equation 78). This is shown in Figure 49 for the reactive flux and in Figure 50 for nonclassical (tunneling) flux. These figures provide an outstandingly detailed view of the dynamics at the separation hypersurface that cannot be obtained by other approaches like the calculation of flux as the time derivative of the population. With our methodology, it is possible to discern when most reaction happens, when recrossing is important, when and how much tunneling (and in what direction) contributes, and also in which geometries. As curiosities: at $t \sim 20$ fs for perprotio or $t \sim 35$ fs for perdeutero, the total flux across $r_1 = 0$ is zero, but this does not mean that everything is still: there are positive and negative fluxes that compensate each other. A lot is happening, but not seen.

With these results we have already obtained the values of the tunneling contribution in the GFP proton wire reaction in this ancient simulation. It has been seen that the tunneling contribution to the flux is not negligible at all. It is a very relevant result, since the simulations[149] of the GFP from which the present study started pointed out that the energy barrier effectively vanishes after promotion to photoactive electronic state.

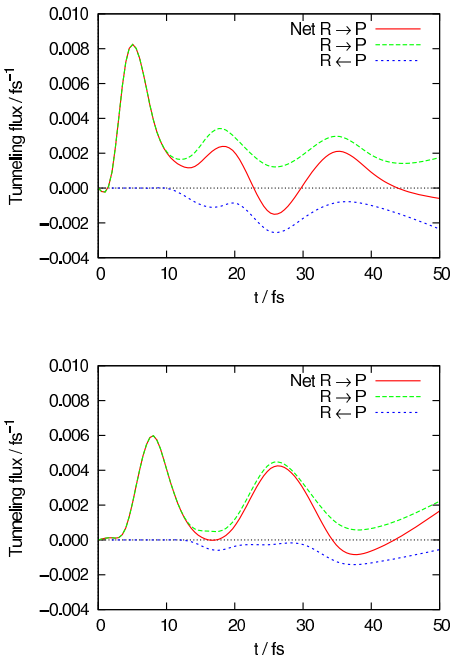


Figure 50: Overall net reactive tunneling flux for the perprotio (top) and perdeutero (bottom) simulations of the proton-wire. In red, the total tunneling flux; in green it is shown the net positive component of this flux (reactants yielding products), and in blue the net negative component of this flux (products yielding reactants).

In a quasiclassical simulation of the same process no system would find itself trapped behind any energy barrier, and hence no tunneling would be quantifiable in this way. Our study has shown that, even in situations in which there is not an effective barrier, an important part of the flux can go across nonclassical regions, and our method can unveil it.

The reasonable agreement between the two ways of calculating the flux supports the accuracy our mathematical develoepment and the implemented numerical methods. It is worth bearing in mind that the aim of this work has not only been the assessment of the tunneling in a given reaction, but the construction of a solid tool of general utility that can allow for any kind of flux filtering in any coordinate system in which a quantum dynamical simulation is solved.

5.7 CONCLUSIONS OF THE METHOD

After explaining both the methodology and its applications to the GFP proton wire, some remarks should be made about these two issues.

Regarding the methodology, a thorough derivation of the current probability equation and its integration over a hypersurface (flux) has been derived, and the dependence of the mass has been eliminated. This derivation is absolutely general and valid for other systems with different Jacobian matrices and different metric (masses), being this general applicability one of the most valuable assets of this work. Finally, a numerical procedure has been designed and adapted for the specific needs of this problem. This has required programming a software that implements the theory in an efficient and satisfactory way.

As for the chemical results on the model potential in absence of protein environment, it has been seen that in the initial stages of the quantum dynamics simulations (at about 6-8 fs after the beginning) the perprotio and perdeutero simulations have a maximum in the total reactive flux, that coincides with a maximum in the nonclassical flux. In absolute terms, tunneling is more important in the perprotio than in the perdeutero dynamics. Since deuterium has a larger mass, their quantum effects are less noticeable. Nevertheless, at these times, the percentage of tunneling is greater in the perdeutero simulation. This can be explained since the deuterium has a lower ZPE, and therefore a lower reactive flux. As a result, having more difficults to access the product region, a larger portion of the system will be forced to go across a nonclassical region. Besides, the analysis of the sampling probability densities have been revealed valuable *per se*, as it enables to know very accurately *how* the chemical reaction is occurring. We have seen that the analysis of the probability densities is in agreement with the chemical intuition.

To conclude this section, it should be said that the mathematical and computational procedures hereby explained furnish an interesting tool that can allow for tracing how (and through which areas) the reactants yield products. This achievement opens a door to other systems and problems that for sure will imply tantalizing challenges in the study of chemical reactivity. This tool provides an extremely detailed view of the events in a chemical reaction, effectively down to atomic resolution.

5.8 STARTING A NEW POTENTIAL ENERGY SURFACE

As it has been mentioned, the study so far explained involved a a proton wire quantum dynamics simulation with potential energy surface that did not take into account the effect of the environment. It has been shown in Section 1.1.1 and in Section 3.3 that, when calculating the energy of a QM system,

the inclusion of the electronic embedding can affect the results, providing more reliable ones in which the polarizing effect of nearby residues (and charges) is accounted for. A symptom of the importance of the environment can be guessed from the results of the QD simulation. An estimate of the $t_{1/2}$ reveals that the process' rate is in the hundreds of femtoseconds, but experimental data suggest rates in the range of several picoseconds. Even though many factors can be held responsible for this deviation ($\sim$ 100 fold), the lack of environment seems the most obvious suspect. Besides, the interpolated potential without environment has a vanishingly small barrier for the process, that cannot account for the KIE of 5 found. For this reason, a new potential energy surface is planned so that a more realistic quantum dynamics simulation can be performed and a similar probability flux analysis can be repeated with more solid conclusions.

5.8.1 *Geometry of the system*

It is important to bear in mind the difference between the model that will be used for the Potential Energy Surface(s), and the model that will be used for the quantum dynamics simulation. The dynamical study will be carried out in a selected group of coordinates (specifically chosen to represent the system, to integrate the time-dependent Schrödinger equation and to analyze the result, let us call it *dynamical set*). The potential energy calculations will be done on the "real" system, which uses a much larger coordinate set to describe the positions of all the atoms in the system (let us call it *geometrical set*). The dynamical set of coordinates needs to be small enough to be able to perform the quantum dynamics simulation, but flexible enough to be able to allow a meaningful analysis.

The proton wire involves the motion of a hydrogen atom from the chromophore to Wat25, another one from Wat25 to Ser205, and another one from Ser205 to Glu222. Considering the donor and acceptor oxygen atoms and the hydrogen atoms transfered, it makes a total of 7 atoms, represented and labeled in Figure 51.

For the PES calculation, excited-state electronic calculations will be carried out, with QM/MM electronic embedding. The coordinates of most atoms are already defined by a *reference* geometry (whose obtention will be explained in Section 5.8.3), and the calculation of the (ground state and excited state) electronic energy surfaces will require the exploration of the positions of the atoms involved in the proton wire. So as to describe

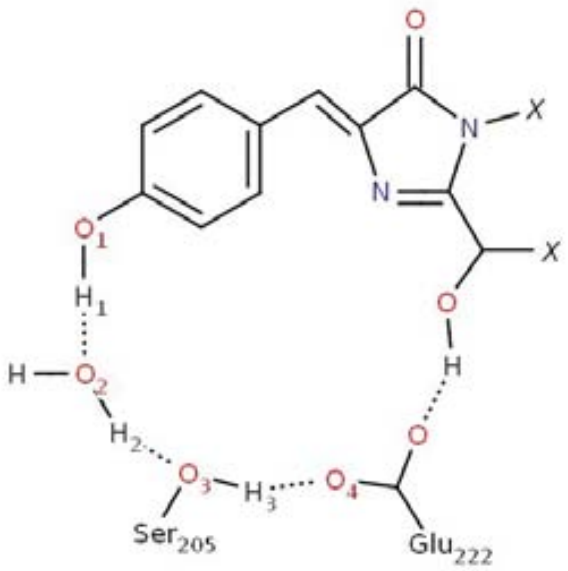


Figure 51: Structure of the GFP proton-wire and labels of the atoms involved.

them, we have chosen a set of 14 coordinates, enough to characterize a 7-atom system in two dimensions (let us call them *standard coordinates*). The set of hydrogen bond links is not collinear. The central hydrogen bond acts as a "hinge" between the two planes, forming an angle β, as can be seen in Figure 52

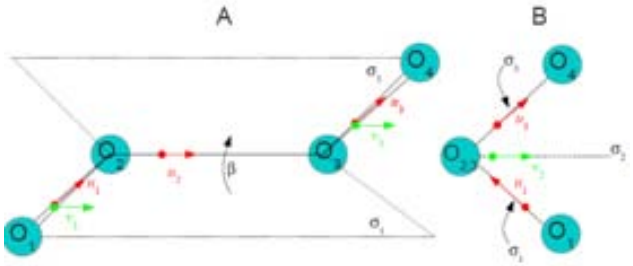


Figure 52: Front (A) and side (B) views of the proton-wire in GFP. Red (u) and green (v) vectors are the basis describing the coordinates for the proton motion.

To deal with coordinate sets for potential energy surfaces and quantum dynamical simulation, a *reference* structure will be needed. Let us see how we are defining the coordinates, depicted in Figure 52:

- Plane σ_1 is defined by oxygens O_1, O_2 and O_3. Plane σ_3 is defined by oxygens O_2, O_3 and O_4. Plane σ_2 is defined by oxygens O_2 and O_3 and bisects angle β, which is the angle between σ_1 and σ_3. Let us denote the 3D position vectors

of these oxygens in the reference structure as $\vec{X}_i^{ref}$, where i identifies the corresponding oxygen atom.

- Plane σ_i will contain proton H_i ($i = 1...3$) in all possible dynamical arrangements (the different planes will be ironed out into one in the dynamics simulation, but not in the potential energy surface calculations).
- $\vec{u}_1$, $\vec{u}_2$ and $\vec{u}_3$ are unit vectors in the direction of the proton transfer (that is, the direction that links donor and acceptor oxygen atoms in a hydrogen bond). These are uniquely defined once the reference positions of the oxygens ($\vec{X}_i^{ref}$) are known and fixed.
- $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$ are unit vectors orthogonal to the respective $\vec{u}_1$, $\vec{u}_2$ and $\vec{u}_3$ vectors. From all possible choices, $\vec{v}_1$ belongs to σ_1 and $\vec{v}_3$ to σ_3. As for $\vec{v}_2$, it belongs to σ_2, which means that it is simultaneusly normal to $\vec{u}_2$ and $\frac{1}{2}(\vec{u}_1 + \vec{u}_3)$

From the above definitions, the $\vec{u}_i$ and $\vec{v}_i$ set of coordinates can be defined from the reference structure as:

- $\vec{u}_1 = \frac{\vec{X}_2^{ref} - \vec{X}_1^{ref}}{|\vec{X}_2^{ref} - \vec{X}_1^{ref}|}$
- $\vec{u}_2 = \frac{\vec{X}_3^{ref} - \vec{X}_2^{ref}}{|\vec{X}_3^{ref} - \vec{X}_2^{ref}|}$
- $\vec{u}_3 = \frac{\vec{X}_4^{ref} - \vec{X}_3^{ref}}{|\vec{X}_4^{ref} - \vec{X}_3^{ref}|}$
- $\vec{v}_1$ is the projection of $\vec{u}_2$ that is orthogonal to $\vec{u}_1$, using Gramm-Schmidt: $\vec{v}_1 = \vec{u}_2 - (\vec{u}_2 \cdot \vec{u}_1)\,\vec{u}_1$, to be normalized
- $\vec{v}_3$ is the projection of $\vec{u}_2$ that is orthogonal to $\vec{u}_3$, using Gramm-Schmidt: $\vec{v}_3 = \vec{u}_2 - (\vec{u}_2 \cdot \vec{u}_3)\,\vec{u}_3$, to be normalized
- $\vec{v}_2$ is a unit vector orthogonal to $\vec{u}_2$ and at the same time, to $\vec{u}_1 + \vec{u}_3$. This can be achieved by using the rule of the vector product: $\vec{v}_2 = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_{2,x} & u_{2,y} & u_{2,z} \\ u_{1,x} + u_{3,x} & u_{1,y} + u_{3,y} & u_{1,z} + u_{3,z} \end{vmatrix}$, to be normalized.

The six vectors above described are actually 3-dimensional vectors, but each pair ($\vec{u}_1$, $\vec{v}_i$) spans a 2-D subspace representing the motion of a proton on a plane containing donor and acceptor oxygen atoms.

5.8.2 *Dynamical Coordinate Set*

Once the generating vectors of the respective proton-transfer coordinates are established, it is possible to define a set of $2-D$ dynamical coordinates for all atoms. For instance, let us take O_1 as an example. The relative position with respect to the reference structure is given by

$$\vec{\tilde{x}}_1 = \vec{X}_1 - \vec{X}_1^{ref} \tag{82}$$

Then, projecting this onto the orthonormal vector set of plane σ_1 we can obtain the dynamical coordinates a_i^0 and b_i^0 (which actually furnish the position of O_1in the plane σ_1):

$$\begin{aligned} a_1^O &= \vec{\tilde{x}}_1 \cdot \vec{u}_1 \\ b_1^O &= \vec{\tilde{x}}_1 \cdot \vec{v}_1 \end{aligned} \tag{83}$$

An equivalent procedure can be followed for the rest three oxygen atoms. If we establish that the motion of O_1 occurs in σ_1, the motion of O_2 and O_3 occurs in σ_2, and the motion of O_4 occurs in σ_3, we have

$$\begin{aligned} \vec{X}_1 &= \vec{X}_1^{ref} + a_1^O\vec{u}_1 + b_1^O\vec{v}_1 \\ \vec{X}_2 &= \vec{X}_2^{ref} + a_2^O\vec{u}_2 + b_2^O\vec{v}_2 \\ \vec{X}_3 &= \vec{X}_3^{ref} + a_3^O\vec{u}_2 + b_3^O\vec{v}_2 \\ \vec{X}_4 &= \vec{X}_4^{ref} + a_4^O\vec{u}_3 + b_4^O\vec{v}_3 \end{aligned} \tag{84}$$

As for the hydrogen atoms, a similar set of coordinates can be defined. Being Y_i^{ref} the position of H_i in the reference system and Y_i its actual position in the PES, the displacement vector is

$$\vec{\tilde{y}}_i = \vec{Y}_i - \vec{Y}_i^{ref} \tag{85}$$

Then, projecting this onto the orthonormal vector set of plane σ_1, we obtain the dynamical coordinates a_i^H and b_i^H.

Assuming that the motion of H_i is restricted to plane σ_i, the absolute position Y_i of a hydrogen atom are:

$$\begin{aligned} \vec{Y}_1 &= \vec{Y}_1^{ref} + a_1^H\vec{u}_1 + b_1^H\vec{v}_1 \\ \vec{Y}_2 &= \vec{Y}_2^{ref} + a_2^H\vec{u}_2 + b_2^H\vec{v}_2 \\ \vec{Y}_3 &= \vec{Y}_3^{ref} + a_3^H\vec{u}_3 + b_3^H\vec{v}_3 \end{aligned} \tag{86}$$

The 7 particles are described by a set of 14 dynamical coordinates: the (a,b) sets.

All coordinates described so far are Cartesian, if somewhat reoriented. Because the coordinates affect a single particle each, the kinetic energy term will consist of 14 Cartesian terms involving the mass of the particle and the second derivative with respect to the a or b variable. In view of a future analysis of the dynamics simulation, if Y_i^{ref} is selected to be the midpoint between donor and acceptor, survival analysis can be done as usual with a Heaviside operator on the a coordinate of each hydrogen.

Even though the coordinate set just described does not allow the system to access all possible configurations, it will cover most (if not all) chemically relevant (that is, moderate energy) configurations, as it confines the H motion to the planes containing donor-acceptor atoms. It is a clear improvement over the 6-internal coordinate system used by Vendrell *et al.*

5.8.3 *Strategy to build the Potential Energy Surface*

To begin with, we have obtained a *reference structure* as follows. Originally, we have considered the A GFP structure, taken from PDB reference 1EMA[159]. It has been solvated with a water droplet of 40 Å diameter. The solvent water molecules have been MM-minimized with CHARMM22, and then the whole system has been QM/MM optimized with ChemShell. The active region has included all residues containing at least one atom at 15 Å or less from the chromophore. The QM region includes 102 atoms: those involved in the proton wire (the chromophore, Wat25, Ser205 and Glu222) and other close residues (Arg96, His148, Thr203). The QM region has been calculated with B3LYP/6-31+G(d), and the MM region with CHARMM22 force field. The QM/MM coupling has been done within the *shift* scheme. The resulting structure corresponds to the A structure from which the rest of the calculation will be done.

The calculations for the PES have been done as follows. The electronic ground and excited energies have been calculated within an electronic embedding QM/MM scheme. The QM and MM regions are the same that have been chosen for the optimization. The calculation of the energy has been done with Gaussian 09[104], with the *charge* option. The MM point charges have been taken according to CHARMM22 force field. The (QM) energy of the system has been calculated with CAM-B3LYP/6-31G(d,p). 10 excited states have been calculated for each configuration.

Afterwards, the grid of points has to be calculated, and fitted into a potential energy function. That function has to pro-

vide the energy of the system –that takes into account the effect of the whole environment– depending on the (a,b) dynamical coordinates. To this end, an Empirical Valence Bond-based method[149] will be used. The underlying idea is that the whole surface cannot be explored because there are many variables and this would imply an unfeasible number of points to evaluate. For this reason, we have evaluated the points corresponding to single proton transfers. The three-proton relay presents twelve different possible paths of one-proton transfer, as seen in Figure 53. Table 14 and Table 15 show, respectively, the paths corresponding to the trasfer of one and two protons. The fitting procedure is a manual and lengthy process, that usually requires trial and error.

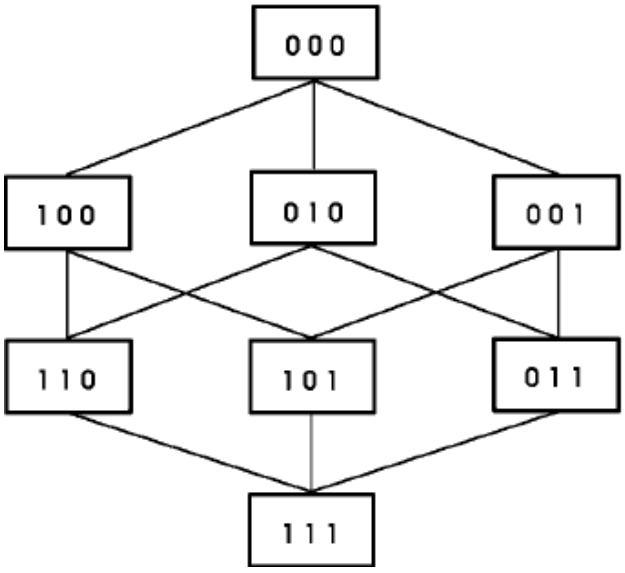


Figure 53: Schematic representation of the different paths connecting the eight possible protonation states of the proton wire.[149] Each figure represents a hydrogen bond (Cro-Wat25, Wat25-Ser205 and Ser205-Glu222): 0 means that the hydrogen atom has not been transfered yet, whereas 1 means that the hydrogen atom has been already transfered.

These paths have been calculated by exploring different positions of the oxygen and hydrogen atoms involved in that proton transfer (oxygen atoms getting close to each other, and the hydrogen atom along the line that separates them).

Also the six different two-proton transfer paths have been considered by evaluating the positions of the hydrogen atoms. The concerted three-proton transfer path has been finally included, also by moving the three hydrogen atoms. In total, 1826 points (configurations) have been evaluated. The energy evaluations have been repeated for the QM system without the MM

Table 14: Twelve different single-proton transfer paths. Each figure represents a hydrogen bond. 0 means that the hydrogen atom has not been transfered yet, and 1 that has been transfered

$000 \rightarrow 100$	$100 \rightarrow 110$	$001 \rightarrow 101$
$000 \rightarrow 010$	$010 \rightarrow 011$	$101 \rightarrow 111$
$000 \rightarrow 001$	$010 \rightarrow 110$	$110 \rightarrow 111$
$100 \rightarrow 101$	$001 \rightarrow 011$	$011 \rightarrow 111$

Table 15: Six different two-proton transfer paths. Each figure represents a hydrogen bond. 0 means that the hydrogen atom has not been transfered yet, and 1 that has been transfered

$000 \rightarrow 110$	$000 \rightarrow 011$	$000 \rightarrow 101$
$100 \rightarrow 111$	$010 \rightarrow 111$	$001 \rightarrow 111$

polarization: this way the effect of the environment will be able to be directly assessed.

With all this, we have obtained the ground and main photoactive state energies for different values of a and b. The fitting of the potential energy surfaces (ground and excited state) has not been completed yet, but an initial observation to our first attempts shows that the energies of A and I are quite similar for both the ground and the excited state. The results are shown in Table 16, which contains the energies of reactants, products and intermediates (they are not necessarily transition states).

It can be seen that reactants are more stable than products in the ground state, and in the excited state the difference shortens, and the barrier seems to decrease. Still, reactants are slightly more stable than products at S_1 (let us recall, though, that the GFP structure comes from an optimization that stabilizes reactants). This does not exclude the ESPT, since the Franck-Condon energy is not taken into account. Finally, we should not forget that the fitting has not been completed, which means that some slight changes might be expected.

Our results are also in accordance with recent calculations performed by Krylov and co-workers[160], that showed that reactants and products were almost isoenergetic ($\sim$ 1 kcal·mole^1) in the ground state.

In any case, it seems that the new PESs including the enviroment and using TDDFT instead of CASSCF/CASPT2 may result in qualitavely different situations (compared to those predicted by Vendrell *et al.*). At the end, a better comprehension of

the GFP proton wire reaction will be possible in the foreseeable future.

Table 16: Ground and excited state electronic energies corresponding to 000 (Cro, Wat25 and Ser205 are protonated, Glu222 is deprotonated), one intermediate (possible TS) and to 111 (Cro is deprotonated, Wat25, Ser205 and Glu222 are protonated). These calculations include the environment. FC means Franck-Condon.

Surface		Energy	($kcal \cdot mole^{-1}$)
S_1	83.9 [FC:84.2]	90.0	84.8
S_0	0.00	8.7	4.5
	000	pos. TS	111

6

CONCLUSIONS

Throughout this thesis, many different problems have been tackled. All of them have in common the need to develop novel theoretical methods combined with the already existing multiscale tools when studying complex systems such as proteins. In all, it has been necessary to consider the quantum nature of the atoms, to take into account the dynamical effects of the systems and to include the environment, among others. These needs are both a *cause* and a *result* of our work: we have undertaken the study of different issues regarding photoactivated proteins that have required a theoretical analysis and the development of several methodologies. At the end, after having gone through a varied collection of problems, we are more aware of the importance of such theoretical multiscale analyses and studies and more confident on the approach that we have adopted. To be more specific, we are going to list the conclusions that we have reached along this fascinating itinerary:

- We have proved that, when studying complex systems as fluorescent proteins, the strategies consisting of: reducing the system to a cluster of few atoms, treating it as a rigid entity or searching for energy minima fail to properly describe its behavior. On the contrary, electrostatic effects and thermal jitter *must* be taken into account.

- We have developed a reliable method to reproduce the absorption spectra of fluorescent proteins by means of combining QM/MM excited state calculations with molecular dynamics simulations. This method consists of selecting random snapshots of the dynamics simulation and perform the electronic excited-states energy calculations on them. This way, the aforementioned electrostatic effects and thermal jitter are accounted for.

- We have succeeded in simulating absorption spectra of RFPs mKeima and LSSmKate2. Regarding the mKeima, it has been proved that the two experimental absorption maxima of 440 nm and 590 nm correspond, respectively, to the neutral and anionic chromophore. As for LSSmKate2, the experimental absorption band centered at 460 nm is correctly reproduced.

- An anharmonic analysis has proved to be essential when dealing with LBHBs. The concerns of Saito and Ishikita[45] have been proved misleading: the energy minimum does *not* necessarily coincide with the expected value of the position of an atom, that is why an quantum mechanical treatment of the atom is required in such cases.

- We have solved the controversy regarding the existence of an LBHB in the PYP. The results point out that there is an LBHB in the crystal protein, but it vanishes when the protein is solvated at room temperature. This conciliating solution has been reached by considering the effects of the electrostatic and thermal effects when needed.

- We have theoretically verified the existence of An LBHB in a protein, the PYP (albeit in crystal), for the first time ever.

- A thorough method to analyze probability flux has been developed. This method is valid for any coordinate system (only the specific Jacobian matrix and the metrics need to be changed). From data obtained in any quantum dynamics simulation, the probability flux can be evaluated. This method has the particularity of analyzing the flux as a surface integral, which implies that the probability flux can be studied over a hypersurface. Thanks to this, magnitudes as the tunnelling contribution, recrossing or, in general, any kind of flux *ltering*, can be studied.

- The surface integral has been calculated with an importance sampling Monte Carlo method, for which probability densities were computed for the sampling. These probability densities have furnished worthy information about *how* the reaction happens: we have been able to practically *map* and classify the different flux contributions.

- The method to compute probability current has been applied to a quantum dynamics simulation of the GFP proton-wire.[74] The flux calculated as a probability current integral and as the time-derivative of the population are in good agreement. Even though the electronic potential energy surface did not present a barrier in the excited-state, there is a non-negligible percentage of tunneling. Particularly, the tunneling is greater in the perprotio simulation than in the perdeutero one. The tunneling contribution was particularly relevant in the initial stages of both the perprotio and perdeutero quantum dynamics simulations

(after 6-8 fs, respectively), that coincided with the maximum in the total reactive flux.

- The recrossing has been analyzed as well, and we conclude that the movement of the wavepacket is highly complex: it goes from reactants to products and vice versa through different regions at different times.
- We have laid the bases of a new GFP proton wire potential energy surface that includes the whole protein environment. From this new potential energy surface, a new model for quantum dynamics simulation has been proposed.
- The GFP proton wire potential energy surfaces so far obtained indicate that reactants and products have similar energies and the barriers are higher, which might make the proton transfer slower in S^1. This constitutes an improvement with respect to the former *in vacuo*[70, 74] calculations, that furnished too large reaction rates.

Part III

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] *The Nobel Prize in Chemistry 2008*. Nobel Media AB, "www.nobelprize.org", 2014.

[2] *The Nobel Prize in Chemistry 2013*. Nobel Media AB, "www.nobelprize.org", 2014.

[3] B. Honig and M. Karplus. *Nature*, 229:558–560, 1971.

[4] A. Warshel and M. Karplus. *J. Am. Chem. Soc.*, 94:5612–5625, 1972.

[5] R. Pariser and R. Parr. *J. Chem. Phys.*, 21:466, 1953.

[6] J. A. Pople. *Trans. Faraday Soc.*, 49:1375, 1953.

[7] A. Warshel and M. Levitt. *J. Mol. Biol.*, 103:227–249, 1976.

[8] J. Gao. *Rev. Comput. Chem.*, 7:119, 1996.

[9] F. Maseras and K. Morokuma. *J. Comput. Chem.*, 16:1170–1179, 1995.

[10] U. C. Singh and P. Kollman. *J. Comput. Chem.*, 7:718–730, 1986.

[11] H. M. Senn and W. Thiel. *Angew. Chem. Int. Ed.*, 48:1198–1229, 2009.

[12] W. W. Cleland and M. M. Kreevoy. *Science*, 264, 1994.

[13] J. A. Gerlt and P. G. Gassman. *J. Am. Chem. Soc.*, 115:11552–11568, 1993.

[14] M. Garcia-Viloca, R. Gelabert, A. González-Lafont, M. Moreno, and J. M. Lluch. *J. Chem. Phys. A*, 101:8727–8733, 1997.

[15] M. Garcia-Viloca, R. Gelabert, A. González-Lafont, M. Moreno, and J. M. Lluch. *J. Am. Chem. Soc.*, 120:10203–10209, 1998.

[16] R. Gelabert, M. Moreno, J. M. Lluch, a. Lledós, V. Pons, and D. M. Heinekey. *J. Am. Chem. Soc.*, 126:8813–8822, 2004.

[17] R. Gelabert, M. Moreno, J. M. Lluch, A. Lledós, and D. M. Heinekey. *J. Am. Chem. Soc.*, 127:5632–5640, 2005.

[18] R. Gelabert, M. Moreno, J. M. Lluch, and A. Lledós. *J. Am. Chem. Soc.*, 119:9840–9847, 1997.

[19] R. Gelabert, M. Moreno, and J. M. Lluch adn A. Lledós. *J. Am. Chem. Soc.*, 120:8168–8176, 1998.

[20] R. Gelabert, M. Moreno, and J. M. Lluch. *Chem. – Eur. J.*, 11:6315–6325, 2005.

[21] F. Hibbert and J. Emsley. *Adv. Phys. Org. Chem.*, 26:255–379, 1991.

[22] J. P. Guthrie. *Chem. Biol.*, 3:162–170, 1996.

[23] A. Warshel and A. Papazyan. *Proc. Natl. Acad. Sci. U.S.A.*, 93:13665–13670, 1996.

[24] P. A. Frey, S. A. Whitt, and J. B. Tobin. *Science*, 264, 1994.

[25] J. A. Gerlt and P. G. Gassman. *Biochemistry*, 32:11943–11952, 1993.

[26] W. W. Cleland. *Biochemistry*, 31:317–319, 1992.

[27] G. L. Kenyon, J. A. Gerlt, G. A. Petsko, and J. W. Petsko. *Acc. Chem. Res.*, 28:178–186, 1995.

[28] S. B. Xiang, S. A. Short, R. Wolfenden, and C. W. Carter. *Biochemistry*, pages 4516–4523, • 1995.

[29] W. W. Cleland and M. M. Kreevoy. *Science*, 269:104, 1995.

[30] P. A. Frey. *Science*, 269, 1995.

[31] A. Warshel, A. Papazyan, and P. A. Kollman. *Science*, 269:102–104, 1995.

[32] B. Schwarz and D. G. Drueckhammer. *J. Am. Chem. Soc.*, 117:11902–11905, 1995.

[33] G. Alagona, C. Ghio, and P. A. Kollman. *J. Am. Chem. Soc.*, 117:9855–9862, 1995.

[34] S. Scheiner and T. Kar. *J. Am. Chem. Soc.*, 117:6970–6975, 1995.

[35] J. B. Tobin, S. A. Whitt, C. S. Cassidy, and P. A. Frey. *Biochemistry*, 34:6919–6924, 1995.

[36] S. O. Shan and D. Herschlag. *Proc. Natl. Acad. Sci. U.S.A.*, 93:14474–14479, 1996.

[37] M. Garcia-Viloca, A. González-Lafont, and J. M. Lluch. *J. Phys. Chem. A*, 101:3880–3886, 1997.

[38] M. Garcia-Viloca, A. González-Lafont, and J. M. Lluch. *J. Am. Chem. Soc.*, 119:1081–1086, 1997.

[39] C. L. Perrin and J. B. Nielson. *Annu. Rev. Phys. Chem.*, 48:511–544, 1997.

[40] W. W. Cleland. *Adv. Phys. Org. Chem.*, 44:1–17, 2010.

[41] C. L. Perrin. *Acc. Chem. Res.*, 43:1550–1557, 2010.

[42] C. L. Perrin. *Science*, 266:1665–1668, 1994.

[43] C. L. Perrin and J. S. Lau. *J. Am. Chem. Soc.*, 128:11820–11824, 2006.

[44] S. Yamaguchi, H. Kamikubo, K. Kurinara, R. Kuroki, N. Niimura, N. Shimizu, Y. Yamazaki, and M. Kataoka. *Proc. Natl. Acad. Sci. U.S.A.*, 106:440–444, 2009.

[45] K. Saito and H. Ishikita. *Proc. Natl. Acad. Sci. U.S.A.*, 109:167–172, 2012.

[46] H. Morise, O. Shimomura, F. H. Johnson, and J. Winant. *Biochemistry*, 13:2656–2662, 1974.

[47] O. Vendrell. *Investigación y Ciencia*, 2008.

[48] O. Shimomura, F. H. Johnson, and Y. Saiga. *J. Cell. Physio.*, 59:223–239, 1962.

[49] O. Shimomura. *FEBS Lett.*, 104:220–222, 1979.

[50] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher. *Science*, 263:802–805, 1994.

[51] R. M. Wachter. *Acc. Chem. Res.*, 40:120–127, 2007.

[52] M. Zimmer. *Chemical Reviews*, 102:759–781, 2002.

[53] S. J. Remington. *Protein Sci.*, 20:1509–1519, 2011.

[54] D. Prasher, V. Eckenrode, W. Ward, F. Prendergast, and M. Cormier. *Gene*, 111:229–233, 1992.

[55] R. Y. Tsien. *Annual Review of Biochemistry*, 67:509–544, 1998.

[56] M. Matz, K. A. Lukyanov, and S. A. Lukyanov. *BioEssays*, 24:953–959, 2002.

[57] K. B. Bravaya, O. M. Subach, N. Korovina, V. V. Verkhusha, and A. I. Krylov. *J. Am. Chem. Soc.*, 134:2807–2814, 2012.

[58] G. S. Baird, D. A. Zacharias, and R. Y. Tsien. *Proc. Natl. Acad. Sci.*, 97:11984–11989, 2000.

[59] R. E. Campbell, O. Tour, A. E. Palmer, P. A. Steinbach, G. S. Baird D. A. Zacharias, and R. Y. Tsien. *Proc. Natl. Acad. Sci. USA*, 99:7877–7882, 2002.

[60] L. A. Gross, G. S. Baird, R. C. Hoffman, K. K. Baldridge, and R. Y. Tsien. *Proc. Natl. Acad. Sci. USA*, 97:11990–11995, 2000.

[61] B. J. Bevis and B. S. Glick. *Nature Biotechnol.*, 20:83–87, 2002.

[62] N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer, and R. Y. Tsien. *Nature Biotechnol.*, 22:1567–1572, 2004.

[63] N. C. Shaner, M. Z. Lin, M. R. McKeown, P. A. Steinbach, K. L. Hazelwood, M. W. Davidson, and R. Y. Tsien. *Nature Methods*, 5:545–551, 2008.

[64] K. D. Piatkevich, V. N. Malashkevich, S. C. Almo, and V. V. Verkhusha. *J. Am. Chem. Soc.*, 132:10762–10770, 2010.

[65] T. Kogure, S. Karasawa, T. Araki, K. Saito, M. Kinjo, and A. Miyawaki. *Nat. Biotechnol.*, 24:577–581, 2006.

[66] F. V. Subach and V. V. Verkhusha. *Chem. Rev.*, 112:4308–4327, 2012.

[67] T. Kogure, H. Kawano, Y. Abe, and A. Miyawaki. *Methods*, 45:223–226, 2008.

[68] M. Chattoraj, B. A. King, G. U. Bublitz, and S. G. Boxer. *Proc. Natl. Acad. Sci. U. S. A.*, 93:8362–8367, 1996.

[69] K. Brejc, T. K. Sixma, P. A. Kitts, S. R. Kain, R. Y. Tsien, M. Ormo, and S. J. Remington. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 1997.

[70] O. Vendrell, R. Gelabert, M. Moreno, and J. M. Lluch. *J. Am. Chem. Soc.*, 128:3564–3574, 2006.

[71] M. A. Lill and V. Helms. *Proc. Natl. Acad. Sci. U. S. A.*, 99:2778–2781, 2002.

[72] D. Stoner-Ma, A. A. Jaye, P. Matousek, M. Towrie, S. R. Meech, and P. J. Tonge. *J. Am. Chem. Soc.*, 127:2864–2865, 2005.

[73] M. Di Donato, L. J. G. W. van Wilderen, I. H. M. Van Stokkum, T. C. Stuart, J. T. M. Kennis, K. J. Hellingwerf, R. van Grondelle, and M. L. Groot. *Phys. Chem. Chem. Phys.*, 13:16295–16305, 2011.

[74] O. Vendrell, R. Gelabert, M. Moreno, and J. M. Lluch. *J. Phys. Chem. B*, 112:5500–5511, 2008.

[75] J. Andrés and J. Bertran, editors. *Theoretical and Computational Chemistry: Foundations, Methods and Techniques*, chapter 4. Publicacions de la Universitat Jaume I, 2007.

[76] C. Randino, M. Moreno, R. Gelabert, and J. M. Lluch. *J. Phys. Chem. B*, 116:14302–14310, 2012.

[77] A. L. Sobolewski and W. Domcke. *Phys. Chem. Chem. Phys.*, 1:3065–3072, 1999.

[78] A. J. A. Aquino, H. Lischka, and C. Hattig. *J. Phys. Chem. A*, 109:3201–3208, 2005.

[79] E. Runge and E. K. U. Gross. *Phys. Rev. Lett.*, 52:997–1000, 1984.

[80] A. Dreuw and M. Head-Gordon. *Chem. Rev.*, 105:4009–4037, 2005.

[81] M. Petersilka, U. J. Gossmann, and E. K. U. Gross. *Phys. Rev. Lett.*, 76:1212–1215, 1996.

[82] S. Hirata and M. Head-Gordon. *Chem. Phys. Lett.*, 314:291–299, 1999.

[83] A. D. Becke. *J. Chem. Phys.*, 98:1372–1377, 1993.

[84] M. Y. Zhang, J. Y. Wang, C. S. Lin, and W. D. Cheng. *J. Phys. Chem. B*, 115:10750–10757, 2011.

[85] Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, and K. Hirao. *J. Chem. Phys.*, 120:8425–8433, 2004.

[86] N. C. Handy. *Mol. Phys.*, 102:2399–2409, 2004.

[87] T. Yanai, D. P. Tew, and N. C. Handy. *Chem. Phys. Lett.*, 393:51–57, 2004.

[88] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, Joseph, D. McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, Wiorkiewicz, J. Kuczera, D. Yin, and M. Karplus. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[89] A. D. J. MacKerell, C. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. *Encyclopedia of Computational Chemistry*, pages 271–277. John Wiley & Sons, Chichester, 1998.

[90] A. D. MacKerell, M. Feig, and C. L. Brooks. *J. Am. Chem. Soc.*, 126, 2004.

[91] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. *J. Comput. Chem.*, 4:187–217, 1983.

[92] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. *J. Comput. Chem.*, 30:1545–1614, 2009.

[93] N. Reuter, H. Lin, and W. Thiel. *J. Phys. Chem. B*, 106:6310–6321, 2002.

[94] D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, and P.A. Kollman. Amber 2014. University of California, San Francisco, 2014.

[95] H. Bekker, H. J. C. Berendsen, E. J. Dijkstra, S. Achterop, R. van Drunen, D. van der Spoel, A. Sijbers, H. Keegstra, B. Reitsma, and M. K. R. Renardus. *Physics Computing 92*. World Scientific, Singapore, 1993.

[96] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. *Comp. Phys. Comm.*, 91:43–56, 1995.

[97] E. Lindahl, B. Hess, and D. van der Spoel. *J. Mol. Mod.*, 7:306–317, 2001.

[98] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. *J. Comp. Chem.*, 26:1702–1718, 2005.

[99] A. Hinchliffe. *Molecular Modelling for Beginners*, chapter 10.4.1. John Wiley & Sons, Chichester, 2008.

[100] *ChemShell, a Computational Chemistry Shell*. www.chemshell.org.

[101] W. Smith and T. R. Forester. *J. Mol. Graphics*, 14:136–141, 1996.

[102] G. G. Hall and C. M. Smith. *Int. J. Quantum. Chem.*, 25:881–890, 1984.

[103] C. M. Smith and G. G. Hall. *Theor. Chem. Acc.*, 69:63–69, 1986.

[104] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino J, G. Zheng G, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and J. D. Fox. Gaussian 09. revision c.1. Gaussian, Inc.:Wallingford CT, 2009.

[105] S. R. Billeter, A. J. Turner, and W. Thiel. *Phys. Chem. Chem. Phys.*, 2:2177–2186, 2000.

[106] R. G. Parr and W. Yang. *Density Functional Theory of Atoms and Molecules*. Oxford University Press, Oxford, U.K., 1989.

[107] A. D. Becke. *J. Chem. Phys.*, 98:5648–5652, 1993.

[108] M. J. G. Peach, T. Helgaker, P. Salek, T. W. Keal, and O. Lutnaes. *Phys. Chem. Chem. Phys.*, 8:558–562, 2006.

[109] Y. G. Wang. *J. Phys. Chem. A*, 113:10873–10879, 2009.

[110] W. Thiel. *MNDO99 Program*. version 7.0. Max-Planck-Institut für Kohlenforschung, Mülheim, Germany, 2005.

[111] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. *J. Am. Chem. Soc.*, 107:3902–3909, 1985.

[112] N. C. Shaner, G. H. Patterson, and M. W. Davidson. *J. Cell Sci.*, 120:4247–4260, 2007.

[113] V. V. Verkhusha and K. A. Lukyanov. *Nat. Biotechnol.*, 22:289–296, 2004.

[114] V. V. Verkhusha, K. D. Piatkevich, J. Hulit, O. M. Subach, B. Wu, A. Abdulla, and J. E. Segall. *Proc. Natl. Acad. Sci. U. S. A.*, 107:5369–5374, 2010.

[115] D. Shcherbo, I. I. Shemiakina, A. V. Ryabova, K. E. Luker, B. T. Schmidt, E. A. Souslova, T. V. Gorodnicheva, L. Strukova, K. M. Shidlovskiy, O. V. Britanova, A. G. Zaraisky, K. A. Lukyanov, V. B. Loschenov, G. D. Luker, and D. M. Chudanov. *Nat. Methods*, 7:827–U1520, 2010.

[116] K. S. Morozova, K. D. Piatkevich, T. J. Gould, J. H. Zhang, J. Bewersdorf, and V. V. Verkhusha. *Biophys. J.*, 99:L13–L15, 2010.

[117] S. Violot, P. Carpentier, D. Bourgeois, and L. Blanchoin. *J. Am. Chem. Soc.*, 131:10356–10357, 2009.

[118] J. N. Henderson, M. F. Osborn, N. Koon, R. Gepshtein, D. Hupper, and S. J. Remington. *J. Am. Chem. Soc.*, 131:13212–13213, 2009.

[119] N. A. Murugan, J. Kongsted, Z. Rinkevicius, and H. Agren. *Phys. Chem. Chem. Phys.*, 14:1107–1112, 2012.

[120] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J.Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson,

F. King, E. Karlsen, M. Sjovoll, A. Fahmi, A. Schafer, and C. Lennartz. *J. Mol. Struc. - Theochem*, 632:1–28, 2003.

[121] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker. *Nucleic Acids Res.*, 32:W665–W667, 2004.

[122] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker. *Nucleic Acids Res.*, 35:W522–W525, 2007.

[123] H. Li, A. D. Robertson, and J. H. Jensen. *Proteins*, 61:704–721, 2005.

[124] S. Olsen and S. C. Smith. *J. Am. Chem. Soc.*, 129:2054–2065, 2007.

[125] E. Sanchez-Garcia, M. Doerr, and W. Thiel. *J. Comput. Chem.*, 31:1603–1612, 2010.

[126] P. Imhof. *J. Chem. Theory Comput.*, 8:4828–4836, 2012.

[127] C. M. Isborn, A. W. Götz, M. A. Clark, R. C. Walker, and T. J. Martínez. *J. Chem. Theory Comput.*, 8, 2012.

[128] X. Lopez, M. A. L. Marques, A. Castro, and A. Rubio. *J. Am. Chem. Soc.*, 127:12329–12337, 2005.

[129] C. Randino, M. Nadal-Ferret, R. Gelabert, M. Moreno, and J. M. Lluch. *Theor. Chem. Acc.*, 132:1327–, 2013.

[130] W. W. Sprenger, W. D. Hoff, J. P. Armitage, and K. J. Hellingwerf. *J. Bacteriol.*, 175:3096–3104, 1993.

[131] K. Saito and H. Ishikita. *Biochim. Biophys. Acta, Bioenerg.*, 1827:387–394, 2013.

[132] I.-R. Lee, W. Lee, and A. H. Zewall. *Proc. Natl. Acad. Sci. USA*, 103:258–262, 2006.

[133] S. Anderson, S. Crosson, and K. Moffat. *Acta Crystallogr.*, 60:1008–1016, 2004.

[134] S. Z. Fisher, S. Anderson, R. Henning, K. Moffat, P. Langan, P. Thiyagarajan, and A. J. Schultz. *Acta Crystallogr.*, 63:1178–1184, 2007.

[135] E. D. Getzoff, K. N. Gutwin, and U. K. Genick. *Natl. Struct. Biol.*, 10:663–668, 2003.

[136] K. Saito and H. Ishikita. *Biochemistry*, 51:1171–1177, • 2012.

[137] P. A. Sigala, M. A. Tsuchida, and D. Herschlag. *Proc. Natl. Acad. Sci. U.S.A.*, 106:9232–9237, 2009.

[138] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in Fortran*. Oxford University Press, Cambridge, U.K., 2nd edition, 1992.

[139] D. T. Colbert and W. H. Miller. *J. Chem. Phys.*, 96:1982–1991, 1992.

[140] R. Gelabert, M. Moreno, and J. M. Lluch. *J. Comput. Chem.*, 15:125–131, 1994.

[141] J. P. Klinman. *Biochemistry*, 52:2068–2077, 2013.

[142] C. L. Perrin and J. D. Thoburn. *J. Am. Chem. Soc.*, 111:8010–8012, 1989.

[143] C. L. Perrin and J. D. Thoburn. *J. Am. Chem. Soc.*, 114:8559–8565, 1992.

[144] C. L. Perrin and J. B. Nielson. *J. Am. Chem. Soc.*, 119:12734–12741, 1997.

[145] C. L. Perrin and Y. J. Kim. *J. Am. Chem. Soc.*, 120:12641–12645, 1998.

[146] C. L. Perrin and B. K. Ohta. *J. Am. Chem. Soc.*, 123:6520–6526, 2001.

[147] C. L. Perrin and B. K. Ohta. *Bioorg. Chem.*, 30:3–15, 2002.

[148] O. Vendrell, R. Gelabert, M. Moreno, and J. M. Lluch. *Chem. Phys. Lett.*, 396:202–207, 2004.

[149] O. Vendrell, R. Gelabert, M. Moreno, and J. M. Lluch. *J. Chem. Theory and Comput.*, 4:1138–1150, 2008.

[150] O. Vendrell, R. Gelabert, M. Moreno, and J. M. Lluch. *J. Phys. Chem. B*, 112:13443–13452, 2008.

[151] H. D. Meyer, F. Gatti, and G. A. Worth, editors. *Multidimensional Quantum Dynamics*. Wiley-VCH, Weinheim, 2009.

[152] G. A. Worth, M. H. Beck, A. Jäckle, and H. D. Meyer. The mctdh package.

[153] H. D. Meyer, U. Manthe, and L. S. Cederbaum. *Chem. Phys. Lett.*, 165:73–78, 1990.

[154] U. Manthe, H. D. Meyer, and L. S. Cederbaum. *J. Chem. Phys.*, 97:3199–3213, 1992.

[155] M. H. Beck, A. Jäckle, G. A. Worth, and H. D. Meyer. *Phys. Rep.*, 324:1–105, 2000.

[156] H. D. Meyer and G. A. Worth. *Theor. Chem. Acc.*, 109:251–267, 2003.

[157] S. Brandt and H. D. Dahmen. *The Picture Book of Quantum Mechanics*. Springer, 4 edition, 2012.

[158] N. Piskunov. *Cálculo diferencial e integral*. Montaner y Simón S.A., Barcelona, 1970.

[159] M. Ormo, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, and S. J. Remington. *Science*, 273:1392–1395, 1996.

[160] B. L. Grigorenko, A. V. Nemukhin, I. V. Polyakov, D. I. Morozov, and A. I. Krylov. *J. Am. Chem. Soc.*, 135:11541–11549, 2013.

Part IV

APPENDIX

A

FIGURES AND TABLES

LIST OF FIGURES

LIST OF TABLES

B

SOFTWARE USED FOR PROBABILITY FLUX CALCULATIONS

As explained in Section 5.5, a specific software has been designed so as to calulculate the (hyper)surface integral of the probability current. This integral has been computed by means of an adaptative Monte Carlo method, based on the Vegas[138] algorithm. Here the code is presented.

```
program mcinteg
implicit none

integer invertible,dimensio,residu
double precision tol
integer i,j,nmostres,ncartesianes,p,maxmostr
double precision b2a,u2au,massatotal
double precision xint(1:5),xlimmin(5),xlimmax(5),mass(7)
double precision xinti1(6),xinti2(6)
double precision xcart(7),xcartmin(7),xcartmax(7),surf,resultat
double precision DensFlx
double complex :: matriucomplex(7,7)
double precision jacobian(7,7),nmostresreal
double complex psi,psiconjg,psipunt,detjacobia,hr1,hcm
double precision norma(7),modulnorma
double complex gp7(7),gp7conjg(7)
double complex gp(6)
double precision jcart(7),jint(7)

! This is for the adaptive MC scheme:
integer nboxes,nsamples,niters,restart
double precision AMCadapt
integer dof,iwhere(5)
integer nrevision,niter,restartpartials
double precision,dimension(:),allocatable :: sumf,sumf2
double precision,dimension(:),allocatable :: sumfu,sumfu2
double precision,dimension(:),allocatable :: sigma,integ
double precision,dimension(:),allocatable :: sigmafu,integfu
double precision,dimension(:,:),allocatable :: rho,AMCcum
double precision,dimension(:,:),allocatable :: rhop,AMCcump
double precision,dimension(:,:),allocatable :: rhon,AMCcumn
double precision,dimension(:,:),allocatable :: rhofp,AMCcumfp
double precision,dimension(:,:),allocatable :: rhofn,AMCcumfn
double precision chi,ibest,sigmabest
```

```
double precision chifu,ibestfu,sigmabestfu
double precision :: bxWidth(5),tmp,density
double precision AMCarea,AMCareap,AMCarean,AMCareafp,AMCareafn
double precision energy,energyhere,modul,modulmax,modulmin
double precision stepfn,nonclassical,pes_s1
integer,dimension(:,:),allocatable :: hitcount
integer ifail
double precision metric(7,7)
double precision dproduct
! end of Adaptive MC variables

!Dummy variables
double precision matriu1(7,7),matriu2(7,7)

!Reading values
read(*,*) niters,nboxes,nsamples
read(*,*) AMCadapt
read(*,*) restart
read(*,*) restartpartials
read(*,*) energy
allocate(rho(5,nboxes),AMCcum(5,nboxes),
& sumf(niters),sumf2(niters),sigma(niters),integ(niters),
& sumfu(niters),sumfu2(niters),sigmafu(niters),
& integfu(niters),
& hitcount(5,nboxes))
allocate(rhop(5,nboxes),AMCcump(5,nboxes))
allocate(rhon(5,nboxes),AMCcumn(5,nboxes))
allocate(rhofp(5,nboxes),AMCcumfp(5,nboxes))
allocate(rhofn(5,nboxes),AMCcumfn(5,nboxes))

!Maximum number of samples
maxmostr=nsamples*nboxes

!Preparing files
open(unit=100,file='statistics.out',access='append')
write(100,*) '# Niters =',niters
write(100,*) '# Nboxes =',nboxes
write(100,*) '# Nsamples =',nsamples,' -> ',maxmostr,
& ' samples per iteration'
write(100,*) '# AMCadapt =',AMCadapt
write(100,*) '# Energy =',energy
close(unit=100)
OPEN(unit=60, file="results.out")

!Explanation of variables:
!xint(1) is R1
!xint(2) is r2
!xint(3) is R2
```

```
  !xint(4) is r3
  !xint(5) is R3
  !cart(1) is the position of O1
  !cart(2) is the position of H1
  !cart(3) is the position of O2
  !cart(4) is the position of H2
  !cart(5) is the position of O3
  !cart(6) is the position of H3
  !cart(7) is the position of O4
  !mass(1), mass(3), mass(5), mass(7) are masses of oxygens
  !mass(2), mass(4), mass(6) are masses of hydrogens

  !Limits for the internal coordinates (this is the
  !range where the wave function is defined). This is in
au.
  !These limits correspond to R1,r2,R2,r3,R3
  !Later on, when dealing with 6 coordinates,
  !xint1(i+1)=xint(i), but we must bear in mind that
  !xlim(1) is the inferior limit of xint(2)

  xlimmin(1)=3.9684245685165038
  xlimmax(1)=6.2360957505259336
  xlimmin(2)=-1.7952396857574662
  xlimmax(2)=1.7952396857574662
  xlimmin(3)=3.9684245685165038
  xlimmax(3)=6.2360957505259336
  xlimmin(4)=-1.7952396857574662
  xlimmax(4)=1.7952396857574662
  xlimmin(5)=3.9684245685165038
  xlimmax(5)=6.2360957505259336

  Starting the adaptive MC
  do dof=1,5
  bxWidth(dof)=(xlimmax(dof)-xlimmin(dof))/nboxes
  if (restart == 0) then
  do i=1,nboxes
  rho(dof,i)=1.0/(nboxes*bxWidth(dof))
  rhop(dof,i)=rho(dof,i)
  rhon(dof,i)=rho(dof,i)
  rhofp(dof,i)=rho(dof,i)
  rhofn(dof,i)=rho(dof,i)
  end do
  else
  open(unit=1000,file='AMC_density_'
  & //CHAR(ICHAR('0')+restart/100)
  & //CHAR(ICHAR('0')+mod(restart/10,10))
  & //CHAR(ICHAR('0')+mod(restart,10))
```

```
  & //'_'
  & //CHAR(ICHAR('0')+dof))
  do i=1,nboxes
  if (restartpartials.EQ.0) then
  read(1000,*) tmp,tmp,rho(dof,i),hitcount(dof,i)
  else
  read(1000,*) tmp,tmp,rho(dof,i),hitcount(dof,i),
  & rhop(dof,i),rhon(dof,i),
  & rhofp(dof,i),rhofn(dof,i)
  end if
  end do
  close(unit=1000)
  end if
  end do

  If this is the first run, without restart, generate the
file with initial sampling densities
  do dof=1,5
  if (restart == 0) then
  open(unit=1000,file='AMC_density_'
  & //'000_'
  & //CHAR(ICHAR('0')+dof))
  rewind(unit=1000)
  do i=1,nboxes
  write(1000,1) xlimmin(dof)+bxWidth(dof)*(i-1),
  & xlimmin(dof)+bxWidth(dof)*(i),rho(dof,i),
  & 0,
  & rhop(dof,i),rhon(dof,i),
  & rhofp(dof,i),rhofn(dof,i)
  end do
  close(unit=1000)
  end if
  end do

  Initialize the average and variance accumulators
  do niter=1,niters
  sumf(niter)=0.0
  sumf2(niter)=0.0
  sumfu(niter)=0.0
  sumfu2(niter)=0.0
  end do
  End adaptative MC

  Atomic masses, in atomic units of mass The mass of hydrogens
2, 4 or 6 will change when analyzing other isotopes
  u2au=1.660538782e-27/9.1093826e-31
  mass(1)= 15.994915*u2au
```

```
  mass(2)=(1.007825*u2au)
  ! mass(2)=(2.014102*u2au) when dealing with a deuteron
  mass(3)=mass(1)
  mass(4)=mass(2)
  ! mass(4)=(2.014102*u2au) when dealing with a deuteron
  mass(5)=mass(1)
  mass(6)=mass(2)
  ! mass(6)=(2.014102*u2au) when dealing with a deuteron
  mass(7)=mass(1)
  massatotal=mass(1)+mass(2)+mass(3)+mass(4)+mass(5) & +
mass(6)+mass(7)
  write(*,*) 'massatotal',massatotal

  !The metric for the Cartesian->MWCartesian
  metric=0.0
  metric(1,1)=1.0/sqrt(mass(1))**2
  metric(2,2)=1.0/sqrt(mass(2))**2
  metric(3,3)=1.0/sqrt(mass(3))**2
  metric(4,4)=1.0/sqrt(mass(4))**2
  metric(5,5)=1.0/sqrt(mass(5))**2
  metric(6,6)=1.0/sqrt(mass(6))**2
  metric(7,7)=1.0/sqrt(mass(7))**2

  !The normal vector describing the flux surface
  norma(1)= -(mass(1)**(-0.5d0))/2.d0
  norma(2)=(mass(2)**(-0.5d0))
  norma(3)=-(mass(1)**(-0.5d0))/2.d0
  norma(4)=0.d0
  norma(5)=0.d0
  norma(6)=0.d0
  norma(7)=0.d0
  modulnorma=sqrt(dproduct(7,metric,norma,norma))
  norma=norma/modulnorma
  !Important: the vector is normalized using MWC metric

  !Jacobian matrix, J = d(qint)/d(qcmw
  jacobian=0.0
  jacobian(1,1) = -((mass(1))**(-0.5d0))/2.d0
  jacobian(1,2) = (mass(2))**(-0.5d0)
  jacobian(1,3) = -((mass(3))**(-0.5d0))/2.d0
  jacobian(2,1) = -(mass(1))**(-0.5d0)
  jacobian(2,3) = (mass(3))**(-0.5d0)
  jacobian(3,3) = -((mass(3))**(-0.5d0))/2.d0
  jacobian(3,4) = (mass(4))**(-0.5d0)
  jacobian(3,5) = -((mass(5))**(-0.5d0))/2.d0
  jacobian(4,3) = -(mass(3))**(-0.5d0)
  jacobian(4,5) = (mass(5))**(-0.5d0)
```

```
  jacobian(5,5) = -((mass(5))**(-0.5d0))/2.d0
  jacobian(5,6) = (mass(6))**(-0.5d0)
  jacobian(5,7) = -((mass(7))**(-0.5d0))/2.d0
  jacobian(6,5) = -(mass(5))**(-0.5d0)
  jacobian(6,7) = (mass(7))**(-0.5d0)
  jacobian(7,1) = ((mass(1))**(0.5d0))/massatotal
  jacobian(7,2) = ((mass(2))**(0.5d0))/massatotal
  jacobian(7,3) = ((mass(3))**(0.5d0))/massatotal
  jacobian(7,4) = ((mass(4))**(0.5d0))/massatotal
  jacobian(7,5) = ((mass(5))**(0.5d0))/massatotal
  jacobian(7,6) = ((mass(6))**(0.5d0))/massatotal
  jacobian(7,7) = ((mass(7))**(0.5d0))/massatotal

  !Initialize the random number generator, using timer as
seed
  call init_random_seed()
  CALL loadpsi(modulmax,modulmin)
  modul=(modulmax*modulmin)**(1./3.)
  open(unit=100,file='statistics.out',access='append')
  write(100,*) '# Discard <=',modul
  close(unit=100)

  !Continuing adaptive MC
  amc : do niter=1,niters
  do dof=1,5
  do i=1,nboxes
  AMCcum(dof,i)=0.0d0
  AMCcump(dof,i)=0.0d0
  AMCcumn(dof,i)=0.0d0
  AMCcumfp(dof,i)=0.0d0
  AMCcumfn(dof,i)=0.0d0
  hitcount(dof,i)=0
  end do
  end do
  end Adaptative MC

  !Starting the sample counter. After this, the subroutine
random_number !gets a 5-tuple of uniform random deviates
and  transform into !internal coordinates
  nmostres=0
  montecarlo : do while(nmostres.lt.maxmostr)
  call random_number(xint)

  !Now we transform xint into xinti1, that is, a system
of 6 coordinates !in which in which the first one is r1=0,
and the others will !take random values between the established
limits
```

```
  xinti1(1)=0.0d0
  !that is, xinti1(1) is r1. r1=0 defines the hypersurface
  nmostres=nmostres+1 nmostresreal=nmostres
  do dof=1,5 AMCarea=0.0 do i=1,nboxes if ((AMCarea.LT.xint(dof)).AND.
& (AMCarea+rho(dof,i)*bxWidth(dof).GE.xint(dof))) then xinti1(dof+1)=(xint(dof)-AM
& +(i-1)*bxWidth(dof) & +xlimmin(dof) iwhere(dof)=i exit
else AMCarea=AMCarea+rho(dof,i)*bxWidth(dof) end if end
do end do

  !Now we have to find the value of the wavefunction at
the given !coordinates
  psipunt=psi(xinti1,6)
  if (real(psipunt)**2+imag(psipunt)**2 >= modul**2) then
  psiconjg=conjg(psipunt)
  call gradpsi(xinti1,6,gp)
  gp7=0.d0
  do p=1,6
  gp7(p)=gp(p)
  enddo
  do p=1,7
  gp7conjg(p)= conjg(gp7(p))
  enddo
  do i=1,7
  jint(i)=dreal((0,-1)*0.5*(psiconjg*gp7(i)-
  & psipunt*gp7conjg(i)))
  enddo
  do i=1,7
  jcart(i)=0.d0
  do j=1,6
  jcart(i)=jcart(i)+jacobian(j,i)*jint(j)
  enddo
  enddo
  DensFlx=dproduct(7,metric,norma,jcart)
  else
  DensFlx=0.0
  end if
  !The factor 1/0.02418884 is for unit change into fs-1
  DensFlx=DensFlx/0.02418884d0
  energyhere=pes_s1(xinti1)
  nonclassical=stepfn(energy-energyhere)
  do dof=1,5
  AMCcum(dof,iwhere(dof))=AMCcum(dof,iwhere(dof))
  & +abs(DensFlx)/rho(dof,iwhere(dof))
  if (DensFlx.GE.0.0) then
  AMCcump(dof,iwhere(dof))=AMCcump(dof,iwhere(dof))
  & +DensFlx/rho(dof,iwhere(dof))
```

```
AMCcumfp(dof,iwhere(dof))=AMCcumfp(dof,iwhere(dof))
& +nonclassical*DensFlx/rho(dof,iwhere(dof))
else
AMCcumn(dof,iwhere(dof))=AMCcumn(dof,iwhere(dof))
& -DensFlx/rho(dof,iwhere(dof))
AMCcumfn(dof,iwhere(dof))=AMCcumfn(dof,iwhere(dof))
& -nonclassical*DensFlx/rho(dof,iwhere(dof))
end if
hitcount(dof,iwhere(dof))=hitcount(dof,iwhere(dof))+1
end do
density=1.0
do dof=1,5
density=density*rho(dof,iwhere(dof))
end do
sumf(niter)=sumf(niter)+DensFlx/density
sumf2(niter)=sumf2(niter)+(DensFlx/density)**2
sumfu(niter)=sumfu(niter)+nonclassical*DensFlx/density
sumfu2(niter)=sumfu2(niter)+
& (nonclassical*DensFlx/density)**2
residu=mod(nmostres,10)
if (residu.eq.0) then
open(unit=60,file='results.out',status='old',access='append')
write(60,*) niter,nmostres,sumf(niter)/nmostres, & sumfu(niter)/nmost
close(unit=60)
endif
end do montecarlo
open(unit=60,file='results.out',status='old',access='append')
write(60,*)
close(unit=60)
do dof=1,5
AMCarea=0.0
AMCareap=0.0
AMCarean=0.0
AMCareafp=0.0
AMCareafn=0.0
do i=1,nboxes
AMCarea=AMCarea+AMCcum(dof,i)*bxWidth(dof)
AMCareap=AMCareap+AMCcump(dof,i)*bxWidth(dof)
AMCarean=AMCarean+AMCcumn(dof,i)*bxWidth(dof)
AMCareafp=AMCareafp +AMCcumfp(dof,i)*bxWidth(dof)
AMCareafn=AMCareafn +AMCcumfn(dof,i)*bxWidth(dof)
end do
do i=1,nboxes
rho(dof,i)=(1.0-AMCadapt)*rho(dof,i)
& +AMCadapt*AMCcum(dof,i)/AMCarea
```

```
rhop(dof,i)=(1.0-AMCadapt)*rhop(dof,i)
& +AMCadapt*AMCcump(dof,i)/AMCareap
rhon(dof,i)=(1.0-AMCadapt)*rhon(dof,i)
& +AMCadapt*AMCcumn(dof,i)/AMCarean
rhofp(dof,i)=(1.0-AMCadapt)*rhofp(dof,i)
& +AMCadapt*AMCcumfp(dof,i)/AMCareafp
rhofn(dof,i)=(1.0-AMCadapt)*rhofn(dof,i)
& +AMCadapt*AMCcumfn(dof,i)/AMCareafn
end do
AMCarea=0.0
AMCareap=0.0
AMCarean=0.0
AMCareafp=0.0
AMCareafn=0.0
do i=1,nboxes
AMCarea=AMCarea+rho(dof,i)*bxWidth(dof)
AMCareap=AMCareap+rhop(dof,i)*bxWidth(dof)
AMCarean=AMCarean+rhon(dof,i)*bxWidth(dof)
AMCareafp=AMCareafp+rhofp(dof,i)*bxWidth(dof)
AMCareafn=AMCareafn+rhofn(dof,i)*bxWidth(dof)
end do
do i=1,nboxes
rho(dof,i)=rho(dof,i)/AMCarea
rhop(dof,i)=rhop(dof,i)/AMCareap
rhon(dof,i)=rhon(dof,i)/AMCarean
rhofp(dof,i)=rhofp(dof,i)/AMCareafp
rhofn(dof,i)=rhofn(dof,i)/AMCareafn
end do
open(unit=1000,file='AMC_density_'
& //CHAR(ICHAR('0')+(restart+niter)/100)
& //CHAR(ICHAR('0')+mod((restart+niter)/ 10,10))
& //CHAR(ICHAR('0')+mod(restart+niter,10))
& //'_'
& //CHAR(ICHAR('0')+dof))
rewind(unit=1000)
do i=1,nboxes
write(1000,1) xlimmin(dof)
& +bxWidth(dof)*(i-1),
& xlimmin(dof)+bxWidth(dof)*(i),rho(dof,i),
& hitcount(dof,i),
& rhop(dof,i),rhon(dof,i),
& rhofp(dof,i),rhofn(dof,i)
end do
close(unit=1000)
end do
```

```
1 format(2f10.6,2x,e12.5,2x,i4,5x,4(e12.5,1x))
integ(niter)=sumf(niter)/nmostres
sigma(niter)=sqrt( sumf2(niter)/nmostres-
integ(niter)**2)
integfu(niter)=sumfu(niter)/nmostres
sigmafu(niter)=sqrt(sumfu2(niter)/nmostres-
integfu(niter)**2)
if (niter.GE.2) then
ibest=0.0
sigmabest=0.0
ibestfu=0.0
sigmabestfu=0.0
do i=1,niter
ibest=ibest+(integ(i)/sigma(i)**2)
sigmabest=sigmabest+1.0/sigma(i)**2
ibestfu=ibestfu+(integfu(i)/sigmafu(i)**2)
sigmabestfu=sigmabestfu+1.0/sigmafu(i)**2
end do
ibest=ibest/sigmabest
sigmabest=1.0/sqrt(sigmabest)
chi=0.0
ibestfu=ibestfu/sigmabestfu
sigmabestfu=1.0/sqrt(sigmabestfu)
chifu=0.0
do i=1,niter
chi=chi+(integ(i)-ibest)**2/sigma(i)**2
chifu=chifu+(integfu(i)-ibestfu)**2/sigmafu(i)**2
end do
chi=chi/(niter-1)
chifu=chifu/(niter-1)
end if
open(unit=100,file='statistics.out',access='append')
if (niter.GE.2) then
write(100,2) niter,integ(niter),sigma(niter),ibest,sigmabest,
& chi,
& integfu(niter),sigmafu(niter),
& ibestfu,sigmabestfu,chifu
else
write(100,3) niter,integ(niter),sigma(niter),
& integfu(niter),sigmafu(niter)
end if
close(unit=100)
2 format(i3,2x,2(2(e12.5,1x),2x,3(e12.5,1x),2x))
3 format(i3,2x,2(2(e12.5,1x),2x,3(13x),2x))
end do amc
```

```
end program

!Subroutines and functions:

double precision function dproduct(ndim,metric,a,b)
implicit none
integer ndim
double precision metric(ndim,ndim),a(ndim),b(ndim)
integer i,j
dproduct=0.0
do i=1,ndim
do j=1,ndim
dproduct=dproduct+metric(j,i)*a(j)*b(i)
end do
end do
return
end

subroutine init_random_seed()
integer :: i, n, clock
integer, dimension(:), allocatable :: seed
call random_seed(size = n)
allocate(seed(n))
call system_clock(count=clock)
seed = clock + 37 * (/ (i - 1, i = 1, n) /)
call random_seed(put = seed)
deallocate(seed)
end subroutine

subroutine loadpsi(modulmax,modulmin)
implicit none
integer i1,i2,i3,i4,i5,i6,i,switch
double complex z,z1
double precision smalimag,smalreal,largimag,largreal
double precision norm,drlit,drlar,dv
double precision bohr,rlitmin,rlitmax,rlarmin,rlarmax
double precision rc(6),avge(6),modulmax,modulmin
integer ic(6)
integer imax(6)
double precision rmin(6),rmax(6)
complex*8 coeff(19,20,19,20,19,20)
common /PSIPNTS/ imax
common /PSIRNGE/ rmin,rmax
common /PSICOEF/ coeff
bohr=0.5291772e0
drlit=2*0.95/bohr/18.0d0
```

```
drlar=1.2/bohr/19.0d0
dv=drlit**3*drlar**3
rlitmin=-0.95/bohr
rlitmax=0.95/bohr
rlarmin=2.1/bohr
rlarmax=3.3/bohr
do i=1,3
rmin(2*i-1)=rlitmin
rmin(2*i)=rlarmin
rmax(2*i-1)=rlitmax
rmax(2*i)=rlarmax
end do
imax(1)=19
imax(2)=20
imax(3)=19
imax(4)=20
imax(5)=19
imax(6)=20
largimag=0.0
largreal=0.0
smalimag=0.0
smalreal=0.0
norm=0.0d0
switch=1
open(unit=92,file='wavefunction.out')
if (switch.EQ.1) then
open(unit=1,file='wavefunction',status='old',
& form='unformatted')
read(1) coeff
close(unit=1)
else
open(unit=1,file='wavefunction',status='old')
write(92,'(a)',advance='no') 'Reading Psi: row:'
do i6=1,imax(6)
write(92,'(i2,a)',advance='no') i6,', '
do i5=1,imax(5)
do i4=1,imax(4)
do i3=1,imax(3)
do i2=1,imax(2)
do i1=1,imax(1)
read(1,*) z1
coeff(i1,i2,i3,i4,i5,i6)=z1
end do
end do
end do
```

```
 end do
 end do
 end do
 close(unit=1)
 end if
 write(92,*) 'Done!'
 write(92,*) 'Coefficients read in.'
 do i=1,6
 avge(i)=0.0e0
 end do
 open(unit=2,file='largest_psi2ex.out',status='unknown')
 write(92,'(a)') 'Computing Statistics...'
 modulmax=0.0
 modulmin=0.0
 do i6=1,imax(6)
 do i5=1,imax(5)
 do i4=1,imax(4)
 do i3=1,imax(3)
 do i2=1,imax(2)
 do i1=1,imax(1)
 !Get the Labs RF coordinates:
 ic(1)=i1
 ic(2)=i2
 ic(3)=i3
 ic(4)=i4
 ic(5)=i5
 ic(6)=i6
 call labcoords(ic,rc)
 !This is to normalize the wave function, taking into account
!the grid size
 z=coeff(i1,i2,i3,i4,i5,i6)/sqrt(dv)
 ! if (real(z).GE.largreal) then
 ! write(2,*) (rc(i),i=1,6),real(z)
 ! end if
 largimag=max(largimag,imag(z))
 smalimag=min(smalimag,imag(z))
 largreal=max(largreal,real(z))
 smalreal=min(smalreal,real(z))
 modulmax=max(modulmax,real(z)**2+imag(z)**2)
 if (z /= (0.0,0.0)) then
 modulmin=max(modulmin,1./ (real(z)**2+imag(z)**2))
 end if
 norm=norm+(real(z)**2+imag(z)**2)*dv
 do i=1,6
 avge(i)=avge(i) +rc(i)*(real(z)**2+imag(z)**2)*dv
```

```
  end do
  !This is to estimate the coefficient of the sinc-DVR expansion
!of Psi direct from the value of the wave function at grid
points
  coeff(i1,i2,i3,i4,i5,i6)=z*dv
  end do
  end do
  end do
  end do
  end do
  end do
  close(unit=2)
  modulmax=sqrt(modulmax)
  modulmin=1./sqrt(modulmin)
  write(92,*) ' Largest Real ',largreal
  write(92,*) ' Smallest Real ',smalreal
  write(92,*) ' Largest Imag ',largimag
  write(92,*) ' Smallest Imag ',smalimag
  write(92,*) ' Norm ',norm
  write(92,*) ' dV ',dv
  write(92,*) ' Largest Module ',modulmax
  write(92,*) ' Smallest NZ Module',modulmin
  write(92,*) 'Average Values of Coordinates:'
  write(92,'(6(2x,f8.5))') (avge(i)/norm,i=1,6)
  close(unit=92)
  end

  double complex function psi(q,n)
  implicit none
  integer n
  double precision q(n)
  integer i,j,i1,i2,i3,i4,i5,i6
  double precision bohr,dv,pi
  double precision dx(n),contr,r(n),b
  double precision arg(n,100)
  integer imax(6)
  double precision rmin(6),rmax(6)
  complex*8 coeff(19,20,19,20,19,20)
  common /PSIPNTS/ imax
  common /PSIRNGE/ rmin,rmax
  common /PSICOEF/ coeff
  if (n.ne.6) then
  write(6,*) 'Error: n=',n
  stop
  end if
```

```
! bohr=0.5291772d0
bohr=1.0
pi=4.0e0*atan(1.0e0)
dv=1.0e0
do i=1,n
dx(i)=(rmax(i)-rmin(i))/(imax(i)-1)
dv=dv*dx(i)
r(i)=q(i)/bohr
end do
do i=1,n
do j=1,imax(i)
b=rmin(i)+(j-1)*dx(i)
arg(i,j)=contr(r(i),b,dx(i))
end do
end do
psi=(0.0e0,0.0e0)
do i6=1,imax(6)
do i5=1,imax(5)
do i4=1,imax(4)
do i3=1,imax(3)
do i2=1,imax(2)
do i1=1,imax(1)
b=arg(1,i1)*arg(2,i2)*arg(3,i3)
b=b*arg(4,i4)*arg(5,i5)*arg(6,i6)
psi=psi+coeff(i1,i2,i3,i4,i5,i6)*b
end do
end do
end do
end do
end do
end do
return
end

subroutine gradpsi(q,n,gp)
implicit none
integer n
double precision q(n),mga
double complex gp(n),gn(n),tmp1,tmp2,psi,wfn
integer i,j,i1,i2,i3,i4,i5,i6
double precision bohr,dv,pi
double precision dx(n),contr,dcontr,r(n),b,b1,b2,b3,b4,b5,b6
double precision arg1(n,100),arg2(n,100)
integer imax(6)
double precision rmin(6),rmax(6)
```

```
  complex*8 coeff(19,20,19,20,19,20)
  common /PSIPNTS/ imax
  common /PSIRNGE/ rmin,rmax
  common /PSICOEF/ coeff
  if (n.ne.6) then
  write(6,*) 'Error: n=',n
  stop
  end if
  ! bohr=0.5291772d0
  bohr=1.0
  pi=4.0e0*atan(1.0e0)
  dv=1.0e0
  do i=1,n
  dx(i)=(rmax(i)-rmin(i))/(imax(i)-1)
  dv=dv*dx(i)
  r(i)=q(i)/bohr
  gp(i)=(0.0e0,0.0e0)
  end do
  arg(i,j) holds the contribution of the ith variable-DVR
function in its jth point
  do i=1,n
  do j=1,imax(i)
  b=rmin(i)+(j-1)*dx(i)
  arg1(i,j)=contr(r(i),b,dx(i))
  arg2(i,j)=dcontr(r(i),b,dx(i))
  end do
  end do
  do i6=1,imax(6)
  do i5=1,imax(5)
  do i4=1,imax(4)
  do i3=1,imax(3)
  do i2=1,imax(2)
  do i1=1,imax(1)
  b1=arg2(1,i1)*arg1(2,i2)*arg1(3,i3)
  b1=b1*arg1(4,i4)*arg1(5,i5)*arg1(6,i6)
  b2=arg1(1,i1)*arg2(2,i2)*arg1(3,i3)
  b2=b2*arg1(4,i4)*arg1(5,i5)*arg1(6,i6)
  b3=arg1(1,i1)*arg1(2,i2)*arg2(3,i3)
  b3=b3*arg1(4,i4)*arg1(5,i5)*arg1(6,i6)
  b4=arg1(1,i1)*arg1(2,i2)*arg1(3,i3)
  b4=b4*arg2(4,i4)*arg1(5,i5)*arg1(6,i6)
  b5=arg1(1,i1)*arg1(2,i2)*arg1(3,i3)
  b5=b5*arg1(4,i4)*arg2(5,i5)*arg1(6,i6)
  b6=arg1(1,i1)*arg1(2,i2)*arg1(3,i3)
  b6=b6*arg1(4,i4)*arg1(5,i5)*arg2(6,i6)
```

```
gp(1)=gp(1)+coeff(i1,i2,i3,i4,i5,i6)*b1
gp(2)=gp(2)+coeff(i1,i2,i3,i4,i5,i6)*b2
gp(3)=gp(3)+coeff(i1,i2,i3,i4,i5,i6)*b3
gp(4)=gp(4)+coeff(i1,i2,i3,i4,i5,i6)*b4
gp(5)=gp(5)+coeff(i1,i2,i3,i4,i5,i6)*b5
gp(6)=gp(6)+coeff(i1,i2,i3,i4,i5,i6)*b6
end do
end do
end do
end do
end do
end do
mga=0.0
do i1=1,6
mga=mga+gp(i1)*conjg(gp(i1))
end do
wfn=psi(q,n)
open(file='gradpsi.out',access='append',unit=19)
write(19,2) q,dreal(sqrt(wfn*conjg(wfn))),sqrt(mga)
close(unit=19)
2 format(6(f6.3,1x),4(5x,e10.3))
return
end

double precision function contr(x,xi,dx)
implicit none
double precision x,xi,dx,pi,sinc
parameter(pi=3.141592654e0)
contr=sinc(pi*(x-xi)/dx)/dx
return
end

double precision function dcontr(x,xi,dx)
implicit none
double precision x,xi,dx,pi,dsinc
parameter(pi=3.141592654e0)
dcontr=pi*dsinc(pi*(x-xi)/dx)/dx/dx
return
end

double precision function sinc(x)
implicit none
double precision x
if (x.eq.0.0d0) then
sinc=1.0e0
else
```

```
sinc=sin(x)/x
end if
return
end

double precision function dsinc(x)
implicit none
double precision x,sinc
if (x.EQ.0.0e0) then
dsinc=0.0e0
else
dsinc=(cos(x)-sinc(x))/x
end if
return
end

subroutine labcoords(ix,ir)
implicit none
integer ix(6)
double precision ir(6)
integer imax(6)
double precision rmin(6),rmax(6)
common /PSIPNTS/ imax
common /PSIRNGE/ rmin,rmax
double precision bohr,x
integer i
! parameter(bohr=0.5291772d0)
parameter(bohr=1.0)
do i=1,6
x=rmin(i)+(rmax(i)-rmin(i))*(ix(i)-1)/(imax(i)-1)
ir(i)=x*bohr
end do
return
end

subroutine producte(a,b,c,n)
implicit none
integer n
double precision, dimension (1:n,1:n),intent(in) :: a,b
double precision, dimension (1:n,1:n),intent(out) :: c
integer :: i,j,k
DO i=1,n
DO j=1,n
c(i,j)=0.0
DO k=1,n
c(i,j)=c(i,j)+(a(i,k)*b(k,j))
```

```
  END DO
  END DO
  END DO
  end subroutine producte

  double precision function stepfn(k)
  implicit none
  double precision k
  if (k > 0.0) then
  stepfn=0.0
  else if (k == 0.0) then
  stepfn=0.5
  else
  stepfn=1.0
  end if
  end
  !This routine expects a 6-element double precision vector
!containing the coordinates (in bohr) of the GFP chromophore's
!QM atoms, in internal, and returns the potential of either
!the ground or photoactive states, depending on the value
of the !_S0_environment flag on compilation. The order expected
is ! q(1) -> r1 ! q(2) -> r2 ! q(3) -> r3 ! q(4) -> R1 !
q(5) -> R2 ! q(6) -> R3

  double precision function pes_s1(q1)
  implicit none
  double precision q(6),q1(6)
  double precision bohr
  parameter(bohr=0.529177249d0)
  integer ndim
  parameter(ndim=8)
  double precision evb(ndim,ndim)
  double precision evbvec(ndim*(ndim+1)/2)
  double precision energy(ndim),evectors(ndim,ndim)
  double precision contact,contactE,kc2au
  double precision morse,lj6
  integer i,j,k
  double precision x(7)
  include the parameters for the PES
  integer np
  parameter(np=189)
  double precision p(np)
  Reorder geometrical coordinates to match the flux program's
  q(1)=q1(1)
  q(2)=q1(3)
  q(3)=q1(5)
```

```
q(4)=q1(2)
q(5)=q1(4)
q(6)=q1(6)
Parameters for the potential for S1:
p(1)= 142.190559945
p(2)= 18.3620386102
p(3)= 3.56170681795
p(4)= 193.635267312
p(5)= -123.234038603
p(6)= 70.4762539303
p(7)= 92.5154271404
p(8)= 168.468388925
p(9)= -37.9670025715
p(10)=104.494192468
p(11)=2.50651066157
p(12)=37.6885656118
p(13)=147.752799451
p(14)=24.8708241035
p(15)=232.20059949
p(16)=1.31220511086
p(17)=0.873663861515
p(18)=65.3176263947
p(19)=2.53534313664
p(20)=0.992872040573
p(21)=113.45728281
p(22)=2.3211793805
p(23)=0.984258541985
p(24)=1.34865222347
p(25)=2.58359668956
p(26)=1.9324420032
p(27)=2.47890401121
p(28)=1.42826493796
p(29)=2.54049713448
p(30)=0.20395231261
p(31)=1.38179573524
p(32)=0.94321272117
p(33)=-1.02993158257
p(34)=0.261246681108
p(35)=1.22120202294
p(36)=-39.7378559304
p(37)=287.227749418
p(38)=1.37921699056
p(39)=-0.275189134845
p(40)=287.227749418
p(41)=1.37921699056
```

```
p(42)=-0.275189134845
p(43)=113.45728281
p(44)=2.3211793805
p(45)=0.984258541985
p(46)=1.34865222347
p(47)=2.58359668956
p(48)=1.9324420032
p(49)=2.47890401121
p(50)=1.42826493796
p(51)=2.54049713448
p(52)=5.75188592986
p(53)=0.959083815615
p(54)=0.94321272117
p(55)=-1.02993158257
p(56)=0.261246681108
p(57)=1.22120202294
p(58)=140.001104762
p(59)=232.20059949
p(60)=1.31220511086
p(61)=0.873663861515
p(62)=52.9773011476
p(63)=1.34939257362
p(64)=0.855038526175
p(65)=52.9773011476
p(66)=1.34939257362
p(67)=0.855038526175
p(68)=1.34865222347
p(69)=2.58359668956
p(70)=1.9324420032
p(71)=2.47890401121
p(72)=1.42826493796
p(73)=2.54049713448
p(74)=0.20395231261
p(75)=1.38179573524
p(76)=0.390468780898
p(77)=-1.17737014446
p(78)=0.261246681108
p(79)=1.22120202294
p(80)=95.6695372753
p(81)=232.20059949
p(82)= 1.31220511086
p(83)=0.873663861515
p(84)=65.3176263947
p(85)=2.53534313664
p(86)=0.992872040573
```

```
p(87)=169.939471656
p(88)=1.5275611813
p(89)=0.876051111445
p(90)=1.34865222347
p(91)= 2.58359668956
p(92)=1.9324420032
p(93)=2.47890401121
p(94)=1.42826493796
p(95)=2.54049713448
p(96)=0.20395231261
p(97)=1.38179573524
p(98)=0.94321272117
p(99)=-1.02993158257
p(100)=0.55824418919
p(101)=1.09351203474
p(102)=-10.0293405614
p(103)=287.227749418
p(104)=1.37921699056
p(105)=-0.275189134845
p(106)=52.9773011476
p(107)=1.34939257362
p(108)=0.855038526175
p(109)=52.9773011476
p(110)=1.34939257362
p(111)=0.855038526175
p(112)=1.34865222347
p(113)=2.58359668956
p(114)=1.9324420032
p(115)= 2.47890401121
p(116)=1.42826493796
p(117)=2.54049713448
p(118)=5.75188592986
p(119)=0.959083815615
p(120)=0.390468780898
p(121)=-1.17737014446
p(122)=0.261246681108
p(123)=1.22120202294
p(124)=-213.736147399
p(125)=287.227749418
p(126)=1.37921699056
p(127)=-0.275189134845
p(128)=287.227749418
p(129)=1.37921699056
p(130)=-0.275189134845
p(131)=169.939471656
```

```
p(132)=1.5275611813
p(133)=0.876051111445
p(134)=1.34865222347
p(135)=2.58359668956
p(136)=1.9324420032
p(137)=2.47890401121
p(138)=1.42826493796
p(139)=2.54049713448
p(140)=5.75188592986
p(141)=0.959083815615
p(142)=0.94321272117
p(143)=-1.02993158257
p(144)=0.55824418919
p(145)=1.09351203474
p(146)=59.10316746
p(147)=232.20059949
p(148)=1.31220511086
p(149)=0.873663861515
p(150)=113.45728281
p(151)=2.3211793805
p(152)=0.984258541985
p(153)=169.939471656
p(154)=1.5275611813
p(155)=0.876051111445
p(156)=1.34865222347
p(157)=2.58359668956
p(158)=1.9324420032
p(159)=2.47890401121
p(160)=1.42826493796
p(161)=2.54049713448
p(162)=0.20395231261
p(163)=1.38179573524
p(164)=0.390468780898
p(165)=-1.17737014446
p(166)=0.55824418919
p(167)=1.09351203474
p(168)=0.0
p(169)=65.3176263947
p(170)=2.53534313664
p(171)=0.992872040573
p(172)=113.45728281
p(173)=2.3211793805
p(174)=0.984258541985
p(175)=169.939471656
p(176)=1.5275611813
```

```
  p(177)=0.876051111445
  p(178)=1.34865222347
  p(179)=2.58359668956
  p(180)=1.9324420032
  p(181)=2.47890401121
  p(182)=1.42826493796
  p(183)= 2.54049713448
  p(184)=5.75188592986
  p(185)=0.959083815615
  p(186)=0.390468780898
  p(187)=-1.17737014446
  p(188)=0.55824418919
  p(189)=1.09351203474
  Input is in au, transfer it to angstrom, and lab. ref.
frame
  x(1)=0.0d0
  x(2)=q(4)/2.0d0+q(1)
  x(3)=q(4)
  x(4)=q(4)+q(5)/2.0d0+q(2)
  x(5)=q(4)+q(5)
  x(6)=q(4)+q(5)+q(6)/2.0d0+q(3)
  x(7)=q(4)+q(5)+q(6)
  do i=1,7
  x(i)=x(i)*bohr
  end do
  kc2au=0.0015936145299079
  Check for close contacts
  contact=0.7
  contactE=150.0d0
  do i=1,3
  if (abs(x(2*i-1)-x(2*i)).LE.contact.OR.
  & abs(x(2*i+1)-x(2*i)).LE.contact) then
  pes_s1=contactE*kc2au
  return
  end if
  end do
  Initialize evb matrix
  do i=1,ndim
  do j=1,ndim
  evb(j,i)=0.0d0
  end do
  end do
  Off-diagonal elements
  evb(1,2)=p(1)
  evb(1,3)=p(2)
```

```
evb(1,4)=p(3)
evb(2,5)=p(4)
evb(2,6)=p(5)
evb(3,5)=p(6)
evb(3,7)=p(7)
evb(4,6)=p(8)
evb(4,7)=p(9)
evb(5,8)=p(10)
evb(6,8)=p(11)
evb(7,8)=p(12)
Protonation states
evb(1,1)=p(14)
& +morse(p(15),p(16),p(17),x(2)-x(1))
& +morse(p(18),p(19),p(20),x(4)-x(3))
& +morse(p(21),p(22),p(23),x(6)-x(5))
& +lj6(p(24),p(25),x(3)-x(1))
& +lj6(p(26),p(27),x(5)-x(3))
& +lj6(p(28),p(29),x(7)-x(5))
& +lj6(p(30),p(31),x(3)-x(2))
& +lj6(p(32),p(33),x(5)-x(4))
& +lj6(p(34),p(35),x(7)-x(6)) evb(2,2)=p(36)
& +morse(p(37),p(38),p(39),x(3)-x(2))
& +morse(p(40),p(41),p(42),x(4)-x(3))
& +morse(p(43),p(44),p(45),x(6)-x(5))
& +lj6(p(46),p(47),x(3)-x(1))
& +lj6(p(48),p(49),x(5)-x(3))
& +lj6(p(50),p(51),x(7)-x(5))
& +lj6(p(52),p(53),x(2)-x(1))
& +lj6(p(54),p(55),x(5)-x(4))
& +lj6(p(56),p(57),x(7)-x(6))
evb(3,3)=p(58)
& +morse(p(59),p(60),p(61),x(2)-x(1))
& +morse(p(62),p(63),p(64),x(5)-x(4))
& +morse(p(65),p(66),p(67),x(6)-x(5))
& +lj6(p(68),p(69),x(3)-x(1))
& +lj6(p(70),p(71),x(5)-x(3))
& +lj6(p(72),p(73),x(7)-x(5))
& +lj6(p(74),p(75),x(3)-x(2))
& +lj6(p(76),p(77),x(4)-x(3))
& +lj6(p(78),p(79),x(7)-x(6))
evb(4,4)=p(80)
& +morse(p(81),p(82),p(83),x(2)-x(1))
& +morse(p(84),p(85),p(86),x(4)-x(3))
& +morse(p(87),p(88),p(89),x(7)-x(6))
& +lj6(p(90),p(91),x(3)-x(1))
```

```
& +lj6(p(92),p(93),x(5)-x(3))
& +lj6(p(94),p(95),x(7)-x(5))
& +lj6(p(96),p(97),x(3)-x(2))
& +lj6(p(98),p(99),x(5)-x(4))
& +lj6(p(100),p(101),x(6)-x(5))
evb(5,5)=p(102)
& +morse(p(103),p(104),p(105),x(3)-x(2))
& +morse(p(106),p(107),p(108),x(5)-x(4))
& +morse(p(109),p(110),p(111),x(6)-x(5))
& +lj6(p(112),p(113),x(3)-x(1))
& +lj6(p(114),p(115),x(5)-x(3))
& +lj6(p(116),p(117),x(7)-x(5))
& +lj6(p(118),p(119),x(2)-x(1))
& +lj6(p(120),p(121),x(4)-x(3))
& +lj6(p(122),p(123),x(7)-x(6))
evb(6,6)=p(124)
& +morse(p(125),p(126),p(127),x(3)-x(2))
& +morse(p(128),p(129),p(130),x(4)-x(3))
& +morse(p(131),p(132),p(133),x(7)-x(6))
& +lj6(p(134),p(135),x(3)-x(1))
& +lj6(p(136),p(137),x(5)-x(3))
& +lj6(p(138),p(139),x(7)-x(5))
& +lj6(p(140),p(141),x(2)-x(1))
& +lj6(p(142),p(143),x(5)-x(4))
& +lj6(p(144),p(145),x(6)-x(5))
evb(7,7)=p(146)
& +morse(p(147),p(148),p(149),x(2)-x(1))
& +morse(p(150),p(151),p(152),x(5)-x(4))
& +morse(p(153),p(154),p(155),x(7)-x(6))
& +lj6(p(156),p(157),x(3)-x(1))
& +lj6(p(158),p(159),x(5)-x(3))
& +lj6(p(160),p(161),x(7)-x(5))
& +lj6(p(162),p(163),x(3)-x(2))
& +lj6(p(164),p(165),x(4)-x(3))
& +lj6(p(166),p(167),x(6)-x(5))
evb(8,8)=p(168)
& +morse(p(169),p(170),p(171),x(3)-x(2))
& +morse(p(172),p(173),p(174),x(5)-x(4))
& +morse(p(175),p(176),p(177),x(7)-x(6))
& +lj6(p(178),p(179),x(3)-x(1))
& +lj6(p(180),p(181),x(5)-x(3))
& +lj6(p(182),p(183),x(7)-x(5))
& +lj6(p(184),p(185),x(2)-x(1))
& +lj6(p(186),p(187),x(4)-x(3))
& +lj6(p(188),p(189),x(6)-x(5))
```

```
  Diagonalize EVB matrix...
  k=1
  do i=1,ndim
  do j=1,i
  evbvec(k)=evb(min(i,j),max(i,j))
  k=k+1
  end do
  end do
  call givens(ndim,-1,ndim,evbvec,energy,evectors)
  Energy has been computed in kcal/mol, turn into hartrees.
Add global shift.
  pes_s1=(p(13)+energy(ndim))*kc2au
  return
  end

  double precision function morse(a,b,c,x)
  implicit none
  double precision a,b,c,x
  morse=1.0d0-exp(-b*(abs(x)-c))
  morse=a*morse*morse
  return
  end

  double precision function dwell(a,b,c,d,e,x)
  implicit none
  double precision a,b,c,d,e,x
  dwell=a+x*(b+x*(c+x*(d+e*x)))
  return
  end

  double precision function lj6(a,b,x)
  implicit none
  double precision a,b,x,t
  t=(b/abs(x))**6
  lj6=4.0d0*a*(t*t-t)
  return
  end
  ! SUBROUTINE GIVENS (NX,NROOTX,NJX,A,B,ROOT,VECT) SUBROUTINE
GIVENS (NX,NROOTX,NJX,A,ROOT,VECT)
  DATA ETA/1.D-15/,THETA/1.D75/
  del1=eta/1.d2
  delta=eta*eta*1.d2
  small=eta*eta/1.d2
  delbig=theta*delta/1.d3
  theta1=1.d3/theta
  emag=eta
```

```
toler=1.d-2*sqrt(eta)
N = NX
FLOATN = DFLOAT(N)
NROOT = IABS(NROOTX)
IF (NROOT.EQ.0) GO TO 1001
IF (N-1) 1001,1003,105
1003 ROOT(1) = A(1)
IF (NROOTX.GT.0) VECT(1,1) = 1.D0
GO TO 1001
105 CONTINUE
NSIZE = (N*(N+1))/2
NM1 = N-1
NM2 = N-2
NP1 = N+1
TRACE = 0.D0
JUMP = 1
DO 1 J=2,NP1
TRACE = TRACE + A(JUMP)
1 JUMP = JUMP + J
TRACE = TRACE/FLOATN
JUMP = 1
DO 2 J=2,NP1
A(JUMP) = A(JUMP) - TRACE
2 JUMP = JUMP + J
SCALE MATRIX TO EUCLIDEAN NORM OF 1. SCALE FACTOR IS ANORM.
FACTOR = 0.D0
DO 70 I=1,NSIZE
70 FACTOR = DMAX1(FACTOR,ABS(A(I)))
IF (FACTOR.NE.0.D0) GO TO 72
DO 78 I=1,NROOT
IF (NROOTX.LT.0) GO TO 78
DO 77 J=1,N
77 VECT(J,I) = 0.D0
VECT(I,I) = 1.D0
78 ROOT(I) = 0.D0
GO TO 1001
72 ANORM = 0.D0
86 SCALE = 1.D0/FACTOR
DO 80 I=1,NSIZE
80 ANORM = ANORM + (A(I)*SCALE)**2
ANORM = ANORM+ANORM
JUMP = 1
DO 81 J=2,NP1
ANORM = ANORM -(A(JUMP)*SCALE)**2
81 JUMP = JUMP + J
```

```
83 ANORM = FACTOR*SQRT(ANORM)
SCALE = 1.D0/ANORM
DO 91 I=1,NSIZE
91 A(I) = A(I)*SCALE
ALIMIT = 1.D0
ID = 0
IA = 1
IF (NM2.EQ.0) GO TO 201
DO 200 J=1,NM2
IA = IA+J+2
ID = ID+J+1
JP2 = J + 2
J1 = J + 1
LIMIT = J1
II = IA
DO 99 I=JP2,N
B(I,5) = A(II)
IF (ABS(B(I,5)).GT.DEL1) LIMIT = I
99 II = II + I
DTEMP = A(ID)
IF (LIMIT.GT.J1) GO TO 110
120 B(J,1) = DTEMP
A(ID) = 0.D0
GO TO 200
110 IDIF = LIMIT -JP2
SUM = DOT(B(JP2,5),B(JP2,5))
SUM = SQRT(SUM + DTEMP**2)
B(J,1) = -SIGN(SUM,DTEMP)
B(J+1,2) = SQRT((1.D0 + ABS(DTEMP)/SUM)*5.D-1)
TEMP = SIGN(5.D-1/(B(J+1,2)*SUM),DTEMP)
II = IA
DO 130 I=JP2,LIMIT
B(I,2) = A(II)*TEMP
130 II = II + I
DAK = 0.D0
IC = ID + 1
LIMLES = LIMIT - 1
DO 188 I=J1,LIMLES
IDIF = I - J1
DTEMP = DOT(B(J1,2),A(IC))
IC = IC + I
178 IP1 = I + 1
JJ = IC + IDIF
DTEMP = DTEMP + DSUMA(B(N,1),A(JJ),IP1,LIMIT)
DAK = DAK + DTEMP*B(I,2)
```

```
188 B(I,1) = DTEMP
IDIF = LIMIT - J1
DTEMP = DOT(B(J1,2),A(IC))
DAK = DAK + DTEMP*B(LIMIT,2)
B(LIMIT,1) = DTEMP
IDIF = LIMIT - J1
IF (LIMIT.EQ.N) GO TO 190
IC = IC + LIMIT
LIMLO = LIMIT + 1
DO 189 I=LIMLO,N
B(I,1) = DOT(B(J1,2),A(IC))
B(I,2) = 0.D0
189 IC = IC + I
190 FACT = -DAK
CALL VECSUM(B(J1,1),B(J1,2))
JJ = ID
DO 160 I=J1,N
A(JJ) = B(I,2)
IF (I.GT.LIMIT) GO TO 161
B2 = B(I,2)
FACT = -B2 - B2
IDIF = I - J1
CALL VECSUM(A(JJ+1),B(J1,1))
161 B1 = B(I,1)
FACT = -B1 - B1
IDIF = MIN0(I,LIMIT) - J1
CALL VECSUM(A(JJ+1),B(J1,2))
160 JJ = JJ + I
200 B(J,6) = LIMIT
201 CONTINUE
B(NM1,1) = A(NSIZE-1)
A(NSIZE-1) = 0.D0
JUMP = 0
DO 320 J=1,NM1
JUMP = JUMP + J
ROOT(J) = A(JUMP)
320 B(J,3) = B(J,1)**2
ROOT(N) = A(NSIZE)
CALL EVQR(ROOT,B(1,3),N,30,SMALL)
DO 325 J=1,N
325 B(J,2) = ROOT(J)
DO 330 I=1,NM1
IP1 = I + 1
DO 330 J=IP1,N
IF (ROOT(I).GE.ROOT(J)) GO TO 330
```

```
TEMP = ROOT(I)
ROOT(I) = ROOT(J)
ROOT(J) = TEMP
330 CONTINUE
IF (NROOTX.LT.0) GO TO 1002
IF (NROOTX.EQ.1) GO TO 807
NTOP = NROOT - 1
DO 400 I=1,NTOP
IF (ABS(ROOT(I+1)-ROOT(I)).LE.TOLER) GO TO 410
400 CONTINUE
GO TO 807
410 JUMP = 0
DO 440 J=1,NM1
JUMP = JUMP + J
ROOT(J) = A(JUMP)
440 B(J,3) = B(J,1)
ROOT(N) = A(NSIZE)
DO 450 I=1,N
DO 445 J=1,N
445 VECT(J,I) = 0.D0
450 VECT(I,I) = 1.D0
CALL QRTN(ROOT,B(1,3),VECT,B(1,2),N,25,EMAG,NJX)
DO 456 I=1,NM1
IP1 = I + 1
DO 455 J=IP1,N
FLIP = VECT(I,J)
VECT(I,J) = VECT(J,I)
455 VECT(J,I) = FLIP
456 CONTINUE
ITOP = NROOT-1
DO 480 I=1,ITOP
IP1 = I+1
DO 480 J=IP1,N
IF (ROOT(I).GE.ROOT(J)) GO TO 480
TEMP = ROOT(I)
ROOT(I) = ROOT(J)
ROOT(J) = TEMP
DO 470 K=1,N
TEMP = VECT(K,I)
VECT(K,I) = VECT(K,J)
470 VECT(K,J) = TEMP
480 CONTINUE
GO TO 940
807 CONTINUE
DO 705 I=1,NROOT
```

```
      DO 15 J=1,N
   15 VECT(J,I) = 1.D0
  705 CONTINUE
      DO 700 I=1,NROOT
  701 AROOT = ROOT(I)
      ELIM1 = A(1) - AROOT
      ELIM2 = B(1,1)
      JUMP = 1
      DO 750 J=1,NM1
      JUMP = JUMP + J + 1
      IF (ABS(ELIM1).LE.ABS(B(J,1))) GO TO 760
      B(J,2) = ELIM1
      B(J,3) = ELIM2
      B(J,4) = 0.D0
      TEMP = B(J,1)/ELIM1
      ELIM1 = A(JUMP) - AROOT - TEMP*ELIM2
      ELIM2 = B(J+1,1)
      GO TO 755
  760 B(J,2) = B(J,1)
      B(J,3) = A(JUMP) - AROOT
      B(J,4) = B(J+1,1)
      TEMP = 1.D0
      IF (ABS(B(J,1)).GT.THETA1) TEMP = ELIM1/B(J,1)
      ELIM1 = ELIM2 - TEMP*B(J,3)
      ELIM2 = -TEMP*B(J+1,1)
  755 B(J,5) = TEMP
  750 CONTINUE
      B(N,2) = ELIM1
      B(N,3) = 0.D0
      B(N,4) = 0.D0
      B(NM1,4) = 0.D0
      ITER = 1
  790 L = N + 1
      DO 780 J=1,N
      L = L - 1
  786 CONTINUE
      ELIM1 = VECT(L,I)-VECT(L+1,I)*B(L,3)-VECT(L+2,I)*B(L,4)
      IF (ABS(ELIM1).GT.DELBIG) GO TO 782
      TEMP = B(L,2)
      IF (ABS(B(L,2)).LT.DELTA) TEMP = DELTA
      VECT(L,I) = ELIM1/TEMP
      GO TO 780
  782 DO 784 K=1,N
  784 VECT(K,I) = VECT(K,I)/DELBIG
      GO TO 786
```

```
780 CONTINUE
GO TO (820,900), ITER
820 ITER = ITER + 1
890 ELIM1 = VECT(1,I)
DO 830 J=1,NM1
IF (B(J,2).EQ.B(J,1)) GO TO 840
850 VECT(J,I) = ELIM1
ELIM1 = VECT(J+1,I) - ELIM1*B(J,5)
GO TO 830
840 VECT(J,I) = VECT(J+1,I)
ELIM1 = ELIM1 - VECT(J+1,I)*TEMP
830 CONTINUE
VECT(N,I) = ELIM1
GO TO 790
900 ELIM1 = 0.D0
DO 904 J=1,N
904 ELIM1 = DMAX1(ABS(VECT(J,I)),ELIM1)
TEMP = 0.D0
DO 910 J=1,N
ELIM2 = VECT(J,I)/ELIM1
910 TEMP = TEMP + ELIM2**2
TEMP = 1.D0/(SQRT(TEMP)*ELIM1)
DO 920 J=1,N
VECT(J,I) = VECT(J,I)*TEMP
IF (ABS(VECT(J,I)).LT.DEL1) VECT(J,I) = 0.D0
920 CONTINUE
700 CONTINUE
940 IF (NM2.EQ.0) GO TO 1002
JUMP = NSIZE - NP1
IM = NM1
DO 950 I=1,NM2
LIMIT = IDINT(B(IM-1,6))
J1 = JUMP
DO 955 J=IM,LIMIT
B(J,2) = A(J1)
955 J1 = J1 + J
IDIF = LIMIT - IM
DO 960 K=1,NROOT
TMP = DOT(B(IM,2),VECT(IM,K))
FACT = -TMP - TMP
CALL VECSUM(VECT(IM,K),B(IM,2))
960 CONTINUE
JUMP = JUMP - IM
950 IM = IM - 1
1002 CONTINUE
```

```
  DO 95 I=1,N
  95 ROOT(I) = ROOT(I)*ANORM + TRACE
  1001 RETURN
  END
  real*8 FUNCTION DSUMA(B,A,IP1,LIMIT)
  IMPLICIT REAL*8(A-H,O-Z)
  DIMENSION B(1),A(1)
  JJ=1
  DSUMA=0.D0
  DO 180 II=IP1,LIMIT
  DSUMA=DSUMA+B(II+1)*A(JJ)
  180 JJ=JJ+II
  RETURN
  END

  SUBROUTINE VECSUM(A,B)
  IMPLICIT REAL*8(A-H,O-Z)
  DIMENSION A(1),B(1)
  COMMON /VECTOR/ FACTOR,LIMIT
  ITOP=LIMIT+1
  DO 10 I=1,ITOP
  10 A(I)=A(I)+FACTOR*B(I)
  RETURN
  END

  SUBROUTINE EVQR(A,B,N,M,TOL)
  IMPLICIT REAL*8(A-H,O-Z)
  DIMENSION A(1),B(1)
  NX = N
  NI = 1
  SH = 0.D0
  K = 0
  IF (NX-2) 50,60,85
  100 K = K + 1
  IF (K.LT.M) GO TO 101
  WRITE (6,1000) K
  1000 FORMAT(35H NO CONVERGENCE OF QR ALGORITHM IN ,I4,11H
ITERATIONS)
  IFAIL=1
  return
  101 AT = A(NX) + A(NX-1)
  ST = AT*5.D-1
  DISC = AT**2 - 4.D0*(A(NX)*A(NX-1)-B(NX-1))
  IF (DISC.LE.0.D0) GO TO 15
  ST = ST - SIGN(SQRT(DISC),ST)*5.D-1
  15 SH = SH + ST
```

```
DO 20 I=1,NX
20 A(I) = A(I) - ST
G = A(NI)
PS = G**2
RS = PS + B(NI)
SX = B(NI)/RS
CXS = 1.D0
CX = PS/RS
U = SX*(G+A(NI+1))
A(NI) = G + U
NTOP = NX - 2
DO 10 I=NI,NTOP
G = A(I+1) - U
IF (CX.GT.TOL) GO TO 12
PS = B(I)*CXS
GO TO 16
12 PS = G**2/CX
16 RS = PS + B(I+1)
B(I) = SX*RS
SX = B(I+1)/RS
CXS = CX
CX = PS/RS
U = SX*(G+A(I+2))
A(I+1) = G + U
10 CONTINUE
A(NX) = A(NX) - U
IF (CX.GT.TOL) GO TO 112
PS = B(NTOP+1)*CXS
GO TO 116
112 PS = ((A(NX))**2)/CX
116 B(NTOP+1) = SX*PS
85 IT = NX
30 IT = IT-1
IF (ABS(B(IT)).LE.TOL) GO TO 40
IF (IT-NI) 100,100,30
40 IF (NX-IT-2) 50,60,70
50 A(NX) = A(NX) + SH
NX = NX-1
K = 1
GO TO 80
60 AL = B(NX-1)
AM = 5.D-1*(A(NX-1)-A(NX))
AMS = AM**2
SAM = SIGN(1.D0,AM)
AN = SQRT(AL+AM**2)
```

```
CX = (AN+ABS(AM))/(2.D0*AN)
SX = B(NX-1)/(4.D0*AN**2*CX)
TA = A(NX-1)
TB = A(NX)
TC = B(NX-1)
I = NX
A(NX-1) = TA*CX+TB*SX+TC*SAM/AN+SH
A(NX) = TA*SX+TB*CX-TC*SAM/AN+SH
B(NX-1) = 4.D0*AMS*CX*SX-ABS(AM)*TC/AN+TC*(CX-SX)**2
K = 1
NX = NX-2
GO TO 80
70 NI = IT + 1
GO TO 85
80 IF (NX.LT.NI) GO TO 90
95 IF (NX-NI-1) 50,60,85
90 IF (NI.EQ.1) RETURN
NI = 1
GO TO 95
END

real*8 FUNCTION DOT(A,B)
IMPLICIT REAL*8(A-H,O-Z)
DIMENSION A(1),B(1)
COMMON /VECTOR/ FACTOR,LIMIT
ITOP=LIMIT+1
DOT=0.D0
DO 10 I=1,ITOP
10 DOT=DOT+A(I)*B(I)
RETURN
END

SUBROUTINE QRTN (A,B,V,EIG,N,M,TOL,NJX)
IMPLICIT REAL*8 (A-H,O-Z)
DIMENSION A(1),B(1),V(1),EIG(1)
NX = N
NNM1 = NJX*(NX-1)
NI = 1
SH = 0
IF (NX-2) 50,60,1
1 K=0
98 ST = EIG(NX)-SH
IT = NX
99 IT = IT-1
IF (ABS(B(IT)).LE.TOL) GO TO 40
IF (IT.GT.NI) GO TO 99
```

```
  100 K = K + 1
  IF (K.LT.M) GO TO 11
  WRITE (6,1000) K
  1000 FORMAT(35H NO CONVERGENCE OF QR ALGORITHM IN ,I4,11H
ITERATIONS)
  IFAIL=1
  return
  CCALL EXIT
  11 IF (K.EQ.1) GO TO 15
  12 AT = A(NX) + A(NX-1)
  ST = AT*5.D-1
  DISC = AT**2-4.D0*(A(NX)*A(NX-1)-B(NX-1)**2)
  IF (DISC.LE.0.D0) GO TO 15
  ST = ST-SIGN(SQRT(DISC),ST)*5.D-1
  15 SH = SH + ST
  DO 20 I=1,NX
  20 A(I) = A(I) - ST
  R = SQRT(A(NI)**2+B(NI)**2)
  S = B(NI)/R
  CS = S
  C = A(NI)/R
  U = (S**2)*(A(NI)+A(NI+1))
  A(NI) = A(NI) + U
  CALL ROTATE(V(NI),C,S,NJX,NNM1)
  NTOP = NX - 2
  DO 10 I=NI,NTOP
  G = A(I+1) - U
  Q = C*A(I+1) - CS*B(I)
  R = SQRT(Q**2+B(I+1)**2)
  B(I) = S*R
  S = B(I+1)/R
  CS = C*S
  C = Q/R
  U = (S**2)*(G+A(I+2))
  A(I+1) = G + U
  CALL ROTATE (V(I+1),C,S,NJX,NNM1)
  10 CONTINUE
  B(NTOP+1) = S*(C*A(NX)-CS*B(NTOP+1))
  A(NX) = A(NX) - U
  85 IT = NX
  30 IT = IT - 1
  IF (ABS(B(IT)).LE.TOL) GO TO 40
  IF (IT-NI) 100,100,30
  40 IF (NX-IT-2) 50,60,70
  50 A(NX) = A(NX) + SH
```

```
      NX = NX - 1
      K = 0
      GO TO 80
   60 AL = -B(NX-1)
      AM = 5.D-1*(A(NX-1)-A(NX))
      AN = SQRT(AL**2+AM**2)
      C = SQRT((AN+ABS(AM))/(2.D0*AN))
      S = SIGN(5.D-1,AM)*AL/(AN*C)
      TA = A(NX-1)
      TB = A(NX)
      TC = B(NX-1)
      CX = C**2
      SX = S**2
      CS = C*S
      A(NX-1) = TA*CX+TB*SX-2.D0*TC*CS+SH
      A(NX) = TA*SX+TB*CX+2.D0*TC*CS+SH
      B(NX-1) = 2.D0*AM*CS+TC*(CX-SX)
      I = NX-1
      S = -S
      CALL ROTATE (V(I),C,S,NJX,NNM1)
      K = 0
      NX = NX-2
      GO TO 80
   70 NI = IT + 1
      GO TO 100
   80 IF (NX.LT.NI) GO TO 90
   95 IF (NX-NI-1) 50,60,98
   90 IF (NI.EQ.1) RETURN
      NI = 1
      GO TO 95
      END

      SUBROUTINE ROTATE(V,C,S,NJX,JTOP)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION V(1)
      JLIM=JTOP+1
      DO 10 J=1,JLIM,NJX
      TA=V(J)
      TB=V(J+1)
      V(J)=TA*C+TB*S
   10 V(J+1)=TB*C-TA*S
      RETURN
      END
```