BASIS SET SUPERPOSITION ERROR EFFECTS, EXCITED-STATE POTENTIAL ENERGY SURFACE AND PHOTODYNAMICS OF THYMINE

David **ASTURIOL BOFILL**

TESI DOCTORAL

Basis Set Superposition Error effects, excited-state Potential Energy Surface and photodynamics of thymine

David Asturiol Bofill

2009

Doctorado Interuniversitario en Química Teórica y Computacional

Dirigida per: Lluís Blancafort San José,

Pedro Salvador Sedano i Miquel Duran Portas

Memòria presentada per a optar al títol de Doctor per la Universitat de Girona

Universitat de Girona

Lluís Blancafort San José, Pedro Salvador Sedano i Miquel Duran Portas professors titulats del Departament de Química de la Universitat de Girona,

CERTIFIQUEM:

Que aquest treball titulat "***Basis Set Superposition Error effects, excited-state Potential Energy Surface and photodynamics of thymine***", que presenta en David Asturiol Bofill per a l'obtenció del títol de Doctor, ha estat realitzat sota la nostra direcció i que compleix els requeriments per poder optar a Menció Europea.

Signatura

Lluís Blancafort   Pedro Salvador   Miquel Duran

Girona, 9 de Novembre de 2009

Summary of the thesis

The study of the photophysics of thymine is the main objective of this thesis. This work has been divided in 4 parts; the first two parts are devoted to find a proper level of theory for the study of thymine, whereas in the third and fourth parts the photohpysics of thymine are studied.

Moran *et al.*[140] found that correlated methods such as the Configuration Interaction with Single and Double excitations (CISD) and Møller-Plesset up to second order (MP2) when used with some of Pople's basis sets, can not describe the planar structure of benzene. In addition, if a planar stationary point is optimized, the frequency analysis shows one or more imaginary frequencies. Given that thymine is a planar aromatic molecule as benzene, a benchmark study has been performed to determine if thymine can also suffer from such pitfalls. For completeness, the study is extended to the rest of the nucleobases, namely uracil, cytosine, adenine, and guanine. Our results show that, when Pople's basis sets are used in conjunction with the MP2 method, minima structures of nucleobases with planar rings present imaginary frequencies. However, the same basis sets studied by Moran *et al.* have been analyzed at the Complete Active Space Self Consistent Field (CASSCF) level for thymine, and no imaginary frequencies have been found in any case. Thus according to our results, it can be concluded that the pitfalls reported for benzene seem to be common to correlated methods describing planar aromatic rings with Pople's basis sets. In contrast, we have shown that the 6-31G* and 6-311G* basis sets, which are of general use in computational studies, can properly describe the minima structures of nucleobases. In addition, we have determined that the CASSCF/6-31G* and CASSCF/6-311G* levels of theory will be used for the study of the photophysics of thymine.

In the first part of the thesis, we have also analyzed the origin of the pitfalls described above. We have shown that they can be explained in terms of intra-molecular Basis Set Superposition Error (BSSE), and that they can be fixed by using a typical BSSE correction technique such as the Counterpoise method (CP), which is implemented in general electronic structure modeling softwares. This method divides the molecule into fragments and this can be a problem as the multiplicity has to be assigned to each fragment. We have shown that independently of the fragments' definition and fragment's multiplicity assignment, the Counterpoise method fixes the imaginary frequencies where present and has no meaningful effects on the descriptions that were already correct. Nevertheless, we stress that one has to take into account that the isolated fragment and the associated ghost orbital calculations must correspond to the same state with the same orientation of singly-occupied

degenerate orbitals, otherwise artifacts might arise during the BSSE removal which can result in a bad description of the molecule.

By using our own code, which allows for a flexible definition of the Counterpoise function, we have been able to fix pitfalls in complicated systems such as the cyclopentadienyl and indenyl anions and naphthalene, in which a negative charge has to be considered and up to five imaginary frequencies were found, respectively. In addition, we have observed that the BSSE has a delocalized nature given that although the imaginary frequencies can be removed by just correcting BSSE for a single fragment, all fragments need to be included in the CP function to recover the frequency values of the correct descriptions.

Experimental studies[23,24,85-88] show that the relaxation of thymine after photon absorption can be described with a biexponential decay. That is, there exist two decay mechanisms, one in the subpicosecond and another one on the picosecond time scale, that lead photoexcited thymine to its initial structure. The existence of a longer component of hundreds of ns has also been reported.[24,85] In the second part of this thesis, the photophysics of thymine have been studied. First, the PES of thymine has been optimized with a high level of theory to determine the decay paths of thymine. This has been carried out with the MS-CASPT2(12,9)//CASSCF(12,9)/6-311G* approach, in which Multi State Complete Active Space Møller-Plesset (MS-CASPT2) single point calculations, which include dynamic correlation, are carried out along minimum energy paths optimized at the Complete Active Space Self Consistent Field (CASSCF) level. Our results show that there exist two paths that after photon absorption can lead to the regeneration of the initial structure (for a better description of the PES we refer to Figure 28 and Figure 31). The first path, Path 1 in Figure 28, leads directly from the Franck-Condon (FC) point to a conical intersection (CI) with the ground state (GS), namely $(Eth)_X$. Due to its barrierless character, this path has been assigned to the subpicosecond decay component determined experimentally. The second path, Path 2 in Figure 28, is indirect and is separated from $(Eth)_X$ by a minimum, $(\pi,\pi^*)_{Min}$, a barrier, $(\pi,\pi^*)_{TS}$, and a CI between the π,π^* and n,π^* states, $(\pi,\pi^*/n,\pi^*)_X$. Two paths connect this CI with the GS. The first path leads directly to $(Eth)_X$ with no further barriers, and the second one leads to a minimum of the n,π^* state and further to a CI with the GS, namely $(n,\pi^*/GS)_X$. Given that the barrier that separates $(Eth)_X$ from the FC structure is only of 0.05 eV, we assign the deactivation through this path to the subpicosecond decay. On the other hand, since the n,π^* state can be accessed via $(\pi,\pi^*/n,\pi^*)_X$, and the CI that leads the population of this state to the GS is not accessible (it lies 0.2 eV above the FC point), we assign the deactivation from this state to the picosecond decay component determined experimentally.

We have carried out quasi-classical dynamics, *i.e.* classical dynamics simulations with a surface hopping algorithm which allows the propagation of nuclei on different surfaces, on the indirect path of thymine. The trajectories have been run at the CASSCF(8,6)/6-31G* level of theory because of their high computational cost. Unfortunately, at this level of theory the FC region is not well described and the barrier is overestimated by 0.2 eV. Thus, the direct path cannot be described because of the lack of dynamical correlation. A previous study[138] showed that trajectories starting on the FC point got trapped at $(\pi,\pi^*)_{Min}$ for more than 500fs, a time range that exceeds our simulations. Thus, we have run quasi-classical dynamics on the indirect path of thymine starting at $(\pi,\pi^*)_{TS}$, from which the behavior at the S_2/S_1 CI and further regeneration of the GS have been sampled. The results confirm the role proposed for this path at the MS-CASPT2 level. That is, part of the population of the indirect path is responsible for the picosecond component since it is funneled to the n,π^* state at $(n,\pi^*/\pi,\pi^*)_X$, whereas the rest of the population decays in the subpicosecond range.

Because of the importance of $(n,\pi^*/\pi,\pi^*)_X$ in the photophysics of thymine, a topological analysis of this seam has been carried out. For consistency with the level of theory used in the dynamics simulations, such analysis has also been carried out at the CASSCF(8,6)/6-31G* level of theory. A structure of *Cs* symmetry has been optimized on the seam of intersection from which a constrained IRC has been performed. This shows that this CI seam presents a sloped-to-peaked topology and that all its parts are energetically accessible from the FC point. This slope-to-peaked topology has also been described in other works,[284-286] where it is observed that the different parts of the seam can determine the photophysics of the molecule. In order to study this possibility, we have carried out quantum dynamics simulations on the indirect path of thymine with a novel method, namely the Direct Dynamics vibrational Multi Configurational Gaussian (DD-vMCG). This method uses diabatic states to propagate a wavepacket which is approximated by a set of Gaussian functions, whose centers move classically. This approximation implies that only local evaluations of the PES calculated on-the-fly are needed to propagate the wavepacket, rather than a precalculated grid of points which represents the PES. As in the semi-classical case, the simulations were started at $(\pi,\pi^*)_{TS}$. The different parts of the seam were analyzed by adding momentum on given coordinates which lead the wavepacket toward the regions of interest. Our results show that the segment of the seam that is reached during the decay has a large influence on the photophysics. In general it is observed that the peaked region of the seam favors the regeneration of the ground state, whereas the sloped one delays the deactivation as that region is responsible for trappings at $(\pi.\pi^*)_{Min}$ and the n,π^* state.

The DD-vMCG method has been applied to the present study of thymine, and all degrees of freedom of the molecule (39) have been taken into account. Such a large amount of degrees of freedom has been taken into account for the very first time with the DD-vMCG method. Problems associated with the use of a small active space, which limits the number of Gaussian centered wavefunctions that form the wavepacket, and the appearance of intruder states that invalidate the diabatic transformation at some points, have been encountered. However, we consider the performance of the method is encouraging as in its initial version it can be used to get a qualitative, mechanistic insight into the photophysics of thymine.

Resum de la tesi

L'estudi de la fotofísica de la timina és el principal objectiu d'aquesta tesi doctoral. Aquest treball ha estat dividit en 4 parts. En les dues primeres, s'ha realitzat un estudi de metodologies per tal de trobar la més adient per dur a terme l'objectiu principal de la tesi. En les altres dues parts de la tesi s'ha estudiat la fotofísica de la timina en detall.

Moran i col·laboradors[140] van publicar un treball en el qual es descrivia que alguns mètodes correlacionats com són el Configuració d'Interaccions amb excitacions Simples i Dobles (CISD) i Møller-Plesset de segon ordre (MP2), al utilitzar-los conjuntament amb bases de Pople per optimitzar l'estructura de mínima energia del benzè, s'obtenien geometries no planes. A més a més, si la optimització es duia a terme forçant la simetria *Cs*, els anàlisis posteriors de freqüències mostraven una o més freqüències imaginàries, tot indicant que no es tractava d'estructures de mínima energia. Degut al fet que la timina, igual que el benzè, és una molècula plana, formada per un anell de sis membres i aromàtica, vam decidir dur a terme una calibratge de mètodes per tal de comprovar si els errors descrits pel benzè es reproduïen en la timina. Per dur a terme un treball més complert, aquest estudi es va ampliar a la resta de bases de l'ADN: uracil, citosina, adenina i guanina.

El nostre estudi mostra que, quan les bases de Pople es fan servir amb el mètode MP2, les estructures de mínima energia de les nucleobases, optimitzades forçant la planaritat de l'anell, presenten freqüències imaginàries. No obstant, si el mateix estudi fet per Moran i col·laboradors es fa amb el mètode d'Espai Actiu Complert de Camp Auto-Consistent (CASSCF) per al cas de la timina, en cap cas apareixen freqüències negatives.

Per tant, en base als nostres resultats, podem concloure que els problemes que presenten els mètodes correlacionats amb les bases de Pople descrits per Moran i col·laboradors pel benzè, semblen ser comuns a les molècules aromàtiques amb anells plans. Sorprenentment, hem observat que les bases 6-31G* i 6-311G*, que són d'ús generalitzat en estudis computacionals, poden descriure perfectament l'aplanament de les estructures de mínima energia de les nucleobases. A més a més, aquest estudi ens ha ajudat a determinar que els nivells de càlcul CASSCF/6-31G* i CASSCF/6-311G* seran els que farem servir per estudiar la fotofísica de la timina.

En la primera part de la tesi, també hem analitzat l'origen d'aquest problemes descrits anteriorment. Hem observat que es poden explicar en termes d'Error de Superposició de Base (BSSE) intra-molecular que, com a tals, es

poden arreglar fent servir tècniques de correcció del BSSE típiques; com per exemple el mètode Counterpoise. Una de les avantatges d'aquest mètode és que està implementat en la majoria de paquets de programari i, per tant, és de fàcil accés. Aquest mètode es basa en la separació de la molècula en fragments, fet que pot suposar un problema ja que, amb la versió actual, s'ha d'especificar la multiplicitat de cada un. La complicació pot ser major en el cas que es treballi amb una sola molècula. En aquesta tesi es demostra que, independentment de la definició dels fragments i de la seva multiplicitat, el mètode Counterpoise arregla les freqüències negatives en els casos on apareixen i que no té efectes apreciables en els que no se n'observen. Cal tenir en compte però, que per tal que el mètode arregli correctament els errors, els càlcul dels fragments aïllats i els dels càlcul corresponent que inclou totes les funcions de base del sistema (ghost orbital), s'han de dur a terme en el mateix estat i amb els orbitals degenerats mono-ocupats igualment orientats.

La utilització d'un programari propi que permet la lliure definició de la funció de Counterpoise, ens ha permès arreglar els errors de sistemes complicats com els anions de ciclopentadiè i indenil i el naftalè. En el cas dels anions, la complicació ve donada pel fet d'haver de tractar amb una càrrega negativa, mentre que pel benzè ve donada pel fet d'haver de corregir 5 freqüències imaginaries. A part d'això, amb l'ajuda d'aquest codi hem observat que el BSSE té un caràcter deslocalitzat, ja que tot i que les freqüències negatives es poden eliminar corregint el BSSE d'un sol fragment, els valors "correctes" no es poden obtenir si no s'inclouen tots els fragments en la funció de Counterpoise.

Un cop trobada la metodologia que es farà servir per estudiar la timina, ens centrarem en el seu estudi. Els estudis experimentals[23,24,85-88] de la timina mostren que el relaxament posterior a l'absorbància d'un electró es pot descriure amb una funció biexponencial. És a dir, que existeixen dos mecanismes de desactivació, un en l'escala de fs i l'altre en la de ps. També s'ha detectat[24,85] la presència d'un altre mecanisme que ajuda a la desactivació de la timina però més lentament (centenars de ns). Per tal de dur a terme l'estudi, primer hem optimitzat la Superfície d'Energia Potencial (PES) de la timina amb un alt nivell de càlcul. Concretament, hem fet servir l'aproximació MS-CASPT2(12,9)//CASSCF(12,9)/6-311G* en la qual càlculs puntuals a nivell de Teoria de Pertorbacions de segon ordre amb una referència d'Espai Actiu Complert (CASPT2) es duen a terme al llarg dels perfils optimitzats a nivell CASSCF. Els nostres resultats mostren que existeixen dos camins de reacció que porten la molècula fotoexcitada cap a la seva estructura inicial (per una millor comprensió del perfil d'aquest camins, es recomana seguir les imatges 26 i 29). El primer camí, Path 1 en la imatge 26, porta directament des del punt d'excitació (FC) a una intersecció cònica amb l'estant fonamental (GS), que l'anomenarem (Eth)$_X$. Degut a la manca de barreres al llarg d'aquest camí,

l'hem assignat a la component de relaxament ultraràpid (fs) determinat experimentalment. El segon camí, Path 2 de la imatge 26, també porta a $(Eth)_X$, però és indirecte. Al llarg del camí hi ha un mínim, $(\pi,\pi^*)_{Min}$, una barrera, $(\pi,\pi^*)_{TS}$, i una intersecció cònica entre els estats π,π^* i n,π^*, $(\pi,\pi^*/n,\pi^*)_X$. Dos camins connecten aquesta intersecció amb l'estat fonamental. El primer camí no té barreres i porta directament al GS a través de $(Eth)_X$. Per altra banda, el segon camí transcorre sobre l'estat n,π^* i porta primer al mínim d'aquest estat, $(n,\pi^*)_{Min}$, i posteriorment a una intersecció amb l'estat fonamental, $(n,\pi^*/GS)_X$. Com que la barrera que separa $(Eth)_X$ del punt FC és de només 0.05 eV, hem assignat el relaxament a través d'aquest camí indirecte i $(Eth)_X$, a la mateixa component ultraràpida d'abans: la de fs. Per altra banda, com que es pot accedir a l'estat n,π^* a través de $(\pi,\pi^*/n,\pi^*)_X$, i la intersecció que permet la desactivació d'aquest estat no és accessible ja que està 0.2 eV per sobre de l'energia del punt FC, hem assignat el relaxament des d'aquest estat a la component de ps.

Un cop explicat l'estudi estàtic de la PES de la timina, procedirem a descriure l'estudi dinàmic. Hem fet simulacions de dinàmiques semi-clàssiques del camí indirecte de relaxament de la timina. És a dir, hem fet servir dinàmiques clàssiques amb un algoritme de salt de superfícies que permet propagar els nuclis en diferents superfícies. Les trajectòries s'han dut a terme al nivell de càlcul CASSCF(8,6)/6-31G* ja que tenen un alt cost computacional. Desgraciadament, a aquest nivell de càlcul, la zona FC no està ben descrita, el que impossibilita la optimització del camí indirecte. A més, la barrera del camí indirecte se sobreestima en 0.2 eV. Un estudi previ semblant al que es vol realitzar, en el qual les trajectòries es van iniciar al punt FC, mostra que totes queden atrapades al $(n,\pi^*)_{Min}$ durant més de 500fs, un temps superior al de les nostres simulacions. Degut a això, i que la zona FC no es pot descriure correctament, hem decidit començar les trajectòries al $(\pi,\pi^*)_{TS}$, des del qual es pot estudiar el comportament de la CI S_2/S_1 i el posterior relaxament cap al GS. Els nostres resultats mostren que la població del camí indirecte és la responsable del component de ps ja que part de la població es pot transferir a l'estat n,π^*, la desactivació del qual es dur a terme en ps. Degut a la importància d'aquesta CI en la fotofísica de la timina, se li ha realitzat un estudi topològic. Per consistència amb metodologia de les dinàmiques, aquest estudi s'ha dut a terme al mateix nivell de càlcul.

Hem optimitzat una estructura amb simetria *Cs* en l'espai d'intersecció de la CI des de la qual s'ha optimitzat un camí de mínima energia restringit a aquest espai. Aquest camí porta directament al punt de mínima energia de la CI i l'estudi dels gradients dels estats al llarg d'aquest camí ens mostra que l'espai d'intersecció té una topologia "sloped-to-peaked". Aquest tipus d'espai d'intersecció ha estat descrit en algun altre treball,[284-286] en els quals s'ha

observat que les diferents parts de l'espai d'intersecció poden determinar la fotofísica de la molècula. Per tal d'estudiar aquesta possibilitat, hem dut a terme simulacions dinàmiques quàntiques al llarg del camí indirecte amb un nou mètode, DD-vMCG. Aquest mètode propaga paquets d'ona sobre, formats per una sèrie de funcions Gaussianes, en estats diabàtics. Això implica que la PES només s'ha d'avaluar localment en el centre de les Gaussianes en comptes d'haver de generar una xarxa de punts que descriguin tota la PES. El punt inicial d'aquestes simulacions és el mateix que per les dinàmiques clàssiques, $(\pi,\pi^*)_{TS}$. Les diferents part de l'espai d'intersecció s'han analitzat dirigint el paquet d'ones cap aquell direcció en concret. Això es pot fer afegint un moment d'inèrcia en la coordenada o coordenades que porten cap a la zona d'interès. Els nostres resultats mostren que la topologia de l'espai d'interacció té una gran influència en el mecanisme de relaxament. En general, s'observa que la regió "peaked" de l'espai d'interacció afavoreix el camí de relaxament que porta directament a l'estat fonamental, mentre que la regió "sloped" retarda la desactivació ja que afavoreix la confinament tant en $(\pi,\pi^*)_{Min}$ com en l'estat n,π^*.

Tal com s'ha dit abans, hem utilitzat el mètode DD-vMCG per dur a terme les dinàmiques quàntiques. Per primer cop amb aquest mètode s'han fet servir 39 graus de llibertat (tots els de la timina). Hem observat problemes associats a l'ús d'un espai actiu reduït, el qual ha limitat el nombre de Gaussianes que formaven el paquet d'ona i també l'aparició d'estats intrusos que invalidaven la transformació diabàtica. No obstant, considerem que el comportament del mètode és satisfactori ja que en la seva primera versió hem obtingut uns resultats qualitatius d'alguns aspectes mecanístics del relaxament de la timina. Tot i això, cal tenir en compte que un mètode com aquest ha de poder oferir dades quantitatives.

Agraiments

No ha sigut fàcil arribar a poder escriure aquestes línies i ara que ho estic fent m'adono que m'entristeix una mica perquè significa que s'acaba l'etapa més meravallosa de la meva vida. Espero ser capaç d'incloure a tothom qui l'ha fet possible en aquest parell de fulls que venen a continuació. En aquest text no hi ha un ordre establert i hi ha faltes d'ortografia, això és degut a què és la part més personal de la tesi i he volgut que sigui així, imperfecte, com jo.

La major part d'aquesta tesi no és mèrit meu, és mèrit de les persones que sempre han estat al meu costat donant-me suport fins i tot quan el rebutjava perquè pensava que no el necessitava. És mèrit de les persones que han patit amb i per mi des de sempre, de les dues persones que fan que em senti afortunat cada cop que les veig. Gràcies i us estimo no són suficients per expressar els meus sentiments, però no se m'acut cap altra manera de fe-rho en un paper. Gràcies papa i mami.

Algú va dir que del que realment ens hem de penedir és de no haver fet alguna cosa, i no pas d'haver-la fet. Si d'algo em penedeixo d'aquest darrers anys, és de no haver passat tant temps com m'hagués agradat amb els meus avis. Tot i que potser no entenen ben bé què he fet durant tot aquest temps, segurament seran ells qui més s'alegrin quan sigui doctor. Només per aquest fet ja em sento orgullós i afortunat, i no els hi puc dir res més que me'ls estimo i que: "avi, àvia, avi Joan; gràcies".

Amb els que sí que he passat moltes estones aquest temps ha estat amb els meus companys de doctorat, i també amics. Vam començar al llegendari 166, edu, albert, david, quim, juanma i mireia. Alguns d'ells van marxar, i la veritat és que s'han trobat a faltar els acudits d'en quim, les vajanades de l'albert, el coneixement i capacitat organitzativa de l'edu i els cotilleos d'en torrente. Lo maco de tot això és que s'ha mantingut el contacte i que en dates senyalades com fires, ens tornem a reunir tots i t'adones que l'amistat segueix sent la que era. Amb el desterrament al parc ens vam ajuntar els dos despatxos, el 166 i 177, tot i que alguns com en Dani, que sempre està disposat a ajudar i comprar gadgets al dealxtreme, es van quedar a la universitat. El trasllat al "parque" va fer que deixés de compartir el despatx amb els meus companys, per fer-ho amb els meus amics. Cada un amb el seu toc freak característic que els fa únics. Gràcies a en juanma, la mireia, la sílvia i en ferran he viscut moments inoblidables amb les activitats "extra-escolars" (partits del barça, pàdels, futbols, bàsquets, voleis, esquí, play, fórmula de, sopars, can mià, lujuria, sessions de photoshop, cremats, karts, etc.), però sobretot amb el dia a dia. Un capítol apart mereixen els congressos viscuts amb tots ells: manchester-liverpool

(la pocha, antrus i pudors), brussel·les (dnis, trepitjades, discoteques, ...), amsterdam (bicicletes, bolets i pastissos), colònia (cuba bars i kebabs), goteborg (pizzes i sueques) i helsinki (sol, insomni i resets). Apart d'ells, també he tingut companys meravallosos com la cristina, en samat, l'oscar, en quantum li, les annes, i més recentment en sergi, en rambo, en ievgeny i l'eloy. Els post-docs (marcel, annapaola i jordi) també mereixen un agraiment, ja que han sigut com els nostres germans grans allà al parc, que sempre que hem necessitat ajuda ens l'han ofert. A tots gràcies.

Bona part del doctorat me l'he passat voltant per aquest món. Tot i que marxar a fora és difícil, he trobat persones que m'han fet sentir com a casa allà on fós. L'istvan i els erasmus a budapest, però sobretot la giulia, en ben, en jacob i en fabri a londres. Sense ells, tot hagués sigut molt més difícil. Gràcies.

Lo bo que tenia estar a la universitat era que et podies relacionar amb tot tipus de gent, fins i tot aquells que van tot el dia amb bata i ulleres i estan envoltats de productes tòxics i al·lucinògens, vaia, els "expis". Sense ells els cafès, sopars, partidillus i sobretot les "jodete" no haguéssin sigut el mateix. Gràcies per les bones estones que em passat.

Lo més important de treballar/estudiar (algú sap com es defineix fer el doctorat?) és estar agust en el lloc on ho fas. El fet que l'iqc sigui un dels millors llocs on fer-ho és degut a les persones que el formen, però sobretot als miquels que són qui han marcat les directrius perquè sigui així. Per això, també vull aprofitar l'ocasió per agrair-los haver-me donat la oportunitat de fer el doctorat aquí i per les facilitats i llibertat que ens han donat per fer tot tipus de coses. No em voldria deixar la resta de l'iqc, josep mª, sergei, sílvia, emili i molt especialment la carme, que en el fons és la mama de tots nosaltres.

No hauria pogut escriure aquesta tesi si no hagués tingut dos jefes com en pedro i en lluís. Desgraciadament no en sabré mai tant com ells però almenys he tingut la oportunitat d'aprendre'n. M'han ensenyat moltes coses, però sobretot els he d'agrair la paciència i disponibilitat que han tingut amb mi per explicar-me les coses que no entenia i també per motivar-me quan no veia les coses gaire clares. Haver voltat pel món també m'ha ensenyat que de jefes com ells, no n'hi ha. Gràcies.

Per sort aquests anys també he tingut vida fora de la universitat. Aquesta "altra" vida ha estat marcada per les meves companyes de pis, martona, txell, anna, i també l'alba. Elles han sigut durant molt de temps la meva família aqui a girona, i m'han ajudat a superar els mals trànguls i a tirar endavant. Juntament amb en pedro, hem passat molt bones estones i en tinc molt bons

records, però de tots ells em quedo amb els sopars i festes a "ca la pacheca". Gràcies.

Tothom necessita una via d'escape, una forma de desestressar-se, i jo l'he trobat amb l'scalextric. Gràcies a en jordi, l'anna, en lluís, en jokin i en sergi he pogut desconnectar dels problemes i passar-m'ho molt bé, no per l'scalextric en si que a vegades pot ser avorrit, sinó per tot lo que l'envolta: viatges a salou, oviedo, león, madrid, igualada, les històries dels dilluns, els sopars al taco-taco, les festes, els carnavals, ..., però sobretot pel dia a dia.

La bisbal també existeix, potser no per tothom però per mi sí. Allà hi ha els meus amics, sempre hi han sigut i sempre hi seran. Tot ells m'han ajudat sempre que han pogut, però lo que més valoro de tot plegat és que sempre hi són. La llista no està completa sense en runaldu i en mau que, tot i que poc, quan ens veiem sempre aconsegueixen que passem una bona estona. Gràcies. No em voldria deixar a la ceaccm, (jo tampoc sé que vol dir), que m'han acollit com un d'ells i també m'han fet costat tot aquest temps. A vosaltres també, gràcies.

Ara que ja estic acabant d'escriure aquestes línies d'agraiment, miro enrera i se'm dibuixa un somriure a la boca perquè veig tot el que m'emporto d'aquest període, tot es bo, però si alguna cosa em fa realment feliç de tot això, ets tu. Tu que has estat amb mi a les bones i dolentes, que m'has fet costat sempre i m'has ajudat en tot. Tu que fas que tot sigui més fàcil i que cada dia sigui especial. Tu que has aconseguit que perdi la por, i m'has fet veure que lo millor que un pot tenir és il·lusió. Tu que dones sentit a frases com "jo més perquè sóc més gran" i "no existeix pq és infinit". Gràcies.

List of publications of this thesis

(1)   Asturiol, D.; Duran, M.; Salvador, P.; *"Intramolecular Basis Set Superposition Error Effects on the Planarity of Benzene and other aromatic molecules: A solution to the problem"; J. Chem. Phys.* **2008**, *128*.

(2)   Asturiol, D.; Duran, M.; Salvador, P.; *"Intramolecular Basis Set Superposition Error Effects on the Planarity of DNA and RNA Nucleobases"; J. Chem. Theor. Comput.* **2009**, *5*, 2574-2581.

(3)   Asturiol, D.; Lasorne, B.; Robb, M. A.; Blancafort, L.; *"Photophysics of the π,π^* and n,π^* States of Thymine: MS-CASPT2 Minimum-Energy Paths and CASSCF on-the-Fly Dynamics"; J. Phys. Chem. A* **2009**, *113*, 10211-10218.

(4)   Asturiol, D.; Lasorne, B.; Robb, M. A.; Blancafort, L.; *"Thymine S_2/S_1 Conical Intersection analysis and quantum dynamics"; J. Phys. Chem. A.* (Submitted).

Publications not included in this thesis

(1)   Asturiol, D.; Duran, M.; Salvador, P.; Torrent-Sucarrat, M. *BSSE-free hardness profiles of hydrogen bond exchange in the hydrogen fluoride dimer; Int. J. Quant. Chem.* **2006**, *106*, 2910-2919.

(2)   Salvador, P.; Asturiol, D.; Mayer, I.; *"A general efficient implementation of the BSSE-free SCF and MP2 methods based on the Chemical Hamiltonian Approach"; J. Comput. Chem.* **2006**, *27*, 1505-1516.

(3)   Asturiol, D.; Salvador, P.; Mayer, I.; *"Dissecting the hindered rotation of ethane"; Chem. Phys. Chem* **2009**, *10*, 1987-1992.

(4)   Kobylecka, M.; Migani, A.; Asturiol, D.; Rak, J.; Blancafort, L.; *"Benign Decay vs. Photolysis in the Photophysics and Photochemistry of 5-Bromouracil. A Computational Study"; J. Phys. Chem. A* **2009**, *113*, 5489-5495.

List of figures

List of tables

List of Acronyms

ANO	Atomic Natural Orbital
BSIE	Basis Set Incompleteness Error
BSSE	Basis Set Superposition Error
CAS	Complete Active Space
CASPT2	Complete Active Space Second Order Perturbation Theory
CASSCF	Complete Active Space Self Consistent Field
CC	Coupled Cluster
CCSD	Coupled Cluster Singles-Doubles
CI	Conical Intersection
CIS	Configuration Interaction Singles
CISD	Configuration Interaction Singles-Doubles
CP	Counterpoise
CP-CISD	Counterpoise corrected CISD
CP-HF	Counterpoise corrected HF
CP-MP2	Counterpoise corrected MP2
DFT	Density Functional Theory
DIIS	Direct Inversion in the Iterative Subspace
DNA	Deoxyribonucleic Acid
E0	Energy of the Ground State
E1	Energy of the first excited state
E2	Energy of the second excited state
EA	Electron Affinity
FC	Franck-Condon
FCI	Full Configuration Interaction
GD	Gradient Difference
GS	Ground State
HF	Hartree-Fock
HOMO	Highest Occupied Molecular Orbital
IC	Interstate Coupling
IRC	Intrinsic Reaction Coordinate
IRD	Initial Relaxation Direction
ISC	Intersystem Crossing
LCAO	Linear Combination of Atomic Orbitals
LHA	Local Harmonic Approximation
LIIC	Linear Interpolation of Internal Coordinates

MC	Multi-Configurational
MCSCF	Multi-Configurational Self Consistent Field
MCTDH	Multi-Configurational Time-Dependent Hartree
MEP	Minimum Energy Path
MMVB	Molecular Mechanics of Valence Bond theory
MO	Molecular Orbital
MP	Møller-Plesset
MP2	Second order Møller-Plesset perturbation theory
MP3	Third order Møller-Plesset perturbation theory
MP4	Fourth order Møller-Plesset perturbation theory
MRCI	Multi-Reference Configuration Interaction
OM2	Orthogonalization method 2
PES	Potential Energy Surface
QM/MM	Quantum Mechanics/Molecular Mechanics
RASSCF	Restricted Active Space Self Consistent Field
REMPI	Resonance-enhanced Multiphoton Ionization
RHF	Restricted Hartree-Fock
RNA	Ribonucleic Acid
S0	Ground state
S1	First Excited State
S1/S0	Conical intersection between S1 and S0
S2	Second Excited State
S2/S1	Conical intersection between S2 and S1
S3	Third Excited State
SCF	Self Consistent Field
SOMO	Single Occupied Molecular Orbital
T1	First Triplet State
T2	Second Triplet State
TD-DFT	Time Dependent Density Functional Theory
TDH	Time Dependent Hartree
TDM	Transition Dipole Moment
TRPES	Time-Resolved Photoelectron Spectra
TS	Transition State
T-T	Thymine-Thymine
UV	Ultra-Violet
ZPE	Zero Point Energy

CONTENTS

1   INTRODUCTION

Probably there is not a more fascinating molecule than the Deoxyribonucleic acid, commonly known as DNA. Maybe its charm lies in its contradictory and uncertain nature, or perhaps, in the fact it might be unique and special. It has been "out there", almost untouched, for thousands of years and amazingly, it was not until 1871 when we first heard[1] of it. Many resources and money have been spent on its study since then. Unfortunately, although many advances have been performed, we still know very little about it. We do not even know for sure who discovered the famous double helix structure of DNA,[2] as its discovery was first credited to James Watson and Francis Crick but it has lately been suggested[3-5] that Maurice Wilkins and Rosalind Franklin should also be recognized for their essential contribution to the discovery.

DNA is an anti-parallel double sequence of nucleobases, namely Adenine, Thymine, Cytosine, and Guanine, as shown in Figure 1. They are coded in genes that contain all the necessary information for the development and functioning of every living being.
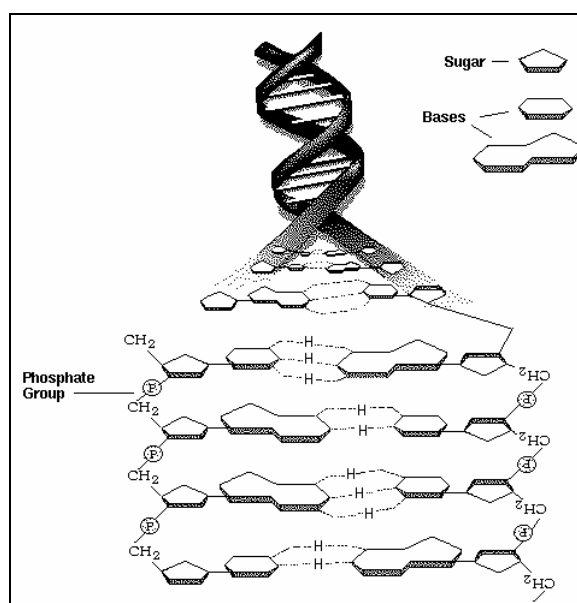
Figure 1. Scheme[6] of the double strand of DNA. Adapted from Access Excellence @ the National Health Museum.

Any variation in the sequence of the bases would translate into gene mutation which would have an unpredictable repercussion in the cell functionality. Taking into account that the importance of the information coded in the DNA, it is not strange that Mother Nature has provided a set of

protection mechanisms to preserve it from external agents. Some of the protection tactics featured by DNA are outlined next.

Among the various protection mechanisms present in nature, one of the most efficient and widespread is isolation. The isolation of nucleobases starts by being kept in the cell's nucleus, which is the most inaccessible part of the cell as two membranes protect it from external agents. In addition, a highly packed conformation of the DNA strand inside the nucleus makes nucleobases even more unreachable. Again, contradiction is present in DNA, since it must be protected to avoid mutations, while at the same time it needs to be easily accessible to perform vital processes for the cell such as replication, transcription, and translation.

Physical barriers can keep (physical) agents away from the coded information, but they cannot protect them from radiation. Nucleobases are the chromophores of DNA, *i.e.* the parts of the DNA that absorb light. If they are exposed to UV light, photons are absorbed bringing the molecule to an excited electronic state where it is prone to react because it has extra energy. Thus, UV light is a potential DNA damaging agent as it can promote photoreactions which, within the DNA strand, can cause gene mutation.

It is obvious that nucleic acid bases must feature several protection tactics against UV radiation, because otherwise the evolution could not have taken place as individuals can not survive to major changes in their DNA. One of the protection tactics of DNA against radiation is external, and corresponds to the ozone layer. Nucleobases have the lowest energy transitions located at the same spectral region as ozone, thus, the most dangerous UV radiation cannot reach the Earth's surface, as it is absorbed by the ozone layer. A more particular shield of DNA is the complex packed conformation it adopts inside the nucleus. It reduces its exposure to light, which hinders photon absorption. However, in spite of these protection tactics, nucleobases are still reached by the UV radiation. Fortunately, DNA also has some tools to minimize the effect of the photoreactions. For instance, the excited states of nucleobases are characterized by an ultra-short lifetime. They get rid of the UV induced extra energy in the sub-picosecond or picosecond time scale, which reduces the probability of photoreaction. In addition, the energy gained in photon absorptions can be redistributed along the DNA structure, which also helps in minimizing the probability of photoreactions.

As seen above, DNA has a large number of protection mechanisms, however, they do not provide 100% of security. In the cases where mutations take place, there are enzymes that can repair[7] DNA mutagenic[8-11] photoproducts

such as cyclobutane pyrimidine dimers[12,13] and 6-4 pyrimidine adducts[14,15] (see Figure 2).
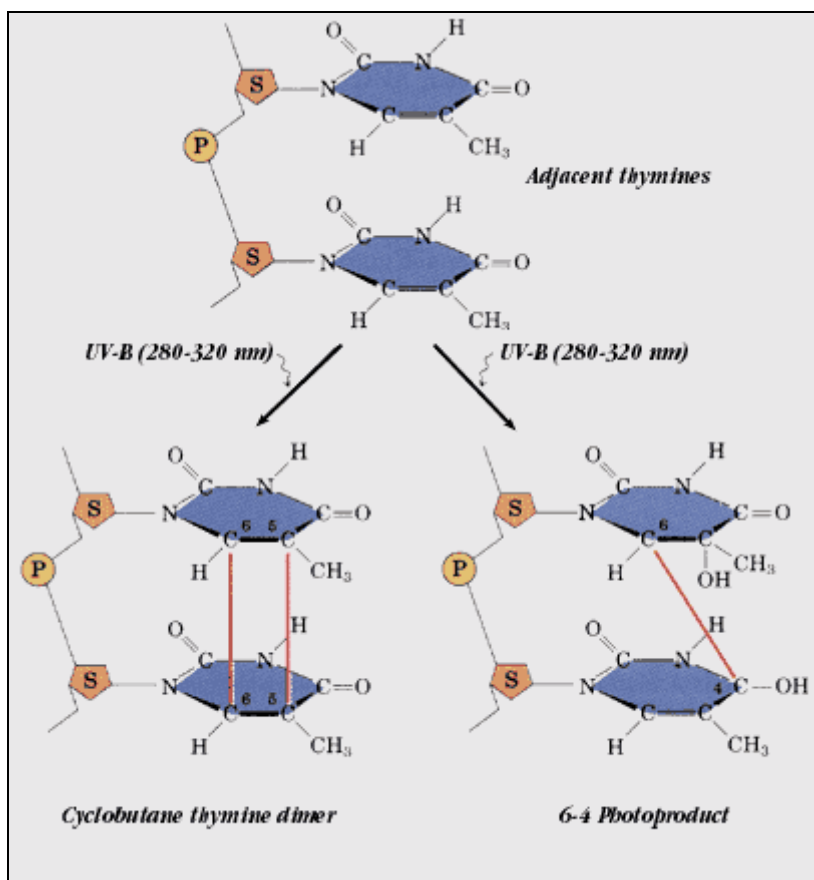
Figure 2. Thymine UV induced photoproducts (adapted from Medical Ecology online resources[16]).

It is obvious that the response of DNA to light is a complex and multivariable process and that its study cannot be faced globally. Here, we will study the photophysics of thymine as a first step towards the full understanding of the DNA protection mechanisms. A broad overview of the most important experimental and theoretical works on this field is presented in section 1.2. However, before reviewing the results present on the literature and explaining the results of this thesis, a brief overview of some useful photochemical concepts will be given.

1.1   Photochemical concepts

Light can interact with molecules changing their properties such as color, structure, stability, reactivity, etc. In this thesis we will study the way nucleobases interact with UV light. As mentioned above, nucleobases are the parts of DNA that absorb and/or emit light. In general, the light a molecule or an object absorbs is indicated by its color. That is, if an object is irradiated with "white" light (light compound of different wavelengths, as that of the sun) depending on what "part" of the light is absorbed, it will adopt one color or another. Actually, the colors of objects do not correspond to the light that they absorb but the one that is reflected. This is because only the light that is reflected reaches our eyes, and therefore, is the one that defines the colors of objects. For instance, it is known that plant leaves absorb "yellow" light because most of the sunlight that reaches the earth is made of "yellow" light and they use it for the photosynthesis. However, most plant leaves are green. This is explained because we only see the light that has been reflected by the plant (the green one), not the absorbed one (the yellow one). Spectrophotometers can determine the wavelength (color) of the light that is absorbed/emitted by a given molecule. The importance of absorption and emission spectra is explained next together with the functioning of spectrophotometers.

1.1.1   Absorption and emission spectra

A spectrophotometer is an apparatus that irradiates samples with white light and records the light that has been absorbed and/or emitted by them, thus, it records absorption and/or emission spectra. An absorption spectrum (see Figure 3) consists of a continuous spectrum (Inset a of Figure 3) with some "lines" which denote the energy of the light that has been absorbed by the sample. These lines, called bands in molecules, appear because the light of that part of the spectrum was used to promote an electron of the sample from a given molecular orbital to another orbital of higher energy. Thus, all irradiated light reaches the detector except that which has been absorbed by the sample (Inset b of Figure 3). On the other hand, the emission spectrum of a given molecule (Inset c of Figure 3) is the light emitted by a molecule which has been previously irradiated. In principle, an emission spectrum should be complementary to the absorption one, nevertheless, usually part of the absorbed energy is transformed, and then the remaining energy is emitted as light, which composes the emission spectrum.
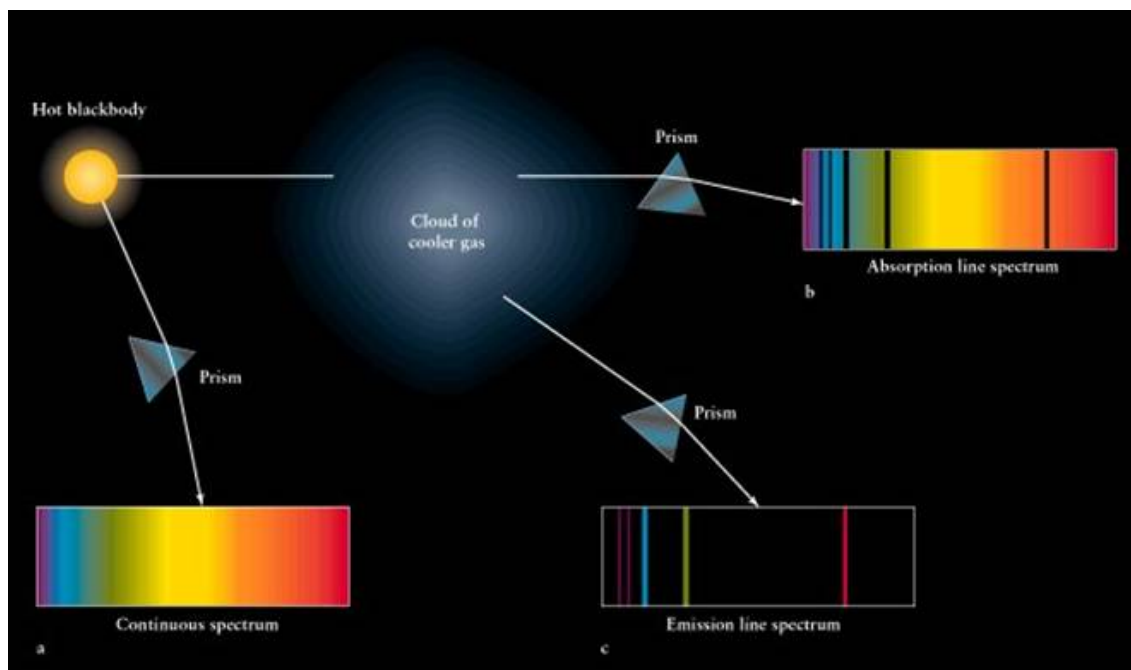
Figure 3. Absorption and emission spectra (adapted from Whitman College's webpage[17]).

Kasha's rule[18] states that photon emissions occur only from the lowest-energy excited state of a molecule. That is, if an electron is promoted to an excited state (blue arrow of Figure 4) and has extra energy to populate higher vibronic levels ($\nu' > 0$ in Figure 4), it will relax to the lowest vibronic level ($\nu' = 0$) from which it will deactivate emitting light (green arrow of Figure 4). Kasha's rule is complementary to the Franck-Condon principle[19,20], which states that an electronic transition is most likely to occur without changes in the positions of the nuclei in the molecular entity and its environment. The resulting state is called a Franck–Condon state and the transition involved a vertical transition. The quantum mechanical formulation of this principle is that the intensity of a vibronic transition is proportional to the square of the overlap integral between the vibrational wavefunctions of the two states that are involved in the transition (see Figure 4).
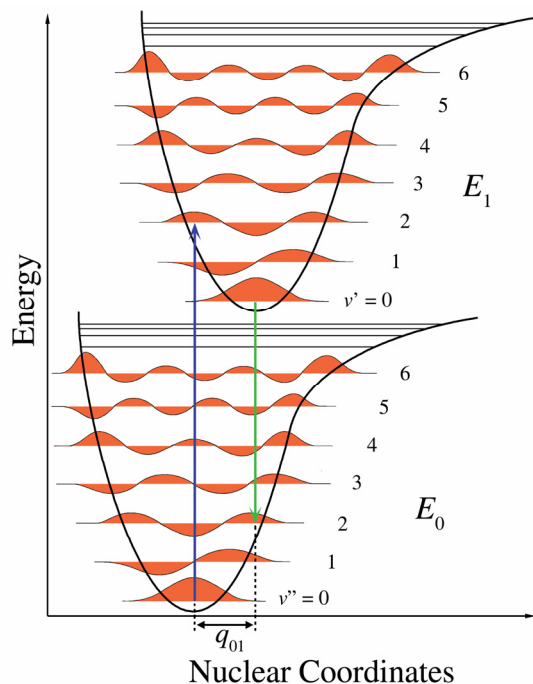
Figure 4. Franck-Condon principle energy diagram. The blue arrow corresponds to the vertical excitation from the ground state (E_0) to the vibrational level of the first excited state (E_1) with highest overlap with the initial state. Similarly, the green arrow denotes the vertical deexcitation (adapted from IUPAC Compendium of Chemical Terminology, 2nd Edition, 1997).

Electronic absorptions and emissions are transitions within different electronic states of a molecule. However, electrons cannot be promoted to whatever orbital since selection rules apply. In short, only transitions between states of the same multiplicity can take place. The fact that two states have the same multiplicity does not necessarily imply that such a transition will appear in the spectrum as it might correspond to a low intensity transition. The intensity of transitions is governed by the *oscillator strength* (f_{ij}), which is a dimensionless quantity that ranges from 0 to 1 and indicates the intensity of transitions, and reads as

$$f_{ij} = \frac{2}{3} \cdot \lambda_{ij} \cdot (\text{TDM}_{ij})^2 \tag{1.1}$$

where λ_{ij} corresponds to the energy of the transition between the states i and j, and TDM stands for Transition Dipole Moment. The electronic structure of a molecule gets modified when a photon is absorbed. Due to the extra energy gained in the absorption, the occupation of the molecular orbitals varies inducing a polarization of the molecule which generates a transition dipole moment. It can be calculated from an integral taken over the product of the

wavefunctions of the initial (i) and final (j) states of a spectral transition and the appropriate dipole moment operator ($\vec{\mathrm{D}}$) of the electromagnetic radiation:

$$\mathrm{TDM} = \sum \left\langle \psi_i \middle| \vec{\mathrm{D}}_\alpha \middle| \psi_j \right\rangle \qquad \alpha = \boldsymbol{x, y, z} \qquad (1.2)$$

The summation is over the coordinates of all charged particles (electrons and nuclei), and its square determines the strength of the transition (IUPAC Compendium of Chemical Terminology, 2nd Edition, 1997). An example of low intensity transitions (allowed transitions with very low probability that do not appear in the spectrum) are those which take place from a lone pair orbital in the molecular plane to an unoccupied orbital of an unsaturated system ($n \rightarrow \pi^*$). These transitions are considered forbidden due to their low probability.

There are various notations to refer to the electronic states of a molecule and the transitions between them. The most widely used nomenclatures are the enumerative and the Kasha's ones. The enumerative nomenclature is based on the energetic order of the states and its multiplicity. The electronic state with the lowest energy defines the ground state, and the adiabatic energies of the other states determine the corresponding labels. Thus, in this notation the singlet ground state is denoted by S_0, whereas the excited states are expressed by the successive numbers S_1, S_2, S_3, ... S_n. A similar formula is used for the triplet states, which are denoted by T_1, T_2 ... T_n. The excitations are expressed by $S_0 \rightarrow S_1$, where the arrow indicates the direction of excitation. It is worth to mention that given that the nature of the states can change along the PES, the the enumerative states can present singularities in their energetic profiles. On the other hand, Kasha's nomenclature only specifies the nature of the state involved in the transition. π, σ, and n characters are used to denote unsaturated, sigma and non-bonding occupied orbitals, respectively. The same characters with an added " * " (π^*, σ^*) are used to refer to the corresponding virtual orbitals. As in the former case, the excitations are expressed as $\pi \rightarrow \pi^*$. With this nomenclature the states present smooth energetic profiles along the PES. We will use both nomenclatures indistinctively throughout this thesis. For the sake of shortness, excitations will either be denoted as π, π^* or $\pi - \pi^*$.

1.1.2   Relaxation mechanisms

Nucleic acids have strong $\pi - \pi^*$ transitions which are responsible for the bands seen in their UV absorption spectra. This means that if a DNA base is exposed to UV light, the photons with enough energy to promote electrons from a π orbital will induce the excitation of these electrons to an unoccupied orbital

of higher energy. Thus, the light with energy (λ) equal to the energy needed to promote a $\pi-\pi^*$ transition, is absorbed. The way the excited molecule returns to its initial state, the ground state (GS), is called the relaxation mechanism.

There are mainly two types of relaxation mechanisms: radiative and non-radiative ones. In the former, the excess of energy is eliminated as light, giving rise to fluorescent and phosphorescent molecules (see Figure 5) depending on the emissive state. Fluorescence involves emissions from the lowest vibrational level of a state of the same multiplicity than the one that has been populated in the excitation to vibrational levels of the ground state. On the other hand, phosphorescence is similar to fluorescence but it is subject to previous intersystem crossings to states of different multiplicity. Due to its forbidden character, except for molecules with large spin-orbit couplings, these transitions are kinetically unfavoured but still occur at longer time scales (ns or longer). If the intersystem crossings are reversible, molecules can exhibit delayed fluorescence. Usually, radiative mechanisms occur in the ns range and are rather slow compared to those which involve CIs and take place in fs. It is known that the longer the excited state lifetime, the higher the photoreaction probability. This implies that molecules with short excited states lifetimes are less prone to suffer photoreactions. It is worth to mention though, that non-radiative mechanisms can also lead to photoproducts in an ultrafast way, although most of the photostable molecules are characterized by returning to their GS in less than 1 ps.
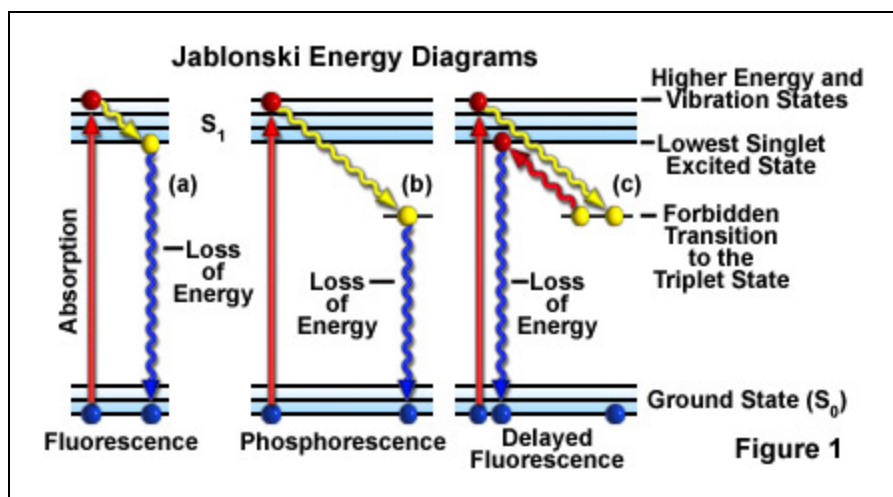
Figure 5. Jablonsky energy diagrams of (a) fluorescence (b) phosphorescence and (c) delayed fluorescence deactivation mechanisms (adapted from Molecular Expressions website[21]).

Geometry distortions favor couplings between the electronic states of the molecules. In regions of the PES where two states are electronically

degenerated, the population can be suddenly transferred to the other states. If the touching states are of the same multiplicity, this decay mechanism is known as internal conversion (IC). However, if the states do not have the same multiplicity, it is called intersystem crossing (ISC). A series of population transfers may lead a photoexcited molecule to its ground state or to a given photoproduct.

In the early literature, DNA bases were commonly described as "non-fluorescent". However, recent advances in experimental techniques have allowed the detection of very low[22] fluorescent quantum yields and enabled the accurate measure of the excited state lifetimes of nucleobases, which are of hundreds of femtoseconds.[23] This indicates that deactivations in nucleobases are non-radiative, although some radiative transitions also occur. In a non-radiative mechanism, the molecule may eventually get trapped in a dark state. A dark state is a state which cannot be accessed by photon absorption. Thus, it does not appear in the spectrum. In photochemistry, dark states are usually used to denote states whose transitions from the GS appear dark in the absorption spectrum due to its forbidden character. Nevertheless, dark states can be accessed via deactivations from higher states, and radiative deactivations can take place but with low intensity and very slow.

It is thought that dark states in DNA are due to the lone electron pairs of heteroatoms and correspond to $n-\pi^*$ excitations. As these transitions are forbidden by the selection rules, it is thought that dark states are responsible for the longer decay timescales seen in DNA.[24] The optimization of PES is one of the procedures used to describe the excited state lifetimes of molecules and their relaxation mechanisms.

1.1.3   Potential Energy Surface(s)

In the Born-Oppenheimer approximation, which will be explained in the next chapter, the Potential Energy Surface (PES) of a molecule can be defined as the surface described by the potential energy function of the molecule with respect to the molecular geometry. The surface dimension depends on the number of atoms of the molecule (2 for diatomic and 3N-6+1 for polyatomic molecules, with N being the number of atoms, and the potential energy the extra dimension). Due to the large number of dimensions of these polyatomic surfaces, they are commonly known as hyper-surfaces. The hyper-surfaces are described by the Schrödinger equation, whose solution gives the molecular energy as a function of the nuclear coordinates. However, the Schrödinger equation has an infinite number of solutions, which correspond to the different electronic states of the molecule, and each state has its own hyper-surface. The

topology of the hyper-surface drives all the processes of the molecule in that particular electronic state. The PES topology of the excited states is different from that of the ground state, and therefore, the regions (geometries) that the molecule can access are different too. Consequently, there are processes that can only take place in excited states.

To study the reactions or processes a molecule may undergo, one has to the study the topology of the PES(s) that can be involved in such processes. In these studies, the stable conformations of molecules are represented by valleys in the 2 or 3 dimensional sketches of PES, and the height of the "mountains" between the valleys indicate the difficulty of the reaction. This so-called static description of the reactivity of molecules can be improved by carrying out dynamics studies, which are based on the solution of time-dependent nuclear Schrödinger equation. The latter, provides the position of the nuclei (potential energy) and the kinetic energy of the molecule at a given time in a given state. The importance of dynamics calculations for this thesis relies on the fact that they can describe non-stationary processes. Those processes are especially relevant in photophysics because photon absorptions provide an extra energy to molecules which starts a sequence of non-stationary events that are commonly known as deactivation process. The most important parameters of the deactivations are the excited state lifetimes and branching ratios. With the potential and kinetic energies, one can determine both the "trajectory" a molecule will follow, *i.e.* the reaction that may take place, and the time needed to do so, *i.e.* the reaction lifetime. If different products are formed, a set of trajectories can determine the probability with which each product can be formed (branching ratio), which includes the probability of regenerating the initial structure. The average of all trajectories time in which the excited state species is consumed, corresponds to the excited state lifetime of the species. In principle, describing all the possible trajectories and regions of the PES that a molecule can access, one gets an accurate description of the reactivity and behavior of the molecule upon photon absorption. Thus, one gets enough data to describe the experimental results.

Dynamics simulations are an indispensable tool for the description of reactions in photophysics, however they are computationally challenging due to its high cost. As it will be shown, in this thesis we have tried to reach a compromise between computational feasibility and results accuracy. This can be done by rationalizing the dynamics simulations, carried out at a computationally feasible level, with a high level PES.

1.1.3.1   PES characterization

Static studies are based on the localization of critical points on the PESs of interest to define all the possible paths that molecules can follow. The characterization of the PESs in photophysics is carried out by locating the critical points of each PES, namely minima, transition states, and conical intersections; and interconnecting them. The determination of the PES provides a set of paths that the molecule may follow in a given chemical process, in other words, it provides the thermodynamic profile of the system. Unfortunately, in general using this procedure, it is not possible to determine the probabilities of accessing the different paths of the PES, since only dynamic studies, which can describe non-stationary events, can provide such information. Nevertheless, the rationalization and study of the different profiles serves to propose (reaction) mechanisms for the case study.

Minima and transition states are common structures to all chemical processes. The former are usually optimized using algorithms that follow the gradient of the PES as it leads to the points of lower energy (minima). Other gradient-based algorithms also permit the optimization of TSs, *i.e.* points in the surface where the energy is minimal for all directions but one. The characterization of these structures is sufficient to study chemical processes where only one PES is considered and where the Born-Oppenheimer approximation is preserved (adiabatic processes). Most of thermo-reactions take place on a single state, usually the ground state. Thus, in those processes the characterization of minima and TSs is sufficient to determine the reaction profile.

In contrast, many photochemical processes do not keep the Born-Oppenheimer approximation since they involve multiple PESs. For instance, in photophysics the Born-Oppenheimer approximation breaks down in radiationless deactivations that lead molecules from one state to another. These processes are called non-adiabatic, and may change the electronic character of the molecule, thus they break the Born-Oppenheimer approximation. This phenomenon occurs in regions of the PES where two or more surfaces touch or nearly touch each other.

1.1.3.2   Touching surfaces regions

The regions of touching or near touching surfaces are commonly known as conical intersections or funnels.[25] When a photon is absorbed (see Figure 6) an

excited species, R*, is generated. This excited species has an extra energy which starts a non-stationary process commonly known as relaxation. Two possibilities are considered, the formation of species P* and subsequent radiative deactivation to the photoproduct P', and the relaxation through the funnel (non-adiabatic process). Funnels allow non-adiabatic transitions between states, and their shape defines the probability of accessing one or other part of the state.
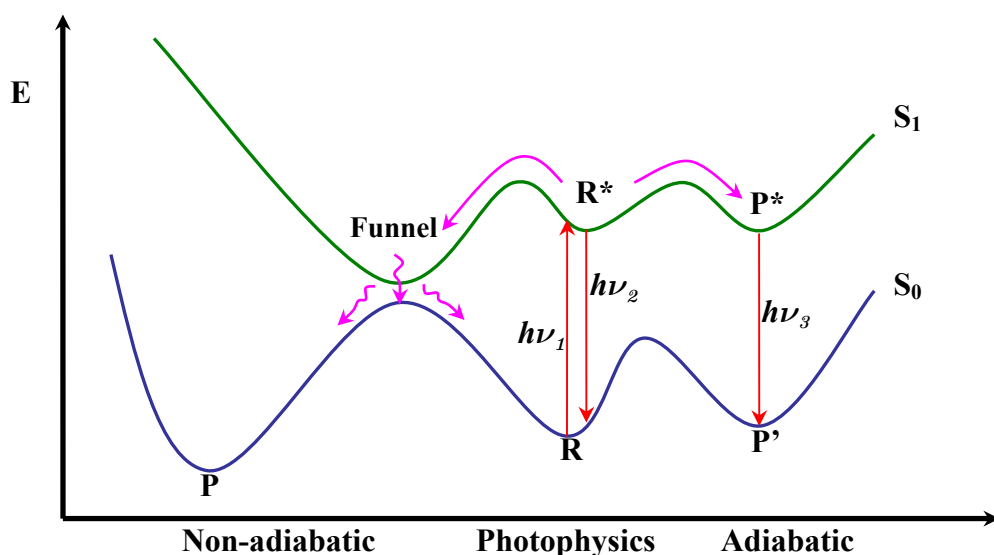
Figure 6. Possible photoprocesses for a molecule: a) Emissive deactivation to the ground state (adiabatic). b) Emissive photoreaction (adiabatic). c) Radiationless deactivation to the initial position (non-adiabatic). d) Internal conversion to a photoproduct (non-adiabatic) (Adapted from Encyclopedia of Computational Chemistry (1998)).

Classical photochemistry[26] assumes that non-adiabatic decays take place via avoided crossings. That is, regions of the space where two or more states nearly touch each other. In this model, if the states are more separated than a few Kcal, the excited species thermalyse and the decay probability is determined by the Fermi's Golden Rule.[27,28] Such a process is supposed to occur in the same time scale than fluorescence. On the other hand, modern photochemistry has shown that non-adiabatic transitions take place in the subpicosecond range via CIs (points of energy degeneracy between surfaces). The latter transitions are much faster than fluorescence, although they might fall in longer time ranges if there exist barriers that hinder the progressive motion towards the CI.

In general, the proximity of the states defines the type and probability of surface crossing. This probability can be approximated by the Landau and Zener relation,[29,30]

$$P = \exp\left(\frac{-\pi^2}{h}\frac{\Delta \boldsymbol{E}^2}{\boldsymbol{v} \cdot \Delta \boldsymbol{S}}\right) \tag{1.3}$$

where P represents the hop probability between two surfaces that are separated by $\Delta \boldsymbol{E}$ energy units, $\boldsymbol{v}$ is the velocity of the nuclear motion along the reaction coordinate, and $\Delta \boldsymbol{S}$ is the difference of surface slopes. It can be readily seen, the hop probability increases for high velocities and differences in surface slopes, and also with the decrease of the energy difference. The limiting case with 100% of hop probability is a CI.

Conical intersections are very important in photochemistry as they allow ultrafast radiationless deactivation processes, which usually lower the photoproducts yield as in DNA.

1.1.3.2.a   Conical Intersections

Crossings of surfaces are known since the early years of photochemistry back in the 1920's. Their existence was first announced by Hund[31] with a very simple argument: *If two curves cross, the electronic states must be degenerated at the point of crossing.* We have seen that this point of crossing is nowadays known as conical intersection.[32] Technological advances in lasers and detectors have allowed the measurement of photoprocesses that take place in tenths of fs. Such fast phenomena are the experimental evidence for the presence of CIs. Nevertheless, the characterization and mechanistic study of CIs has been carried out on the theoretical side. Von Neumann and Wigner provided[33] the mathematical condition for the existence of CIs. The authors showed that only two parameters need to be adjusted to find a point of crossing between two surfaces of a polyatomic molecule, 3 if heavy atoms are present and magnetic forces cannot be neglected. These two parameters were assigned to nuclear coordinates and their corresponding vectors, namely the gradient difference and interstate coupling vectors, by Teller in 1937.[32] These two vectors form the so-called branching space, and the rest of coordinates compose the intersection space. If the energy of the states is plotted against the branching space coordinates, a double cone, or diabolo, shape after which conical intersections were named is obtained.

If we write the energy matrix of a two state (A and B) problem, $\mathbf{M}$, as a function of two coordinates that belong to the branching space, Q_{x_1} and Q_{x_2}, and an extra coordinate which belongs to the intersection space, Q_{x_3}, one obtains

$$\mathbf{M}(\boldsymbol{Q}) = (\lambda_1 Q_{x_1} + a Q_{x_1}^2 + b Q_{x_2}^2 + c Q_{x_3}^2)\mathbf{I} + \begin{vmatrix} \dfrac{1}{2}\alpha Q_{x_1} & \beta Q_{x_2} \\ \beta Q_{x_2} & -\dfrac{1}{2}\alpha Q_{x_1} \end{vmatrix} \qquad (1.4)$$

with $\mathbf{I}$ being the identity matrix. The matrix $\mathbf{M}$ has the form of a Taylor expansion of the energy along the 3 coordinates, whose eigenvalues read as:

$$E_{A,B}(\boldsymbol{Q}) = (\lambda_1 Q_{x_1} + a Q_{x_1}^2 + b Q_{x_2}^2 + c Q_{x_3}^2) \pm \frac{1}{2}\sqrt{(\alpha Q_{x_1})^2 + 4(\beta Q_{x_2})^2} \qquad (1.5)$$

The important part of these equations is the second summand of eq. (1.4) , as it is the one that defines the energy degeneracy. The point of degeneracy corresponds to that in which the diagonal terms of the matrix $\mathbf{M}$ are equal and the off-diagonal terms are 0. That is, $\alpha Q_{x_1} = -\alpha Q_{x_1}$ and $\beta Q_{x_2} = 0$.
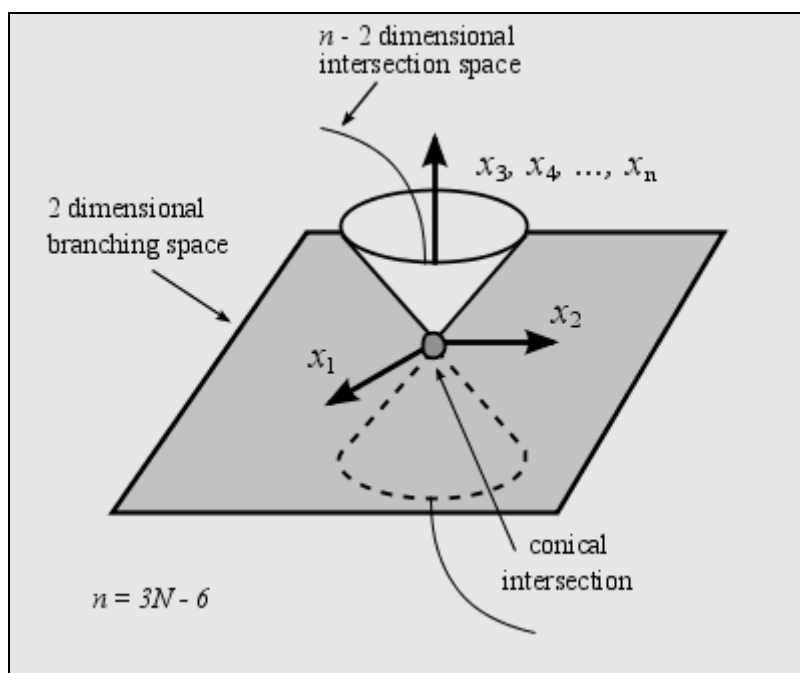
Figure 7. Plot of the potential energy surface as a function of the branching space $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ (IUPAC Compendium of Chemical Terminology, 2nd Edition, 1997)

However, this degeneracy point is not unique, as according to eq. (1.5) any displacement along $\boldsymbol{x}_3$ has no effect on the energy degeneracy. Thus, CIs are not single points of degeneracy, but hyperlines of 3N-6-2 dimensions of energy degeneracy. This degeneracy is only lifted along the Gradient Difference and Interstate Coupling vectors, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ respectively. Along the former the difference in energy of the two coinciding surfaces is maximum, whereas along

the direction of the latter vector the mixing between the wavefunctions of the upper state, ψ_2, and the lower state, ψ_1, is maximum. If we consider a CI of two states of different symmetries, $\boldsymbol{x_2}$ is the coordinate along which the symmetry is lowered in order to allow them to mix. The gradient difference and interstate coupling vectors respectively read as:

$$\boldsymbol{x_1} = \frac{\partial \left(\boldsymbol{E_1} - \boldsymbol{E_2} \right)}{\partial \boldsymbol{q}} \qquad \boldsymbol{x_2} = \left\langle \psi_1 \left| \frac{1}{\partial \boldsymbol{q}} \right| \psi_2 \right\rangle \qquad (1.6)$$

As mentioned above, these vectors are in fact coordinates which indicate the direction of energy degeneracy lifting. Their length is indicated by the parameters α and β (see eq. (1.5)), which correspond to the length of the branching space vectors at the point of energy degeneracy:

$$\alpha = \boldsymbol{x_1} \bigg|_{\boldsymbol{x}=0} \qquad \beta = \boldsymbol{x_2} \bigg|_{\boldsymbol{x}=0} \qquad (1.7)$$

The shape of the cone plays an important role in the decay as it influences the transition to the lower state and can favor certain directions. Two types of conical intersections, namely sloped and peaked CI, were defined by Ruedenberg[34] depending on the shape of the cone. These two shapes are shown in Figure 8. In sloped CIs (cartoon a), the two surfaces have similar gradients (slopes) near the point of intersection, that is why they are called sloped. In contrast, peaked CIs (cartoon b) are characterized by having gradients with opposite sign at the point of crossing.
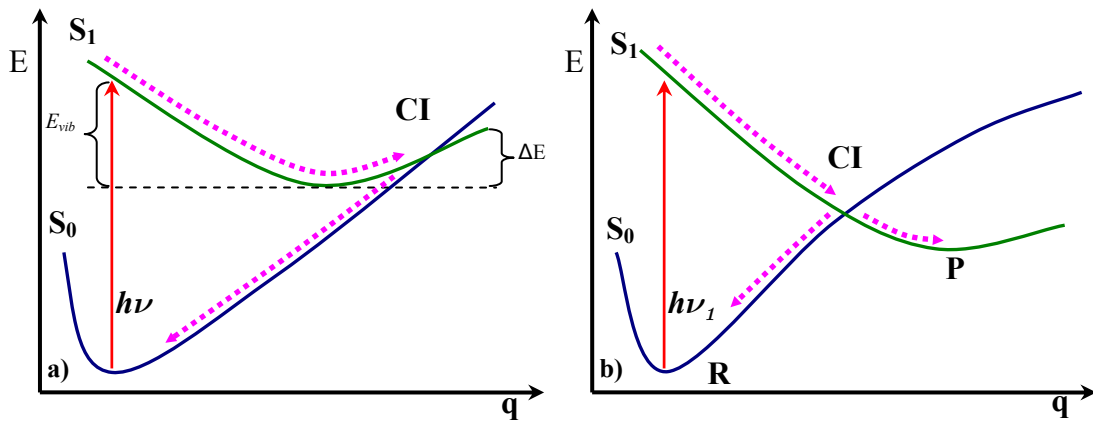
Figure 8. Sloped and peaked crossings as defined by Ruedenberg.

The mathematical conditions for peaked and sloped topologies depend on the relative magnitudes of λ and α, with the former being:

$$\lambda = \frac{1}{2} \frac{\partial \left(E_A + E_B \right)}{\partial \boldsymbol{q}} \Bigg|_{\boldsymbol{x}=0} \tag{1.8}$$

According to eq. (1.8), values of $\lambda > \dfrac{\alpha}{2}$ denote sloped CIs, whereas peaked CIs correspond to $\lambda < \dfrac{\alpha}{2}$.

A relaxation path with a sloped crossing usually leads to the initial GS geometry if the crossing is close to the FC structure. On the other hand, if the crossing lies far away from the initial structure the relaxation can easily yield a structure on the GS different from the initial geometry. If the reaction path involves a peaked crossing, both the initial geometry and a product $\boldsymbol{P}$ can be reached. The probability with which they are reached depends on the *branching ratio*. Cases where the path that regenerates the initial structure is preferred are called *aborted* conical intersections, because the population that can follow the path to product $\boldsymbol{P}$, to which the CI gives access, is diverted to reactant $\boldsymbol{R}$. Population transfers are usually faster in peaked CIs, provided that enough kinetic energy is acquired, because its shape favors non-adiabatic transitions. On the other hand, the shape of sloped CIs reduces the kinetic energy and favors oscillations on the seam, which decrease the momentum favoring adiabatic population transfers. As it will be shown in the results, the topology of the CI seam is very important for the description of the processes given that its character can change. If the seam has different topologies, each part of the seam can favor the generation of a different species, and this can be used to control the decay by driving the molecule towards a certain part of the seam.

We have just seen that the shape of the CI plays a major role in the decay to the lower state. In addition, a CI seam can have different topologies along it. The CI seam has been defined as rectilinear displacement along the intersection space. This picture is only approximate because it does not take into account second order effects. In the next section, the importance of the second order effects and the shape of the CI seam are explained.

1.1.3.3   Second order effects at CIs

In contrast to the development of the previous section, it is known that linear displacements along the intersection space coordinates, Q_{x_3} in eq. (1.4), can lift the degeneracy of the states. This is exemplified in Figure 9a, where the displacement along the intersection space vector, $\boldsymbol{x}_3$, lifts the degeneracy of S_1 and S_0. This effect is caused by the difference of the second derivatives

(Hessians) of the two states along the intersection space. This implies that the CI seam has a curved shape (see Figure 9b). Thus, the interstate and branching spaces become curved, *i.e.* they mix continuosly along the seam of intersection.
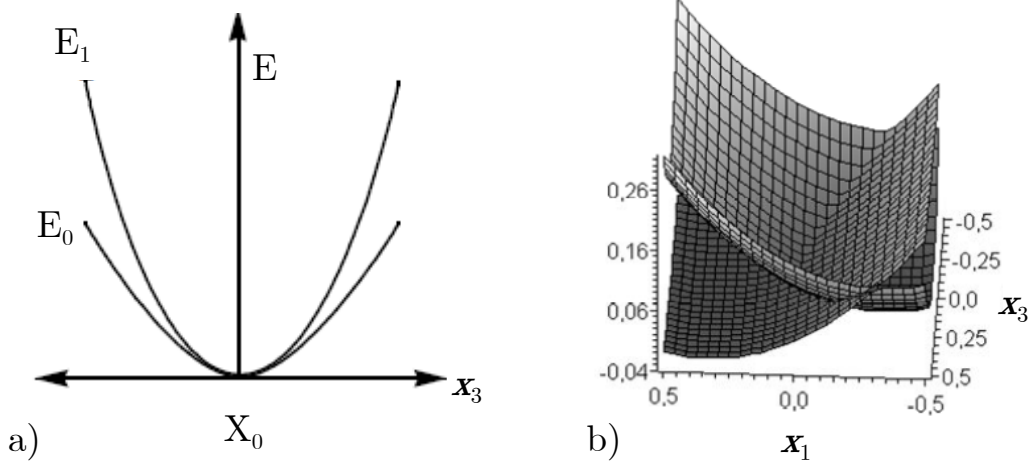
Figure 9. Second order conical intersection picture. a) 3-coordinate model potential energy surface along x_3. b) 3-coordinate model along x_1 and x_3. Adapted from Ref. 35.

In order to take this effect into account,[36-38] second order terms along the intersection space coordinates must be added to eqs. (1.4) and (1.5). Thus, the potential energy matrix reads as:

$$\mathbf{M}(\boldsymbol{Q}) = (\lambda_1 Q_{x_1} + a Q_{x_1}^2 + b Q_{x_2}^2 + c Q_{x_3}^2)\mathbf{I}$$

$$+ \begin{bmatrix} \dfrac{1}{2}\left(\alpha Q_{x_1} + \delta\gamma Q_{x_3}\right) & \beta Q_{x_2} \\ \beta Q_{x_2} & -\dfrac{1}{2}\left(\alpha Q_{x_1} + \delta\gamma Q_{x_3}\right) \end{bmatrix} \quad (1.9)$$

where the term $\delta\gamma$ corresponds to the difference between the second derivatives of the two intersecting states along Q_{x_3}. Similar terms appear for the branching space coordinates but they have been neglected for simplicity. It follows that now the energy of the two states (eigenvalues of eq. (1.9)) reads as:

$$E_{A,B}(\boldsymbol{Q}) = (\lambda_1 Q_{x_1} + a Q_{x_1}^2 + b Q_{x_2}^2 + c Q_{x_3}^2)$$

$$\pm \frac{1}{2}\sqrt{(\alpha Q_{x_1} + \delta\gamma Q_{x_3}^2)^2 + 4(\beta Q_{x_2})^2} \quad (1.10)$$

It can be readily seen that the energy degeneracy is preserved along combined displacements of Q_{x_1} and Q_{x_3}, and that the condition for energy

degeneracy becomes $\alpha Q_{x_1} + \delta\gamma Q_{x_3}^2 = -\left(\alpha Q_{x_1} + \delta\gamma Q_{x_3}^2\right) = 0$, and $\beta Q_{x_2} = 0$. This constrain describes the curvature of the seam since it allows the definition of energy degeneracy by the explicit equation $Q_{x_1} = -\dfrac{\delta\gamma}{\alpha} Q_{x_3}^2$, which is a parabola. Only one degree of freedom, f_3, is needed to characterize this parabola. For simplicity, it is convenient to define this parameter such that it corresponds to Q_{x_3}. This way the seam is defined as:

$$Q_{x_1}(f_3) = -\frac{\delta\gamma}{\alpha} f_3^2 \tag{1.11}$$

Substituting f_3 into eq. (1.10) and keeping the energy degeneracy condition ($\beta Q_{x_2} = 0$), the energy of the seam becomes:

$$\mathrm{E}_{A,B}(f_3) = \mathrm{E}_{\text{seam}}(f_3) = -a\left(\frac{\delta\gamma}{\alpha}\right)^2 f_3^4 + c f_3^2 \tag{1.12}$$

The curvature of the branching and intersection spaces is illustrated in Figure 10. In this figure, $\mathbf{t}(f_3)$ and $\mathbf{n}(f_3)$ are vectors tangent and perpendicular to the seam, respectively, which in the origin correspond to the interstate and gradient difference coordinates. It is clear that $\mathbf{t}(f_3)$ and $\mathbf{n}(f_3)$ change gradually along the seam, and that away from the origin they correspond to a linear combination of the original $\mathbf{x}_1$ and $\mathbf{x}_3$ vectors.
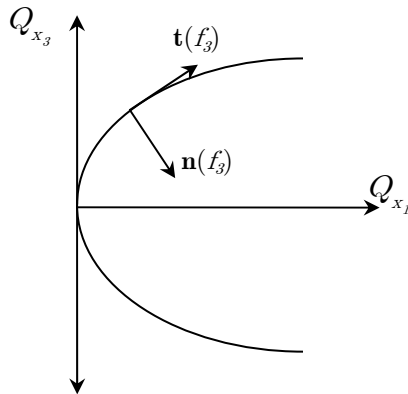
Figure 10. Projection of a seam of intersection on the $\mathbf{x}_1,\mathbf{x}_3$ plane including second order effects. Adapted from Ref. 39.

The second-order picture has two important consequences for the present thesis. First, the topology can change along the seam because of its curvature. It will be shown (see section 4.4) that this is the case for the S_2/S_1 seam of

intersection of thymine, which changes from sloped to peaked. Second, the analysis of the local branching space along the seam is a useful tool to characterize the CI seam.

It has just been shown that two CI with different branching space coordinates can belong to the same seam of intersection. Thus, the CI seam can be considered as an analogue of a Born-Oppenheimer surface, where minima, reaction paths, and transition states can be optimized. The CI seam can be mapped by running an IRC with an additional constrain that assures the energy degeneracy along the path.[40] This is used in the present thesis to determine the energy profile of the seam of intersection of thymine. In addition, minima and transition states can also be located on the seam, but they need to be characterized by evaluating the Hessian. This requires the calculation of a "special" Hessian which can provide the second derivatives in the subspace of the intersection space.

1.1.3.3.a   Intersection space Hessian

In this section the so-called intersection space Hessian will be briefly overviewed. We will first consider the generalization of eq. (1.10) to the complete number of coordinates of a given system. In this case, the energy around the intersection in intersection-adapted coordinates[39] reads as:

$$
\begin{aligned}
\mathrm{E}_{\mathrm{A,B}}(\boldsymbol{Q}) = & \left(\lambda_1 Q_{x_1} + \lambda_2 Q_{x_2} + \sum_{i=3}^{n} \lambda_i Q_{x_i} \right) + \frac{1}{2} \sum_{i,j=1}^{n} \omega_{ij} Q_{x_i} Q_{x_j}) \\
& \pm \frac{1}{2} \sqrt{ \left(\alpha Q_{x_1} + \frac{1}{2} \sum_{i,j=1}^{n} \delta\gamma_{ij} Q_{x_i} Q_{x_j} \right)^2 + 4 \left(\beta Q_{x_2} + \frac{1}{2} \sum_{i,j=1}^{n} \eta_{ij}^{\mathrm{AB}} Q_{x_i} Q_{x_j} \right)^2 }
\end{aligned} \tag{1.13}
$$

where n is the number of coordinates, λ_i are the projections of the average gradient (see eq. (1.8)) of the two states (A and B) along the branching space (i = 1,2), and intersection space coordinates (i = 3, ..., n), ω_{ij} and $\delta\gamma_{ij}$ are the elements of the average and difference Hessians, and η_{ij}^{AB} are the second-order coupling elements. At an optimized point of the intersection the terms $\lambda_i \left(i \geq 3 \right)$ are zero. Consequently, the Neumann-Wigner conditions of energy degeneracy become:

$$
\begin{aligned}
\alpha Q_{x_1} + \frac{1}{2} \sum_{i,j=1}^{n} \delta\gamma_{ij} Q_{x_i} Q_{x_j} = 0 \\
\beta Q_{x_2} + \frac{1}{2} \sum_{i,j=1}^{n} \eta_{ij}^{\mathrm{AB}} Q_{x_i} Q_{x_j} = 0
\end{aligned} \tag{1.14}
$$

Following the previous development, we define the curvilinear coordinates that fulfill the energy degeneracy conditions. For simplicity, the second-order terms that involve the branching space modes, $\delta\gamma_{ij}$ and η_{ij}^{AB} for $i,j < 3$, will be neglected. It follows that we are considering n-2 curvilinear coordinates which belong to the intersection space, namely f_i. According to eq. (1.14), these coordinates fulfill the following conditions:

$$Q_{x_1} = -\frac{\displaystyle\sum_{i,j=3}^{n} \delta\gamma_{ij} f_i f_j}{2\alpha}$$
$$Q_{x_2} = -\frac{\displaystyle\sum_{i,j=3}^{n} \eta_{ij}^{AB} f_i f_j}{2\beta}$$

(1.15)

Substituting these coordinates into eq. (1.13) and truncating to second order, the energy of the seam reads as:

$$\mathrm{E}_{\mathrm{seam}}(\boldsymbol{f}) = \sum_{i,j=3}^{n} -\frac{\lambda_1}{4\alpha}\delta\gamma_{ij} f_i f_j - \frac{\lambda_2}{4\beta}\delta\eta_{ij}^{AB} f_i f_j + \frac{\omega_{ij}}{4} f_i f_j$$

(1.16)

This expression gives the energy of the seam along the intersection space. The characterization of the critical points, namely minima and TS, is given by the matrix of second derivatives of eq. (1.16) with respect to the curvilinear coordinates (f_i $i \geq 3$), i.e. the intersection-space Hessian (H^{IS}):

$$H_{ij}^{IS} = \left(\frac{\partial^2 \mathrm{E}_{\mathrm{seam}}}{\partial f_i \partial f_j}\right)_0 = \frac{1}{2}\left(\omega_{ij} - \frac{\lambda_1}{\alpha}\delta\gamma_{ij} - \frac{\lambda_2}{4\beta}\delta\eta_{ij}^{AB}\right)$$

(1.17)

The eigenvectors of the intersection-space Hessian are the curvilinear normal modes, which are a set of vectors parallel to the curvilinear coordinates at the expansion point. The eigenvalues of the Hessian are the second derivatives of the energy along the curvilinear coordinates. From these eigenvalues one can obtain the intersection-space frequencies. Similarly to the frequencies of a Born-Oppenheimer PES, the imaginary eigenvalues of the intersection-space Hessian denote points of the seam connected to other points, in this case CIs, of lower energy along that coordinate. This analysis has been used in the present thesis to characterize the CI seam of thymine (see section 4.4.2.1).

More information on second order effects and intersection-space Hessians, as well as the full mathematical development of the equations presented above can be found in Refs. 38 and 35.

It is essential to describe CIs in detail to study the photoprocesses of molecules, since the CIs shapes play an important role in ultrafast relaxations as it will be seen in the Results of this thesis. However, the first thing that must be considered when analyzing a CI is its energetic accessibility, given that population transfers will only occur if it is energetically accessible. A way to determine if a CI is accessible is defining its minimum energy point. That is, optimizing the point in the intersection seam with the lowest energy. A way to optimize CIs is overviewed next. Note second order effects are not taken into account in the following section, see Ref. 40 for CI optimization with second order effects.

1.1.3.4   Optimizing conical intersections

CIs are hyperlines of n-2 dimensions (being n the number of geometric variables) where two PES are degenerate. In order to locate the CI minimum, the lowest-energy point on the n–2-dimensional hyperline has to be optimized. There are various algorithms for CI optimization[41-44] and the one[41] implemented in the commercial package Gaussian 03[45] is explained in this section because it has been used in this thesis.

It has been shown that CIs are hyperlines of energy degeneracy of n-2 dimensions (n-1 dimensions if they have different multiplicity). The optimization of the minimum of this hyperline can be carried out with one[42,44] or two constraints.[43] The former option is explained next. Given that at the CI seam x_1 and x_2 lift the degeneracy, the minimum of the CI has to be optimized on the remaining coordinates, *i.e.* the intersection space. At the minimum energy point of the CI seam, the gradient is 0. Thus, the algorithm consists on an energy minimization in the 3N-6-2 dimensional space orthogonal to the x_1,x_2 plane with the energy degeneracy constrain, $\boldsymbol{E_\alpha} = \boldsymbol{E_\beta}$.

The gradient of the optimization has two components ($\boldsymbol{g}$ and $\boldsymbol{f}$), and reads as:

$$\overline{\boldsymbol{g}} = \boldsymbol{g} + \boldsymbol{f} \qquad\qquad (1.18)$$

The condition of minimizing the energy difference between the crossing states ($\boldsymbol{E_\alpha}$-$\boldsymbol{E_\beta}$ minimization) is accounted for gradient $\boldsymbol{f}$ and is defined as:

$$f = \frac{\partial}{\partial \boldsymbol{q}_{\alpha}} \left(\boldsymbol{E}_{\beta} - \boldsymbol{E}_{\alpha} \right)^{2} = 2 \left(\boldsymbol{E}_{\beta} - \boldsymbol{E}_{\alpha} \right) \boldsymbol{x}_{1} = 0 \tag{1.19}$$

being $\boldsymbol{x}_1$ the gradient difference vector. The length of $\boldsymbol{x}_1$ has no significance, only its direction is important. However, since $\boldsymbol{x}_1$ is large if the CI is peaked and small if it is sloped, and this will affect $\boldsymbol{f}$, the vector is normalized so as the gradient $\boldsymbol{f}$ only depends on the energy difference. Thus, the gradient $\boldsymbol{f}$ has the form:

$$f = 2 \left(\boldsymbol{E}_{\beta} - \boldsymbol{E}_{\alpha} \right) \frac{\boldsymbol{x}_{1}}{\left| \boldsymbol{x}_{1} \right|} = 0 \tag{1.20}$$

It can be readily seen that $\boldsymbol{f}$ will go to zero when the difference in energy is small, independently of the magnitude of $\boldsymbol{x}_1$.

On the other hand, gradient $\boldsymbol{g}$ denotes the constrained optimization on the plane orthogonal to $\boldsymbol{x}_1, \boldsymbol{x}_2$.

$$g = \boldsymbol{P} \frac{\partial \boldsymbol{E}_{\beta}}{\partial \boldsymbol{q}_{\alpha}} \tag{1.21}$$

where $\boldsymbol{P}$ is the projector of the gradient of $\boldsymbol{E}_{\beta}$ into the (n-2)-dimensional space orthogonal to the $\mathbf{x}_1, \mathbf{x}_2$ plane, $\boldsymbol{P} = \mathbf{I} - \boldsymbol{x}_1 \boldsymbol{x}_1^{\mathrm{T}} - \boldsymbol{x}_2 \boldsymbol{x}_2^{\mathrm{T}}$. This gradient has to be slightly modified in the case that the two surfaces are of different multiplicity. Then, $\boldsymbol{x}_2$ is zero and the projection is carried out on the (n-1)-dimensional orthogonal complement space.

The gradient to be used for the optimization of the CI minimum is, therefore,

$$\bar{g} = \boldsymbol{P} \frac{\partial \boldsymbol{E}_{\beta}}{\partial \boldsymbol{q}_{\alpha}} + 2 \left(\boldsymbol{E}_{\beta} - \boldsymbol{E}_{\alpha} \right) \frac{\boldsymbol{x}_{1}}{\left| \boldsymbol{x}_{1} \right|} \tag{1.22}$$

One of the advantages of the method is that once $\boldsymbol{f}$ goes to 0, the minimum optimization is assured by the normal Hessian updating/evaluation procedure. This efficiency relies on the fact that $\boldsymbol{g}$ and $\boldsymbol{f}$ are orthogonal, which implies that the CI minimum can not be optimized if both $\boldsymbol{g}$ and $\boldsymbol{f}$ are not 0.

State crossing optimizations can sometimes be problematic, especially if the gradients of the crossing sates are nearly parallel and of similar length. That is because the gradient difference ($\mathbf{x}_1$) becomes very small, and this causes inaccuracies in the projection onto the constrained (n-2)-dimensional space. If that shall happen, a step size decrease could solve the problem. Eventually, one would have to settle for less strict convergence criteria.

The optimization of crossing points between surfaces has been explained, however there might be different paths which lead from an excited state to the GS. That is, not only the characterization of minima and TS is essential to study the decay mechanism but the paths that interconnect these structures are crucial too. A way to interconnect stationary points is explained next.

1.1.3.5   Interconnecting stationary points

The critical points of a PES represent the reactants, products and intermediates of a reaction, whereas the pathways that interconnect them describe the mechanisms of the reaction. An infinite number of paths can connect two points in a surface, however, in static studies only the one that requires the least increase in energy to get from reactants to products is considered as it depicts the easiest way to go from one point to another. Such a path is commonly known as Minimum Energy (Reaction) Path (MEP) and it usually follows the steepest descent path. Unfortunately, the steepest descent path of a surface can vary upon the coordinate system. There is a coordinate system that has special significance from a dynamic point of view. The steepest descent step in mass-weighted Cartesian coordinates corresponds to the one followed by a molecule with no kinetic energy. The steepest descent path in mass-weighted Cartesian coordinates is also known as Intrinsic Reaction Coordinate (IRC).[46] Note an IRC does not account for a classical trajectory since in a classical trajectory, the kinetic energy deviates the molecule from the ideal MEP giving access to other parts of the PES. For this reason, IRCs cannot substitute dynamics simulations, although they are essential to rationalize the dynamics results as they depict bottom line paths that serve as reference.

There are different algorithms to calculate MEPs. They are defined as first-order, second-order and higher-order methods. We next explain the González-Schlegel second-order method[47,48], which is implemented in Gaussian 03, and has been used in this thesis.

1.1.3.5.a   González-Schlegel IRC algorithm

The IRC procedure is summarized in Figure 11. It is based on the use of 3 points to generate an arc that depicts the MEP. From a starting point $\mathbf{q}_k$, with gradient $\mathbf{g}_k$, a new point $\mathbf{q}_{k+1}$ is optimized so that the path between the two points is an arc of a circle, and so that the gradients of the two points ($\mathbf{g}_k$ and $\mathbf{g}_{k+1}$) are tangent to this arc. In practice, the algorithm starts at $\mathbf{q}_k$ and first generates a pivot point, $\mathbf{q}^*_{k+1}$, at $(1/2)s$ distance along the gradient $\mathbf{g}_k$. This pivot point is, therefore, defined as:

$$\mathbf{q}^*_{k+1} = \mathbf{q}_k - (s/2)\frac{\mathbf{g}_k}{\left|\mathbf{g}_k\right|} \tag{1.23}$$

Then, the pivot point is taken as the center of a hypersurface of radius $s/2$ from which a constrained optimization is performed so as to obtain the new point $\mathbf{q}_{k+1}$. This point assures that the energy gradient along the arc defined by $\mathbf{g}_k$ and $\mathbf{g}_{k+1}$ is maximum. Because of the constrain, the gradient of the new point, $\mathbf{g}_{k+1}$, is parallel to $\mathbf{q}_{k+1} - \mathbf{q}^*_{k+1}$. Hence, $\mathbf{q}_k$ and $\mathbf{q}_{k+1}$ lie on an arc of a circle of tangents $\mathbf{g}_k$ and $\mathbf{g}_{k+1}$.
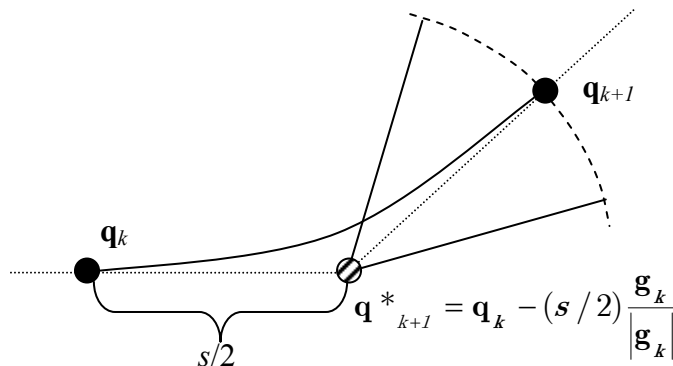
Figure 11. González-Schlegel IRC algorithm

IRC algorithms yield a set of discrete points on a PES that altogether depict a MEP. Depending on the length of s, the position and the energy profile along the path may vary. In extreme cases, different values of s can lead to completely different products of a reaction.

In general, IRCs need a starting direction to optimize the MEP. This initial direction is perfectly defined at TSs, where the imaginary frequency vector indicates the direction towards which the energy decreases. Thus, IRCs starting at TSs use the imaginary frequency as initial direction (forward or

reverse). Nevertheless, IRCs can also be started in regions where the steepest descent line is not predefined as is the case of CIs or FC structures. In these structures, there is not an imaginary frequency that indicates the minimal energy direction. Thus, an initial relaxation direction (IRD) must be provided to the IRC algorithm. A simple technique to obtain IRDs is explained next.

1.1.3.5.b   IRD

The major problem of the IRC algorithm is that the placement of the pivot point is crucial, and if it is not placed on the steepest descent path, it might lead to a MEP which is not the true MEP (such an effect is illustrated in Figure 12). Thus, one of the major utilities of the IRD procedure is to define the proper direction towards which the pivot point is to be set. The general procedure to calculate the initial reaction direction is based on the analysis of the vicinity of the starting point and is shown for the particular case of the FC structure in Figure 12. It is based on the optimization of the energy of a hyperspherical cross-section of the hypersurface centered (red line in Figure 12) at the starting point. That is, the point of minimum energy is optimized on a hypersphere of radius d, centered at the FC. The energy profile of the optimization can show different minimum energy points which correspond to different paths. The vectors that join the starting point with these minimum energy points define the different IRDs.
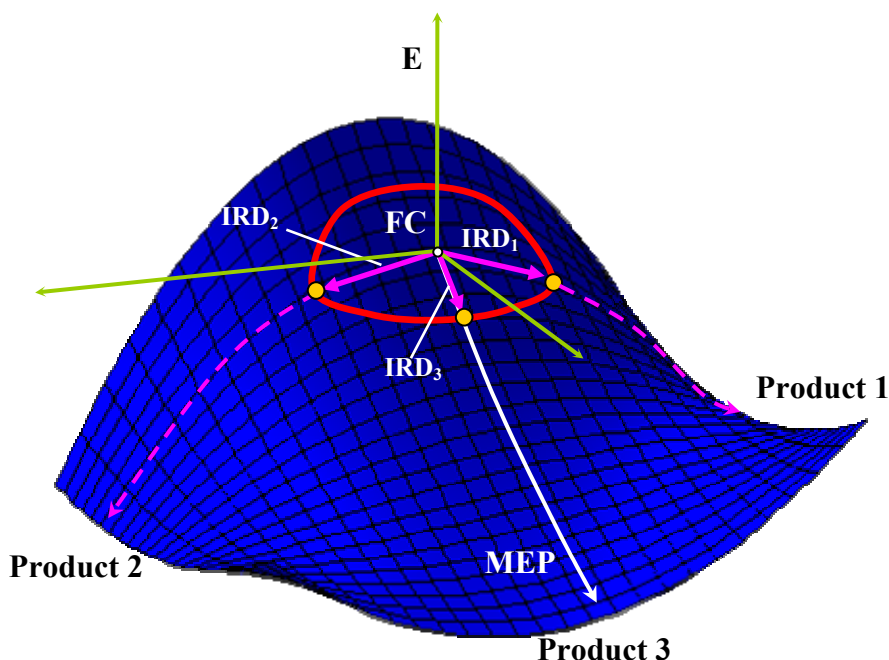
Figure 12. IRDs calculated from a circular cross-section and corresponding MEPs from the FC structure.

The radius of the hypersphere is usually chosen to be small (0.25 a.u.) so as to define the steepest descent direction on the vicinity of the starting point. Various MEPs can be obtained by following the IRDs. This is also illustrated in Figure 12. In this figure, the MEP follows IRD_3, however, there are also 2 other IRDs that define competitive paths which lead to products 1 and 2 respectively. Thus, the IRD procedure can be used to explore different parts of the PES which might be important for the photophysics of the molecule as they may lead to crossings with other states, generation of various photoproducts, or could explain competing deactivation paths.

The IRD procedure can also be used to map the shape of the PES in the vicinity of the starting point. Given that the IRD algorithm optimizes the energy of the hypersphere, a sort of contour plot map can be generated by successive IRD calculations. Keeping the starting point, the accuracy of the plot is determined by the increase of the radius of the hypersphere. Such a procedure is depicted in Figure 13.
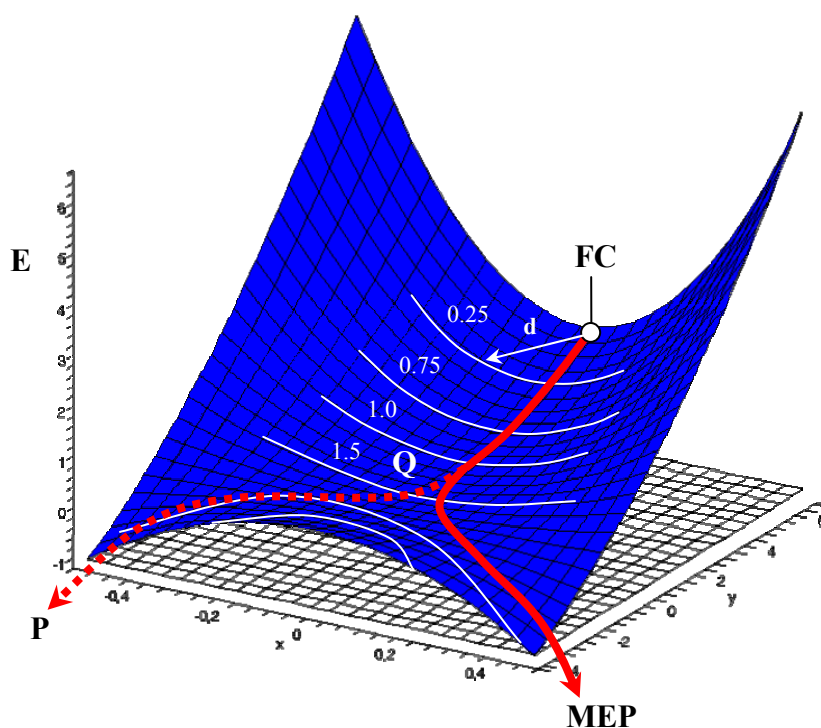
Figure 13. Characterization of a PES with successive IRD calculations of increasing radius (d).

Similar to the IRC, this type of IRD calculations give the profile of the PES. It is observed that the MEP leads from the FC structure to the point Q, where two deactivation paths emerge. In this example, the MEP corresponds to the full red line. However the use of the IRD shows that there exists a

competing deactivation path (dashed red line). A similar scenario is observed when the starting point is a CI (see Figure 14). In this case different IRDs serve to map the vicinity of the CI and show that different products can be reached from the cone tip.
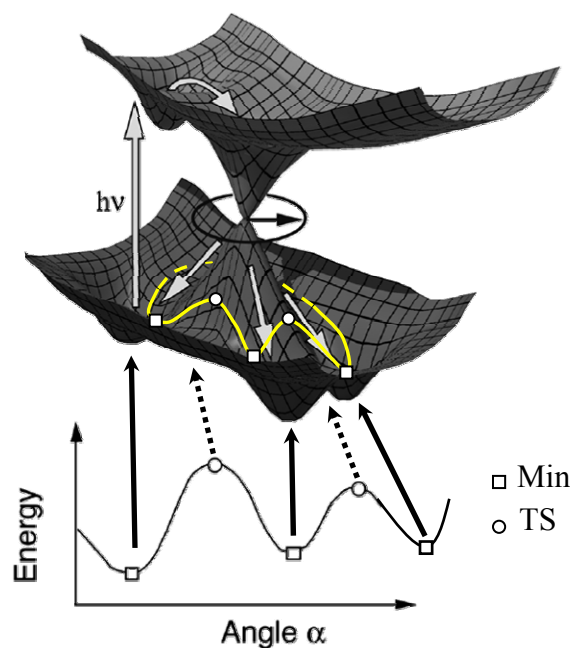
Figure 14. Deactivation paths from a CI and energy profile along a circular cross-section centered on the CI point of radius d. Note: the general IRD procedure extends the energy minima search to an (n-1)-dimensional spherical cross-section (hypersphere) rather than a mono-dimensional cross-section as depicted above. Adapted from Ref. 49.

It has been shown that the IRD procedure allows the extensive study of PESs. Static calculations[50-56] are very useful to determine all the competing relaxation paths of a molecule.[57-59] However, as mentioned before, experimental lifetimes and branching ratios can only be determined with the help of dynamics simulations. A brief introduction to dynamics simulations is explained next.

1.1.4   Dynamics simulations

Static calculations serve to describe the PES of a molecule, which can be used to rationalize the mechanisms of reaction. In simple thermal reactions, the preferred reaction mechanism is usually determined by comparison of energy barriers. The path which has the lowest barrier determines the preferred reaction mechanism. In photochemistry, however, not only the height of barriers

is important but the excited state lifetimes and branching ratios are also
important to describe photoinduced reactions. These properties cannot be
determined by static models. Nevertheless, molecular dynamics can reproduce
the time scales of the relaxation paths of a molecule and non-adiabatic
properties such as branching ratios. In addition to that, dynamics simulations
can be very helpful in photophysics studies since they represent the theoretical
approach to ultrafast pumb-probe experiments.

There are different types of molecular dynamics simulations, and they can
be classified depending on the way the nuclear motion is treated. Two main
groups, namely classical and quantum dynamics methods, define the boundaries
of the molecular dynamics framework. These two methodologies present some
limitations especially when dealing with non-adiabatic processes of large
systems. A large variety of methods, which allow the study of such systems,
have been developed. Next, the full classical and quantum dynamics methods
are explained first together with some of their limitations. Later, different
methods with a mix classical-quantum character, which approach quantum
dynamics are explained.

Pure or full quantum dynamics methods are based on the quantum
solution of the nuclear and electronic parts of the Schrödinger equation.
However, they can only be applied to small molecules of no more than 4 atoms,
because the so-called many-body problem arises for larger molecules. The
movement of many (>4) interacting systems can not be analytically solved.
Thus, the use of approximations to solve the nuclear problem is necessary for
most of the systems. One of the most extended approaches is based on the use
of a wavepacket to describe the nuclear motion. In spite of this approximation,
such molecular dynamics are still considered "full" quantum dynamics, although
strictly speaking they are not. In these full quantum dynamics of nuclear motion
methods, both the electronic and nuclear parts are solved by means of quantum
mechanics. One of the strong points of these methods is that the gradient and
Hessian are calculated analytically. This requires an analytical expression for
the PES. Unfortunately, because of its complexity, in most cases it is impossible
to directly obtain an analytic expression for the PES. The solution to this
problem represents a main problem of the method, since the analytical
expression is usually obtained from mathematical fitting to a previously
calculated grid of points on the region of interest. The main problem is that
analytical fittings can only be carried out for a reduced number of degrees of
freedom. Thus, one must reduce the number of degrees of freedom before
generating the grid. This is usually accomplished by freezing those coordinates
which are less relevant for the dynamics. Such an approximation is usually a
good solution for ultrafast processes as only a few coordinates are activated
during the reaction. However, if complicated systems or larger timescales are

considered, the reduced analytical expression of the PES may not describe the system properly.

On the one hand, the molecular dynamics approach is based on the use of classical mechanics to determine the movement of molecules, which are described by particles. In contrast to quantum dynamics, this allows to treat a large (thousands) number of atoms. Most of the classical dynamics methods do not use an analytical expression for the PES from which the Hessian and gradients are obtained. In one hand, this represents a disadvantage with respect to full quantum dynamics, because instead the energy and Hessian need to be calculated (locally) at each step of the trajectory (on-the-fly) and it might be rather expensive, especially if *ab initio* methods are used. However, there exist Hessian updating techniques which allow for a fast estimation of the Hessian and can be used in those cases where the second derivatives calculation is too expensive. On the other hand, at the same time the on-the-fly calculation procedure represents an advantage vs. the full quantum dynamics since it avoids the bottleneck that represents the calculation of the grid of points (previous to any full quantum dynamics simulation) that is used to obtain the analytic expression for the PES. In addition it allows to take into account a larger number of degrees of freedom. One of the problems of this methodology is that a large number of trajectories must be run in order to get a random Gaussian distribution of trajectories, which emulates the nuclear wavepacket – only a broad sampling assures all the geometries relevant for the dynamics are explored –. However, the biggest inconvenient of classical dynamics is that it cannot be used to treat non-adiabatic processes such as bond breaking, ultrafast excited states relaxations or charge transfer problems, as they require a quantum electronic structure description that goes beyond the Born-Oppenheimer approximation.

The inclusion of quantum dynamics effects into classical methods defines the quasi-classical methods. Classical dynamics with trajectory surface hopping is one of the most extended quasi-classical approaches. This allows the treatment of non-adiabatic processes by means of classical dynamics. In most of these methods the electronic wavefunctions are computed on-the-fly as solution of the Schrödinger equation, and the nuclei are propagated classically with the surface and velocity of the electronic part. Multiconfigurational methods such as CASSCF, assure the proper description of the different states, and the surface hopping algorithm determines at which state the Hessian has to be calculated, *i.e.* allows the description of non-adiabatic processes such as relaxations that go through CIs. This is the reason why a semi-classical dynamics method with a surface hopping algorithm has been used in this thesis. There are other approaches, namely mixed-quantum-classical dynamics, which have a dual quantum and classical character. In these methods, some degrees of freedom are

treated classically and others quantum mechanically. Another approach to full quantum dynamics of nuclear motion is provided by some "quantum" methods which use Gaussian functions to approximate the nuclear wavepacket. One of these methods is the Direct Dynamics variation Multi-Configurational Gaussian (DD-vMCG) and has been used in the present thesis to study the S_2/S_1 CI of thymine (see section 2.2). One of the main advantages of this method is that, unlike the full quantum dynamics methods, the PES and Hessian are evaluated locally and on-the-fly, which allows the treatment of larger numbers of degrees of freedom. The number of Gaussian functions that expand the wavepacket determine the degree of approximation to full quantum dynamics. In the limit of expansion (infinite Gausians), the DD-vMCG method turns to be a full quantum dynamics method of nuclear motion. Another approach to quantum dynamics is the so-called Multi-Configurational Time-Dependent Hartree method, from which the DD-vMCG is derived. This method is considerd a full quantum dynamics method of nuclear motion, since as defined before it uses a wavepacket to approach the nuclear motion and a reduced analytical expression of the PES.

We have just seen how the improvement on the description of the nuclear motion serves to converge to full quantum dynamics. However, the description of non-stationary processes is not only dependent on the nuclear part, but also on the electronic one. In this sense, mainly three different approaches are used, namely semi-empirical, DFT, ab initio, and QM/MM methods if solvent is taken into account. Since the aim of this thesis is not to give a review on molecular dynamics methods, the reader is referrered to literature[60-64] if more information on dynamics methods is desired.

The most relevant advances and the most important works on the photophysics of DNA are briefly explained next. More detailed information on thymine is given in sections 1.2.1 and 1.2.2.

1.2   Experimental and Computational Background

The response of DNA to light has been object of study since the early 1960s. However, due to the femtosecond character of DNA relaxation mechanisms, only recent technological advances in femtosecond pumb-probe experiments have allowed the direct study of such ultrafast phenomena. In this section, we will review the most important advances concerning the excited states of DNA.

The study of DNA deactivation mechanisms has been one of the major applications of the time-resolved absorption experiments. The precision and accuracy of these studies, however, has been limited by the laser and detector technologies. Nowadays, femtosecond time-resolved absorption experiments carried out on nucleobases try to determine the deactivation mechanisms by direct determination of the lifetimes of the different species formed upon excitation. Such a precision was unreachable in early experiments, which were limited to fluorescence analysis studies. However, in spite of the resolution of actual experiments, decay mechanisms cannot be directly identified, but only a general picture can be drawn from the experimental data. The decay mechanisms are assigned following some general hints. For instance, ultrashort (hundreds of fs) excited state lifetimes indicate reactions are unlikely to take place during the decay path and that relaxation proceeds via internal conversions. On the other hand, long decay timescales (ps-ns) do not necessarily imply photoreactions occur in the excited states, but they can indicate quenchings of other states or the presence of barriers or trappings. Nevertheless, it is generally accepted that long excited state lifetimes favor photolesions because more intermediates and states are involved in the decay, and therefore, there are more possibilities of obtaining different structures.

Absorption experiments are usually carried out in solvent due to its simplicity and also because the solvent mimics the environment the DNA has in the cell nucleus. However, it is difficult to know if the aqueous solution used in the experiments can emulate that of the nucleus or if, on the contrary, the differences with the real environment can affect the decay processes. Some experiments[65-68] reveal that the DNA hydrophobic interactions exclude water from inside the DNA, which reduces its interaction with the environment. It has also been argued that the solvent might wash away the properties of the nucleobases as it can cause geometric disorders, change in the conformations or

tautomerism. Such an uncertainty makes both experiments in the gas phase and in solvent necessary, given that the effects that the solvent can have in the decay process can only be determined by comparison of solvent and gas phase experiments.

On the theoretical side, most of the studies are carried out in the gas phase, as is the case in this thesis, because of the high computational cost of the calculations which take into account the effects of the solvent. Thus, we will generally focus on gas phase experiments, although the most important works under solvent conditions will also be explained.

The first time-resolved experiments for the study of the excited states of nucleobases were carried out by Nikogosyan[69] in the early 80's. Those and other studies[70,71], reported excited state lifetimes for the nucleobases of the order of ~4 ps. Lasers with femtosecond excitation pulses were introduced[72,73] later in the 1990s, and should have substantially improved the excited state lifetimes resolution. However, too high pump intensities led to multiphoton absorptions and multiple overlapping signals which reduced the precision of the technique. Only Häupl[74] *et al.* provided shorter excited state lifetimes measures of the order of 1ps with an excitation pulse resolution of 200fs, although these lifetimes were of the same order ore even shorter than the instrumental response function which was estimated to be of ~4ps, and a further study showed[75] that such a time resolution was not sufficient to capture the subpicosecond excited decays. The most important finding of the decade was probably the announcement of a bi-exponential decay character for guanine, which was reported by Fujiwara *et al.*[76] in an experiment carried out at different pH values. That bi-exponential decay character was the first indication of the existence of different excited state deactivation channels for nucleobases, what would lately become a focus of discussion.

Later on, Pecourt *et al.* reported the excited states lifetimes of a number of nucleosides,[75,77] which lied on the subpicosecond range, which supposed a breakthrough in the photochemistry of DNA. Gustavsson *et al.* reported a bi-exponential character deactivation for thymine[78,79] and adenine,[25] which was also reported[80] for the rest of DNA nucleobases in another study. In those studies, the shortest deactivation channel was assigned to the direct radiationless deactivation from the lowest π,π^* state. Whereas, the origin of the long one remained rather uncertain until recently Hare *et al.*[81] found it corresponds to a dark state (n,π^*), which also acts as a gateway to populate a triplet state $(^3\pi,\pi^*)$. In the same work, the authors described that not only the bases were involved in the deactivation process but the phosphate group too as they observed pyrimidine nucleotides exhibited longer decay lifetimes than the corresponding nucleosides. Such an effect was not observed in purines.

Turning to the gas phase (see section 1.2.1 for a more detailed explanation of the gas phase experiments on thymine), the first experiments on nucleobases were carried out in supersonic jet expansions,[82,83] although jet-cooled with resonance-enhanced multiphoton ionization (REMPI) or laser-induced fluorescence (LFI) studies proved, later, to be much more efficient in providing information on deactivation processes. The first pump-probe transient ionization experiments[23,84] on jet-cooled bases carried out by Kang *et al.* represented a breakthrough on the field. The authors observed monoexponential ultrafast decay timescales (0.8-6.4 ps) from the first singlet excited states of nucleobases, with the exception of thymine, which also showed a longer deactivation channel, firstly assigned to a triplet state and subsequently reassigned by He *et al.*[24,85] to a dark state (n,π*). A three step relaxation mechanism was also proposed by Ullrich *et al.* in 2004.[86,87] In contrast, Canuel *et al.*[88] using a high resolution femtosecond laser (80fs), proposed a two step deactivation mechanism for all DNA and RNA nucleobases which included ultrafast conversions to n,π* states. These latest works proved that the early experiments[69,73,89] used too low resolutions to detect the ultrafast relaxation decay lifetimes of DNA. Nevertheless, the study of the photophysics of DNA is not a resolved issue as technological advances give rise to new and more accurate studies which complement the proposed picture and can provide different or competing decay mechanisms. This is the case of adenine, whose relaxation was generally accepted to proceed via a two step mechanism that involved the π,π* and n,π* states before regenerating the GS, and a recent work[90] has proved that a π,σ* dissociative channel plays a major role in the electronic relaxation.

In general, regarding the relaxation of DNA nucleobases, studies in the gas phase and aqueous solution reach similar conclusions. In spite of some discrepancies in the ordering of the lifetimes of the nucleobases, both agree two deactivation channels exist for nucleobases. An ultrafast internal conversion (hundreds of fs) leading to the ground state has been assigned to a barrierless path. The excited state with a lifetime of tens to hundreds of ps has been assigned to a relaxation from an n,π* state that acts as intermediate in the relaxation from the π,π* state.
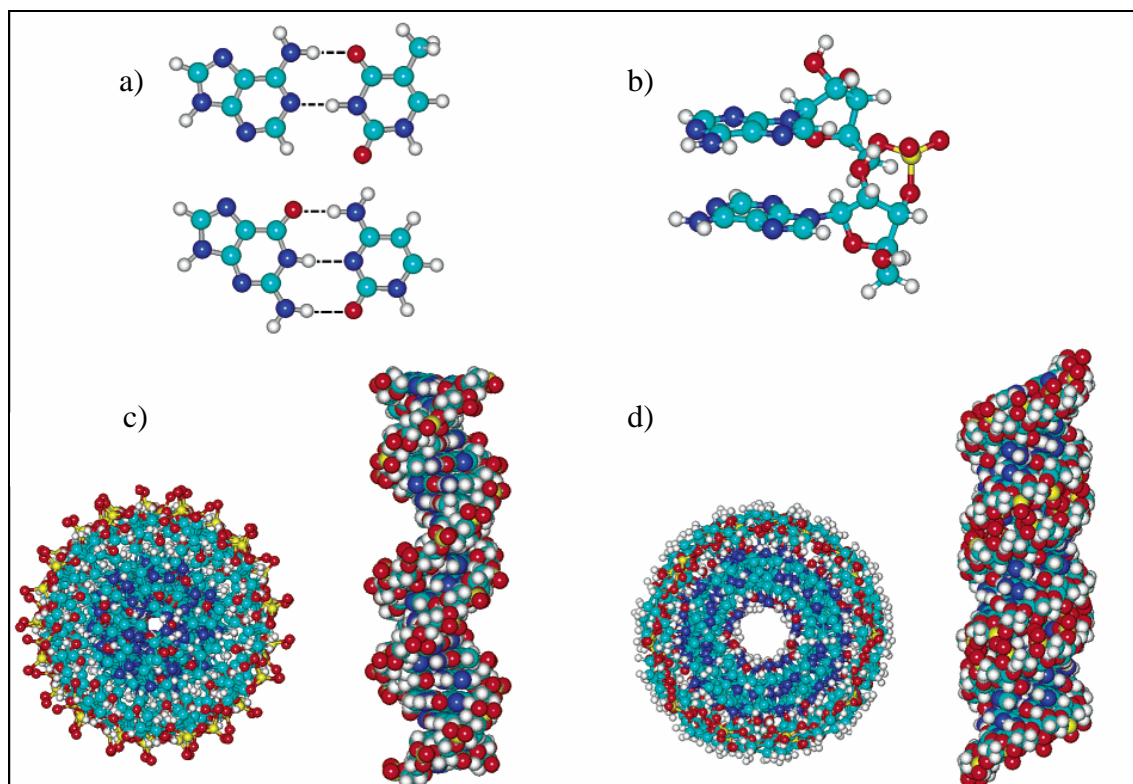
Figure 15. Representative structures of base multimers from Ref. 91: (a) Watson and Crick base pair, A-T (top) and G-C (bottom); (b) base-stacked form of the dinucleoside monophosphate ApA; (c) B-form double-stranded DNA, views down the helical axis (left) and from the side (right); (d) A-form double-stranded DNA, views down the helical axis (left) and from the side (right).

Excited state dynamics of multiple monomers (multimers) in the form of dimers, base pairs, polymers and double strand chains (see Figure 15) are much more complicated than those of single bases. The proximity of the bases in such structures leads to new states that are not present in the isolated bases.

Nucleobases interact in 3 different ways: 1) a phosphate group (sugar) covalently links one nucleobase to each other. 2) Van der Waals interactions between the π orbitals of the atoms in the ring dispose the bases in vertical stacks (π stacking). 3) H-bonds between adjacent bases. The result of these interactions is the well known α-double helix structure of DNA. Such structure plays an important role in the deactivation process, since the extra energy gained in photon absorption is distributed among the bases. When absorptions take place in a multimer, different species can form. An excimer[92-94] is an emissive excited state formed by an excited molecule somehow coupled to another identical base which remains in the GS (*i.e.* the excitation is localized in a single nucleobase). When the excitation is delocalized between different bases, an exciplex is formed. That is, upon photon absorption one base approaches

another one, which remains in the GS, and couple forming the exciplex. Once the relaxation has taken place, the exciplex is no longer stable and it decomposes into its initial components. A more general term is denoted by exciton,[95,96] which is used to name delocalized excitations in the multimer. Because the aim of the present work is not the study of excimers' photophysics, only some general features about these excited state species will be explained. Nevertheless, an extensive discussion about formation of excimers and other related species can be found in an excellent review by Crespo-Hernández *et al.*[91]

The excited state dynamics of DNA dimers or base pairs are rather similar to those of the single monomers. They both have been described[97-99] to proceed via ultrafast relaxation channels involving n,π* states as intermediates, and quenchings to π,σ* states in adenine.[98] Long-lived species such as excimers[100] and H transfer reactions have also been suggested to take place in dimers after photon absorption.[99,101-103] In spite of the great chemical interest of such structures, recent investigations have pointed out that dimers seem not to be appropriate simplified models of DNA, as their excited state dynamics differ from those of the multimers.[104-109] Thus, the study of larger models of DNA, such as oligomers, is starting to focus the attention of the research community.

As explained above, the multimers have different emissive species from monomers and dimers. Oligomers are characterized by having long-lived excited states with red-shifted fluorescence compared with that of the simpler models.[110] These longer fluorescent decay times do not lead to more fluorescent species, but they are thought[111-113] to act as DNA photolesion precursors. The formation of these species and some photoproducts were found to depend on the conformation[114] adopted by the multimer at the instant of absorption[115] and on the base sequence of the multimer.[109] One of the first time-resolved studies on oligomers was carried out by Plessow *et al.*[108] The authors reported biexponential decay lifetimes with a short component of hundreds of ps and a long one of the order of ns that was assigned to excimer fluorescence. They also got evidence that adenine-adenine and adenine-thymine stacks were responsible for most of the long-lived emissions and that adenine had a greater tendency to form excimers. Markovitsi *et al.*[106] performed the first femtosecond fluorescence study on oligomers and they observed that the majority of the fluorescence decayed in a few ps. Surprisingly, no long-lived channels on the ps-ns range were observed, but the authors agreed that longer-lived channels, which exceeded the detection limit, could not be excluded.

Crespo-Hernández *et al.* shed some light on the deactivation channels of some oligomers,[100,105] not without controversy,[104,107,116] when they claimed that every excitation that takes place in a base stack decays to an excimer which may subsequently regenerate the GS or trigger a photoproduct. As it was first

reported by Plessow *et al.*, this mechanism was mostly observed in adenine oligomers, whereas no excimers were detected in thymine ones. Thymine oligomers showed similar excited state lifetimes to those of the monomers, clearly indicating that excitations and subsequent deactivations take place in single nucleobases. However, the fact that thymine photodimers were also detected, may suggest that excimers formed in the thymine oligomer might lead to T-T photodimerization. A similar conclusion was drawn for adenine-thymine duplexes, as both phenomena were observed. However, it has been stated in a recent study by Takaya *et al.*[117] that the long-lived excited states seen in long DNA model systems[100,114,116,118,119] are formed in stacks of two bases and correspond to exciplex states created by interbase charge transfer. The authors also point out that these exciplex states are not formed initially but are populated after Frenkel excitons[120] (excitations that travel from one atom to another, i.e. traveling excited states). Frenkel excitons can be described as. Although exciplexes seem to be responsible, or at least contribute to the decay mechanisms, there are still many questions to be answered, and thus more studies are needed to get a definitive mechanism for the DNA deactivation. traveling

Despite the great amount of studies performed on DNA models, there is very few available data[72,121,122] on native DNAs. Since it seems that the base sequence plays a major role in the relaxation transient and that it varies along the DNA double helix, it is quite difficult to assign decay time scales or relaxation mechanisms within a DNA double strand as many shall take place at the same time. Therefore, only deep technological advances can shed light on such an issue as up to now only the study of simplified models can be afforded.

In the next section the most relevant works on thymine, mainly on the gas phase, are reviewed.

1.2.1   *Thymine experimental studies*

There is not a general agreement on the relaxation of thymine when decay mechanisms involving two and three steps are described[23,24,87,88] depending on the type of experiment and precision of the method.

Three intrinsic lifetimes ranges have been reported: A short component of hundreds of femtoseconds, a larger one of tenths of picoseconds and a long-living dark state of hundreds of nanoseconds. Chronologically, Kang *et al.*[23], for the very first time in the gas phase, reported a double exponential decay for thymine in a 267 nm pump-probe ionization experiment.[84] In this experiment, the molecule is pumped to its excited state by a third harmonic pulse of a

Ti:sapphire and then, after a certain time delay, the population of the excited state is probed by multiphoton ionization using fundamental light. The short component was consumed in 6.4ps and a larger one, which was assigned to a triplet state, in 100ps. Later on, a resonantly enhanced pump-probe multiphoton ionization experiment[24,85] (REMPI) showed the presence of a long living dark state for different methyl substituted uracil, and thymine nucleobases with excited state lifetimes of 23 to 209 ns depending on the excitation energy and the substituents. The authors suggested that the dark state should correspond to a low lying n,π^* state, which might be coupled to a π,π^* state by out-of-plane vibrational modes via conical intersections. It was observed that when the excitation energy increased the dark state lifetime decreased, giving evidence for the existence of an energy barrier which hinders the effective deactivation from that state through a CI with the ground state. In the same work, He *et al.* claim that the large deactivation component, which was assigned to deactivation from the triplet state by Kang *et al.*, could also correspond to a trapping in a dark state.

Three channels of $<$50fs, 490fs and 6.4ps[23] were reported in the deactivation of thymine in a Time-Resolved Photoelectron Spectroscopy (TRPES)[123] experiment at 250 nm.[87] In this experiment, a high temperature glass nozzle is used to introduce the sample into a magnetic bottle photoelectron spectrometer where the molecular beam is pump-probed. The authors pointed out that they could not obtain reliable global fits due to the poor signal-to-noise ratio of the long time range TRPES spectra. Nevertheless, they could reproduce other literature values within 0.5 ps. of accuracy.

More recently, an experiment[88] with a mass-selected resonant ionization technique[124]– two laser pulses are generated in a Ti:sapphire. One pulse pumps the molecule to its excited state and the other pulse goes to the delay line and probes the population of the excited molecule in the extraction region of a linear time-of-flight mass spectrometer – with a resolution of 80fs at 267nm suggested a two component decay for the nucleobases with components of 105fs and 5.12ps for thymine. Although only adenine was studied in detail in this work, a common decay behavior was described for all nucleobases. The short components were assigned to the $\pi,\pi^*/n,\pi^*$ transition, whereas the longer ones to the deactivation from the n,π^* state.

Summing up, experimental studies of thymine photophysics report a multi-exponential dynamics pathway with one or two short components of $<$500fs followed by a longer one of tenths of picoseconds (5.12 – 6.4ps). There is also evidence for the presence of a long living (hundreds of ns) dark state of n,π^* character lying between the π,π^* state and the GS. That dark state is probably connected to the GS through a CI which is hardly reachable due to the presence

of an energy barrier on the same state. The presence of a triplet state in the same energy range was also reported.

Taking into account that technology has not yet evolved enough to study photophysical processes in great detail and that different experiments reach different conclusions, it arises that computational chemistry turns out to be essential to shed some light on the photophysics of nucleobases. Theoretical calculations are meant to reproduce experimental results giving further explanations to unresolved experimental data. The advantage of the computational studies relies on the fact that they can easily identify the character of the states and can provide valuable information on the decay mechanisms. On the other hand, large and complicated systems can only be studied using simpler models, what can make the experimental assignment difficult.

1.2.2   Computational studies

Due to the fact that the photophysics of nucleobases is one of the hot topics of this decade, some of the results explained in this section have appeared during the realization of the present thesis and have overlapped our owns. In these cases, only the results of the literature are presented.

Initial computational studies[125-127] on DNA predicted planar ground state minima structures for the nucleobases, with the exception of those with a NH_2 group, which was optimized slightly out of the ring plane.

Non-planar minima for the two lowest excited states of thymine, uracil and cytosine were firstly reported by Shukla *et al.* in a CIS study.[128] The presence of these minima indicated that the relaxation of the photo excited pyrimidines would probably take place in the ps-ns range. Nevertheless, a computational study[129] within the MRCI framework provided the first evidence for a barrierless path on uracil. It revealed the presence of a gradient-minimized direct pathway from the Franck-Condon (FC) region to an ethylenic S_1/S_0 conical intersection, characterized by a C-H bond lying almost perpendicular to the ring plane. In the same work, a decay to a dark state after a S_2/S_1 internal conversion and a CI between the n,π^* and ground states were also described. Subsequent studies have located the ethylenic S_1/S_0 intersection at the complete active space self consistent field (CASSCF) and complete active space second order perturbation (CASPT2) levels of theory both for thymine and uracil,[130,131] and evidence has been also found at the completely renormalized equation of motion coupled cluster (CR-EOM-CCSD(T)) level.[132] The intersection is an analogue of the twisted ethylene conical intersection.[133-136]

A similar barrierless decay mechanism for thymine was suggested by Perun *et al.* in a later work[130] which gave notice of a series of CIs and non-planar minima for the two lowest excited states of thymine at the CASSCF and CC2 levels of theory. In the same work, the fs and ps channels reported in Ref. 88 were assigned to the barrierless relaxation and the deactivation from low vibronic levels of the n,π^* state, respectively. It was also argued that the relaxation lifetime of the n,π^* state depends on the relative energy of the minimum with respect to a barrier that separates it from a CI with the ground state. The authors suggested that the deactivation from that state could also account for the ns component reported in Ref. 85.

The barrierless decay from the FC region was also studied at the CASPT2 level of theory and generalized to all pyrimidine nucleobases by Merchan *et al.* In that work,[131] the n,π^* state was found to lie above the π,π^* one all along the relaxation channel. In such a case, the population of the n,π^* state is unlikely, and the fs channel was assigned to the barrierless passage from the FC to the $CI(S_1/S_0)$, whereas the ps component was proposed to arise from the deactivation of a high-lying planar minimum optimized on the same π,π^* state. In a subsequent work,[137] the same group found two CIs along the barrierless path which lead to the triplet manifold. The CI with higher energy was located in the FC vicinity and could lead to either the singlet or triplet states ($^{1,3}n,\pi^*$) as they overlap in that region. On the other hand, a crossing with the $^3\pi,\pi^*$ state was found at the same region of the S_1/S_0 CI, which can be reached without a barrier. On the basis of these findings, the ps channel was reassigned to the deactivation of the triplet manifold, and the consumption of the high-lying planar minimum was estimated to take place in ~9 ns.

Another explanation for the decay lifetimes of thymine was proposed by Hudock et. al.[138] on the basis of a mixed *ab initio* molecular dynamics and time-resolved photoelectron spectroscopy (TRPES) study. The dynamics calculations were carried out at the CASSCF level and the PES was optimized using the CASPT2//CASSCF approach. The fs passage was assigned to the vibrational relaxation from the FC point to a non-planar minimum optimized on the π,π^* state. This minimum lies on S_2 above the optically dark n,π^* state. The same authors attributed the picosecond component to further deactivation from that minimum. Note, these authors performed a full-dimension direct quantum dynamics study, but their result was rather surprising as most of the trajectories were started in the FC region and got trapped in a minimum of the π,π^* state. Due to the fact that the trajectories were stopped at 500 fs, no GS regeneration was observed, and the study could not provide a full picture of the radiationless mechanisms nor decay timescales that lead to GS regeneration.

Very recently, a semiempirical OM2//MRCI dynamics with a trajectory surface hopping algorithm provided a different explanation for the decay lifetimes of thymine. In this work,[139] two relaxation mechanisms were observed in the sub-picosecond range. The fastest mechanism corresponds to direct decay from the FC structure to the GS via the ethylenic CI, whereas the second mechanism corresponds to decay from the π,π^* state to the n,π^* state and further to the ground state via a CI. The two paths lie in the subpicosecond range and explain the decay components of <50 fs (single step mechanism) and 490 fs (two step mechanism) reported by Ullrich et al.[87]

To summarize, all computational studies in the literature agree that thymine ultrafast relaxation proceeds through an ethylene-like[133-136] conical intersection between the first excited state, S_1, and the ground state, S_0. However, there is no consensus (see Figure 16) in the paths that may lead excited thymine to GS regeneration. The reason for such a disagreement is that PESs are highly dependent on the methodology and the level of calculation. All these results show that there is need of a high level PES which provides a better description of the global picture and helps to understand the deactivation mechanism.
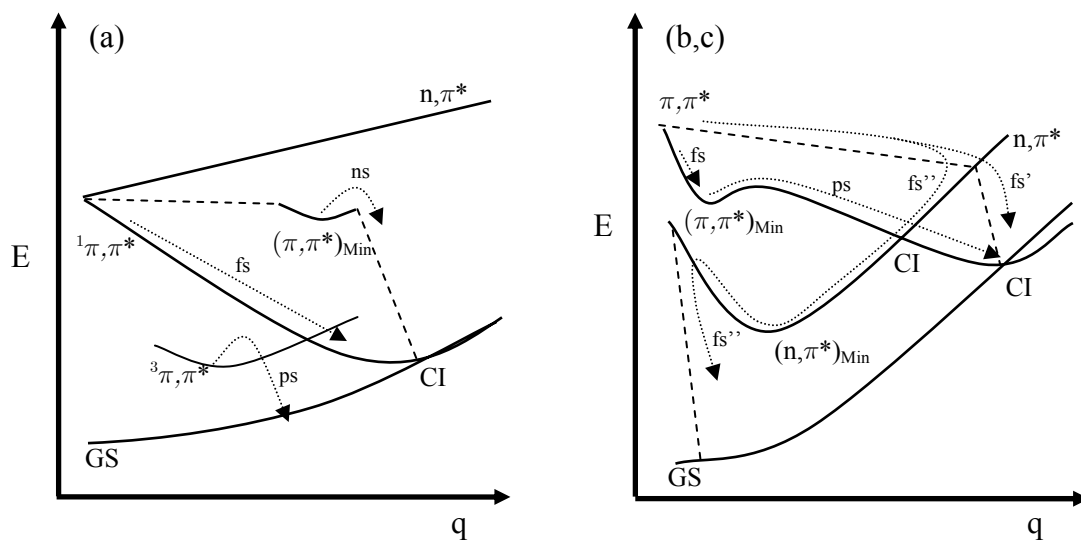
Figure 16. Summary of thymine's relaxation models found in the literature. a) 3 components corresponding to deactivations from the π,π^* state, $(\pi,\pi^*)_{Min}$, and $^3\pi,\pi^*$ state were reported[131,137] to lie in the fs, ps and ns time ranges, respectively. b) 2 components corresponding to relaxation from FC to $(\pi,\pi^*)_{Min}$, and further deactivation from that minimum were assigned[138] to the fs and ps components, respectively. c) 2 fs components, fs' (<50 fs) and fs'' (490 fs), were reported[139] for two different two-step mechanisms corresponding to π,π^*-GS, and π,π^*-n,π^*-GS, respectively.

1.3   Computational methodology

When facing the computational study of a system, one must first consider the cost and accuracy of the calculations that have to be performed. Accurate calculations can be computationally very expensive or even unfeasible. The use of simplified models can help in such a situation, although sometimes these models cannot describe all the properties of the system and therefore, cannot be used. Thus, the first stage of a computational study should be the analysis of the system to be studied and the choice of the methodology and level(s) of theory that will be used throughout. There are mainly two types of methodologies: the specific and general ones. Specific methods can only be applied to a limited number of systems for which they usually provide excellent results. On the other hand, general methodologies can be applied to a vast number of cases, but their results are not so accurate as in the former. It is a general procedure to perform an initial calibration before using a specific method in order to check if the methodology is valid for the system that has to be studied. However, when general methods are used, we tend to assume they will be applicable to the case study and forget that they may fail unexpectedly. Some failures of such general-use methods in describing GS minima of ordinary molecules are presented next.

1.3.1   Failures of general computational methods in ring planarity description

An accurate structure optimization is essential for the study of the photophysics of a molecule. For instance the GS minimum is the structure that determines the energy of the excited states from which the relaxation process starts. Thus, it is important to make sure that the methods that will be used to perform the study of a system, can optimize structures properly. This is exemplified for benzene, which is one of the most studied molecules because of its singularity and "simplicity". It is a planar, aromatic, six-memebered ring with no substituents. Its high symmetry, which translates into relatively low cost computations, makes benzene the perfect target of many studies. Moreover, it is the perfect model for a lot of organic molecules composed of rings, and in many cases it is used as a reference. In spite of the knowledge accumulated over the years on benzene and the different techniques applied for its study, recently, Moran *et al.*[140] reported a broad number of *ab initio* calculations on benzene and other planar arenes at different correlated levels of theory. Surprisingly, some of

the frequently used Pople's basis sets yielded non-planar minima rather than planar geometries. The planar optimized structures turned out to be transition states exhibiting one (or more) large imaginary frequency. More intriguing is the fact that usual high level (correlated) calculations, namely Møller-Plesset[141] (MP2) and Coupled Cluster[56,142,143] with single, double and triple excitations (CCSD(T)), predicted wrong structures for benzene and other planar arenes, whereas single-determinant based methods like Hartree-Fock[144] (HF) and Density Functional Theory methods (DFT) such as BLYP[145,146] and B3LYP[146,147], which are computationally cheaper, led to the expected planar minima and hence no imaginary frequencies.

Other authors had already reported failures of electronic structure methods for the correct description of low-lying out-of-plane vibrational frequencies of benzene,[148,149] several planar arenes,[150-154] and other non-rigid molecules.[155] Similar pitfalls were found more recently by Shabahzian[156] for the challenging $(B_6C)^{2-}$ anion, particularly at the MP2 level of theory.

Rather than to any deficiency of the post-HF electronic structure methods including electron correlation, the origin of the problem has been suggested to be rooted on (atom-centered) basis set deficiencies. Simandiras *et al.*[155] found that the use of f-type basis functions was necessary to obtain accurate bending frequencies. Martin *et al.*[149] dealt in detail with the benzene case and concluded that out-of-plane bending modes are pathologically basis set dependent, suggesting that a basis set superposition could be at the origin of the problem, based on the work of Sellers and Almlöf.[157] In fact, Jensen[158] found that imaginary frequencies can also appear at the single-determinant level for a double-zeta quality basis set including d-type diffuse functions (aug-pc-1) for a rather narrow range of diffuse function's exponent values.

Perhaps, the most important error introduced by the use of truncated atom-centered basis sets is the so-called basis set superposition error (BSSE).[159] It has been recognized for years[160] that BSSE introduces some spurious extra binding in the *ab initio* calculations and that its correction is essential to properly describe intermolecular interactions. In the last decade it has been also shown that BSSE does not affect merely the interaction energy, but also the topology of the PES of these systems, which translates into geometrical and vibrational effects (see section 2.1.3.2).

1.3.1.1   Intramolecular BSSE

Much less attention has been paid to BSSE effects in single molecules, what is referred to as intramolecular BSSE.[157,161-164] Noteworthy exceptions are

recent studies on aromatic-backbone intramolecular interactions[165-168] on peptide models, that have put forward that the accurate determination of relative stabilities of conformers is heavily affected by BSSE. In fact, the first CP calculation by Jansen and Ros[159] was carried out on a single cation (HCO^+). In single molecules there is not a problem a priori with the fact that one atom or fragment can use the basis functions centered on other atoms or parts of the molecule. Actually, this helps to naturally describe polarization or charge transfer effects. The problem arises when the use of external basis functions is merely due to a lack of flexibility of the fragment's basis set; *i.e.* when the basis set incompleteness error (BSIE) is strongly geometry dependent, as spurious geometry changes can be induced by intramolecular BSSE effects. Unbalanced descriptions of PES can also emerge. This translates directly into poor vibrational frequencies and likely changes in the topology of the stationary points, as found by Moran et. al.[140] In that revealing report, the authors showed that in those cases where a non-planar optimized structure was found for benzene, the two-electron BSIE tends to dramatically increase for geometries away from planarity. In fact, even for the levels of theory where no imaginary frequencies are observed, BSIE still shows moderate dependence on the geometry. Thus, the two-electron BSIE diagnostic seems indeed a valuable tool to detect possible spurious geometries induced by BSSE effects but it cannot fix the problem. A recent study[169] based on the use of chemical hardness profiles is also able to detect such spurious structures, but again, it does not offer a solution to the problem.

There are a few examples in the literature where BSSE accounts for large geometric effects, namely the hydrogen fluoride dimer[170,171] or weak C-H···O[172] interactions. We propose BSSE correction (see section 2.1.3.2 for more details) as a way to fix the anomalous behavior of normal-use electronic methods such as those reported by Moran *et al.* Such an issue takes special relevance in the present work as we want to get a high quality PES for thymine which, as will be seen later on, can also suffer from such pitfalls. Therefore, we will investigate the performance of the Counterpoise method in fixing the failures reported for benzene, to later extend it to other systems. In addition, more complicated cases for which the fragments' definition is not clear, such as the cyclopentadienyl and indenyl anions or naphthalene, will be considered.

1.3.2   Pitfalls on DNA and RNA nucleobases

Failures on geometry optimizations is an issue that must be always kept in mind, but especially in photophysics studies as the GS minimum corresponds to the FC point, which is the starting point of all relaxation pathways and the geometry at which the vertical spectrum is calculated. In order to get an

accurate FC point, many studies[130,138,173] optimize the ground state minimum
using methods which include dynamic correlation. Therefore, the use of MP2 is
rather general for this purpose[125,170,174-176] due to its simplicity and relatively low
computational cost. It has been seen that great attention must be paid when
using correlated methods with Pople basis sets, as one may come across
unexpected failures in minima optimizations. Failures in geometry optimization
can affect the vertical spectrum of molecules. This can translate into a bad
description of the ordering of the states, which is essential to describe the
deactivation paths. This is illustrated in Figure 17, where a small geometrical
change in the position of the minimum translates into a rather different
deactivation scenario.

The full line depicts a FC point from which a wavepacket can hardly
deactivate. Thus, the deactivation of this species would be much slower than
that described by the dashed blue line, in which the wavepacket that describes
the excited species has enough energy to surmount the TS, and therefore, to
regenerate the GS with less difficulties than in the former case. The difference
between the two descriptions is a small variation of the position of the
minimum of the PES. Thus, this shows that the optimization of the minimum is
very important in photophysics.
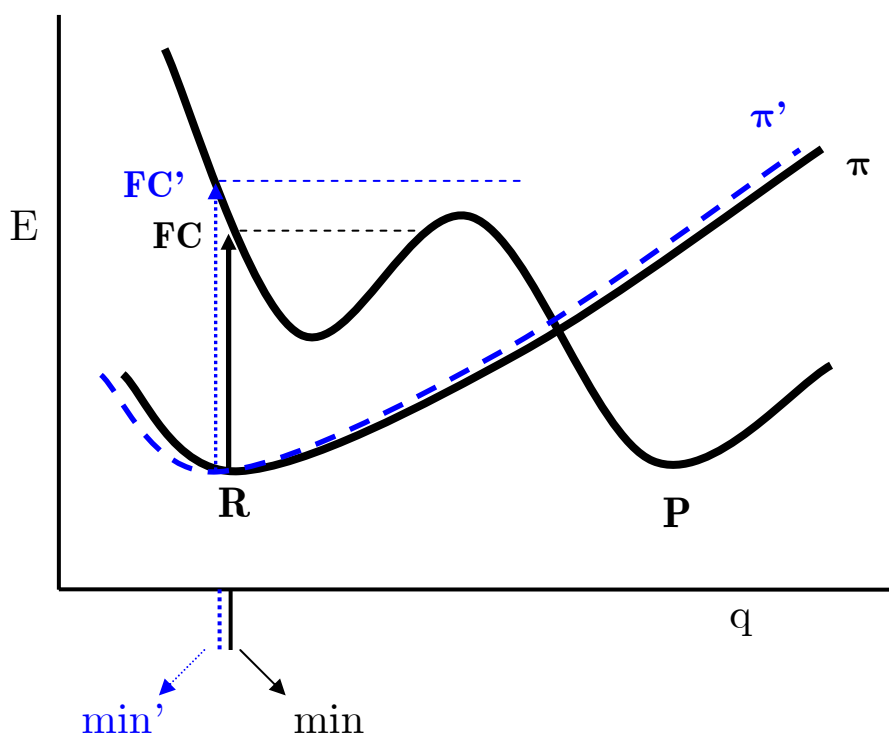
Figure 17. Effects of geometrical optimization anomalies in the description of
deactivation paths.

According to the aim of this thesis, a MP2 and CASSCF benchmark study has been carried out for thymine in order to choose the best methodology to optimize the FC point and perform the photophysics study. It has also been studied if thymine can suffer similar pitfalls to those reported for benzene. The study is extended to the other nucleobases at the MP2 level for four selected basis sets of general-use, namely 6-31G*, 6-31+G*, 6-311G* and 6-311+G*, and the strategy followed in the case of benzene is applied here too.

Another point of interest is to determine to which extent such basis set deficiencies are localized in a region of the molecule and whether it would be sufficient to correct for intramolecular BSSE only locally, that is, using a specific Counterpoise function that would take into account only a subset of atoms of the system. This would likely be the case for intermolecular hydrogen bond formation or for the interaction between the two ends of a long chain-like molecule. In such cases a local treatment may be of use not just because BSSE correction would be irrelevant in most parts of the molecule, but also to avoid spurious effects of the CP-correction itself. In this respect, the definition of fragments with a strong overlap (*i.e.* when breaking a chemical bond) can lead to different electronic states for the fragment and ghost-orbital calculations, causing spurious CP corrections. This possibility is studied, and a solution based on the multiplicity assignment is applied.

2   THEORETICAL METHODS

Various methodologies and levels of theory of electronic structure methods have been used in this thesis. Given that these methodologies have only been used at user level and that we are more interested in the chemistry of the results rather than in the theory behind them, the methodologies will not be explained in great detail but will be plainly overviewed. A further insight will be given into those parts which are more relevant for the understanding of the results or have implications for the choice of methodology.

The present chapter is divided in four parts. General considerations about electronic structure methods are explained in the first part. The theoretical background concerning the nature of basis sets and one of the simplest methods to solve the Schrödinger equation are explained in the second part. The general theory of the *ab initio* methods that have been used in this thesis will be explained in the third part. Finally, the molecular dynamics methods used in this thesis are overviewed in the fourth section. Some problems that arise from the use of some of these methodologies as well as a simple procedure to correct these anomalies are also overviewed on the second section.

The *ab initio* methods will be explained following the development of the subjects in the books by Szabo,[177] Helgaker,[178] Roos and Widmark,[179] Hehre *et al.*,[180] Klessinger and Michl,[59] Kutateladze,[49] Yarkony,[181] and Domcke *et al.*[64]

2.1   Ab initio methods

Ab initio methods can, in principle, obtain the exact structure and energy of molecules without carrying out any experiment or taking into account any experimental parameter. This fact is indicated by the term "*ab initio*", which literally means start from scratch. This term is used to identify the methods that are based on the solution of the non-relativistic time-independent Schrödinger[182] equation to predict the properties and energy of stationary states of molecules. The Schrödinger equation can only be explicitly solved for the H atom, but fortunately some approximations to the explicit solution of the Schrödinger equation have been developed. Some of these approximations are explained next.

2.1.1 Schrödinger equation

The non-relativistic time-independent Schrödinger equation reads as:

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

where $\hat{H}$ is the Hamiltonian operator that provides the total energy (E) of the system represented by the wavefunction Ψ, which depends on the coordinates of all particles and also on their spin coordinates.

In atomic units, the Hamiltonian operator for a system of N electrons and M nuclei is:

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{\alpha=1}^{M}\frac{\nabla_\alpha^2}{M_\alpha} - \sum_{\alpha=1}^{M}\sum_{i=1}^{N}\frac{Z_\alpha}{\left|\boldsymbol{r}_{i\alpha}\right|} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{\left|\boldsymbol{r}_{ij}\right|} + \sum_{\alpha>\beta}^{M}\frac{Z_\alpha Z_\beta}{\left|\boldsymbol{R}_{\alpha\beta}\right|} \tag{2.2}$$

where the first two terms correspond to the electronic and nuclear kinetic energy operators, respectively. The third term corresponds to the electron-nuclei potential energy, the fourth one to the electron-electron repulsion energy, and the last one to the nuclei-nuclei potential energy. $\boldsymbol{r}$ variables are used for the electronic coordinates and $\boldsymbol{R}$ for the nuclear ones. N and M correspond to the number of electrons and nuclei, respectively. The following notations are also used: $\left|\boldsymbol{r}_{i\alpha}\right| = \left|\boldsymbol{r}_i - \boldsymbol{R}_\alpha\right|$, $\left|\boldsymbol{r}_{ij}\right| = \left|\boldsymbol{r}_i - \boldsymbol{r}_j\right|$ and $\left|\boldsymbol{R}_{\alpha\beta}\right| = \left|\boldsymbol{R}_\alpha - \boldsymbol{R}_\beta\right|$.

The Hamiltonian can be expressed in a shorter notation as:

$$\hat{H}\left(\boldsymbol{r},\boldsymbol{R}\right) = \hat{T}_e\left(\boldsymbol{r}\right) + \hat{T}_N\left(\boldsymbol{R}\right) + \hat{V}_{N\text{-}e}\left(\boldsymbol{r},\boldsymbol{R}\right) + \hat{V}_{e\text{-}e}\left(\boldsymbol{r}\right) + \hat{V}_{N\text{-}N}\left(\boldsymbol{R}\right) \tag{2.3}$$

where the operators are written in the same order as in eq. (2.2) as a function of the electronic ($\boldsymbol{r}$) and/or nuclear coordinates ($\boldsymbol{R}$).

As stated before, such equations can only be explicitly solved for the H atom, since only two interacting particles are present. For larger systems, the solution to the Schrödinger equation can only be approximated. One of the most widely used approximations is the so-called Born-Oppenheimer approximation.

2.1.2 The Born-Oppenheimer approximation

The Born-Oppenheimer approximation[183] is central to quantum chemistry. It relies on the fact that nuclei are more than one thousand times heavier than electrons. It is assumed that nuclei move much slower than electrons, and that when a nuclei moves the electrons immediately reorder themselves around it, *i.e.* the electrons move in a field of fixed nuclei. Thus, the second term of eq. (2.2) is zero, and the last term, the repulsion between the nuclei, is constant. This approximation uncouples the electronic and nuclear motions and allows for a separation of the Hamiltonian into two parts, the electronic and nuclear ones. The electronic Hamiltonian reads as:

$$\hat{H}_e = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{\alpha=1}^{M}\sum_{i=1}^{N}\frac{Z_\alpha}{\left|\boldsymbol{r}_{i\alpha}\right|} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{\left|\boldsymbol{r}_{ij}\right|} \tag{2.4}$$

The eigenfunctions of the electronic Hamiltonian correspond to the electronic states of the molecule, and the eigenvalues to the electronic energy of the system. This energy depends parametrically on the position of the nuclei, thus the electronic Schrödinger equation is:

$$\hat{H}_e(\boldsymbol{r},\boldsymbol{R})\Psi_e(\boldsymbol{r};\boldsymbol{R}) = E_e(\boldsymbol{R})\Psi_e(\boldsymbol{r};\boldsymbol{R}) \tag{2.5}$$

The fact that the electronic energy depends parametrically on $\boldsymbol{R}$, implies that solving the electronic Schrödinger equation for all the nuclear coordinates one will obtain a set of energies. If the repulsion energy between the nuclei is added to this set of electronic energies ($\boldsymbol{E}_e$), the adiabatic potential energy surface of the system ($\boldsymbol{E}_{tot}$) is obtained:

$$\boldsymbol{E}_{tot} = \boldsymbol{E}_e(\boldsymbol{R}) + \sum_{\alpha>\beta}^{M}\frac{Z_\alpha Z_\beta}{\left|\boldsymbol{R}_{\alpha\beta}\right|} \tag{2.6}$$

This potential energy surface are the eigenvalues of the function Ψ_e, which is an eigenfunction of the electronic Hamiltonian. However, the Hamiltonian has infinite eigenfunctions, each one corresponding to an electronic state. The eigenfunction which provides the lowest energy is called the ground state, and the rest are commonly known as excited states. The determination of the lowest potential energy surfaces of thymine is one of the centrals part of this thesis (see Results). Thus, the electronic wavefunction is actually a linear combination of electronic functions, one for each electronic state:

$$\Psi = \sum_k c_k \Psi_k \tag{2.7}$$

Each eigenfunction, or electronic state, depends parametrically on $\boldsymbol{R}$, thus each state has its corresponding adiabatic potential energy surface.

Once the electronic part is solved, one can proceed to solve the total Schrödinger equation by taking the same assumptions as before. If the electrons move much faster than nuclei, it is reasonable to change the electronic coordinates of eq. (2.2) by their average values, *i.e.* we are assuming that the nuclei move in a field of electrons. This averaged field of electrons is actually the adiabatic potential energy obtained solving the electronic Schrödinger equation, and thus the nuclear Hamiltonian is expressed as:

$$\hat{H}_N = -\frac{1}{2} \sum_{\alpha=1}^{M} \frac{\nabla_\alpha^2}{M_\alpha} + E_e(\boldsymbol{R}) \tag{2.8}$$

Accordingly, the nuclear Schrödinger equation reads as:

$$\left(\hat{T}_N(\boldsymbol{R}) + E_e(\boldsymbol{R})\right)\Psi_N = E\Psi_N \tag{2.9}$$

where E includes the electronic, translational, vibrational and rotational energies of the system and is the Born-Oppenheimer approximation to the total energy (E of eq. (2.1)). It is worth to mention that the Born-Oppenheimer approximation is only applicable if the electronic states are well separated, that is when infinitessimal nuclear displacements do not change the nature of the electronic state.

2.1.2.1   Born-Oppenheimer approximation breakdown

It has been seen that the Hamiltonian can be splitted into nuclear and electronic parts, and that their corresponding Schrödinger equations yield a set of nuclear and electronic eigenfunctions. Thus, the total wavefunction can be expressed as:

$$\Psi(\boldsymbol{r}, \boldsymbol{R}) = \sum_k^K \Psi_e^k(\boldsymbol{r}; \boldsymbol{R})\Psi_N^k(\boldsymbol{R}) \tag{2.10}$$

where k is the number of states. If we want to solve the total electronic Schrödinger equation and the electronic part is already solved, the solution turns into a set of coupled eigenvalue equations depending on the nuclear coordinates:

$$\left(\boldsymbol{T}_N(\boldsymbol{R}) + \boldsymbol{H}_e(\boldsymbol{R}) \right) \Psi_N = E \Psi_N \tag{2.11}$$

where E is the Born-Oppenheimer approximation to the total energy, and $\boldsymbol{H}_e(\boldsymbol{R})$ and $\boldsymbol{T}_N(\boldsymbol{R})$ correspond to the electronic Hamiltonian and nuclear kinetic energy matrices, respectively. By definition, the electronic Hamiltonian matrix is diagonal, since $\left\langle \Psi_e^{k'} \middle| \Psi_e^k \right\rangle = \delta_{k'k}$ with k indicating the electronic state. On the other hand, the elements of the nuclear kinetic energy matrix correspond to the expected value of the nuclear kinetic energy operator:

$$(\boldsymbol{T}_N(\boldsymbol{R}))_{k'k} = \delta_{k'k}\hat{T}_N + \sum_A \frac{1}{M_A} \left\langle \Psi_e^{k'} \middle| \hat{P}_A \Psi_e^k \right\rangle \hat{P}_A + \left\langle \Psi_e^{k'} \middle| \hat{T}_N \Psi_e^k \right\rangle \tag{2.12}$$

where $\hat{P}_A$ is the nuclear momentum operator, which is Hermitian and purely imaginary. Its diagonal elements are zero and the off-diagonal ones satisfy:

$$\left\langle \Psi_e^{k'} \middle| \hat{P}_A \Psi_e^k \right\rangle = \frac{\left\langle \Psi_e^{k'} \middle| \left[\hat{P}_A, H_e \right] \middle| \Psi_e^k \right\rangle}{E_k - E_{k'}} \tag{2.13}$$

If the Born-Oppenheimer approximation is preserved, that is when the electronic states k and k' are well separated, the second and third terms of eq. (2.12) are neglected. However, if the Born-Oppenheimer approximation breaks down, that is the electronic states are close, the off-diagonal elements cannot be neglected. This represents that the electronic states are coupled to each other, which difficults the solution of the Schrödinger equation. In order to avoid the solution of the coupled equations, a diabatic transformation[184,185] can be performed. It consists on a rotation of the adiabatic states to generate diabatic ones, whose off-diagonal terms are zero. Thus, the transformation consists on a diagonalization of the nuclear kinetic energy matrix, which "reduces" the problem to the finding of the so-called mixing angle, provided only 2 states are coupled. In the diabatic representation, the coupling of the states is due to the electronic energy and is a scalar quantity which is much more easy to estimate numerically. One of the problems of this approximation is that in general the diabatic angle does not exist, therefore the exact transformation from adiabatic

to diabatic states can only be performed approximately (see section 2.2 for further details).

The solution of the nuclear Schrödinger equation is essential to describe processes such as tunneling, scattering, and non-adiabatic transitions. In contrast, the solution of the electronic part is sufficient to describe most of the thermal processes. Thus, the nuclear part is usually neglected unless a more accurate description of the system is needed. From this point on the subscripts that have been used to differentiate between the nuclear and electronic parts will be omitted and only the latter will be considered.

2.1.3   Molecular orbital theory

The Born-Oppenheimer approximation allows the separation of the electronic and nuclear parts of the Hamiltonian, which allows for an approximated solution of the Schrödinger equation. One of the possible solving procedures is based on the use of the solution of the electronic part to resolve the nuclear one, which provides the total energy of the system. The electronic Hamiltonian only depends on the electronic coordinates. Thus, its eigenfunctions (wavefunction) consists on a function of the spatial coordinates of electrons, $\psi(x,y,z)$. This functions are called orbitals, and its square moduli corresponds to the probability of finding the electron around a small volume around the central position of the atom, *i.e.* the atomic orbitals. In order to completely describe the electrons of a system, an infinite set of spatial functions is needed. Since this is obviously unaffordable, the infinite set of spatial functions is approximated by a reduced set of functions that is usually expressed as a linear combination of atomic orbitals (see next section for further details). The spatial function is not sufficient to describe electrons, but an extra function is needed, the so-called spin function. The total spin function is expanded by two functions, $\alpha(\xi)$ and $\beta(\xi)$, which depend on a spin coordinate, ξ, that is unknown. The expected values of the spin angular momentum operator along the axis z of these functions are $+1/2$ and $-1/2$, *i.e.* spin up and spin down, for α and β respectively. These two functions form a complete set and are orthonormal to each other:

$$\left\langle \alpha \middle| \alpha \right\rangle = \left\langle \beta \middle| \beta \right\rangle = 1 \quad , \quad \left\langle \alpha \middle| \beta \right\rangle = \left\langle \beta \middle| \alpha \right\rangle = 0 \tag{2.14}$$

Thus, the wavefunction of an electron consists on a product of two functions, the spatial and spin functions, which results in a spinorbital function, $\chi(x,y,z,\xi)$. It is worth to mention that for a set of K spatial functions or

orbitals, a 2K set of spinorbitals can be obtained since two spin functions, $\alpha(\xi)$ and $\beta(\xi)$, can be combined with each spatial function to form a spinorbital.

The spatial and spin coordinates are usually expressed collectively as $\boldsymbol{x}$. Accordingly, a wavefunction for a N-electron system is written as $\Psi\left(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N\right)$.

If one takes an approximated electronic Hamiltonian that does not take into account the interaction between electrons or that includes it in an averaged way, *i.e.* neglects the last term of eq. (2.4) or makes it constant, the electronic Hamiltonian can be expressed as a sum of monoelectronic operators:

$$\hat{H} = \sum_{i=1}^{N} \hat{h}(i) \tag{2.15}$$

The eigenfunctions of this monoelectronic operators are the spinorbitals of the system, and the corresponding total wavefunction is, therefore, a product of these spinorbitals:

$$\Psi\left(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N\right) = \chi_i\left(\boldsymbol{x}_1\right)\chi_j\left(\boldsymbol{x}_2\right)...\chi_k\left(\boldsymbol{x}_N\right) \tag{2.16}$$

This approximated wavefunction has the form of a Hartree product.[144] However, this is not a proper wavefunction since it does not fulfill the Pauli exclusion principle,[186] which states that a proper wavefunction of a polyelectronic system must be antisymmetric with respect to the interchange of coordinates ($\boldsymbol{x}$) of any two electrons. A way to fulfill the antisymmetry principle is expressing the wavefunction of the N-electron system as a Slater determinant[187,188] of spinorbitals:

$$\Psi(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_i(\boldsymbol{x}_1) & \chi_j(\boldsymbol{x}_1) & \cdots & \chi_n(\boldsymbol{x}_1) \\ \chi_i(\boldsymbol{x}_2) & \chi_j(\boldsymbol{x}_2) & \cdots & \chi_n(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(\boldsymbol{x}_n) & \chi_j(\boldsymbol{x}_n) & \cdots & \chi_n(\boldsymbol{x}_n) \end{vmatrix} \tag{2.17}$$

where χ_j denotes the different spinorbitals and ($\boldsymbol{x}_i$) the electronic coordinates of the system. N is the total number of electrons, which is used to normalize the wavefunction.

2.1.3.1   Basis Sets

The construction of a wavefunction as a Slater determinant of spinorbitals has been explained so far. We have seen that each spinorbital is formed by a spatial function, atomic orbital, and a spin function. When two atomic orbitals approach each other, they combine forming a molecular orbital. Molecular orbitals are expressed as linear combinations of atomic orbitals (LCAO). The combination of n atomic orbitals results in n molecular orbitals, which are usually orthonormalized.

As mentioned before, only an infinite expansion of the spatial part of the atomic orbitals can provide an exact wavefunction. Since this is not possible, atomic orbitals, and therefore the wavefunction, can only be approximated. In practice, the atomic orbitals are usually expressed as a linear combination of basis functions

$$\psi_i = \sum_{\mu=1}^{N} c_{\mu i}\phi_{\mu} \tag{2.18}$$

where ϕ_{μ} are the basis functions, $c_{\mu i}$ the expansion coefficients, and N the number of basis functions used to approximate the atomic orbital ψ_i. In principle any function can be used to expand the atomic orbitals, however for a better approximation is convenient to use basis functions which have the symmetry properties of the exact atomic orbitals. Two of the most used types of basis functions that follow these premises are the Slater Type Orbitals and Gaussian-type atomic functions. The former have the form:

$$S(\boldsymbol{r},\upsilon,\varphi) = N\boldsymbol{r}^{n-1}e^{-\zeta r}Y_l^m\left(\upsilon,\varphi\right) \tag{2.19}$$

where N is a normalization factor, n is the principal quantum number of the orbital, ζ is a constant related to the effective charge of the nucleous, and $Y_l^m\left(\upsilon,\varphi\right)$ defines the angular momentum part with the corresponding orbital (l) and magnetic (m) quantum numbers. The general form of the Gaussian-type atomic functions in Cartesian coordinates (x,y,z) is:

$$G_{ijk}^{\alpha,\boldsymbol{R}}(\boldsymbol{r}) = N_{ijk}^{\alpha}\left(x-R_1\right)^i\left(y-R_1\right)^j\left(z-R_1\right)^k e^{-\alpha\left(\boldsymbol{r}-\boldsymbol{R}\right)^2} \tag{2.20}$$

where N_{ijk}^{α} is a normalization factor, and $\boldsymbol{R}$ and α are called the center and exponent of the Gaussian function, respectively. $L = i + j + k$, and defines the type of Gaussian function, s-type (L=0), p-type (L=1), d-type (L=2), ...,

and so on. The main difference between the Slater and Gaussian types of functions is that in the latter, the exponential of r is squared. This leads to a Gaussian function which present a maximum at the center of the atom, rather than a cusp with finite gradient as in the Slater-type case. In principle, this makes the Gaussian function less suitable for molecular representation. In contrast, they allow for the calculation of integrals analytically, which makes the calculations much faster. Fortunately, an accurate representation of the atom and analytical integrals can be obtained at the same time by using a basis set of *contracted Gaussian functions*. It consists on the construction of each basis function as linear combination (contraction) of Gaussian functions (primitives). As a result, one can obtain functions which are much similar to the Slater type ones, but with the advantages of the Gaussian type functions. However, this represents an extra step as one needs to choose the exponents of the primitives and the contraction coefficients in order to obtain the desired basis function(s). The accuracy and performance of a calculation depend on the quality of the basis set. Therefore, the basis set choice is a key point of the study of a system, although some authors[189] claim it is more art than science.

The main contributions to the definition and optimization of basis functions were provided by the groups of Huzinaga,[190-193] Pople[194-200] and Dunning.[201-204]

2.1.3.1.a   *Minimal Basis Set*

The so-called minimal basis set is the basis set that has the least number of basis functions per atom required to describe the occupied orbitals of that atom. In many cases, the energy and other properties cannot be properly determined with such a small basis set, therefore in practice they are not used.

One of the smallest basis sets that can be used is the STO-LG family.[199] In this family each basis function is obtained as a linear combination of L primitive Gaussian functions. The contraction coefficients and exponents are chosen so that the final basis function approximates a Slater function. The STO-LG family is expanded up to 6 primitives although it has been observed that the use of 3 primitive functions is enough to reproduce a Slater function. Thus, the STO-3G has become the practical minimal basis set.

2.1.3.1.b   *Double Zeta Basis Sets*

Minimal basis sets have a low variational flexibility among other limitations. A way to improve the minimal basis set is to double the number of core and valence functions that are used in the minimal basis set, *i.e.* to use a

double-ζ (zeta) basis set. The objective is to have a function whose best orbital exponents are slightly higher than the optimal exponents of the minimal basis set, and another function with somewhat lower exponents. This allows the system to perform an expansion or contraction of the basis functions by just varying some linear parameters. Thus, in the wavefunction optimization procedure, the method will optimize the energy of the system by varying these coefficients. This variational flexibility translates into a better basis set and better description of the system. Hence, if we have a basis set which has twice as many basis functions as a minimal basis set, this basis set is called double-ζ (zeta), DZ. In principle a DZ basis set for the H atom has two functions, and 10 for a C atom. However, there are some basis sets such as the 3-21G, 4-31G, and 6-31G, which are commonly known as double-ζ but which, strictly speaking, are not. Actually they are double-ζ in the valence part, not in the core. These basis sets are called "split-valence" basis sets. The notation used for this basis sets means that, for instance in the 3-21G case, the core orbitals are described by a contraction of 3 primitive Gaussian functions (**3**-21G), while the valence part is described by 2 functions (3-**21**G), one made of 2 primitives (3-**2**1G) and the other one made of a single function (3-2**1**G).

2.1.3.1.c   Polarization functions

The next step in improving a basis set could be to go to triple or quadruple-ζ. This can lead to unbalanced basis sets such as the 6-311G, which might trigger spurious descriptions of electronic structure parameters such as the vibrational frequencies (see Results). An alternative to this, is to add polarization functions to the basis set, *i.e.* adding functions of immediately higher angular momentum. If we take as reference a DZ basis set, this implies adding *p*-type functions to H and He, *d*-type functions to the first row atoms, and so forth.

The reason why these functions are called polarization functions is simple and very illustrative. Consider the H atom in its GS, which has only the 1s orbital occupied and is of spherical symmetry. If this H atom is affected by an electric field, such as that induced by a neighboring atom, the electronic cloud is deformed, *i.e.* the density cloud is no longer symmetric but polarized. In terms of basis sets, such a distribution can not be described by spherical (1s) molecular orbitals, one needs *p*-type functions to do so. Hence, the final solution can be considered a hybridized orbital. A similar effect is observed for the atoms of the first row, but *d*-type functions are added instead of *p* ones.

The most famous split-valence double-ζ plus polarization basis set is Pople's 6-31G*. Following the notation introduced above, the core orbital is

described by a single function formed of 6 primitives while the valence part is described by two functions constructed from 3 and 1 primitives respectively. The "*" denotes the addition of polarization functions to non-H atoms. The 6-31G** basis set, is the same as 6-31G* but including polarization functions on the H atoms, too.

Another type of valence-polarized double-ζ basis sets are the so called "correlated consistent" basis sets, cc-pVXZ. These basis sets were introduced by Dunning, and are optimized for correlated wavefunctions (see Multiconfigurational methods section). X determines the number of basis functions that are used in each basis set with respect to a minimal basis, X=D, T, Q, 5, 6, … The functions of the different basis sets are added in shells. Thus, cc-pVDZ for the C atom would include 3s2p1d, cc-pVTZ 4s3p2d1f, cc-pVQZ 5s4p3d2f1g, etc. This systematic basis set enlargement can be used to extrapolate the basis set limit.

2.1.3.1.d   Diffuse functions

The addition of polarization functions might not be sufficient to describe systems like anions or certain non-bonding interactions. Anions tend to have their orbitals expanded compared to neutral atoms, therefore the usual basis sets may not describe them properly. In order to describe these special cases, one must add diffuse functions, which in Pople's basis sets are functions with small exponents (large radial extension). These extra functions are denoted by a "+" sign, *i.e.* 6-31+G*. This is the same basis set as 6-31G* but with an extra set of s- and p-type functions for the atoms of the first row. As in the polarization case, a second "+", 6-31++G*, is used to express that diffuse functions are added to H atoms too.

The inclusion of diffuse functions in Dunning's basis sets is indicated by the prefix aug. Thus, Dunnig's double-ζ plus valence-polarization and diffuse functions basis set is indicated by aug-cc-pVDZ.

2.1.3.1.e   ANO-type basis sets

The Atomic Natural Orbital[205] type of basis sets are especially indicated for highly correlated calculations. This family of basis sets provides some improvements over the ones mentioned above by addressing some of their shortcomings. Some of the problems of Gaussian basis sets arise from the fact of using segmented functions. This means that a primitive Gaussian can contribute to only one contracted function. This can cause large BSSE effects, incorrect nodal structure, and need of polarization functions to describe complicated

systems as anions. The ANO basis sets are made of the natural orbitals with highest occupations from SCF and Configuration Interaction atomic calculations with uncontracted basis. The natural orbitals of these calculations are truncated by their occupation numbers to form a contracted set, and the expansion coefficients of these natural orbitals, in terms of the primitive basis functions, form the coefficient matrix for the general contraction for each atom.

One of the advantages of these basis sets is that they lower the BSSE and have the correct nodal structure. In addition, the basis sets can be easily expanded by lowering the occupation number cutoff that defines the contraction, with the advantage that the new and larger basis set contains the smaller one as a subset.

The construction of such basis sets can be problematic since the description of the atom is crucial. The main problem is that it is not clear whether the optimized basis for a particular state, usually the GS, can describe other states, specially if they have different configurations. Another problem is that the ANO basis sets may need larger number of basis functions for those cases which focus on the description of diffuse regions.

2.1.3.2   Basis Set Superposition Error

The easiest way to define the BSSE is considering a system (AB) formed by two interacting fragments, A and B. The stabilization energy of the system can be easily calculated by subtracting the energy of the fragments from the energy of the total system:

$$\Delta E(\text{AB}) = E_{AB}^{\alpha \cup \beta}(\text{AB}) - E_A^{\alpha}(\text{A}) - E_A^{\beta}(\text{B}) \tag{2.21}$$

where the subindices indicate the fragment to which the energy corresponds, the parenthesis the geometry at which the energy is computed, and the superindices the basis functions that are used for each calculation. This procedure is known as the supermolecular approximation, and although there are other ways[206,207] to get the interaction energy directly, this is the most widely used due to its simplicity. Unfortunately, there is an error associated to such an approximation. Taking into account that we cannot get the exact energy of any of the parts (except for the H atom), and that we are forced to use basis sets with finite number of basis functions, *i.e.* incomplete basis sets, only approximate energies are obtained for each of the fragments. In addition to that, in the calculation of the total energy of the system, E_{AB}, the molecular orbitals of the system expand on the subspace formed by the AB basis functions, whereas, in the calculation of the energy of the fragments, E_A and E_B, the molecular orbitals of the fragment

A can only expand on the A basis functions and the same for fragment B. Thus, the calculation of E_{AB} is more accurate than E_A or E_B as, it has been obtained using a different basis set. As a consequence, E_A and E_B are underestimated compared to E_{AB}, and the interaction energy becomes overestimated. It is precisely this mismatch between E_{AB} and E_A and E_B what is known as Basis Set Superposition Error. It was first described by Jansen and Ros,[159] although the name was coined by Liu and McLean.[208]

There are different ways to deal with BSSE. One option is to increase the basis set quality, since in the complete basis set limit the BSSE should vanish. Whereas this could be accomplished at the HF or DFT levels of theory, for post-HF methods the BSSE has been shown to converge to zero very slowly with basis set improvement.[209,210] On the other hand, there are several strategies[211-214] to correct for the BSSE, with the Counterpoise method[211] (CP) being the simplest and most widely used one.

2.1.3.2.a   *Counterpoise method*

The Counterpoise method is an *a posteriori* procedure to correct for BSSE. It consists on the improved calculation of the fragments by using all the basis functions of the complex to obtain the energy of each fragment, so as to calculate fragments and complex at the same level of theory. The derivation of the Counterpoise equations is presented next.

It is convenient to express the stabilization energy as sum of interaction (E_{int}) and relaxation energies (E_{rel}),

$$\Delta E(\text{AB}) = \Delta E_{int}(\text{AB}) + \Delta E_{rel}(\text{A,B}) \tag{2.22}$$

The interaction energy is the difference between the energy of the complex and the energy of each fragment at the geometry it adopts in the complex. That is:

$$\Delta E_{int}(\text{AB}) = E_{AB}^{\alpha \cup \beta}(\text{AB}) - E_{AB}^{\alpha}(\text{A}) - E_{AB}^{\beta}(\text{B}) \tag{2.23}$$

Thus, the interaction energy only depends on the geometric parameters of the system. The relaxation energy corresponds to the difference between the energy of the isolated fragments at their optimum geometries ($E_A^{\alpha}(\text{A})$ and $E_A^{\beta}(\text{B})$) and the energy of the fragments in the geometry of the complex ($E_{AB}^{\alpha}(\text{A})$ and $E_{AB}^{\beta}(\text{B})$), *i.e.* it measures the geometrical relaxation that the fragments undergo in the complex:

$$\Delta E_{rel}(\text{A,B}) = E_{AB}^{\alpha}(\text{A}) - E_A^{\alpha}(\text{A}) + E_{AB}^{\beta}(\text{B}) - E_B^{\beta}(\text{B}) \qquad (2.24)$$

It is worth to mention that the relaxation energy is naturally positive and that it depends on the geometric parameters of both the complex and the isolated fragments. According to the Counterpoise philosophy, only the interaction energy will contribute to BSSE, since in the relaxation energy the energies of each fragment are calculated with the same basis set. The BSSE can be minimized if all the terms of the interaction energy are calculated with the same basis set. Thus, the Counterpoise corrected interaction energy reads as:

$$\Delta E^{CP}(\text{AB}) = \left(E_{AB}^{\alpha \cup \beta}(\text{AB}) - E_{AB}^{\alpha \cup \beta}(\text{A}) - E_{AB}^{\beta \cup \beta}(\text{B}) \right) + \\ + \left(E_{AB}^{\alpha}(\text{A}) - E_A^{\alpha}(\text{A}) + E_{AB}^{\beta}(\text{B}) - E_B^{\beta}(\text{B}) \right) \qquad (2.25)$$

This corrected interaction energy can be expressed with a correction term (δ_{AB}^{BSSE}) by grouping the terms corresponding to individual fragments as:

$$E_{AB}^{\alpha}(\text{A}) + E_{AB}^{\beta}(\text{B}) - E_{AB}^{\alpha \cup \beta}(\text{A}) - E_{AB}^{\alpha \cup \beta}(\text{B}) = \delta_{\text{AB}}^{\text{BSSE}} \qquad (2.26)$$

Then, the Counterpoise corrected interaction energy reads as:

$$\Delta E^{CP}(\text{AB}) = \Delta E_{AB}^{\alpha \cup \beta}(\text{AB}) + \delta_{\text{AB}}^{\text{BSSE}} \qquad (2.27)$$

It is worth to mention that this correction term tends to zero with the basis set improvement, is always a positive correction for variational methods, and is dependent on the geometry of the complex. This last property implies that the correction term is not constant along the PES, thus the shape of a corrected PES does not necessarily coincide with that of an uncorrected one as illustrated in Figure 18.
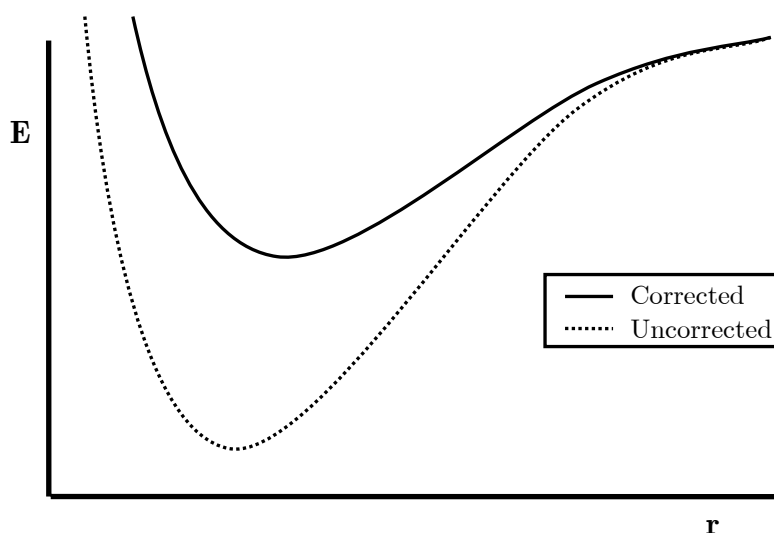
Figure 18. BSSE corrected and uncorrected PES for a given system.

This implies that to determine the interaction energy of a system, a BSSE corrected PES must be used. In this sense, an equivalent description of the BSSE correction is obtained if the CP correction term is directly assigned to the energy of the PES rather than to the variation of energies:

$$E^{CP}(\mathrm{AB}) = E_{AB}^{\alpha \cup \beta}(\mathrm{AB}) + \delta_{\mathrm{AB}}^{\mathrm{BSSE}} \qquad (2.28)$$

It has been argued that BSSE is only present in the interaction energy calculations, but it is also present when not dealing with interaction energies. For instance, the total and interaction energies only differ in one term, which does not depend on BSSE, but only on the geometry and fragment's definition.

By adding the CP correction to the total energy of the system,[215] it is possible to obtain BSSE-corrected energies and any molecular property that may be stated as a derivative of the energy, which includes optimized geometries or vibrational frequencies. Experience gathered on the BSSE correction of intermolecular interactions over the last years shows that the results obtained with BSSE-corrected calculations using moderate basis sets are generally close to those that one can obtain with much larger basis sets with or without correction, provided that a) the basis set is flexible enough to describe the true physical interactions in the system, and b) BSSE effects on the geometry are also taken into account by optimization on the BSSE-corrected PES.

2.1.4   The Hartree-Fock method

Since the exact solution of the Schrödinger equation is not possible for large molecules, finding and describing approximate solutions to the electronic Schrödinger equation has been the major preoccupation for quantum chemists since the birth of quantum mechanics. In this sense, the Hartree-Fock (HF) method,[144,216,217] is essential to quantum chemistry, either as a calculation tool or as a starting point toward more accurate approximations.

2.1.4.1   Hartree-Fock equations

It has been shown in section 2.1.2 that the Hamiltonian operator can be split into electronic and nuclear parts, which allows for an approximate solution of the Schrödinger equation. We are interested in the solution of the Schrödinger equation which better approaches the exact solution. According to the variational principle, this "best" solution is provided by the spinorbitals, which give the lowest energy. Thus, since the energy of the system depends on the wavefunction, and the wavefunction can be approximated as a Slater determinant of spinorbitals, one can obtain the "best" spinorbitals by minimizing the energy of a closed-shell system, which is given by:

$$E_0 = \left\langle \Psi_0 \left| \hat{H} \right| \Psi_0 \right\rangle = 2\sum_a \left\langle \chi_a \left| \hat{h} \right| \chi_a \right\rangle + 2\sum_{a,b} \left\langle \chi_a \chi_b \left| \chi_a \chi_b \right\rangle - \left\langle \chi_a \chi_b \right| \chi_b \chi_a \right\rangle \quad (2.29)$$

where the wavefunction, $\left| \Psi_0 \right\rangle$, is expressed in bracket notation and corresponds to a Slater determinant. $\hat{H}$ is the electronic Hamiltonian, and the subindex "0" indicates it corresponds to the ground state. The expression of the energy has been separated into monoelectronic and bielectronic terms. The elements of the first term are:

$$\left\langle \chi_a \left| \hat{h} \right| \chi_a \right\rangle = \int d\boldsymbol{r}_1 \psi_a(\boldsymbol{r}_1) \left(-\frac{1}{2}\nabla_1^2 - \sum_{\alpha=1}^{M} \frac{Z_\alpha}{\left| \boldsymbol{r}_{1\alpha} \right|} \right) \psi_a(\boldsymbol{r}_1) \quad (2.30)$$

where the different spinorbitals are denoted by the subindex *a*. Since the spin part of the spinorbital is not affected by the operator it has been omitted, and only the spatial part, ψ_a, is considered. The elements of this term correspond to the average kinetic and nuclear attraction energy of electrons. The second term of eq. (2.29) is the so-called Coulomb integral:

$$\left\langle \chi_a \chi_b \middle| \chi_a \chi_b \right\rangle = \int d\boldsymbol{r_1} d\boldsymbol{r_2} \left| \psi_a(\boldsymbol{r_1}) \right|^2 \frac{1}{r_{12}} \left| \psi_b(\boldsymbol{r_1}) \right|^2 \tag{2.31}$$

This integreal corresponds to the repulsion between the charge clouds $\left| \psi_a(\boldsymbol{r_1}) \right|^2$ and $\left| \psi_b(\boldsymbol{r_1}) \right|^2$, which in general is expressed as $J_{ab} = \left\langle ab \middle| ab \right\rangle$. The other bielectronic term is the so-called exchange integral and does not have any physical meaning:

$$\left\langle \chi_a \chi_b \middle| \chi_b \chi_a \right\rangle = \int d\boldsymbol{r_1} d\boldsymbol{r_2} \psi_a^*(\boldsymbol{r_1}) \psi_b(\boldsymbol{r_1}) \frac{1}{r_{12}} \psi_b^*(\boldsymbol{r_2}) \psi_a(\boldsymbol{r_2}) \tag{2.32}$$

This integral is usually expressed as $K_{ab} = \left\langle ab \middle| ba \right\rangle$. Thus, the Hartree-Fock energy can be written in general form as:

$$E_0 = 2\sum_a h_{aa} + \sum_{a,b} 2J_{ab} + K_{ab} \tag{2.33}$$

As mentioned above, the minimization of this energy with respect to the spinorbitals will provide the best approximation to the exact energy in the Hartree-Fock framework. Since the wavefunction is a Slater determinant, a constrain which assures the orthogonality of the spinorbitals throughout the optimization must be added. This condition is:

$$\left\langle \chi_a \middle| \chi_b \right\rangle = \delta_{ab} \tag{2.34}$$

The minimization of E_0 with respect to the spinorbitals and including the orthogonality constrain, is obtained by minimizing the Lagrangian functional of the spinorbitals defined as:

$$\hat{L}\left[\chi_a\right] = E_0\left[\chi_a\right] - \sum_{ab} \varepsilon_{ba} \left(\left\langle \chi_a \middle| \chi_b \right\rangle - \delta_{ab} \right) \tag{2.35}$$

with ε_{ba} being the undetermined Lagrange multipliers, E_0 the expected value of $\left| \Psi_0 \right\rangle$ (see eq. (2.29)), and the remaining term the spinorbital orthogonality constrain.

Setting the first variation to zero, $\delta \hat{L}\left[\chi_a\right] = 0$, and working through some algebra, the so-called canonical Hartree-Fock equations are obtained:

$$\left[\hat{h}(\boldsymbol{x}_1) + \sum_{b=1}^{N} \hat{J}_b(\boldsymbol{x}_1) - \hat{K}_b(\boldsymbol{x}_1) \right] \chi_a(\boldsymbol{x}_1) = \varepsilon_a \chi_a(\boldsymbol{x}_1) \tag{2.36}$$

where, ε_a are the energy eigenvalues associated to the spinorbitals χ_a, and the Coulomg and Exchange operators are defined as:

$$\hat{J}_b(\boldsymbol{x}_1)\chi_a(\boldsymbol{x}_1) = \left[\int d\boldsymbol{x}_2 \chi_b^*(\boldsymbol{x}_2) \frac{1}{r_{12}} \chi_b(\boldsymbol{x}_2) \right] \chi_a(\boldsymbol{x}_1)$$

$$\hat{K}_b(\boldsymbol{x}_1)\chi_a(\boldsymbol{x}_1) = \left[\int d\boldsymbol{x}_2 \chi_b^*(\boldsymbol{x}_2) \frac{1}{r_{12}} \chi_a(\boldsymbol{x}_2) \right] \chi_b(\boldsymbol{x}_1) \tag{2.37}$$

The so-called Fock operator, $\hat{f}(a)$, corresponds to sum of operators in square brackets of eq. (2.36), and its eigenfunctions and eigenvalues are the spinorbitals and spinorbital energies, respectively.

If the spinorbitals are expressed as linear combinations of basis functions (see eq. (2.18)), the minimization of the energy with respect to the spinorbitals turns into a minimization with respect to the orbitals coefficients (C_μ), which in matrix form is written as:

$$\boldsymbol{Fc} = \boldsymbol{Sc\varepsilon} \tag{2.38}$$

These are the so-called Roothan-Hall[218] equations, where $\boldsymbol{c}$ are the expansion coefficients, ε is the Lagrangian multiplier, $\boldsymbol{S}$ is the overlap matrix., and the elements of $\boldsymbol{F}$ are:

$$F_{ab} = \left\langle \psi_a \left| \hat{f} \right| \psi_b \right\rangle \tag{2.39}$$

The fact that the operator that is used to minimize the energy depends intrinsically on the same spinorbitals, makes an iterative method necessary to obtain the spinorbitals which correspond to the "best" energy. For this reason, the HF method is also known as the Self Consistent Field (SCF) method. This iterative procedure needs a set of trial spinorbitals to build the initial operator. This trial spinorbitals are successively modified until convergence is achieved.

2.1.4.1.a   HF limitations

All the problems of the Hartree-Fock method arise from the fact of choosing a single Slater determinant as a wavefunction. Such a wavefunction cannot take into account all electronic correlation, therefore it leads to a total

electronic energy which is different from the exact solution of the non-relativistic Schrödinger equation within the Born-Oppenheimer approximation. Due to the HF is a variational method, the lowest energy that can be obtain with this method is always above the exact energy. The energy difference between the exact solution (within the B.O. approximation) and the HF one is commonly known as correlation energy, a term coined by Löwdin.

$$E_{corr} = E_{exact} - E_{HF} \tag{2.40}$$

A certain amount of electronic correlation is already considered within the HF approximation. It is found in the exchange term of eq. (2.29), which describes the correlation between electrons with parallel spin. This basic correlation prevents two parallel-spin electrons from being found at the same point of the space and is often called the Fermi correlation.

The paradigm of the shortcomings of HF is the H_2 molecule dissociation. Within the restricted HF methodology (alpha and beta electrons are restricted to have the same spatial function), the H_2 wavefunction is 50% ionic and 50% covalent for all bond lengths, which yields a bad description of the two neutral atoms at the dissociation limit. This problem can be solved by using the unrestricted HF methodology; however, with such a methodology (each spinorbital has its own spatial function) spin contamination can appear. If spin contamination is present, the wavefunction is no longer an eigenfunction of the total spin operator. This implies that the wavefunction is somewhat contaminated by other states, what can artificially raise the energy of the system. The HF methodology has also other limitations which are outlined next:

- Stretched bond energies are too large, what affects transition state structures by overestimating their activation energies.

- Equilibrium bond lengths are too short at the RHF level (*i.e.* the potential well is too steep). The HF method 'overbinds' the molecule. This an also affect the vibrational frequencies.

- The wavefunction overestimates the 'ionic' character leading to too large dipole moments (and also atomic charges).

In summary, the HF method is a good method to compute closed shell systems in equilibrium structures, but it has some limitations when optimizing TSs and treating open shell systems.

2.1.5   Multi-Configurational Methods

The HF method is a very simple method that allows the solution of the Schrödinger equation. It uses a monodeterminantal wavefunction, which is optimized to obtain the best description of the GS. Such a description can be improved by using methods that account for more electron correlation such as Møller-Plesset or Coupled Cluster, or in a different way, Density Functional Theory. Nevertheless, post-HF methods present difficulties in describing complex electronic-structure problems such as bond breakings or excited states. In these systems, the wavefunction is often dominated by more than one electronic configuration, therefore, a multiple-determinant approach such as the Configuration Interaction (CI)[219-221] or Multi-Configuration methods is needed.

The main difference between HF and the multi-configurational methods is that the latter use linear combinations of Slater determinants as wavefunctions, while the HF method only uses a single Slater determinant. That is:

$$\Psi_{MC} = a_0 \Psi_{HF} + \sum_{i=1} a_i \Psi_i \qquad (2.41)$$

Since the HF wavefunction corresponds to the GS, the other Slater determinants forming the multi-configurational wavefunctions correspond to excited states. The HF method can be understood as a particular case of a multi-configurational method (MC) where $a_i = 0$ and $a_0 = 1$. In many cases the HF method gives a good description of the wavefunction, and this is reflected on the composition of the wavefunction, where the HF configuration accounts for more than 90% of the Multiconfigurational Self Consistent Field (MCSCF) wavefunction ($a_0 \approx 0.9$).

The most complex and accurate MCSCF description is provided by the Full Configuration Interaction method (FCI). A FCI wavefunction is a linear combination of all possible excitations, which extended to an infinite basis set yields the exact solution of the Schrödinger equation. Unfortunately, a FCI calculation even for a finite basis is computationally extremely expensive and only possible for small systems. Nevertheless, some approaches to the FCI method with a reasonable computational cost, such as the FCI truncated at single and doubles excitations (CISD) or the Complete Active Space Self Consistent Field (CASSCF) approach, have been developed.

These methods will be overviewed in the next section, while other methods including the so-called dynamic correlation will be explained in the next chapter

2.1.5.1   The Configuration-Interaction method

The CI method is not multi-configurational but will serve as a reference for the CASSCF method that will be explained in the next section. The CI wavefunction consists of a linear combination of Slater determinants, the expansion coefficients of which are variationally determined. This method is flexible and can give highly accurate wavefunctions for small systems.

The first step in a CI calculation is the generation of MOs by the SCF procedure. Different configuration states are generated taking as reference the HF configuration, and are linearly combined to form the CI wavefunction. This procedure leads to a wavefunction whose configuration states have the same symmetry as the reference state. The CI configuration states are classified as monoexcitations, double-excitations, triple-excitations, …, depending on the number of electrons that are promoted with respect to the reference state. The number of configuration functions with appropriate symmetry increases rapidly with the number of electrons and basis functions. For a full CI calculation the number of Slater determinants generated by a system with n orbitals and $2k$ electrons is given by $N_{det} = \begin{pmatrix} n \\ k \end{pmatrix}^2$. Since this number becomes easily untreatable, truncated CI methods can only be used. For instance, the CISD method only takes into account configuration states made of single (CIS) and/or double (CISD) excitations. CIS calculations are usually not carried out because, according to Brillouin's theorem,[222,223] they do not improve the HF description. This point is shown next. The HF approximation to the exact wavefunction of the ground state is a single determinant, $\left| \Psi_0 \right\rangle$, made of the N spinorbitals with the lowest orbital energies. The CIS method, uses more than one determinant to describe the GS, but these "extra" determinants can only include single excitations, $\left| \Psi_a^r \right\rangle$, that is changing an occupied spinorbital (χ_a) for a virtual one (χ_r). Thus, the resulting CIS wavefunction reads as:

$$\Psi_0^{\text{CIS}} = c_0 \Psi_0 + \sum_{ra} c_a^r \Psi_a^r \tag{2.42}$$

The fact that $\left| \Psi_0 \right\rangle$ and $\left| \Psi_a^r \right\rangle$ only differ in one spinorbital, implies that the two determinants only differ in one row or column. If a determinant only differs in one row or column from the ground state one, any linear combination of the two can be expressed as one determinant. Thus, any combination of $\left| \Psi_0 \right\rangle$ and $\left| \Psi_a^r \right\rangle$ can be expressed as a single determinant.

One drawback of these truncated CI methods is the lack of size-extensivity.[224] A size-extensive method is such that the energy calculated thereby scales linearly with the number of particles. An important advantage of a size-extensive method is that it allows straightforward comparisons between calculations involving variable numbers of electrons, e.g. ionization processes or calculations using different numbers of active electrons. Lack of size-extensivity, however, implies that errors from the exact energy increase as more electrons enter the calculation.

Another problem of the truncated CI methods is that the orbitals which form the reference configuration state are generated self-consistently in the field of a single electronic configuration, which does not assure their relevance for a multi-configurational system. An obvious solution to this problem is to carry out a CI calculation where the orbitals are variationally optimized for each state simultaneously with the coefficients of the electronic configurations. This way, the orbitals of the optimized wavefunction can describe any of the states of the case study and do not introduce a bias towards a particular configuration. This approach is known as Multi-configurational Self-Consistent Field (MCSCF). The Complete Active Space Self-Consistent Field method (CASSCF), which belongs to the MCSCF family is explained next.

2.1.5.2   CASSCF

In the CASSCF approach, the total orbital space (all the orbitals that form the wavefunction) is partitioned in 3 subsets, namely *occupied*, *active*, and *virtual*. The occupied orbitals usually correspond to core orbitals and are defined as doubly occupied or *inactive*. In contrast, the virtual orbitals correspond to unoccupied orbitals. The active orbitals form the so-called active space. The active space usually includes all the orbitals which are relevant to describe the processes of interest of the case study. For instance, if we want to study a bond breaking process, the active space must include the orbitals that are involved in bond, usually sigma or pi orbitals. In photophysics studies, usually π orbitals are considered as most of the electronic transitions take place between π orbitals. Nowadays, because of computational cost, the CAS calculations are limited to fourteen active electrons distributed among the same number of active orbitals.

The CASSCF method can only account for the electronic correlation of the orbitals of the active space as only excitations within the active space are considered in the CASSCF wavefunction. The energy difference between the exact energy (FCI calculation with complete basis set) and the HF energy is usually considered as the total electron correlation, although, strictly speaking

this is not the "true" electron correlation because some correlation energy is already included in the HF energy. On the other hand, the correlation energy is usually partitioned into *dynamic* and *static* correlation energies. The former is related to the movement of the electrons and can be accounted by perturbation methods such as MP2. Unfortunately, these perturbation methods are not variational. They provide correlated energies but no new wavefunctions, which can be iteratively optimized to yield exact energies, can be obtained. In contrast to this, the static correlation is important in cases where the GS has to be described with more than one configuration, such as non-adiabatic reactions. It is usually defined as the energy difference between MCSCF methods and the HF method.

As MCSCF, and consequently CASSCF, theory is highly complex and the aim of this work is not to achieve a complete MCSCF theoretical development, only a general overview of the MCSCF energy and wavefunction optimization will be done. If further details are desired, see the books by Helgaker,[178] Yarkony, [170] and Roos and Widmark.[179,181] However, it is worth to mention that several strategies are used within the CASSCF methodology to increase the computational efficiency and to allow a multi-reference state optimization. These features are essential in MCSCF calculations and make the MCSCF methods a powerful tool to study excited states systems.

The CASSCF wavefunction is formed of a linear combination of the ground state wavefunction and the determinants that correspond to excitations within the active space. The size of the total orbital space, and especially the size of the active space determine the number of determinants that form the wavefunction and, therefore, the cost of the calculations. In MCSCF methods, both electronic configurations and MO coefficients are variationally optimized. An iterative method is used to get the optimal wavefunction, but this represents a difficult nonlinear task which restricts the length of the MCSCF expansion.

2.1.5.2.a   *CASSCF wavefunction optimization*

As stated above, MCSCF wavefunctions are expressed as linear combinations of configuration state functions or determinants ($\left| i \right\rangle$):

$$\left| \boldsymbol{c}, \boldsymbol{a} \right\rangle = \mathrm{e}^{(-\hat{a})} \sum_i c_i \left| i \right\rangle \qquad (2.43)$$

where $\boldsymbol{c}$ are the configuration state functions coefficients, $\boldsymbol{a}$ correspond to the orbital coefficients, and the operator $\mathrm{e}^{(-\hat{a})}$ carries out unitary transformations among the spinorbitals as in the HF case. The MCSCF wavefunction expression

is identical to the CI model except for the presence of the orbital-rotation operator, which is the operator that performs the MO optimization for each state.

The ground state MCSCF wavefunction is obtained by minimizing the energy with respect to the parameters of the orbital-rotation operator (a) and the CI coefficients, c, simultaniously:

$$E = \min_{c,a} \frac{\left\langle c,a \left| \hat{H} \right| c,a \right\rangle}{\left\langle c,a \middle| c,a \right\rangle} \tag{2.44}$$

At convergence, the wavefunction is expanded into an orbital basis where $\hat{a}$ is zero, adopting the form of a standard CI wavefunction, but with optimal orbitals. Due to the multi-configurational character of the wavefunction, the optimization can be performed for any chosen reference state, whose energy will be minimized.

The wavefunction optimization procedure follows the same strategy as the HF case. A Lagrangian equation is defined to minimize the energy. In this case, since the minimization is performed with respect two parameters, this equation has the form of a Taylor expansion of the energy truncated to second order:

$$Q\left(\lambda\right) = E^{(0)} + \boldsymbol{E}^{(1)T}\boldsymbol{\lambda} + \frac{1}{2}\boldsymbol{\lambda}^{T}\boldsymbol{E}^{(2)}\boldsymbol{\lambda} \tag{2.45}$$

where $\boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{c} \end{pmatrix}$ is a vector that represents the general MCSCF energy of the reference state, $Q(\boldsymbol{\lambda})$ corresponds to the second-order MCSCF energy and $E^{(0)}$ is the zeroth order MCSCF energy: $E^{(0)} = \left\langle 0 \left| \hat{H} \right| 0 \right\rangle$.

$\boldsymbol{E}^{(1)}$ and $\boldsymbol{E}^{(2)}$ are the electronic gradient and electronic Hessian respectively, which can also be written in block form as:

$$\boldsymbol{E}^{(1)} = \begin{pmatrix} {}^{c}\boldsymbol{E}^{(1)} \\ {}^{a}\boldsymbol{E}^{(1)} \end{pmatrix} \qquad \boldsymbol{E}^{(2)} = \begin{pmatrix} {}^{cc}\boldsymbol{E}^{(2)} & {}^{ca}\boldsymbol{E}^{(2)} \\ {}^{ac}\boldsymbol{E}^{(2)} & {}^{cc}\boldsymbol{E}^{(2)} \end{pmatrix} \tag{2.46}$$

where the left superscripts denote the differentiation with respect to configuration (c) or orbital (a) coefficients respectively:

$$
{}^{c}E_i^{(1)} = \left(\frac{\partial E}{\partial c_i} \right)_{\lambda=0} \qquad {}^{a}E_i^{(1)} = \left(\frac{\partial E}{\partial a_{pq}} \right)_{\lambda=0}
$$

$$
{}^{aa}E_{pq,rs}^{(2)} = \left(\frac{\partial^2 E}{\partial a_{pq}\, \partial a_{rs}} \right)_{\lambda=0} \quad {}^{cc}E_{i,j}^{(2)} = \left(\frac{\partial^2 E}{\partial c_i \partial c_j} \right)_{\lambda=0} \quad {}^{ac}E_{pq,i}^{(2)} = \left(\frac{\partial^2 E}{\partial a_{pq}\, \partial c_i} \right)_{\lambda=0}
$$

(2.47)

In order to determine the different components of the Hessian, it is convenient to define the energy of the system, eq. (2.44), as:

$$
E(\boldsymbol{c}, \boldsymbol{a}) = \frac{W(\boldsymbol{c}, \boldsymbol{a})}{S(\boldsymbol{c}, \boldsymbol{a})} \tag{2.48}
$$

where the numerator holds the configuration weights:

$$
W\left(\boldsymbol{c}, \boldsymbol{a} \right) = \left(\langle \boldsymbol{c} | \hat{P} + \langle 0 | \right) \mathrm{e}^{(\hat{a})} \, \hat{H} \, \mathrm{e}^{(\hat{a})} \left(|0\rangle + \hat{P} | \boldsymbol{c} \rangle \right) \tag{2.49}
$$

with $\hat{P}$ being the projector operator, which projects out the ground state component of the reference state $\left(|0\rangle \right)$ from $\left(|\boldsymbol{c}\rangle \right)$, and where the denominator contains the overlap between these states:

$$
S\left(\boldsymbol{c}, \boldsymbol{a} \right) = \left(\langle \boldsymbol{c} | \hat{P} + \langle 0 | \right) \mathrm{e}^{(\hat{a})} \mathrm{e}^{(\hat{a})} \left(|0\rangle + \hat{P} | \boldsymbol{c} \rangle \right) = \langle 0 | 0 \rangle + \langle \boldsymbol{c} | \hat{P} | \boldsymbol{c} \rangle \tag{2.50}
$$

Considering that the overlap does not depend on the orbital coefficients, we can rewrite the MCSCF energy expression as:

$$
S(\boldsymbol{c}) E(\boldsymbol{c}, \boldsymbol{a}) = W(\boldsymbol{c}, \boldsymbol{a}) \tag{2.51}
$$

And differentiating this expression with the assumption of real orbitals and the fact that $S^{(0)}=1$, yields to:

$$
\boldsymbol{E}^{(1)} = \boldsymbol{W}^{(1)} - E^{(0)} \boldsymbol{S}^{(1)}
$$

$$
\boldsymbol{E}^{(2)} = \boldsymbol{W}^{(2)} - \boldsymbol{E}^{(1)} \boldsymbol{S}^{(1)^{\mathrm{T}}} - \boldsymbol{S}^{(1)} \boldsymbol{E}^{(1)^{\mathrm{T}}} - \boldsymbol{E}^{(0)} \boldsymbol{S}^{(2)}
$$

(2.52)

Now, by direct differentiation of the last expressions and applying some simplifications, one can deduce the compact equations of the gradient and Hessian respectively:

$$^{c}E_{i}^{(1)} = 2\left\langle \Psi_{i} \left| \hat{P}\hat{H} \right| \Psi_{0} \right\rangle = 2\left\langle \Psi_{i} \left| \hat{H} \right| \Psi_{0} \right\rangle - 2c_{i}^{(0)}E^{(0)}$$

$$^{o}E_{pq}^{(1)} = \left\langle \Psi_{0} \left\| \left[E_{pq}^{-},\hat{H} \right] \right\| \Psi_{0} \right\rangle$$

$$^{cc}E_{i,j}^{(2)} = 2\left\langle \Psi_{i} \left| \hat{P}\left(\hat{H} - E^{(0)} \right)\hat{P} \right| \Psi_{j} \right\rangle = 2\left\langle \Psi_{i} \left| \hat{H} - E^{(0)} \right| \Psi_{j} \right\rangle - c_{i}^{(0)}E_{j}^{(1)} - c_{j}^{(0)}E_{i}^{(1)} \qquad (2.53)$$

$$^{ac}E_{pq,i}^{(2)} = 2\left\langle \Psi_{i} \left| \hat{P}\left[E_{pq}^{-},\hat{H} \right] \right| \Psi_{0} \right\rangle = 2\left\langle \Psi_{i} \left\| \left[E_{pq}^{-},\hat{H} \right] \right\| \Psi_{0} \right\rangle - 2c_{i}^{(0)}E_{pq}^{(1)}$$

$$^{aa}E_{pq,rs}^{(2)} = \frac{1}{2}\left(1 + P_{pq,rs} \right)\left\langle \Psi_{0} \left\| \left[E_{pq}^{-}\left[E_{rs}^{-},\hat{H} \right] \right] \right\| \Psi_{0} \right\rangle$$

where E_{pq}^{-} corresponds to the antisymmetric combinations of excitation operators:

$$E_{pq}^{-} = E_{pq} - E_{qp}, \qquad (2.54)$$

and $P_{pq,rs}$ the permutation of the pair indices *pq* and *rs*.

From the differentiation of the Lagrangian expression with respect to the reference CI vector ζ, the standard CI eigenvalue problem is obtained:

$$\hat{H}\boldsymbol{\zeta} = E\boldsymbol{\zeta} \qquad (2.55)$$

A Fock operator which is separated into inactive, virtual and active sub-Fock operators is obtained from these equations, too. For further details see Ref. 178.

The energy minimization procedure involves two iterative procedures (orbitals and coefficients) that are performed simultaneously. This optimizations must be performed carefully as they can carry convergence problems. One of the problems of the wavefunction optimization are the redundant orbital rotations. This rotations can yield Hessian singularities or optimization failures. Thus, such rotations must be found and avoided. Convergence problems or bad descriptions might also arise when two states are almost degenerate. In such a situations it is a general procedure to optimize two or more states simultaneously, *i.e.* use a reference state which is a linear combination of states. In such an optimization, the orbitals are not optimized separately for each state, but are iteratively transformed to minimize the average energy of all states, *i.e.* the so-called state-average optimization. This is accomplished by introducing an

energy function which is composed of a weighted linear combination of states. That is, the energy to be optimized is:

$$E = \sum_{\mathrm{I}} w_{\mathrm{I}} E_{\mathrm{I}} \qquad (2.56)$$

where the weights for each state are denoted by the fixed parameters, w_{I}. The weights are usually identical for each state, but in some occasions can be chosen different. This is the case of some dynamic calculations where at the starting point the states are well separated and no state-average is necessary, therefore one state is given a weight of one. If along the simulation the states get successively closer up to degeneracy, the weight of the starting states is gradually decreased along the simulation until the two states are equally state-averaged at degeneracy.

One of the problems of such optimizations is that the orbitals are not optimal for each separate state, and that the description of each state is not so accurate. On the other hand, the state-averaged procedure provides a set of orthonormal, non-interacting electronic states, which simplifies subsequent CI or response-theory calculations.

2.1.5.2.b   CASSCF limitations

The major drawback of the CASSCF method is the lack of dynamic correlation, as this can result in a bad description of the energy of the system. The lack of dynamic correlation is caused by the use of a limited active space, currently up to 14 electrons distributed along 14 orbitals, (14,14). The active space determines the number of configurations that will compose the wavefunction. Some tricks can be applied to ignore states of different multiplicity or symmetry in the calculations and reduce the size of the active space. This reduction of the active space may allow the study of a system which needs larger active spaces to be described. In contrast, the fact that the active space can be defined as desired for each calculation implies that the CASSCF method is not a black box method. This free choice can bias the descriptions of the systems and makes the method prone to suffer from artifacts.

Although it will not be explained here, it is worth to mention that there exists another MCSCF scheme, namely the Restricted Active Space Self-Consistent Field (RASSCF),[225,226] which allows more electrons and orbitals to be included in the active space and accounts for some dynamic correlation. However, this larger active space does not provide full dynamic correlation either. Both static and dynamic correlations can be recovered if MCSCF methods are used in conjunction with perturbation theory methods. One of

these methods, is the so-called CASPT2 approach, which uses a CASSCF reference wavefunction to perform a second order Møller-Plesset correction, which gives correlated energies and allows using larger active spaces. This, and other methods that account for dynamic correlation are explained next.

2.1.6   *Including Dynamic correlation*

The HF method, which is the simplest way to solve the Schrödinger equation, has been seen in the previous section. Because of its simplicity, it can not be used to study excited states where more than one state is needed. In order to study excited states, one can use a multideterminantal method such as CI, CASSCF, ... However, it might occur that the description provided by these methods is not accurate enough because of the lack of dynamic correlation. There are many methodologies which include dynamic correlation, namely Møller-Plesset perturbation theory, Density Functional Theory, Coupled Cluster, Multi-Reference CI, etc.. Since most of the calculations in the present thesis have been carried out at the CASSCF level, only the second order Møller-Plesset perturbation theory[141] applied to HF and CASSCF methodologies, respectively, will be overviewed. In addition, this is the most reliable methodology to describe processes in the excited states, specially non-adiabatic transitions with the GS. Such transitions take place at rather distorted geometries where the GS is multi-referential.

2.1.6.1   **Møller-Plesset perturbation theory**

The use of Møller-Plesset perturbation theory must be understood from the point of view that it represents an easy and relatively fast way to improve HF results by inclusion of dynamic correlation. The key point is that MP can systematically approach the exact solution to the Schrödinger equation order by order of expansion of the energy, where the zeroth order expansion is the HF solution. Thus, the MP2 approach is a second order correction to the HF solution, MP3 is a third order, ..., and so forth. Unfortunately MPn (where n is the order of the expansion) approaches have some limitations. The lack of convergence of some expansions is probably the most important drawback of MP. Another important problem is the fast increasing computational cost of the calculations along with the orders of expansion. This reduces the applicability of high MP orders of correction, which become as expensive as other methodologies that can approach better the exact solution to the Schrödinger equation. For instance, the MP4 level of theory is as computationally expensive as CCSD. In addition, experience has shown that MP3 level usually does not improve MP2

(in general, odd corrections do not perform better than even ones). All this translates into an unfrequent use of high order MPn (n>2) expansions.

2.1.6.1.a   MP2

Perturbational methods are based on the partition of the Hamiltonian operator in two parts: an unperturbed part (zeroth-order Hamiltonian), $\hat{H}_0$, and the perturbation $\hat{V}$ operator, as:

$$\hat{H} = \hat{H}_0 + \lambda\hat{V} \tag{2.57}$$

where λ represents the perturbation variation. In the solving procedure, the perturbation is gradually added to the unperturbed Hamiltonian until the exact solution is achieved ($\lambda = 1$).

Within the Rayleigh-Schrödinger perturbation theory, which is a particular case of MP perturbation theory, the initial Hamiltonian is expressed as an infinite sum of terms and the exact wavefunction as an infinite sum of functions:

$$\Psi_i = \sum_{k=0}^{\infty} \lambda^{(k)} \Psi_i^{(k)} \tag{2.58}$$

Consequently, the energy of this wavefunction is also expressed as an infinite sum of terms:

$$E_i = \sum_{k=0}^{\infty} \lambda^{(k)} E_i^{(k)} \tag{2.59}$$

In principle, any unperturbed Hamiltonian can be used. However, convergence is faster if the Hamiltonian contains the main features of the exact Hamiltonian and provided that the perturbation is smaller than the unperturbed Hamiltonian. In the original Møller-Plesset theory the zeroth-order wavefunction is an exact wavefunction of the Fock operator ($\hat{f}$), which is used in the unperturbed Hamiltonian, and the perturbation operator is the correlation potential:

$$\hat{V} = \hat{H} - \hat{f} - \left\langle \Psi_0 \left| \hat{H} - \hat{f} \right| \Psi_0 \right\rangle \tag{2.60}$$

A shifted Fock operator is used as unperturbed Hamiltonian so as to make the HF energy be the expected value of the zeroth-order expansion. Thus,

$$\hat{H}_0 = \hat{f} + \left\langle \Psi_0 \left| H - \hat{f} \right| \Psi_0 \right\rangle \tag{2.61}$$

Provided that $\hat{f}\Psi_0 - \left\langle \Psi_0 \left| \hat{f} \right| \Psi_0 \right\rangle \Psi_0 = 0$, it is readily seen that the expected value of the unperturbed Hamiltonian reduces to

$$\hat{H}_0 \Psi_0 = \left\langle \Psi_0 \left| H \right| \Psi_0 \right\rangle \Psi_0 \tag{2.62}$$

which is the HF energy. According to Brillouin's theorem (section 2.1.5.1) the first order correction, $\left\langle \Psi_0 \left| \hat{V} \right| \Psi_0 \right\rangle$, is zero. Thus, the first improvement on the HF energy takes place at the second order correction because the single excited determinants cannot improve the GS description.

In contrast to the original MP formulation, a different partitioning of the Hamiltonian is usually used in computational chemistry. An unshifted Fock operator is used as zeroth order Hamiltonian and the perturbation operator corresponds to the electronic correlation. Therefore, the expected value of the zeroth order Hamiltonian is no longer the HF energy but the sum of occupied orbital energies. The HF energy is then obtained at the first order correction of the energy. This alternative partitioning has no other implications. In both formulations, the first improvement of the HF energy takes place at the second-order correction term, and the subsequent orders of correction are the same.

The MP2 approach is the most widely used one in the framework of MP theory because it offers a good compromise between computational cost and quality of the results. It is expressed as a correction term to the HF energy which reads as:

$$E_0^{(2)} = \left\langle \Psi_0^0 \left| \hat{V} \right| \Psi_0^1 \right\rangle = \sum_{k \neq 0} \frac{\left| \left\langle \Psi_0^0 \left| \hat{V} \right| \Psi_k^0 \right\rangle \right|^2}{E_0^{(0)} - E_k^{(0)}} \tag{2.63}$$

where the superscripts indicate the correction order and the subscripts the electronic state. It can be proved that only those terms which involve double excitations between occupied and virtual molecular orbitals contribute to the second order correction term, which then reduces to

$$E_0^{(2)} = \sum_{\substack{a<b \\ r<s}} \frac{\left| \left\langle \Psi_0 \left| \sum_{i<j} r_{ij}^{-1} \right| \Psi_{ab}^{rs} \right\rangle \right|^2}{E_a + E_b - E_r - E_s} = \sum_{\substack{a<b \\ r<s}} \frac{\left| \left\langle ab \| rs \right\rangle \right|^2}{E_a + E_b - E_r - E_s} \tag{2.64}$$

where in the bielectronic integrals a,b and r,s denote the occupied and virtual orbitals respectively. The energies of the denominator correspond to the orbital energies.

As mentioned above, experience shows that in most cases third-order correction does not improve MP2, as it only uses double excitations too. Thus, the MP4 approach must be used to improve MP2. However, the use of such a level of theory is not advisable since it is computationally very expensive. Hence, if more accurate results are needed, one should use the CCSD level of theory since it has a somewhat lower computational cost than MP4, and analytical gradients and frequencies are available for most of softwares.

2.1.6.2   CASPT2

We have just seen the use of MP2 to recover dynamic correlation from a HF reference function. In the present section a similar approach is described, namely CASPT2, but in this case the reference wavefunction is a CASSCF function. This method[50,51] has turned out to be one of the most accurate *ab initio* approaches, which has been successfully applied to the study of a broad number of systems, especially those related to electronic spectroscopy and photophysics. This is the main reason why this method has been used in the present thesis.

The zeroth-order Hamiltonian can be expressed as:

$$\hat{H}_0 = \hat{P}_0 \hat{F} \hat{P}_0 + \hat{P}_X \hat{F} \hat{P}_X \tag{2.65}$$

where $\hat{P}_0 = \left| \Psi_0 \right\rangle \left\langle \Psi_0 \right|$ is a projection operator onto the reference function, $\hat{P}_X$ is the corresponding projection operator for the rest of the configurations, and $\hat{F}$ is the Fock operator.

Within the CASPT2 method, the reference wavefunction is a CAS function and the configuration space in which it is expanded is assumed to be diagonal. It is partitioned into 4 subspaces: V_0, V_K, V_{SD}, and V_X. V_0 is the one dimensional space spanned by the reference function for the state under consideration, V_K is the space spanned by the orthogonal complement to the reference wavefunction used to build the CAS wavefunction (*i.e.* the rest of the CAS CI space), V_{SD} is the space spanned by all singly and doubly excited states (with respect to the CAS reference), and V_X is the rest of the CI space. Note that only the V_{SD} space will interact with the reference wavefunction via the total Hamiltonian. Consequently, the zeroth-order Hamiltonian is formulated in

such a way that only the subspace SD contributes to the expansion of the first order wavefunction and reads as:

$$\hat{H}_0 = \hat{P}_0 \hat{F} \hat{P}_0 + \hat{P}_K \hat{F} \hat{P}_K + \hat{P}_{SD} \hat{F} \hat{P}_{SD} + \hat{P}_X \hat{F} \hat{P}_X \tag{2.66}$$

In the MP2 case, the Fock operator was chosen so that its eigenvalues corresponded to the sum of occupied orbital energies, that is in diagonal form. Here this is no longer possible. In the MCSCF matrix the F operator is defined in a form such that, for inactive and virtual orbitals, its diagonal elements correspond to the orbital energies. Thus, it shall be written as the one-electron operator:

$$\hat{F} = \sum_{p,q} f_{pq} \hat{E}_{pq} \tag{2.67}$$

where ,

$$f_{pq} = h_{pq} + \sum_{r,s} D_{rs} \left[\left(pq \middle| rs \right) - \frac{1}{2} \left(pr \middle| qs \right) \right] \tag{2.68}$$

$D_{rs} = \left\langle \Psi_0 \middle| \hat{E}_{rs} \middle| \Psi_0 \right\rangle$, and $\hat{E}_{rs}$ is the excitation operator.

Considering that the reference wavefunction is a CASSCF wavefunction, one can apply the generalized Brillouin's theorem which states that f_{pq} is zero when one of the indices corresponds to an inactive orbital and the other to an external orbital. Then, the operator F can be simplified by defining a new set of orbitals by diagonalization of each of the three diagonal blocks of the matrix f.[227] Such a transformation is only possible for a CASSCF reference function since the CAS CI space is invariant with respect to this rotation and rotations among each subset of orbitals. That is, only rotations between orbitals of different subsets can change the wavefunction(s). This is used to simplify the f matrix and make it diagonal in each of the three subsets. In order to do so, the orbitals are dived into inactive, active, and external; and $\hat{F}$ is written as:

$$\begin{aligned} \hat{F} = \sum_i \varepsilon_i \hat{E}_{ii} + \sum_t \varepsilon_t \hat{E}_{tt} + \sum_a \varepsilon_a \hat{E}_{aa} \\ + \sum_{i,t} f_{ti} \left[\hat{E}_{it} + \hat{E}_{ti} \right] + \sum_{i,a} f_{ai} \left[\hat{E}_{ia} + \hat{E}_{ai} \right] + \sum_{t,a} f_{at} \left[\hat{E}_{ta} + \hat{E}_{at} \right] \end{aligned} \tag{2.69}$$

where the labels $i, t,$ and a correspond to inactive, active and external orbitals respectively.

As mentioned above (eq. (2.66)), only those configurations which interact directly with the CAS reference wavefunction have to be included in the first order wavefunction. All these configurations belong to the SD space, and consequently the first order wavefunction can be written as:

$$\Psi_1 = \sum_{p,q,r,s \notin active} C_{pqrs} \left| pqrs \right\rangle \tag{2.70}$$

where $\left| pqrs \right\rangle = \hat{E}_{pq}\hat{E}_{rs} \left| \Psi_0 \right\rangle$. Moreover, the indices p, q, r, and s cannot correspond to active electrons since the resulting wavefunction would belong to the CAS CI space (K subspace in eq. (2.66)).

This way, each term of the first order expansion consists of linear combinations of functions with coefficients from the CAS reference function, Ψ_0, which in matrix form reads as:

$$\left(\boldsymbol{F} - E_0\boldsymbol{S}\right)\boldsymbol{C} = -\boldsymbol{V} \tag{2.71}$$

where, $\boldsymbol{F}$ is the Fock matrix in the SD subspace, $\boldsymbol{S}$ is the overlap matrix, $\boldsymbol{C}$ is the coefficient vector of eq. (2.70) and $\boldsymbol{V}$ is the vector that represents the interaction between the SD subset and the reference function, that is $V_{pqrs} = \left\langle \Psi_0 \left| \hat{H} \right| pqrs \right\rangle$. However, the resulting states are not orthogonal to each other neither linearly independent, thus eq. (2.71) cannot be solved directly. Instead, first the $\boldsymbol{S}$ matrix is diagonalized and transformed into a somewhat smaller matrix which has no longer linear dependencies. Then, in a second step the resulting equation is diagonalized by some iterative method, usually the robust Davidson approach.[228] The main problem is the non-diagonality of $\boldsymbol{F}$, whose elements have the general form:

$$\left\langle pqrs \left| \hat{F} \right| p'q'r's' \right\rangle = \sum_{\alpha,\beta} \boldsymbol{f}_{\alpha,\beta} \left\langle \Psi_0 \left| \hat{E}_{sr}\hat{E}_{qp}\hat{E}_{\alpha\beta}\hat{E}_{p'q'}\hat{E}_{r's'} \right| \Psi_0 \right\rangle \tag{2.72}$$

As a consequence, the first, second, third and fourth order density matrices of the CAS reference function are needed to compute the matrix elements. The fifth order density matrix is not needed because at least one of the four indices p, q, r or s needs to be either inactive or external and therefore can be contracted out. The calculation of these matrices is very expensive and this step is the most time consuming part of the CASPT2 approach.

2.1.6.2.a   Intruder states and Level Shift

In this sub-chapter, the common problem of the intruder states in CASPT2 will be overviewed, and a solution[229] to the problem is presented. The intruder states are states of the V_{SD} space which have a zeroth-order energy that is close to the reference energy (E_0), which causes the denominator to be near zero. Although it can also occur in the ground state, intruder states usually appear in excited states calculations. This is usually related to weakly occupied orbitals which have energies close to those of external orbitals, or strongly occupied orbitals with energies close to those of internal orbitals. With small energy denominators, the corresponding first order coefficient terms become large, and the second order perturbation theory, which assumes that perturbations must be small, does not hold anymore. This can be observed in the form of the normalized CASPT2 wavefunction,

$$\left| \Psi \right\rangle = \sqrt{\omega} \left| \Psi_0 \right\rangle + \sqrt{1-\omega} \left| 1 \right\rangle \tag{2.73}$$

where $\omega = \dfrac{1}{1+S}$ is the weight of the CAS reference function, S is the overlap and $\left| 1 \right\rangle$ is the normalized first order wavefunction. The presence of an intruder state is detected when the value of ω in a specific calculation is much smaller than a given value of approximately 0.45. If there is a strong intruder state, the only solution that can be taken is to increase the active space in order to include the intruder state in the CAS CI space. If the active space can not be increased because it is already on the limit of applicability and there is a "strong" intruder state, then the second order perturbation theory is not valid to treat that system and another level of theory must be used.

Since the presence of intruder states depends on the geometry, it can cause singularities in the potential energy function and lead to divergence of the perturbation series. Nevertheless, if the intruder state is weakly interacting with the reference states, there is the possibility to use a level shift technique which can be used throughout, does not introduce any bias and does not imply an active space enlargement.

The level shift technique consists in the addition of a small constant, hereafter ε, to the zeroth order Hamiltonian, which yields the following first order equation:

$$\left(\hat{H}_0 - E_0 + \varepsilon \right) \tilde{\Psi}_1 = \left(E_1 - \hat{H}_1 \right) \Psi_0 \tag{2.74}$$

The tilde on Ψ_1, the first order perturbation to the wavefunction, indicates that it is affected by the level shift. The first order perturbation is expanded by Φ,

$$\tilde{\Psi}_1 = \sum_i^M \tilde{C}_i \Phi_i \tag{2.75}$$

which is an eigenfunction of the zeroth-order Hamiltonian, $\hat{H}_0 \Phi_i = \varepsilon_i \Phi_i$. The space spanned by $\left\{ \Phi_i \right\}_{i=1}^M$ is identical to the first order interacting space, $\hat{Q}\hat{E}_{pq}\hat{E}_{rs}\left| \Psi_0 \right\rangle$, where $\hat{Q}$ is a projector onto the complement of the CAS wavefunction space and $\left\{ \Phi_i \right\}_{i=1}^M$ is obtained diagonalizing the zeroth order Hamiltonian after orthonormalization and removal of the linear dependencies of the functions obtained by the excitation operators.

The coefficients of the expansion read as:

$$\tilde{C}_i = -\frac{\left\langle \Phi_i \left| \hat{H}_1 \Psi_0 \right\rangle\right.}{\varepsilon_i - E_0 + \varepsilon} = -\frac{V_i}{\Delta_i + \varepsilon} \tag{2.76}$$

and, therefore the second order energy depends on ε:

$$\tilde{E}_2 = -\sum_{i=1}^M \frac{\left| \left\langle \Phi_i \left| \hat{H}_1 \Psi_0 \right\rangle \right. \right|^2}{\varepsilon_i - E_0 + \varepsilon} = -\sum_{i=1}^M \frac{\left| V_i \right|^2}{\Delta_i + \varepsilon} \tag{2.77}$$

This expression corrects the effect of intruder states. It corrects the energy at regions close to singularities, where the denominator is very small and can lead to large perturbations. However, when the denominator is not small the level shift could have a non-negligible effect on the energy. In order to avoid a large effect of the level shift to the energy, Roos *et al.*[229] pondered its contribution to the energy with the weight of the CASSCF reference wavefunction, $\tilde{\omega}$:

$$E_2 = \tilde{E}_2 - \varepsilon\left(\frac{1}{\tilde{\omega}} - 1\right) \tag{2.78}$$

Such an approach leads to a second order energy which is much less sensitive to the level shift parameter, and allows for more robust calculations. Another approach, namely an imaginary level shift,[230,231] which is similar to this one but is more robust is starting to be used instead of the (real) level shift

technique. However, for consistency, only the real level shift approach has been used in this thesis.

2.1.6.3   MS-CASPT2

Despite the good performance of the CASPT2 method in describing most of the systems, there are some cases in which its use can result inadequate. Some problems related to the use of CASPT2 as well as a method which can prevent them are explained in the present section.

CASPT2 inaccuracies may arise in situations where the CASSCF function is not a good reference function such as avoided crossings or regions with strong mixing of valence and Rydberg states. A typical CASPT2 failure is the description of the ionic and neutral LiF dissociation.[232] In such a reaction (Figure 19), the CASPT2 method describes a double crossing region rather than an avoided crossing.
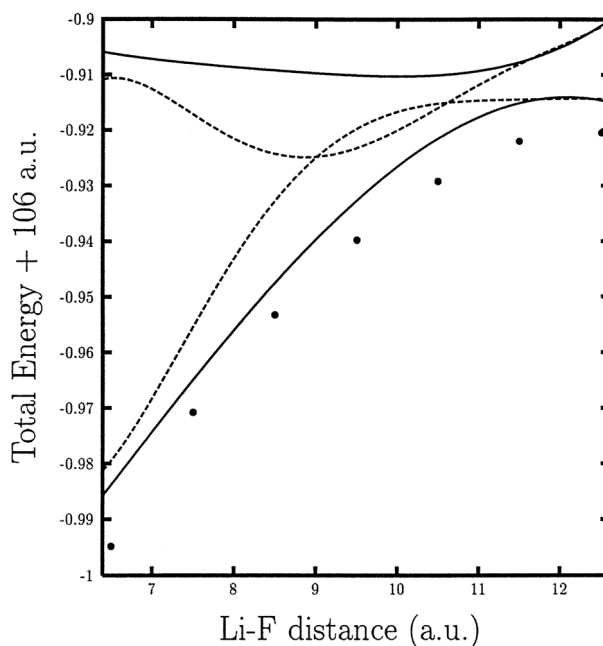
Figure 19. Potential energy surface for the two lowest $1\Sigma^+$ states of LiF. Dashed lines represent CASPT2 energies. Solid lines correspond to MS-CASPT2 calculations and the dots correspond to FCI values (Adapted from Ref. 232).

Dashed lines correspond to the CASPT2 energies of the two lowest excited states of LiF and clearly describe two regions of CI at 9 and 10.5 a.u. before dissociation. It can be observed that the CASPT2 curves differ substantially from the FCI profile (dotted line), specially in the CI region. Such a bad behavior can also affect the vertical spectrum of molecules, as it was shown[233]

for ethane, where the $^1B_{1u}$ state is predicted to lie 0.4 eV above its "correct" value. Luckily, there exists a simple method, namely multi-state CASPT2 (MS-CASPT2), which is capable of fixing these pitfalls. In Figure 19, the solid lines correspond to MS-CASPT2 energy profiles and describe the correct PES for the dissociation of LiF.

The problems described for CASPT2 are related to the quality of the reference (CASSCF) function. The main difference between the CASPT2 and MS-CASPT2 methods is therefore the construction of the reference state. In the former method the reference space (P) corresponds to a single state CASSCF function,

$$P = \left| \alpha \right\rangle \left\langle \alpha \right| \tag{2.79}$$

which an eigenfunction of the zeroth-order Hamiltonian,

$$H_0 \left| \alpha \right\rangle = E_0^\alpha \left| \alpha \right\rangle \tag{2.80}$$

On the other hand, the reference state of the MS-CASPT2 approach is spanned by two or more state-averaged CASSCF states,

$$P = \sum_{\alpha=1}^{d} \left| \alpha \right\rangle \left\langle \alpha \right| \tag{2.81}$$

where d is the dimension of the reference space and the reference states are eigenfunctions of their zeroth-order Hamiltonian:

$$H_0^\alpha \left| \alpha \right\rangle = E_0^\alpha \left| \alpha \right\rangle \; \left(\alpha = 1, \; 2, \; ..., \; d \right) \tag{2.82}$$

In the MS-CASPT2 approach an effective Hamiltonian, H^{eff}, is computed perturbatively and diagonalized within the reference space, which allows for a mixing of the states contained in the state-average function. At the second order perturbation, the diagonal elements of H_{2nd}^{eff} are the corresponding CASPT2 energies and the off-diagonal elements account for the mixing of the different states.

The MS-CASPT2 method is, in fact, a generalization of the CASPT2 one. In most cases the two methods provide similar results, although the CASPT2 method presents difficulties when dealing with wavefunctions which are dominated by more than one configuration. In such cases, given that the MS-CASPT2 method accounts for the mixing of the states, it represents a solution

to the problem. However, it must be used carefully because if the active space is not good enough the MS-CASPT2 can overcouple the states and change the energy profile of the system.

2.2   Molecular Dynamics

Some approaches capable of describing the electronic structure of molecules with increasing accuracy have been explained so far. Starting with HF, which is one of the simplest methods and can only describe one electronic state at a time, up to the MS-CASPT2 approach, which can treat multiple states and accounts for dynamic correlation. Despite the complexity of some of these methods, they can only provide a static picture. These methods do not take into account the time variable involved into the motion of nuclei, *i.e.* the evolution of the molecular geometry as the reaction proceeds in time. Static computations might not be sufficient to get reliable data on properties such as quantum yields or excited-state lifetimes, or to describe some non-adiabatic processes where time is essential and specific quantum effects affect the motion of the nuclei.

Two types of molecular dynamics that have been applied in this thesis to the study of the photophysics of thymine are explained next. In the first section (2.2.1), a quasi-classical dynamics method that allows the description of non-adiabatic events, that is classical dynamics with surface hopping algorithm[234], is overviewed. In the second part (section 2.2.3), a highly efficient approach to quantum dynamics, which allows to use a large number of degrees of freedom in the propagation by using a set of Gaussian functions to approach the wavepacket, namely the Direct Dynamics variation Multi-Configurational Gaussian (DD-vMCG) is explained.

2.2.1   Quasi-classical dynamics

When an electron of a molecule is promoted, the molecule changes the quantum state. The evolution of this non-stationary state is described by the time-dependent Schrödinger equation. As mentioned in section 2.1.2, the Born-Oppenheimer approximation allows for a separation of the electronic and nuclear parts of the Hamiltonian. Actually, it separates the electronic and nuclear motions, allowing the system evolution to be described by a function of the nuclei. This function is commonly known as wavepacket and moves on the PES created by the adiabatic motion of electrons. Quasi-classical dynamics methods are characterized by treating the electronic part of the Hamiltonian with quantum mechanics and the propagation of the system with Newton's equations of motion. The classical description of the nuclear motion implies that a rather large number of trajectories, namely a swarm of trajectories, are needed

to simulate the evolution of the wavepacket. In practice, the wavepacket is approximated by a vector whose elements correspond to the populations of the trajectories in the initial ensemble.

In our case, the problem of simulating the wavepacket is "reduced" to the solution of Newton's equations of motion (and the corresponding PES and derivatives of the electronic part at each step) for a number of different initial conditions. These initial conditions are usually generated with a random sampling procedure, namely microcanonical sampling. In photochemistry, the initial conditions are usually generated on the GS and translated to the excited state. The reason for that is that photon absorption by molecules at 0 K can be modeled by a so-called Franck-Condon wavepacket that results from the promotion of the molecule from the vibrational ground state in the electronic ground state, to a superposition of vibrational states in the excited electronic state by a laser pulse of no duration in time, which is white light (all frequencies).

Heisenberg's uncertainty principle states that the position and velocity of molecules cannot be determined exactly at the same time. This principle holds even at zero Kelvin, a temperature at which molecules are not supposed to move. Thus, if there is no momentum at 0 K, the determination of the position implies that both, position and velocity, are known at the same time because velocity is 0. This would break Heisenberg's uncertainty principle. In fact, momentum at 0 K is not null and the geometry does not correspond to the equilibrium position either. These are only the expectation values of two Gaussian distributions, which correspond thus to a residual energy, namely the zero-point energy (ZPE), according to the uncertainty principle. The ZPE is used in the microcanonical sampling to generate a Gaussian distribution of initial geometries within the boundaries defined by the turning points that correspond to the ZPE. The typical procedure uses a given reference geometry (q_0), usually the GS minimum or a given point of interest on the surface, whose ZPE is distributed into potential (V) and kinetic energies (T), see Figure 20. The reference geometry is randomly distorted leading to a certain position on the PES, $q_0 + \Delta q_0$, with its corresponding V. In a second step the remaining energy (ZPE - V) is randomly assigned to the different atoms, leading to a random distribution of velocities for the different atoms.
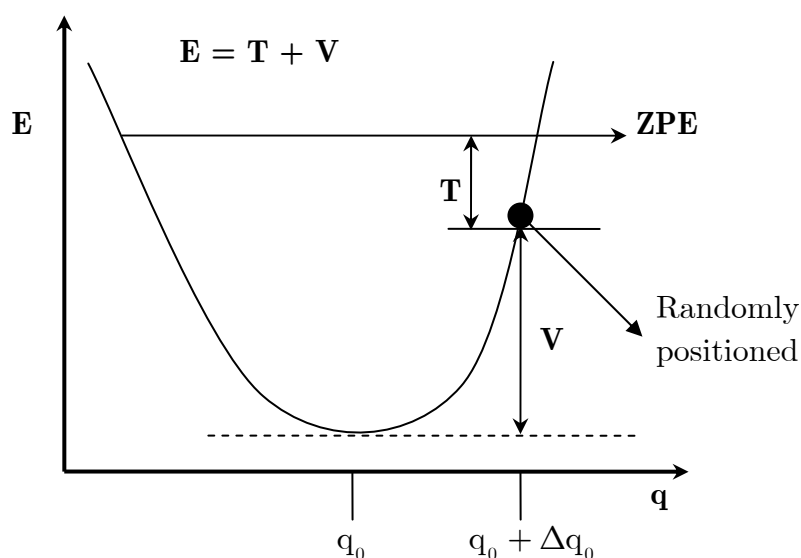
Figure 20. Typical microcanonical sampling procedure. Adapted from Ref. 49.

The sampling procedure also takes into account the possibility of distributing or adding kinetic energy in a pre-determined way. This scheme is widely used in mechanistic studies as one can either decide within which modes T will be distributed or to which mode(s) a certain amount of T will be added. This is very useful in cases where dynamics are started at TSs because this allows us to define which direction will be explored. In addition, there is no possible sampling along the transition mode (negative curvature, hence no contribution to the ZPE), so that the energy added along this mode must be given explicitly. Otherwise, most of the computational effort might be lost in trajectories evolving towards undesired directions. As we will see in this thesis, such a scheme can also be used to study the shape of a CI by exciting those modes which lead to a determined part of the CI, which in other cases can be used to carry out relaxation direction control. Obviously, in these cases the time constants lose significance but one gets valuable mechanistic data.

Once the initial conditions are defined, the molecular dynamics procedure reduces to the propagation of the system for each of the initial conditions until it gets trapped around a stationary point or the region of interest is no longer accessible.

2.2.2   Quasi-classical dynamics propagation

In the classical limit of the Schrödinger equation, the evolution of the nuclear wavepacket can be written as an ensemble of pseudo-particles evolving classically,

$$M\ddot{\boldsymbol{R}} = -\nabla V \qquad (2.83)$$

where M is a diagonal matrix containing the masses of the nuclei and $\ddot{\boldsymbol{R}}$ is the second derivatives vector (acceleration). The components of this vector are defined by R_α. ∇ is the derivative vector operator with respect to the nuclear coordinates,

$$\nabla_\alpha = \frac{\partial}{\partial R_\alpha} \qquad (2.84)$$

and V is the potential which depends on the position of the nuclei, $\boldsymbol{R}$.

The evolution of the pseudo-particles is driven by the force each pseudo-particle "feels" at time t. Thus, the electronic part of the Schrödinger equation needs to be solved at each step in order to provide the gradient (force),

$$F_\alpha(\boldsymbol{R}) = -\nabla_\alpha V(\boldsymbol{R}) = -\nabla_\alpha \left\langle \psi\left(\boldsymbol{r};\boldsymbol{R}\right) \middle| H_{el} \middle| \psi\left(\boldsymbol{r};\boldsymbol{R}\right) \right\rangle, \qquad (2.85)$$

that is used to propagate the nuclei a step forward.

The idea of the propagator is simple but computationally expensive because the electronic part requires the use of the SCF procedure. Thus, given that the electronic part represents the bottleneck of the method, most efforts in providing cheaper molecular dynamics have focused on reducing the number of times the electronic part is solved. A way to reduce the number of quantum calculations is to enlarge the size of the propagation steps. Early molecular dynamics methods used cheap integration schemes[235,236] that allow for small time-steps but only require the evaluation of the first derivatives of the PES and the forces. However, most of the "modern" methods use the Helgaker-Chen algorithm.[237,238] This algorithm is based on the approximation of the nearby PES at each point by a harmonic model. Its trivial integration provides the trust radius which is used by the equations of motion to take a step forward from R_0 to R_t'. In a second step the "true" information on R_t' is used to fit a fifth-order polynomial that is used to improve the surface and carry out a corrector step that leads to R_t. Despite the necessity of calculating second derivatives, the overall algorithm is cheaper than the first-order one because larger steps are allowed.

Another improvement on the efficiency of molecular dynamics is the Hessian updating algorithm,[239] which performs an updating of the Hessian matrix rather

than Hessian evaluation at each step. This method allows for much faster dynamics and critical points optimization algorithms. Thus, the Hessian updating algorithm allows the study of larger molecules and the use of more accurate quantum mechanics methodologies in dynamics calculations. More information on these models can be found in Refs. 239-241.

2.2.2.1   Surface hopping

The way molecular classical dynamics simulates the propagation of a wavepacket on a single PES has been overviewed. Quasi-classical dynamics are based on the same algorithm, but allow the wavepacket to be propagated on several coupled PES through an approximate treatment of the non-adiabatic couplings. Given that many quantum methods can be used to solve the electronic part of the Schrödinger equation, the study of multiple surfaces with quasi-classical dynamics can be carried out at different levels of theory, namely CC, TD-DFT, or CASSCF. None of this methods is capable of treating non-adiabatic events, thus they are approached by a surface hopping algorithm.

Trajectory surface hopping algorithms, first introduced by Bjerre and Nikitin[242] and Tully and Preston,[234] determine what adiabatic state is to be used to compute the energy and gradient at touching surface regions. One of the first surface hopping algorithm was introduced by Tully,[234,243] however, an algorithm[244] which uses the occupancy or population of the adiabatic states to determine the hop probability has been used in this thesis and is explained next. The occupation is given by projection of the time-dependent wavefunction for the electronic motion (time-dependent configuration interaction vector),

$$\mathrm{a}(t) = \begin{pmatrix} \mathrm{C}_1(t) \\ \vdots \\ \mathrm{C}_\mathrm{k}(t) \\ \vdots \end{pmatrix}, \tag{2.86}$$

on the adiabatic states, $\left\langle \mathrm{a}(t) \middle| \Psi^k \right\rangle$, with $\mathrm{C_k}$ being the complex coefficient which gives the contribution of the different k states. This vector can be propagated in time by solving

$$\boldsymbol{a}(t+\tau) = \exp(-\mathrm{i}\mathrm{H}\tau)\boldsymbol{a}(t)\,, \tag{2.87}$$

where H is the electronic Hamiltonian in the state function basis, and where $\hbar$ has been set to 1. The squares of the vector elements give the occupation of the electronic states as a function of time. The occupation of each state is monitored along the trajectory, and when the occupation of the starting adiabatic state is found to be below a user-defined threshold (for example 0.2), a surface hop takes place. That is, beyond that point the other surface is used to compute the energy and gradients of the trajectory. In order to reproduce the non-adiabatic processes better, in a hop the energy is conserved by correcting[245] the momentum in the direction of the derivative coupling vector, $F_{ij} = \left\langle \psi_i \middle| \nabla \psi_j \right\rangle$. To reduce the computational cost, the surface hopping algorithm is not used at each step, but just when the energy difference between the states is lower than a pre-defined threshold. In addition to this, to avoid multiple switches near a CI, no further hops are allowed in the first 5 fs after a given hop.

More information on surface hopping algorithms can be found in Ref. 244, including those which use the Landau-Zener model to evaluate the surface hopping probability.

2.2.3   Quantum dynamics

Quantum molecular dynamics can be carried out in the time-independent picture by diagonalization of the Hamiltonian, or in the time-dependent picture by propagation of a wavepacket. The two options are equivalent but we will focus on the latter as it is seems to be the most natural way of describing the motion of particles and it offers some advantages especially if approximations are used. In the early days of quantum molecular dynamics the time-dependent picture mentioned above was largely ignored. This, started to change with the appearance of the first paper on Gaussian wavepacket propagation by Heller[246] in 1975. New numerical techniques, which allow the numerically exact solution of the time-dependent Schrödinger equation, and powerful integrators such as the split-operator,[247-249] Lanczos[250] or Chebyshev,[251] have been developed since then.

The standard approach for solving the time-dependent Schrödinger equation is the numerically exact propagation of a wavepacket represented in time-independent basis set. The corresponding wavefunction has the form:

$$\Psi\left(Q_1,...,Q_f,t\right) = \sum_{j_1=1}^{N_1}\cdots\sum_{j_f=1}^{N_f} C_{j_1\cdots j_f}(t)\prod_{\kappa=1}^{f}\chi_{j_\kappa}^{(\kappa)}\left(Q_\kappa\right) \tag{2.88}$$

where $Q_1, ..., Q_f$ are the nuclear coordinates, f denotes the number of degrees of freedom, $C_{j_1...j_f}$ are the time-dependent expansion coefficients, $\chi_{j_\kappa}^{(\kappa)}$ are the time-independent basis functions for degree of freedom κ, and N_κ correspond to the number of basis functions.

The equations of motion for the expansion coefficients can be derived from the Dirac-Frenkel variational principle[252,253]

$$\left\langle \delta\Psi \left| H - i\partial_t \right| \Psi \right\rangle = 0 \tag{2.89}$$

where ∂_t denotes the partial derivative with respect to time and yields a system of coupled linear first-order ordinary differential equations, which can be solved by integrators like those mentioned above. However, the computational cost of such calculations becomes prohibitive with the increasing number of degrees of freedom and limits its use to systems of six degrees of freedom.

Fortunately, there are different methods that approximate the solution of the time-dependent Schrödinger equation. Most of these methods are based on approximating the wavefunction as is the case of the Time Dependent Hartree (TDH) method.[252,254] This method is the basis of the one that has been used in the present thesis and is explained next.

2.2.3.1   TDH

The TDH method tries to overcome the limitations that represent the numerically exact solution of the Schrödinger equation by approximating the wavefunction as a Hartree product of one-dimensional functions, the so-called *single-particle functions* or *orbitals*. Thus, the wavefunction is written as:

$$\Psi(x, y, t) = a(t)\varphi_1(x, t)\varphi_2(y, t) \tag{2.90}$$

where a is a time-dependent complex number and φ_1 and φ_2 are the single-particle functions. This wavefunction of is not uniquely defined because phase and normalization factors might be shifted from one term to another. Constraints that fix the phases are defined in differential form as:

$$\left\langle \varphi_1 \left| \dot\varphi_1 \right\rangle = \left\langle \varphi_2 \right| \dot\varphi_2 \right\rangle = 0 \tag{2.91}$$

These constraints also guarantee that the norms of the functions will not change throughout the propagation if they are normalized initially.

The equations of motion that drive the evolution of the wavepacket are derived applying the Dirac-Frenkel variational principle (see eq. (2.89)). Thus, substituting by the wavefunction described in eq. (2.90), the variation with respect to a yields:

$$\left\langle \varphi_1 \varphi_2 \left| i\dot{a}\varphi_1\varphi_2 + ia\dot{\varphi}_1\varphi_2 + ia\varphi_1\dot{\varphi}_2 - Ha\varphi_1\varphi_2 \right\rangle = 0 \right. \tag{2.92}$$

which taking into account the constraints imposed in eq. (2.91), reduces to:

$$i\dot{a} = \left\langle H \right\rangle a \tag{2.93}$$

with, $\left\langle H \right\rangle$ being $\left\langle H \right\rangle = \left\langle \varphi_1\varphi_2 \left| H \right| \varphi_1\varphi_2 \right\rangle$. The corresponding variations of the single-particle functions are:

$$i\dot{\varphi}_1 = \left(H^{(1)} - \left\langle H \right\rangle \right)\varphi_1 \quad \text{and} \quad i\dot{\varphi}_2 = \left(H^{(2)} - \left\langle H \right\rangle \right)\varphi_2 \tag{2.94}$$

with the mean-field operators being $H^{(1)} = \left\langle \varphi_2 \left| H \right| \varphi_2 \right\rangle$ and $H^{(2)} = \left\langle \varphi_1 \left| H \right| \varphi_1 \right\rangle$ respectively.

Writing the three parts together, the TDH wavefunction can be considered as being propagated by an effective Hamiltonian,

$$i\dot{\Psi} = H_{eff}\Psi \tag{2.95}$$

composed of $H_{eff} = H^{(1)} + H^{(2)} + \left\langle H \right\rangle$.

The advantage of the TDH method is that the equations of motion of the wavepacket correspond to a set of coupled one-dimensional functions that compared to the numerically exact propagation case, represent a substantial reduction of the computational effort needed to solve the equations. This allows a larger number of degrees of freedom to be considered. However, with such an approximation the correlation between the degrees of freedom is no longer treated correctly. In addition, similarly to the HF approach in *ab initio* calculations, due to the TDH method only takes into account a single configuration. This implies that this method is not appropriate to study non-adiabatic processes such as the photophysics of thymine. A generalization of this method, namely the Multi-Configurational Time Dependent Hartree (MCTDH),[255] which takes multiple configurations into account and can treat non-adiabatic processes, is explained next.

2.2.3.2   MCTDH

The MCTDH method is a generalization of the TDH one. In the limiting case where only one configuration is used, the equations are identical to those of the TDH. The increase of configurations converges the wavepacket propagation to the numerically exact one (eq. (2.88)). Another particularity of the method is that it can be used to treat multiple configurations. In such a case the wavefunction is expresses as a sum of wavefunctions each one corresponding to a given electronic state. A brief overview of the working equations of method for a single electronic state is given next. The generalization to more states is straightforward and is given afterwards. The full derivation of the following equations has been explained elsewhere.[255] If more technical details are desired, see Ref. 256.

The wavefunction that will be used to obtain the equations to propagate the wavepacket of a system with f degrees of freedom, is written as a linear combination of Hartree products:

$$\Psi(Q_1,...Q_f,t) = \sum_{j_1=1}^{n_1}\cdots\sum_{j_f=1}^{n_f} A_{j_1\cdots j_f}(t)\prod_{\kappa=1}^{f} \varphi_{j\kappa}^{(\kappa)}\left(Q_\kappa,t\right) \tag{2.96}$$

with $Q_1,...,Q_f$ being the nuclear coordinates, $A_{j_1\cdots j_f}$ the MCTDH expansion coefficients, and $\varphi_{j\kappa}^{(\kappa)}$ the n_κ one-dimensional expansion functions (single-particle functions) for each degree of freedom (κ). Note eq. (2.90) is obtained if n_κ is set to 1. It is worth to mention that in this equation not only the expansion coefficients are time-dependent, but also the basis functions. This is one of the reasons of the efficiency of the method, because as the basis evolve with the wavepacket only a few single-particle functions (which are optimized at each time step) are needed. It is also important to note that a different number, n_κ, of single-particle functions can be used for each degree of freedom.

For simplicity, the expansion coefficients will now be expressed by the composite index J, and the Hartree products of single-particle functions as the function Φ_J. Thus, the coefficients and Hartree products will be $A_J = A_{j_1\cdots j_f}$ and $\Phi_J = \prod_{\kappa=1}^{f}\varphi_{j\kappa}^{(\kappa)}$, respectively. This way the expression of the wavefunction reduces to:

$$\Psi(Q_J,t) = \sum_J A_J\phi_J \tag{2.97}$$

As in the TDH case, the MCTDH wavefunction (eq. (2.96)) is not uniquely defined in this form and constraints must be introduced to get rid of the redundancies and guarantee that the single-particle functions remain orthonormal for each degree of freedom (κ) along the propagation. These constraints are:

$$\left\langle \varphi_j^{(\kappa)}\left(0\right)\middle|\varphi_l^{(\kappa)}\left(0\right)\right\rangle = \delta_{jl} \tag{2.98}$$

and

$$\left\langle \varphi_j^{(\kappa)}\left(t\right)\middle|\dot{\varphi}_l^{(\kappa)}\left(0\right)\right\rangle = -i\left\langle \varphi_j^{(\kappa)}\left(t\right)\middle|g^{(\kappa)}\middle|\dot{\varphi}_l^{(\kappa)}\left(0\right)\right\rangle \tag{2.99}$$

where $\dot{\varphi}_l^{(\kappa)}$ is the derivative with respect to time of the l single particle function of the κth degree of freedom, and $g^{(\kappa)}$ is an Hermitian constraint operator. The choice of this operator is arbitrary and does not affect the quality of the wavefunction. One of the advantages of this freedom, is that it can be used to simplify the integration of the equations of motion and increase the performance of the method. It is convenient to choose this operator so that it only acts on the κ-th degrees of freedom, since this allows the Hamiltonian operator to be split up into this hermitian operator, h_κ, and the "rest" of the Hamiltonian, say H_R, which includes the correlation between the degrees of freedom:

$$\hat{H} = \sum_{\kappa=1}^{f} h_\kappa + H_R \tag{2.100}$$

More information on constraint operators can be found in Ref. 39.

In order to derive the MCTDH equations of motion, the *single-hole functions* are introduced. These functions, $\Psi_l^{(\kappa)}$, correspond to a linear combination of Hartree products of all the single-particle functions except that of the coordinate Q_κ. These single-hole functions allow the definition of the mean field operators and density matrices as:

$$\left\langle H_R \right\rangle_{jl}^{(\kappa)} = \left\langle \Psi_j^{(\kappa)}\middle|H_R\middle|\Psi_l^{(\kappa)}\right\rangle \quad \text{and} \quad \rho_{jl}^{(\kappa)} = \left\langle \Psi_j^{(\kappa)}\middle|\Psi_l^{(\kappa)}\right\rangle = \sum_{J}{}^\kappa A_{J_j^\kappa}^* A_{J_j^\kappa} \tag{2.101}$$

Note that diagonalizing the density matrix the natural populations and natural orbitals are obtained. This populations can be used to check the quality of the wavefunction.[256]

Using these definitions and the expressions mentioned above, one can express Ψ, $\dot{\Psi}$, and the variation $\delta\Psi$ as:

$$
\Psi = \sum_J A_J \Phi_J = \sum_{j=1}^{n_\kappa} \varphi_j^{(\kappa)} \Psi_j^{(\kappa)}
$$

$$
\dot{\Psi} = \sum_{\kappa=1}^{f} \sum_{j=1}^{n_\kappa} \dot{\varphi}_j^{(\kappa)} \Psi_j^{(\kappa)} + \sum_J \dot{A}_J \Phi_J \qquad (2.102)
$$

$$
\frac{\delta\Psi}{\delta A_J} = \Phi_J \quad \text{and} \quad \frac{\delta\Psi}{\delta\varphi_j^{(\kappa)}} = \Psi_j^{(\kappa)}
$$

Applying the variational principle (eq. (2.89)) and the constraints stated above, the variation of the coefficients yields:

$$
i\dot{A}_J = \sum_L \left\langle \Phi_J \left| H_R \right| \Phi_L \right\rangle A_L \qquad (2.103)
$$

where $h_{jl}^{(\kappa)} = \left\langle \varphi_j^{(\kappa)} \left| h^{(\kappa)} \right| \varphi_l^{(\kappa)} \right\rangle$ and $A_J = A_{j_1 \ldots j_f}$

The corresponding variation with respect to the single-particle functions after some algebra leads to the final equations of motion that follow:

$$
i\dot{A}_J = \sum_L \left\langle \Phi_J \left| H_R \right| \Phi_L \right\rangle A_L
$$

$$
i\dot{\varphi}^{(\kappa)} = \left[h^{(\kappa)} \mathbf{1}_{n_\kappa} \boldsymbol{\varphi}^{(\kappa)} + \left(1 - P^{(\kappa)}\right)\left(\boldsymbol{\rho}^{(\kappa)}\right)^{-1} \left\langle \boldsymbol{H}_R \right\rangle^{(\kappa)} \right] \boldsymbol{\varphi}^{(\kappa)} \qquad (2.104)
$$

The single-particle functions have been expressed as a vector, $\boldsymbol{\varphi}^{(\kappa)} = \left(\varphi_1^{(\kappa)}, \ldots, \varphi_{n_\kappa}^{(\kappa)}\right)^T$, the density matrix as $\boldsymbol{\rho}^{(\kappa)}$, the mean field operator matrix as $\left\langle \boldsymbol{H}_R \right\rangle^{(\kappa)}$, and the unit matrix as $\mathbf{1}_{n_\kappa}$. The projector on the space spanned by the single-particle functions has the form $P^{(\kappa)} = \sum_{j=1}^{n_\kappa} \left| \varphi_j^{(\kappa)} \right\rangle \left\langle \varphi_j^{(\kappa)} \right|$. One of the advantages of these equations is that the matrix elements of $\left\langle \boldsymbol{H}_R \right\rangle^{(\kappa)}$ and the mean-fields $\left\langle \Phi_J \left| H_R \right| \Phi_L \right\rangle$ are evaluated using only the residual part of the Hamiltonian (see eq. (2.100)).

The equations of motion explained so far consider that the nuclei move on a single PES. This picture is valid for many systems but not for the study of photophysics, in which the nuclei can move on two or more PES. Two possibilities can be used to treat non-adiabatic systems with the MCTDH

method. The first is the so-called *single-set formulation*, in which all the electronic states use the same set of single-particle functions. In such a formulation, an extra degree of freedom is added to the equations to represent the electronic state. One advantage of this formulation is that the equations of motion derived before remain unchanged, and that only an extra degree of freedom (κ_e) which runs up to the number of states (σ), *i.e.* $Q_{\kappa_e} = 1, 2, ..., \sigma$, is needed. accordingly, the number of single particle functions for such an electronic mode is set to the number of states, $n_\kappa = \sigma$. In contrast, in the second formulation, namely *multi-set formulation*, each state uses a different set of single-particle functions as the wavefunction and the Hamiltonian are expanded in the set of electronic states, $\left|\alpha\right\rangle$, as follows:

$$
\begin{aligned}
\left|\Psi\right\rangle &= \sum_{\alpha=1}^{\sigma} \Psi^{(\alpha)} \left|\alpha\right\rangle \\
H &= \sum_{\alpha,\beta=1}^{\sigma} \left|\alpha\right\rangle H^{(\alpha\beta)} \left\langle\beta\right|
\end{aligned}
\tag{2.105}
$$

where each state function has the form of eq. (2.96).

With this formulation the equations of motion need to include an extra label which denotes the state under consideration. Thus, in the multi-set formulation, if the simplest constraint operator is chosen so that $g^{(\alpha,\kappa)} = 0$, the equations of motion read as:

$$
\begin{aligned}
i\dot{A}_J^{(\alpha)} &= \sum_{\beta=1}^{\sigma} \sum_L \left\langle \Phi_J^{(\alpha)} \left| H^{(\alpha\beta)} \right| \Phi_L^{(\beta)} \right\rangle A_L^{(\beta)} \\
i\dot{\varphi}^{(\alpha,\kappa)} &= \left(1 - P^{(\alpha,\kappa)}\right)\left(\boldsymbol{\rho}^{(\alpha,\kappa)}\right)^{-1} \sum_{\beta=1}^{\sigma} \left\langle \boldsymbol{H} \right\rangle^{(\alpha\beta,\kappa)} \boldsymbol{\varphi}^{(\beta,\kappa)}
\end{aligned}
\tag{2.106}
$$

More detailed information on the derivation of these expressions can be found in Ref. 257.

The MCTDH is a method for propagating wavepackets on a PES in which the PES must be known globally. A big effort is needed to generate such a grid which describes the entire PES of a system. Thus, the difficulty of representing the PES is a major problem of this method. However, there are approximations which allow the use of quantum dynamics methods for the study of large systems. One of this approximations was first introduced by Heller[246] and is based on the use of a set of Gaussian functions to describe the wavepacket. The use of Gaussian functions implies that not all the PES must be known in order

to propagate the wavepacket, but just the regions where the functions move. Since Gaussian functions stay localized around their centres, it is thus sufficient to calculate a the PES and Hessian locally on-the-fly around the centre of each function at each time step to propagate the functions. Such a scheme is given by the direct dynamics variation multiconfiguration Gaussian wavepacket method (DD-vMCG),[258,259] which is explained next.

2.2.3.2.a   DD-vMCG

This method uses a superposition of Gaussian functions to approximate the wavepacket. The use of a Gaussian wavepacket is advantageous in many senses. Besides of allowing the use of on-the-fly techniques to locally calculate the PES (direct dynamics), it simplifies the solution of the equations of motion and gives a classical flavor to the method since, according to Ehrenfest's theorem, the center of a single Gaussian wavepacket moves classically. This classical picture can help in interpreting the results. It is worth to mention that Gaussian functions do not move classically but are coupled to each other, unless a single multidimensional Gaussian is used. This non classic behavior allows the description of quantum effects such as tunneling.[260]

The DD-vMCG is faster than MCTDH and needs less Gaussian functions than similar approaches such as the multiple spawning method[135] to converge to the "full" quantum dynamic picture of nuclear motion. The main features of this method are presented next.

The vMCG wavefunction reads as:

$$\Psi_{s}(x,t) = \sum_{j} A_{j}^{(s)}(t) g_{j}^{(s)}(x,t) \qquad (2.107)$$

As it can be readily seen, the single-particle functions of all degrees of freedom (see MCTDH) have been replaced by 1D Gaussian functions. Thus, with this method the wavepacket is represented as a superposition of normalized Gaussian functions ($g_{j}(x,t)$), which are defined by two real parameters, namely the mean position $x_{j}(t)$ and mean momentum $p_{j}(t)$ of the centre of the function.

$$g_{I}(\boldsymbol{q}_{I},t) = \exp\left(\sum_{rs} q_{r} a_{Irs}(t) q_{s} + \sum_{r} \xi_{Ir}(t) q_{r} \eta_{I}(t)\right) \qquad (2.108)$$

where $\mathbf{a}$, $\boldsymbol{\xi}$, and η are time-dependent complex parameters, which are contained in the matrix $\boldsymbol{\Lambda}$. Combining the degrees of freedom in Gaussian

functions reduces the effort required to propagate the wavefunction because the number of configurations scales exponentially with the number of degrees of freedom, and Gaussian functions scale much better. This implies that the variational evolution of each Gaussian is not carried out in terms of an expansion of a set of stationary basis functions but is performed by optimization of a rather small number of time-dependent parameters.

The equations of motion of the vMCG method are derived in the same way as those of the MCTDH one, but in this case only variations with respect to the expansion coefficients (A_j) and the Gaussian function parameters are considered. The derivation of these expressions[258,259] will not be explained here as it is out of the scope of the present thesis. However, the final form of the equations of the DD-vMCG method, which correspond to a set of coupled dynamical equations, is shown next since they have been used in the present thesis. They are presented in matrix nomenclature[259] and read as:

$$i\hbar\dot{\boldsymbol{A}} = \boldsymbol{S}^{-1}\Big[\boldsymbol{H} - i\hbar\boldsymbol{\tau}\Big]\boldsymbol{A}$$
$$i\hbar\dot{\boldsymbol{\Lambda}} = \boldsymbol{C}^{-1}\boldsymbol{Y} \tag{2.109}$$

where $\boldsymbol{H}$ is the Hamiltonian matrix in the Gaussian wavepacket basis set, and $\boldsymbol{C}$, $\boldsymbol{Y}$ and $\boldsymbol{\tau}$ are the matrices and vectors involving overlap matrices and higher moments of the Hamiltonian.

As mentioned above, the equations of motion require the evaluation of the potential energy surface. In this case, the Local Harmonic Approximation (LHA) has been used locally, where the potential (V) is expanded to second order around the center (x_0) of each Gaussian function (i) as it moves:

$$V_i(\boldsymbol{x}) = V(\boldsymbol{x}_0) + \boldsymbol{q}_i^T\boldsymbol{V'} + \boldsymbol{q}_i^T\boldsymbol{V''}\boldsymbol{q}_i \tag{2.110}$$

The first ($\boldsymbol{V'}$) and second derivatives ($\boldsymbol{V''}$), that are needed for the expansion, are calculated on-the-fly by an interfaced quantum chemistry program, which in our case is Gaussian 03.

The procedure of the vMCG method in a two state system is explained next. A wavepacket is generated in each state. The initial wavepacket on the upper state is the GS vibrational function. That is, a Gaussian function centered at the starting position, usually FC, the width of which is determined by the fundamental frequency of vibration. A number of Gaussian functions of equal width are then generated around this initial function. They are positioned so as to keep the overlap between them below 0.5. A similar wavepacket is generated on the lower state. At time 0, the wavefunction is only composed of

the initial function, *i.e.* the Franck-Condon wavepacket, because the expansion coefficients of the rest of the Gaussians functions are set to 0, but the initial one which is set to 1. The positions and occupancies of the functions are self-adjusted by means of the equations of motion as the propagation goes on. The wavepackets describe quantum mechanical trajectories except when only one function is used to approximate the wavepacket, since then, according to Ehrenfest's theorem, the center of the function describes a classical trajectory. As mentioned above, the functions move on PES which are approximated by the LHA at each step. However, the vMCG software includes a database which stores the position, energy, and first and second derivatives of each function at each step. This database is used in successive calculations. The algorithm, before performing the Hessian calculation, which is the time consuming part of the method, first checks the database for previous information on that point. If that part of the PES is present in the database, it is not calculated and the information is taken from the database. This feature reduces the computational effort of the method and makes the simulation faster.

2.2.4 Non-adiabatic events with quantum dynamics

The time dependent Schrödinger equation within the Born-Oppenheimer approximation reads as:

$$\left[T_{\mathrm{N}} + \boldsymbol{E}_e(\boldsymbol{R})\right]\Psi_{\mathrm{N}}(\boldsymbol{R},t) = \mathrm{i}\hbar\frac{\partial}{\partial t}\Psi_{\mathrm{N}}(\boldsymbol{R},t) \tag{2.111}$$

where $\boldsymbol{E}_e(\boldsymbol{R})$ is the matrix of eigenvalues of the electronic Hamiltonian, *i.e.* the adiabatic PES (the corresponding eigenvectors being the adiabatic states), and T_{N} the kinetic energy operator. In this representation, the non-adiabatic terms of the kinetic energy operator are neglected because it is assumed that the different states are not coupled with the variation of the nuclear coordinates. However, in regions of degeneracy, these terms cannot be neglected, and the Schrödinger equation that has to be used is:

$$\left[T_{\mathrm{N}} + \boldsymbol{E}_e(\boldsymbol{R}) + \boldsymbol{\Lambda}\right]\Psi_{\mathrm{N}}(\boldsymbol{R},t) = \mathrm{i}\hbar\frac{\partial}{\partial t}\Psi_{\mathrm{N}}(\boldsymbol{R},t) \tag{2.112}$$

where $\boldsymbol{\Lambda}$ corresponds to the non-adiabatic coupling operator. In this case the Hamiltonian is:

$$H(\boldsymbol{R}) = \hat{T}_N\mathbf{I} \text{ - } \boldsymbol{\Lambda}(\boldsymbol{R}) + \boldsymbol{E}_e(\boldsymbol{R}) \tag{2.113}$$

where $\mathbf{I}$ is the unit matrix. $\boldsymbol{\Lambda}$ is related to the non-adiabatic coupling vector, F, whose components are given by:

$$\hat{F}_{ij} = \frac{1}{V_j - V_i}\left\langle \psi_i \left| \frac{\partial \hat{H}_e}{\partial \boldsymbol{R}} \right| \psi_j \right\rangle \tag{2.114}$$

where ψ_i and ψ_j are the electronic functions of states i and j, respectively, and V_i and V_j the adiabatic energies of these states. It is readily seen that in regions of degeneracy, the terms of this vector will tend to infinity, leading to singularities on the PES such as CI. To avoid this problem, an alternative scheme, which is capable of dealing with degenerate surfaces, namely the diabatic representation, is usually used in quantum dynamics to treat non-adiabatic events. This representation has been used in this thesis and is explained next.

2.2.4.1   Diabatic representation

As it stands, the diabatic representation uses diabatic PES to propagate the wavepackets. These surfaces are smooth since the singularities that arise in the adiabatic picture, when two surfaces approach each other, are not present. This is accomplished by choosing wavefunctions so that the non-adiabatic coupling operator matrix ($\boldsymbol{\Lambda}$) of eq. (2.113) almost vanishes. Nevertheless, the couplings that determine the population transfers are still present, but now are represented by the off-diagonal terms of a potential ($\boldsymbol{W}$) in the electronic Hamiltonian. Thus, in the diabatic representation, the nuclear Hamiltonian is written in matrix form as:

$$H(\boldsymbol{R}) = \hat{T}_N\mathbf{I} + \boldsymbol{W}(\boldsymbol{R}) \tag{2.115}$$

The eigenfunctions of $\boldsymbol{W}(\boldsymbol{R})$ correspond to the adiabatic states, and as mentioned above, the coupling between these states is provided by the off-diagonal elements of this matrix.

The diabatic and adiabatic (eq. (2.113)) representations are equivalent and are related by unitary transformations ($\boldsymbol{S}(\boldsymbol{Q})$) such that:

$$\boldsymbol{E}_e(\boldsymbol{Q}) = \boldsymbol{S}(\boldsymbol{Q})^\dagger \boldsymbol{W}(\boldsymbol{Q})\boldsymbol{S}(\boldsymbol{Q}) \quad \text{and} \quad \phi(\boldsymbol{R}) = \boldsymbol{S}(\boldsymbol{R})\psi(\boldsymbol{R}) \tag{2.116}$$

where $\boldsymbol{S}(\boldsymbol{R}) = \begin{pmatrix} \cos\alpha(\boldsymbol{R}) & \sin\alpha(\boldsymbol{R}) \\ -\sin\alpha(\boldsymbol{R}) & \cos\alpha(\boldsymbol{R}) \end{pmatrix}$, and the rotation (mixing) angle is:

$$\alpha(\boldsymbol{R}) = \frac{1}{2}\arctan\frac{2W_{12}}{W_{22}-W_{11}} \tag{2.117}$$

where the terms W_{ij} correspond to the elements of the 2x2 potential energy matrix. In the ideal case, the mixing angle should fulfill the equation:

$$\nabla\alpha(\boldsymbol{R}) = \left\langle \psi_1(\boldsymbol{R}) \middle| \nabla\psi_2(\boldsymbol{R}) \right\rangle \tag{2.118}$$

to make the derivative couplings strictly zero in the diabatic basis, however it hardly happens and only states which approximate this relation are considered.

In the vMCG method, the diabatic representation is combined with on-the-fly dynamics, thus the potential is calculated at every step *ab initio*, as an adiabatic potential. This implies that the transformation (2.112) must be carried out at every step of the trajectory, before the propagation. In our case, the transformation is based on a simple form of the regularized diabatic states method of Köppel.

2.2.4.1.a   Regularized diabatic states

Köppel et al. showed[261-263] that only the linear derivative coupling terms of the Taylor expansion of the energy around a conical intersection are responsible for the singularities of the adiabatic basis. In views of this, the authors proposed a transformation from the adiabatic basis to the quasi-diabatic one by only removing the leading terms of the Taylor expansion, *i.e.* the construction of regularized diabatic states. In such a transformation only the linear coupling terms are removed and the rest are neglected because they are small. Following this scheme, the mixing angle is defined up to first order and the diabatic to adiabatic transformation reads as:

$$\boldsymbol{S}^{(1)\dagger}\boldsymbol{W}^{(1)}\boldsymbol{S}^{(1)} = \begin{pmatrix} \Delta V^{(1)} & 0 \\ 0 & -\Delta V^{(1)} \end{pmatrix} \tag{2.119}$$

where $\boldsymbol{S}^{(1)} = \begin{pmatrix} \cos\alpha^{(1)} & \sin\alpha^{(1)} \\ -\sin\alpha^{(1)} & \cos\alpha^{(1)} \end{pmatrix}$, $\boldsymbol{W}^{(1)} = \begin{pmatrix} kQ_{x_1} & \lambda Q_{x_2} \\ \lambda Q_{x_2} & -kQ_{x_1} \end{pmatrix}$ and $\Delta V^{(1)} = \sqrt{k^2 Q_{x_1}^2 + \lambda^2 Q_{x_2}^2}$ (the dependences on the nuclear positions have been

neglected for simplicity). Q_{x_1} and Q_{x_2} are rectilinear displacements from a given point of degeneracy along the so-called tuning and coupling modes. In general, these modes are chosen to be the gradient difference and interstate coupling vectors, respectively.

If this (first order) transformation is applied to a general adiabatic potential energy matrix, one gets the working equation for the potential energy matrix in the basis of regularized diabatic states. Thus, only the kinetic energy operator, which is diagonal, needs to be added to get the electronic Hamiltonian in the regularized diabatic basis:

$$\boldsymbol{H}^{reg} = \hat{T}_N \mathbf{I} + \boldsymbol{W}^{reg} = \left(\hat{T}_N + \sum\right)\mathbf{I} + \frac{\Delta}{\Delta^{(1)}}\boldsymbol{W}^{(1)} \tag{2.120}$$

where $\Delta = \left(V_1 - V_2\right)/2$ is the half-difference of the adiabatic potential energy surfaces, and $\Delta V^{(1)}$ is the same term of eq. (2.119). It is worth to mention that with this scheme all the quantities needed to define the transformation are obtained from the adiabatic states, *i.e.* the *ab initio* calculations.

In practice, we proceed the following way: the coordinates of a given conical intersection geometry, which defines the transformation, are contained in the vector $\boldsymbol{R}_0$. $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ contain the components of the half-gradient difference and derivative coupling vectors, which are calculated at the reference point. These vectors are used to calculate the elements of $\boldsymbol{W}^{(1)}$ as:

$$\begin{aligned} kQ_{x_1} &= \left(\boldsymbol{R} - \boldsymbol{R}_0\right)\cdot\boldsymbol{x}_1 \\ \lambda Q_{x_2} &= \left(\boldsymbol{R} - \boldsymbol{R}_0\right)\cdot\boldsymbol{x}_2 \end{aligned} \tag{2.121}$$

where $\boldsymbol{R}$ are the coordinates of a given point $((\pi,\pi^*)_{TS}$ in our case, see section 4.4). Finally, to make the adiabatic and quasi-diabatic states coincide, the tuning and coupling modes are rotated to generate new, $\boldsymbol{x}'_1$ and $\boldsymbol{x}'_2$, modes such that Q_{x_2} is zero at the starting point $\boldsymbol{R}$, $\left(\boldsymbol{R} - \boldsymbol{R}_0\right)\cdot\boldsymbol{x}'_2 = 0$.

The main problem of the transformation is the assumption that the basis of diabatic states at the starting point is sufficient to describe the wavefunction along the whole propagation. Thus, the dimension of the matrices of the adiabatic and diabatic potential energies is finite (2x2 in our case), and it is assumed that the diabatic and adiabatic spaces can be truncated to this size. The correspondence between diabatic and adiabatic states is exact at the dynamics starting point. However, the appearance of further states in the

wavefunction during the propagation (*i.e.* a change in the character of the wavefunction) can cause the transformation to fail. The validity of this assumption for thymine will be discussed in the results section.

3 OBJECTIVES

The main objective of the present thesis is to gain further insight into the photophysics of thymine and provide a mechanistic description of the processes involved in the excited state relaxation observed at the picosecond range and below. Such an objective is not straightforward and requires a careful choice of the methodology. Thus, the study has been divided into two parts. The first part is devoted to find a methodology which is capable of describing the ring planarity properly. In addition, a method based on BSSE correction techniques that fixes the pitfalls encountered by Moran *et al.*, is presented and analyzed. Thus, the objectives of this part is:

- Find a proper basis set to describe the photophysics of thymine.

- Determine if nucleobases suffer from pitfalls on the ring planarity description.

- Analyze if BSSE is at the origin of the pitfalls and its local character.

- Fix the pitfalls on the ring planarity description by means of the Counterpoise method.

- Determine the effect of fragments' definition and fragments' multiplicity assignment on the correction of BSSE.

The second part focuses on the study of the photophysics of thymine and the simulation of the excited state decay, where the following objectives are considered:

- Optimize a high level *ab initio* PES to explain the decay components found experimentally.

- Simulate the decay of the excited state of thymine with quasi-classical and quantum dynamics.

- Characterize the region of the surface responsible for the population of the n,π^* state observed experimentally.

- Check the performance of a novel quantum dynamics code, namely direct dynamics vMCG.

4   RESULTS

This chapter is divided into 4 parts. In the first part, the CP method is applied to fix the anomalies in the description of the planarity of benzene and other planar arenes. In the second part, a benchmark of the methods and basis sets that will be used for the subsequent study of the photophysics of thymine is performed. In addition, the study carried out in the first part is extended to DNA and RNA nucleobases. The last two sections focus on the photophysics of thymine. In the third one, the PES of thymine is calculated at a high level of theory, and in a second step, the deactivation of thymine is investigated by means of quasi-classical dynamics simulations. Finally, the influence of the shape of the seam of CI in the relaxation mechanism has been analyzed with quantum dynamics.

4.1   BSSE effects on the planarity of benzene and other planar arenes: A solution to the problem

We show that the pitfalls reported by Moran *et al.*[140] can be interpreted in terms of intramolecular BSSE effects and, as such, the calculations can be corrected using the BSSE correction techniques typically applied to intermolecular interactions, particularly the Counterpoise method. For this purpose, CP-corrected optimizations and subsequent CP-corrected frequency calculations for a number of arenes (see Figure 21) at the HF, B3LYP, MP2 and CISD levels of theory with several basis sets are carried out. It will be shown that, with no exception, correcting for intramolecular BSSE fixes the anomalous behavior of the correlated methods, whereas no significant differences are observed in the single-determinant case.
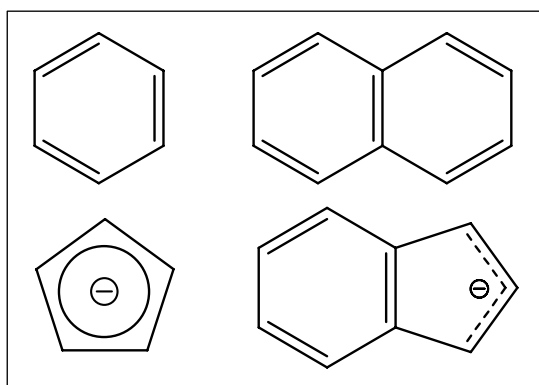
Figure 21. Arenes considered in this study.

4.1.1   Computational details

As detailed in Appendix I, we have used both Gaussian 03 and our own code that allows us to exploit the symmetry of the molecules by minimizing the number of fragment calculations in both the CP-corrected optimization and frequency calculations. This is particularly important for the case of CISD, as in any ghost-orbital calculation, symmetry must be turned off. The case of the cyclopentadienyl anion is a special one because, even though the 6 π electrons are fully delocalized in the ring, in principle, the extra electron must be assigned to one of the five C-H fragments, thus breaking the symmetry of the CP-corrected calculation. Since our program allows us for a flexible definition of the CP correction term, we have determined it using both neutral and negatively charged C-H moieties. The final CP correction term we have used has been simply,

$$\delta^{BSSE}(\vec{R}) = \frac{1}{5}\delta^{BSSE}_{(CH)^-}(\vec{R}) + \frac{4}{5}\delta^{BSSE}_{(CH)}(\vec{R}) \tag{4.1}$$

thus, preserving the symmetry of the calculation. The case of indenyl anion is also problematic. Note that up to *five* imaginary frequencies are obtained at the MP2/6-311G level of theory, the largest one exceeding $1000i$ cm^{-1}. For this system, a Counterpoise function that keeps the C_{2v} symmetry of the molecule and equally shares the electron charge among the three C-H moieties of the five-member ring has been used.

4.1.2   Fragments' definition

A key point when estimating intramolecular BSSE effects is the proper definition of the fragments forming each molecule in order to apply eq. (2.28). A natural and unambiguous way to proceed could be to take atomic fragments. The problem with this approach is that the number of necessary extra calculations to obtain the CP correction would be 2N, being N the number of atoms. It would be desirable to be able to define larger fragments while maintaining the molecular symmetry of the system. In this sense, it is essential to note that in all reported cases the problems are associated with out-of-plane bending low-lying modes. This tells us that in these cases the intramolecular BSSE does not seem to affect appreciably bonds or bond angles (otherwise stretching and other bending modes would have been affected). Taking all this

into account, we can infer that proper moieties that can be used as fragments are the C-H units constituting the arenes.

In order to assess the effect of different fragment definitions, we have first determined the single-point CP-corrected energy profile along the b_{2g} vibrational mode of benzene at the MP2/6-311G level of theory, for which an imaginary frequency of $722i$ cm^{-1} is obtained. The uncorrected and CP-corrected results using twelve atomic fragments, six C-H fragments, as well as three (C-H)$_2$ fragments are displayed in Figure 22.
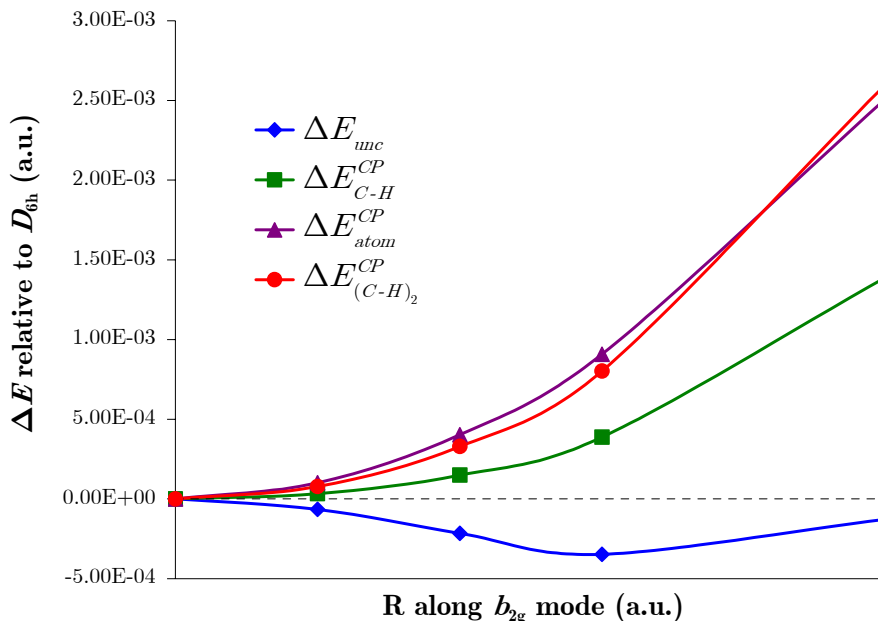
Figure 22. CP-corrected energies along the b_{2g} vibrational mode in benzene.

It can be readily seen that even though the results obviously depend upon the fragments' definition, all CP-corrected energy profiles properly describe the distortion as a vibration, in contrast to the uncorrected profile.

4.1.3   Vibrational frequencies

The lowest vibrational frequencies for benzene, along with the symmetry of the vibrational mode, both uncorrected and CP-corrected, are gathered in Table 1. We have used three basis sets ranging from qualitatively good behavior (6-31+G*) to dramatic BSSE effects (6-311++G), and applied the same levels of theory as in Ref. 140 for comparison.

	MP2		MP2 CP-corrected		CISD		CISD CP-corrected	
6-31+G*	b_{2g}	*182*	e_{2u}	384	e_{2u}	418	e_{2u}	425
	e_{2u}	379	b_{2g}	473	b_{2g}	591	e_{2g}	639
	e_{2g}	618	e_{2g}	613	e_{2g}	643	a_{2u}	713
6-311G	b_{2g}	*722i*	e_{2u}	382	b_{2g}	*189i*	e_{2u}	382
	e_{2u}	333	b_{2g}	442	e_{2u}	397	b_{2g}	442
	a_{2u}	620	e_{2g}	625	e_{2g}	652	e_{2g}	625
6-311++G	b_{2g}	*1852i*	e_{2u}	409	b_{2g}	*1384i*	e_{2u}	455
	e_{2u}	*468i*	e_{2g}	631	e_{2u}	281	e_{2g}	653
	e_{2g}	470	a_{2u}	703	e_{2g}	650	a_{2u}	760

Table 1. Lowest uncorrected and CP-corrected vibrational frequencies (cm^{-1}) of benzene for different levels of theory and basis sets (spurious frequency values in italics).

Noticeably, the CP correction fixes the anomalous behavior of the correlated methods in all cases. Consequently, benzene is found to be planar for all levels of theory. No significant differences are observed for higher frequencies or for the already qualitatively correct single-determinant calculations like Hartree-Fock or B3LYP. Also, the CP-corrected optimized frequencies do not differ appreciably from the uncorrected ones (Tables with all CP-corrected frequencies for all levels of theory are given in Appendix I). The CP-corrected frequencies are also in better agreement with both experimental[264] and best theoretical estimates[149] than the uncorrected ones, even ignoring the lowest five vibrational modes (see Table 2).

	CP-HF		HF		CP-B3LYP		B3LYP	
6-31+G*	7.9	(7.4)	8.6	(7.9)	1.0	(1.0)	1.1	(1.1)
6-311G	9.0	(8.1)	9.6	(8.4)	2.0	(1.9)	2.4	(1.9)
6-311++G	8.2	(7.6)	9.8	(8.6)	2.4	(2.3)	2.4	(1.9)
	CP-MP2		MP2		CP-CISD		CISD	
6-31+G*	4.5	(3.0)	7.2	(3.9)	4.4	(4.4)	5.4	(4.7)
6-311G	5.1	(3.2)	16.1	(5.2)	5.1	(3.2)	10.6	(4.9)
6-311++G	3.7	(3.9)	31.1	(13.2)	9.7	(7.4)	18.7	(20.7)

Table 2. Average error (in %) for the computed harmonic frequencies of benzene with respect to experiment.[264] In parenthesis the error computed without the lowest five frequencies.

A similar behavior is observed in the case of naphthalene, as shown in Table 3. At the modest MP2/6-31G level of theory the out-of-plane b_{2g} vibrational mode is poorly described, exhibiting an imaginary frequency of $402i$ cm^{-1}. The correction of the intramolecular BSSE leads to a value of 570 cm^{-1}, whereas no meaningful differences are observed for the rest of low frequencies.

	MP2 (6-31G)	CP-correct.
b_{3u}	159	159
a_u	177	181
b_{1u}	369	362
b_{1g}	359	368
b_{3u}	419	446
b_{2g}	426	450
b_{2g}	*402i*	570

Table 3. Lowest MP2 uncorrected and CP-corrected vibrational frequencies of naphthalene (cm^{-1}) for different basis sets (spurious frequency values in italics).

Finally, the results for the charged species are gathered in Table 4. It is remarkable that the CP-correction fixes even the pathological case of the indenyl anion at the MP2/6-311G level. The intramolecular BSSE effects are so large in this system that the matching between uncorrected and corrected out-of-plane normal modes is dubious. In the case of the cyclopentadienyl anion, the imaginary frequency associated with a degenerate $e_{2''}$ normal mode also disappears upon CP-correction, whereas large effects on other low-frequency normal modes are also observed. Hence, the involvement of the Counterpoise functions in these systems has proven to be sufficient to handle the extra electron and to keep overall molecular symmetry to cope with the extra electron and keep overall molecular symmetry, have proved to be successful.

For testing purposes, we have also applied to these systems a simpler Counterpoise function ignoring the extra charge (*i.e.*, neutral C-H and C-C moieties). The frequencies obtained (not reported) are again real in all cases, while the differences with the frequencies reported on Table 4 do not exceed 25%. Hence, the charge (and electronic state) of the fragments do not seem to be critical to correct for BSSE for these systems. This had also been observed before in the case of charged intermolecular complexes[164] and it is at the heart of the general success of the Counterpoise method. The BSSE effects are noticeable when the BSIE is strongly geometry-dependent. The latter, affects the description of the occupied molecular orbitals and also the virtual, being the

reason why post-Hartree-Fock methods are more prone to suffering from strong BSSE effects. Therefore, the BSSE is not directly related to the number of electrons of the system. This also explains the fact that hardness profiles, which have been successfully applied to the present problem,[169] are much more robust to level of theory and basis set effects than energy profiles (the chemical hardness is computed as the difference between the vertical ionization potential and electron affinity). Indeed, invoking the Counterpoise philosophy, an energy difference-based measure will perform better than a single energy value.

Indenyl anion			Cyclopentadienyl anion		
MP2/6-311G		**CP-correct.**	**MP2/6-311G**		**CP-correct.**
a_2	*1072i*	331	$e_{2''}$	*401i*	614
b_1	*555i*	1058	$e_{1''}$	284	640
a_2	*458i*	604	$a_{2''}$	523	680
b_1	*200i*	823	$e_{2''}$	632	848
a_2	*125i*	848	$e_{2'}$	860	854
b_1	199	197			
a_2	305	221			
b_2	391	390			

Table 4. Lowest MP2 uncorrected and CP-corrected vibrational frequencies (cm^{-1}) of indenyl and cyclopentadienyl anions (spurious frequency values in italics).

4.2   Global and local BSSE effects on nucleobases

Frequencies on planar minimum geometries of the GS of thymine are calculated for a series of Pople's basis sets at the MP2 and CASSCF levels. In those cases in which imaginary frequencies are observed, the Counterpoise procedure described before is applied. We show that the pitfalls reported for benzene and other planar arenes are common to other planar aromatic rings such as DNA and RNA nucleobases. Similarly to benzene, nucleobases present spurious imaginary frequencies associated with out-of-plane bending modes for planar stationary points obtained at the MP2 level. In this case, however, the study has been limited to four selected Pople's basis sets of general use. We also determine to which extent such basis set deficiencies are localized in a region of the molecule and whether it would be sufficient to correct for intramolecular BSSE only locally, that is, using a specific Counterpoise function that would take into account only a subset of atoms of the system. It is also shown that when the overlap between fragments is strong (*i.e.* when breaking a strong chemical bond) the electronic state of the fragment and that of the ghost-orbital calculation might differ, causing spurious CP corrections. A solution to this problem is presented too.

4.2.1   Computational details

All *ab initio* calculations have been carried out with the Gaussian 03[265] program. Standard CP-corrected geometry optimizations and vibrational frequency calculations have been performed using the automatic procedure as implemented in Gaussian 03 at the MP2 level of theory (frozen core). For special Counterpoise function definitions we have also used our own code, which allows us to exploit symmetry if any, and permits the use of different specific Gaussian keywords for each fragment calculation (with the Counterpoise keyword the process is automatized in Gaussian 03 but all fragment calculations share the same options).
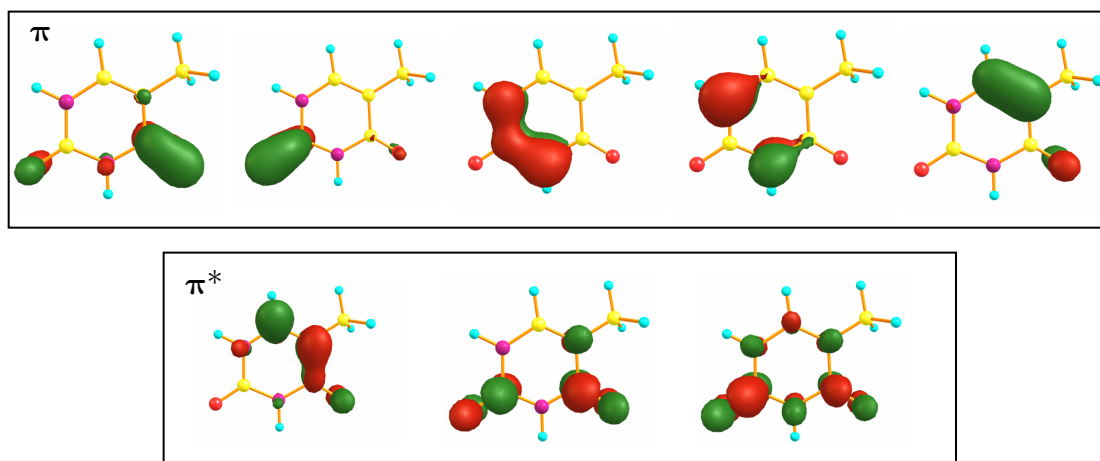
Figure 23. Orbitals used in the CASSCF calculations for thymine.

Thymine and uracil nucleobases were optimized within *Cs* symmetry. No symmetry constraint other than planar ring was used for cytosine, guanine and adenine. An active space including all π orbitals (10 electrons in 8 orbitals) was used for the CASSCF calculations of thymine (see Figure 23).

4.2.2   Thymine benchmark

We have performed geometry optimizations and frequency calculations at the MP2 and CASSCF levels of theory for the same group of basis sets used by Moran *et al.* in the benzene case (over 24 basis sets featuring Pople's 3-21G, 6-31G and 6-311G families and Dunning's cc-pVXZ basis). At the MP2 level we have obtained imaginary frequencies for the planar optimized structures for 12 of the basis sets used (see Table 5). For particularly unbalanced basis sets such as 6-311++G and 6-311+G, up to three imaginary frequencies have been found. The results are slightly worse than in the case of benzene, as for thymine also the MP2/6-31+G* and MP2/6-31+G** lead to spurious results. Again, these spurious imaginary frequencies were found in correspondence to out of plane bending low-lying modes.

3-21G			3-21+G			3-21++G		
CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2
121	113	111	131	117	106	137	117	106
148	131	145	180	155	150	207	170	146
189	164	160	210	177	154	216	173	151
6-31G			**6-31G***			**6-31G****		
CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2
120	110	111	107	108	106	107	107	106
156	145	150	149	138	139	144	136	138
185	151	158	162	148	153	173	145	151
6-31+G			**6-31+G***			**6-31+G****		
CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2
119	*140i*	100	112	*79i*	71	104	*106i*	69
171	93	126	155	100	109	146	97	109
185	124	145	183	140	154	165	138	156
6-31++G			**6-31++G***			**6-31++G****		
CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2
118	*180i*	99	108	*100i*	64	103	*127i*	62
161	90	124	160	99	107	145	95	108
187	124	144	179	136	151	165	135	158
6-311G			**6-311G***			**6-311G****		
CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2
121	97	106	106	103	104	103	100	102
167	128	147	143	139	140	144	134	137
181	137	151	173	146	147	160	151	152
6-311+G			**6-311+G***			**6-311+G****		
CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2
112	*581i*	73	97	*159i*	75	93	*162i*	56
159	*217i*	99	138	84	108	130	81	105
172	*133i*	144	167	138	152	165	147	156
-	172	154	-	157	-	-	153	162

Table 5. Lowest harmonic vibrational frequencies of thymine (cm^{-1}) at the CASSCF, MP2 and Counterpoise-corrected MP2 levels of theory. Basis sets in black indicate benzene is not planar at the corresponding MP2 level. Imaginary frequencies are displayed in italics.

Table 5 Cont.

6-311++G			6-311++G*			6-311++G**		
CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2
112	*628i*	64	108	*185i*	69	102	*180i*	46
163	*243i*	96	160	73	106	143	71	103
171	*185i*	145	179	137	150	160	146	149
cc-pVDZ			cc-pVTZ			aug-cc-pVDZ		
CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2	CASSCF	MP2	CP-MP2
-	110	101	-	111	-	-	106	104
-	147	134	-	150	-	-	139	141

Table 5. Lowest harmonic vibrational frequencies of thymine (cm^{-1}) at the CASSCF, MP2 and Counterpoise-corrected MP2 levels of theory. Basis sets in black indicate benzene is not planar at the corresponding MP2 level. Imaginary frequencies are displayed in italics.

It is remarkable that at the CASSCF level, no imaginary frequencies have been found in any case. This shows that the CASSCF methodology is perfectly valid to study this system, and also that the planarity problems could be associated to two-electron excitations to high energy virtual orbitals with large diffuse character. Dunning basis sets have not been calculated at the CASSCF level because of the high computational cost that represent the optimization and frequency calculations. In addition, we do not expect that those results can provide any relevant data which has not been observed for the other basis, since the MP2 values are already correct for these basis sets. We have also performed single-point calculations at higher levels of theory with the 6-31+G* and 6-311+G* basis sets along the (uncorrected) vibrational mode associated to the imaginary frequency at the MP2 level of theory (see Figure 24). MP3, MP4(DQ) and CCSD produce the correct profile in both cases. The inclusion of the triples in the MP4 energy expression leads to a wrong profile and the inclusion of singles leads to wrong description only for the large basis set.
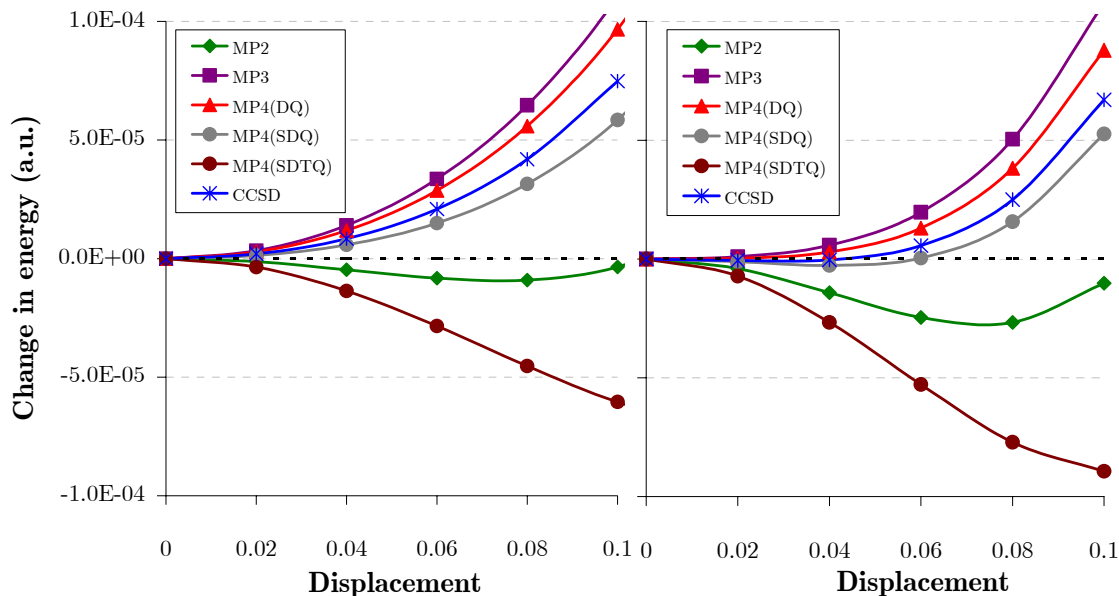
Figure 24. CCSD and MPn energies along the vibrational mode associated to the imaginary frequency for thymine at the MP2 level with the 6-31+G* (left) and 6-311+G* (right) basis sets.

From this data it is difficult to infer that the problem at the MP2 level might be due to a convergence problem of the MPn series. To answer this question a much more systematic study would be required, which is beyond the scope of the present work.

4.2.3   BSSE removal on nucleobases

Motivated by these similarities with the benzene case, we have attempted an analogous approach for the BSSE removal based upon (mainly) diatomic fragments. Due to the heteroatomic character of the nucleobases (see Figure 25) we encounter, aside of C-H moieties, other units such as C=O, N-H or C-CH$_3$ fragments. Lewis structures suggested the use of doublet and singlet multiplicities for C-H and C=O fragments, and numerical evidence recommended to use triplet and quadruplet for N-H and C-CH$_3$ fragments, respectively. The reasons behind this choice will be made clear later on.
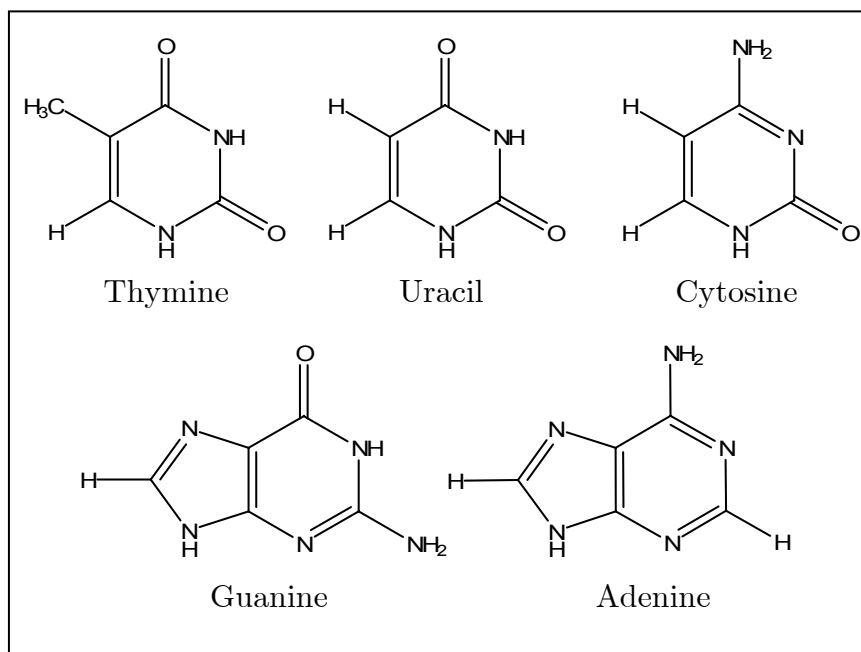
Figure 25. Nucleobases considered in this study

The structures were reoptimized according to the total CP-corrected energy and CP-corrected frequency calculations were performed on the CP-optimized planar stationary structures. In Table 6 we present the three lowest vibrational frequencies for thymine, uracil and guanine obtained for four selected basis set cases. With the abovementioned fragment definition the CP procedure provided excellent results in all cases. The imaginary frequencies were removed in the problematic cases and no significant effect was observed for those which showed proper behavior.

Nucleobases present less symmetry constraints than benzene and other arenes considered in the previous work. Thus, in the present case, one has more freedom to choose proper fragments. In fact, if the basis set deficiencies would be rather localized, one could use a Counterpoise function including only those fragments that would be needed to correct for such deficiencies. Accordingly, we have explored several fragment definitions and Counterpoise functions for this system. Some of our findings are described next.

	Thymine		Uracil		Cytosine	
	MP2	CP-correct.	MP2	CP-correct.	MP2	CP-correct.
6-31G*	107	106	134	135	128	128
	138	139	159	156	203	202
	148	153	371	369	357	358
6-31+G*	*80i*	71	*20i*	86	96	105
	100	109	132	138	190	193
	140	153	315	335	337	349
6-311G*	103	104	137	136	123	124
	139	141	151	152	203	201
	147	151	370	370	360	359
6-311+G*	*160i*	75	*113i*	93	67	108
	84	108	103	137	189	193
	140	151	280	342	309	359

Table 6. CP-corrected and uncorrected frequencies in cm^{-1} of optimized planar structures of pyrimidine nucleobases. Imaginary frequencies are displayed in italics.

First of all, the use of different multiplicity on the N-H fragments has a dramatic effect on the low lying out-of-plane mode. The reason is that, in the singlet case, conventional ghost orbital calculations lead to a qualitatively different state than that of the isolated fragment calculation. The explanation is simple in terms of molecular orbital occupations. In the singlet case the HOMO corresponds to one of the p_x, p_y degenerate orbitals in the isolated fragment calculation. However, in the ghost-orbital calculation, this degeneracy is broken and the in-plane p orbital is stabilized by the presence of ghost-orbitals (mainly of s symmetry) of the neighboring atoms. Energetically speaking there is no apparent problem in the energy difference between the isolated and ghost-orbital calculations. However, if the HOMO in the isolated fragment calculation does not happen to have the same orientation as in the ghost-orbital calculation artificial effects appear beyond energy correction, namely first and second-order derivatives. This can be visualized by comparing the difference between the two densities obtained with and without ghost orbitals at the Hartree-Fock level, as shown in Figure 26. The position of the ghost-atoms is represented by semitransparent blue spheres. The isosurface value in the triplet case is set to 0.0005. Thus, the differences are very small and partially localized in the closest carbon ghost-atoms. Because of BSSE-like basis set extensions, the density is redistributed in the ghost-orbital calculation, increasing in the vicinity of the

closest atoms from which the basis functions are used and slightly decreasing in the region close to the nuclei. However, in the case of the singlet calculation, the density difference between the ghost-orbital calculation and the isolated fragment is much larger (isosurface value is set to 0.005 for clarity) and localized in the N-H unit. The typical polarized picture suggests that the two densities correspond to two rotated electron distributions. The inclusion of such energy (and specially energy derivative) differences in eq. (2.28) leads to an essentially wrong CP-correction, which has no correcting effect on the out-of-plane molecular distortions and introduces spurious effects on the stretching modes associated to the N-H moieties. A similar effect has been observed for the rather unchemical C-CH$_3$ fragment arising from the methyl substituent in the heterocycle. The fragment in the doublet state exhibits a double bond between the carbon atoms whose orientation is again strongly affected by the presence of ghost-orbitals. Such a fragment definition leads to meaningless CP-corrected frequencies. Of course, these problems could be solved simply by proper rotation of the orbitals of isolated fragment calculation, but this might not be easily achieved in automatized procedures such as implied by the use of Counterpoise keyword in Gaussian 03, for instance. This just shows that one must be very careful in these cases when carrying out routine ghost-orbital calculations to quantify basis set extension effects.
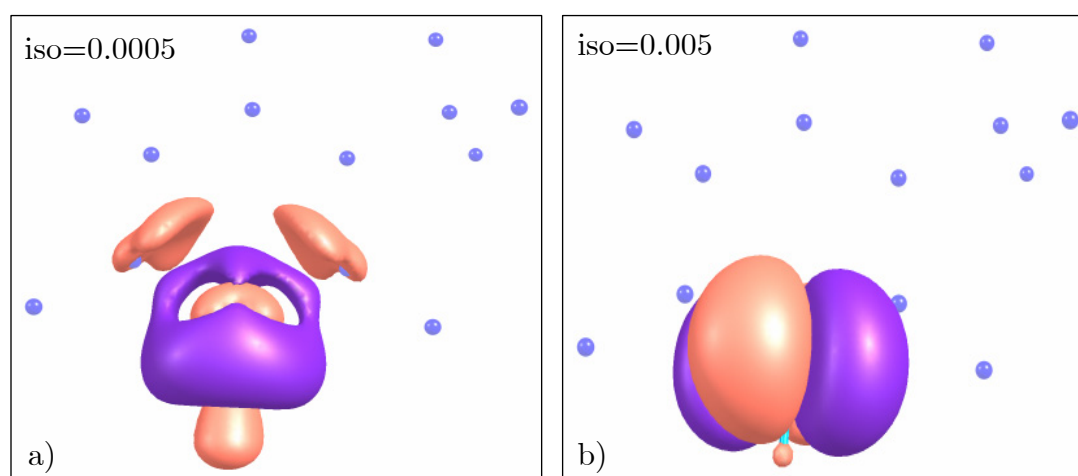
Figure 26. Density difference plot between ghost-orbital and isolated calculations for an N-H fragment in thymine for a) triplet and b) singlet electronic states. The position of the ghost-atoms is shown with semitransparent blue spheres.

We also explored the effect of using multiplicity specification for the C-H and C=O fragments and no noticeable differences were observed. In the case of the C-H fragment one might foresee similar problems associated with the partial occupation of degenerate p orbitals in the low spin case. It seems rather

b)

fortunate that the conventional ghost-orbital calculation lead to a similar orientation of the SOMO orbital.

4.2.3.1   Local BSSE

Another point of interest is to determine to which extent the intramolecular BSSE exhibited by this system is a local effect or not. For this purpose we have considered separately each fragment contribution to the Counterpoise correction and obtained the corresponding CP-corrected frequencies. These results are summarized in Table 7. It is worth to mention that the CP-optimization does not lead to meaningful geometry changes with respect to the conventional MP2 calculation. For instance, the largest deviation on the internal coordinates of thymine at the MP2/6-31+G* level induced upon CP-optimization were just 0.009 Å and 0.3° in bond distance and angles, respectively. That means that one can reasonably obtain frequency corrections with partial Counterpoise functions on the same CP-optimized geometry obtained with the full Counterpoise correction, which largely simplifies the following analysis.

CP-func.	Freq.	CP-func.	Freq	CP-func.	Freq.	CP-func.	Freq.
Unc.	*79i*	1,3	*24i*	1,2,3	38	1,2,3,4	54
1	*65i*	2,3	39	1,3,4	38	1,2,3,5	51
2	*51i*	2,4	31	1,4,5	25	1,2,4,5	41
3	*40i*	3,4	37	2,3,4	51	1,3,4,5	50
4	*53i*	4,5	28	2,3,5	48	2,3,4,5	60
5	*57i*			3,4,5	47	1,2,3,4,5	63

Table 7. Lowest vibrational frequency value in cm^{-1} (Freq.) of various partial CP-corrected calculations. The numbers of the fragments included in the CP-function are defined in Figure 27. The first value corresponds to the uncorrected calculation. Imaginary frequencies are displayed in italics.

At the MP2/6-31+G* level, the lowest lying out-of-plane vibration for planar thymine shows an imaginary frequency of 79*i*. The use of a CP-correction including contributions from just one of the six fragments depicted in Figure 27 does not lead in any case to a change in the topology of the stationary point. The value of the imaginary frequency decreases in all cases, reaching a highest value of 40*i* in the best case, for the N-H fragment in *ortho* position with respect to the C-H group (number 4 in Figure 27). Already when considering two adjacent fragment's contributions at a time one can observe a change in the

topology. Using fragments 4-5 and 3-4 in the Counterpoise function the lowest frequency assumes values of 39 and 37 cm^{-1}, respectively. Nevertheless, a similar value (35 cm^{-1}) is obtained including distant fragments in *para* position like 1 and 4. Other combinations involving the C-H fragment also provide corrections in the proper direction. With the progressive inclusion of more fragment contributions to the CP-correction the value of the lowest frequency increases monotonically up to the reported value of 71 cm^{-1} when using all six fragments. For instance, using the three adjacent fragments 3,4,5 the value is 51 cm^{-1} and including all contributions except that of the C-CH$_3$ fragment, an almost converged value of 63 cm^{-1} is obtained. With this analysis we can conclude that the intramolecular BSSE effects were to some extent localized around the N-H moiety in *ortho* position with respect to the C-H group. However, its removal is not enough to produce a change of topology and contributions from adjacent C-H and C=O groups must also be taken into account. We have also seen that contributions from distant fragments with little chemical significance such as the C-CH$_3$ could be safely ignored from the CP function if necessary. It arises from our results that BSSE effects seem to be quite delocalized on the heterocycle and accordingly in order to get a proper BSSE removal all fragment contributions should be taken into account. Nevertheless, we don't expect this to be a general trend for intramolecular BSSE problems. Further studies on the determination of the local character of BSSE effects in intramolecular hydrogen bonding situations are currently under work.
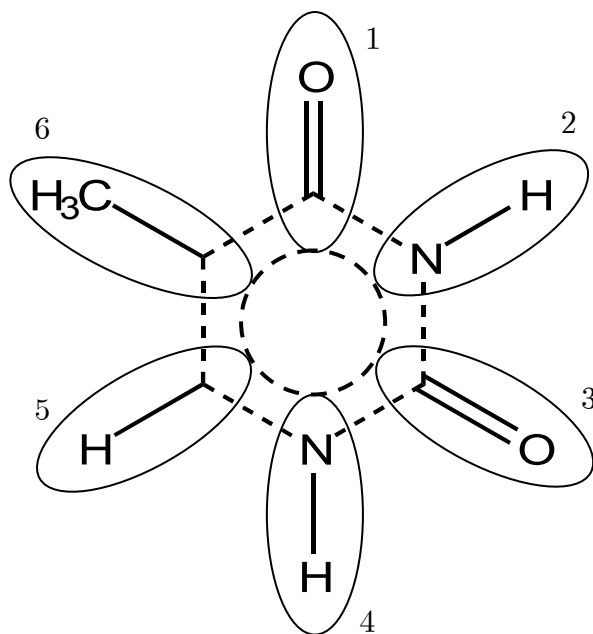
Figure 27. Intramolecular fragments used for the CP-correction in thymine.

After analyzing the thymine case in deep detail, in the following we will describe the results obtained for the rest of nucleobases that we have considered.

4.2.4   BSSE effects on nucleobases

The three lowest vibrational frequencies obtained for all DNA and RNA nucleobases at the MP2 level with four representative basis sets are given in Table 6 and Table 8. The CP-corrected values obtained after proper intramolecular BSSE removal are also provided. Thymine and uracil structures were optimized with *Cs* symmetry, whereas guanine, adenine and cytosine present a planar ring with a NH_2 group slightly out of plane due to pyramidalization, which breaks the symmetry. CP-corrected results obtained suggest that geometrical reoptimizations might not be necessary as only very minor changes are observed between the uncorrected and CP-corrected geometries. Nevertheless, all CP-corrected frequency calculations were carried out upon CP-optimized geometries.

Fortunately, all optimized structures using the 6-31G* and 6-311G* basis sets, which are among the most used basis sets in the literature, were characterized as true minima. The CP-correction (see below) did not change this situation and in fact, the values of the three lowest frequencies were obtained within a deviation of 3% with respect to the corresponding conventional calculation.

When using diffuse functions the intramolecular BSSE effects can be very important. Concerning the CP-correction, pyrimidine derivatives present no special difficulties for a proper fragment definition. They are characterized by a six member ring and six substituents, (except for cytosine, which has one unsubstituted position in the ring). Uracil is very similar to thymine as they only differ by a methyl group, which in the case of uracil is a C-H unit. Accordingly, the results obtained for uracil follow the same tendency as those observed for thymine. Imaginary frequencies associated to the low lying out-of-plane mode are found for the 6-31+G* and 6-311+G* basis sets. When no diffuse functions are included in the basis set, the planar structures correspond to true minima. Again, the CP-correction using the analogous fragment definitions as in the case of thymine is able to account for this pitfall in both cases, with very little effect on the already correct descriptions.

Interestingly, no imaginary frequencies have been found in the case of cytosine, despite his similarity with thymine and especially with uracil (see Figure 25). Yet, again for the 6-31+G* and 6-311+G* basis sets the values of

the lowest vibrational frequencies are somewhat too small compared with the results obtained with more balanced basis sets. Hence, there are some intramolecular BSSE effects but not to the extent of changing the topology of the planar stationary points. In fact, cytosine shares with the other two nucleobases the same three substituents that were observed to contribute more to the BSSE effects in thymine. This shows again that for such heterocyclic systems the BSSE effects are rather subtle and delocalized. For completeness we have performed also CP-corrected optimizations and frequency calculations for this system. The absence of substituent in *ortho* position with respect to the C=O group leads to the difficulty of dealing with a single-atom fragment; highly symmetric in its isolated state but not in the presence of ghost-orbitals. In order to avoid the problems described before we have considered the N atom as a fragment in high spin state. As alternative, we have explored also the use of a larger fragment involving two adjacent positions of the ring, namely NC-NH$_2$. Both fragment definitions lead to very similar results (maximum deviation of 8 cm^{-1} in the third frequency at the MP2/6-31+G* level), but only the results from the first option are reported in Table 8. Even though no imaginary frequencies were observed, the CP-corrected frequencies are more similar among the four basis sets used than the uncorrected ones. For instance, the somewhat too low values of 67 and 310 cm^{-1} obtained with the 6-311+G* basis set are blue-shifted by 40 and 50 cm^{-1} to a value in much better agreement with the one obtained with the 6-311G* basis set.

Finally, the results obtained for adenine and guanine basis sets are collected in Table 8. Adenine seems to be the less prone to intramolecular BSSE effects. No imaginary frequencies have been observed for the planar optimized structures and only a slight drop of about 20-25 cm^{-1} in the value of the lowest vibrational frequency when including diffuse functions has been observed. The situation in the case of guanine is different as again difficulties are observed specially with the 6-31+G* basis set, for which an imaginary frequency of 55i is obtained. The value of 8 cm^{-1} obtained with the 6-311+G* basis set can also be considered as spurious.

Purine bases are characterized by having a heterocyclic six-membered ring fused to an imidazole ring. Therefore, the definition of fragments to account for intramolecular BSSE must be somewhat different to that of the pyrimidine bases discussed above. The Lewis structures for each molecule show the presence of a double bond involving a C-C pair in the edge of the two fused rings which probably should not be broken. Following this premise we ended up with C=O, N-H, C=C diatomic fragments and larger N=C-NH$_2$ and N=CH fragments involving the unsaturated N atoms, which were considered in high spin state. Once again, the results obtained are very satisfactory. For guanine, the two wrong vibrational frequencies obtained with the 6-31+G* and 6-

311+G* basis are efficiently removed with the CP-correction. Even in the case of adenine, where the BSSE effects were less pronounced, the CP-correction induced a slight blue-shift in the lowest vibrational frequencies to final values much closer to those obtained with more balanced basis sets. Several other fragment definitions were also tested, for instance involving a central NC=CN fragment. The CP-corrected results were proved to be very similar, provided situations like those described in detail in the case of thymine were not present.

	Adenine		**Guanine**	
	MP2	**CP-correct.**	**MP2**	**CP-correct.**
6-31G*	159	158	129	127
	207	205	151	152
	272	272	192	194
6-31+G*	126	151	*55i*	116
	185	192	128	133
	275	262	165	178
6-311G*	160	161	131	133
	213	208	156	155
	273	274	195	194
6-311+G*	139	150	8	113
	192	195	127	131
	274	276	164	175

Table 8. CP-corrected and uncorrected frequencies of optimized planar structures of purine nucleobases. Imaginary frequencies are displayed in italics.

4.3   Photophysics of the π,π^*

and n,π^* states of thymine

Based on the results of the previous section which shows that the CASSCF methodology and 6-31G* and 6-311G* basis sets do not suffer from spurious minima optimizations, they will be used in this section to study the photophysics of thymine. First, the relaxation pathways present on the literature are recalled to make the starting scenario clear (see section 1.2.2 for further details) .

4.3.1   Starting scenario

In general, computational studies agree that the relaxation of thymine proceeds via a CI[129-132,138,139,266,267] between the ground state and the lowest π,π^* state (see section 1.2.2). This CI has an ethylene-like structure and was first described for uracil. It is characterized by a short C_4-O_8 and a methyl group lying perpendicular to the molecular plane because of the pyramidalization of C_5. The geometry of this structure is shown in Figure 29 and is labeled, hereafter, $(Eth)_X$. Two paths to access the intersection have been described with CASSCF and CASPT2 calculations. The corresponding one-dimensional energy profiles are shown as insets in Figure 28, together with a two-dimensional sketch of the potential energy surface in the vicinity of the Franck-Condon region which shows the two paths (see also Fig. 8 of Ref. 138). Path 1 is a barrierless, direct path from the Franck-Condon structure to $(Eth)_X$ on the lowest-lying, spectroscopically active π,π^* state.[131] Path 2 is indirect and goes through a non-planar minimum of the π,π^* state, namely $(\pi,\pi^*)_{Min}$. This minimum is separated from $(Eth)_X$ by a barrier and a crossing with the n,π^* state.
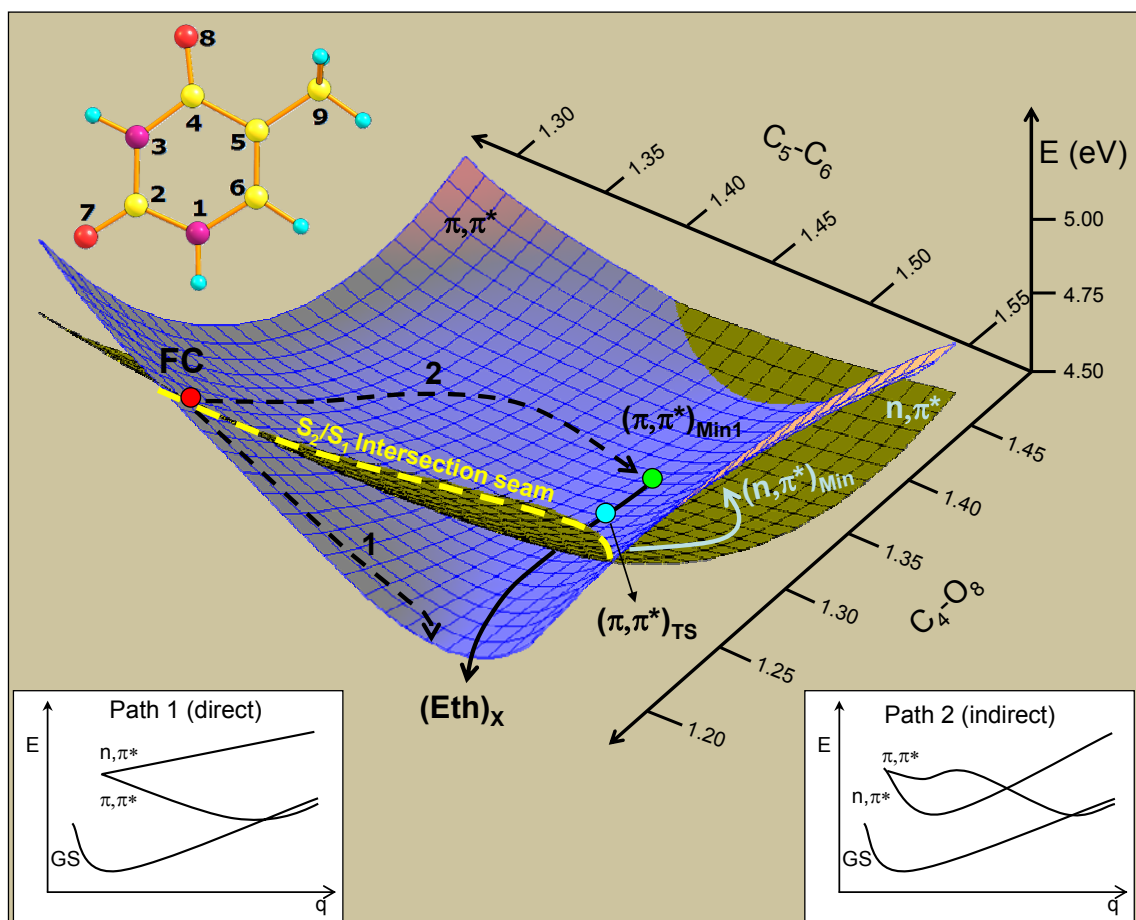
Figure 28. Two-dimensional sketch of the two lowest excited state potential energy surfaces (S_1 and S_2) of thymine in the vicinity of the Franck-Condon region. Insets: FC structure with atom numbering and energy profiles for the paths contained in the two-dimensional sketch.

We have recalculated the direct and indirect paths in a uniform manner, using the MS-CASPT2//CASSCF approach, which is one of the approaches of reference for accurate excited-state calculations. Moreover, we have carried out semi-classical, on-the-fly, dynamics calculations along the indirect path. In the dynamics calculations, the nuclei are propagated classically on a CASSCF potential energy surface, and the passage through the conical intersections is treated with a trajectory surface hopping algorithm (see section 2.2.1). This methodology has been applied before to several photochemical and photophysical problems,[268-271] and a similar implementation at the multireference configuration interaction (MRCI) level has been applied to study the decay of adenine.[272] The starting scenario for our dynamics is the two-dimensional surface of Figure 28. Similar models have been previously proposed in Refs. 138 and 131. For simplicity, the surface is sketched along the C_4-O_8 and C_5-C_6 stretching coordinates (see Figure 28 for the numbering), although the actual paths are more complicated and involve other coordinates that correspond to out-of-plane

modes, such as C_6 pyramidalization and C_5-C_6 bond torsion. The two paths coexist on the surface and merge as they approach $(Eth)_X$. The CASSCF trajectories started at the FC structure in Ref. 138 follow the indirect path and get trapped in $(\pi,\pi^*)_{Min}$ for a time that exceeds the time range of our studies, which is of a few hundreds of femtoseconds. Therefore we have studied the second step of this path (full line in Figure 28) starting our trajectories from the transition state $(\pi,\pi^*)_{TS}$ that separates this minimum from $(Eth)_X$. Because of the limitations of the CASSCF level of theory used for the dynamics, the information that we can obtain from the trajectories is mainly mechanistic. For this reason, and due to the computational cost of the trajectories, we have run a reduced number of trajectories, where we have sampled the main paths encountered after $(\pi,\pi^*)_{TS}$ with 14 trajectories. This combined static and dynamical approach shows that both paths provide a rapid deactivation route to the ground state through the ethylenic intersection, $(Eth)_X$, while the indirect path is also a route to populate the n,π^* state, where the system can get trapped for a longer time.

4.3.2   Computational Details

MS-CASPT2//CASSCF potential energy surface: The static calculations were carried out at the MS-CASPT2//CASSCF level of theory, using the 6-311G* basis set. Optimizations and minimum energy path calculations have been performed at the CASSCF level using Gaussian 03,[265] with no symmetry restrictions. The minimum energy paths were obtained with the intrinsic reaction coordinate[47] and initial relaxation decay techniques[273] in mass weighted Cartesian coordinates. The displacements are given in bohr times square root of atomic mass unit (atomic units, a. u.).

A (12,9) active space consisting of the 8 π orbitals and the O_8 lone pair was used for the optimization of structures $(\pi,\pi^*)_{Min}$, $(n,\pi^*)_{Min}$, $(n,\pi^*/GS)_X$, and $(n,\pi^*/\pi,\pi^*)_X$, and the intrinsic reaction coordinate calculations involving these structures. Equal state average over the two higher states has been used in the optimizations whenever possible. In the cases with 3 states where convergence could not be achieved, the three states were averaged with equal weights. In the state-averaged optimizations using 9 active orbitals, solution of the state-averaged coupled perturbed multi-configurational SCF (SA-CPMCSCF) equations is not feasible, and the orbital rotation contributions to the gradients were neglected.[274] Structure $(Eth)_X$ and the minimum energy path from the FC point to that structure were optimized with a smaller (10,8) active space, where the oxygen lone pair was removed because its occupation was close to 2.0. Structure $(\pi,\pi^*)_{TS}$ was also optimized with a (10,8) active space to solve the SA-CPMCSCF equations and calculate the orbital rotation contributions to the

gradient, thus allowing for an analytical frequency calculation. In this case, the π orbital with highest occupation was removed from the active space. The minimum energy path calculations from this transition structure were also carried out with the (10,8) active space.

MS-CASPT2 single point calculations have been carried out with Molcas5.4.[275] All energies in eV discussed here are relative to the MS-CASPT2 ground state energy (CASSCF(10,8)/6-311G* optimized geometry). The MS-CASPT2 calculations were carried out with a CASSCF(12,9)/6-311G* reference wavefunction over the six lowest roots, because close-lying higher states in some regions of the potential energy surface make the inclusion of six roots necessary to obtain smooth profiles for the lowest π,π^* and n,π^* states. In all cases state-averaging over all states with equal weights and a real level shift[276] parameter of 0.3 were used.
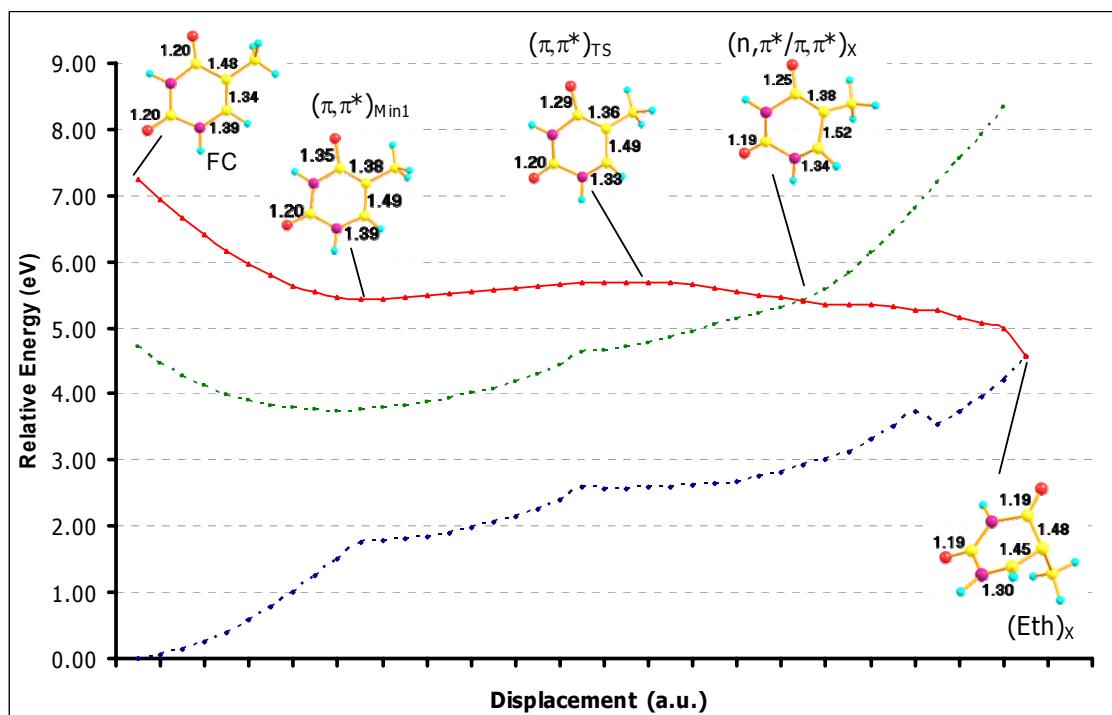
Figure 29. LIIC CASSCF(8,6)/6-31G* of the indirect path (FC-$(\pi,\pi^*)_{Min}$-$(\pi,\pi^*)_{TS}$-$(n,\pi^*/\pi,\pi^*)_X$-$(Eth)_X$)

For a better comparison between the MS-CASPT2 static calculations and the CASSCF dynamics, the critical points were reoptimized at the CASSCF(8,6)/6-31G* level (see below for the active space orbitals). In this case, the connection between the critical points at the CASSCF level was confirmed by linear interpolations in internal coordinates between the structures and the resulting PES is shown in Figure 29 (see Figure 31 for the MS-CASPT2(12,9)/6-311G* and CASSCF(8,6)/6-31G* single-point energies).

CASSCF Semiclassical dynamics: We have carried out semiclassical direct dynamics trajectories[237,238,277] where the nuclear gradient and Hessian are calculated "on-the-fly" in the full space of $3N$-6 coordinates (where N is the number of atoms) and the nuclei are propagated classically. The calculations have been carried out with a development version of Gaussian 03[278] at the CASSCF(8,6)/6-31G* level. The trajectories where started on S_2, and when they approached a region of S_2/S_1 degeneracy (threshold of 8 mhartrees), the CASSCF surface hopping algorithm was activated. In this algorithm, the time-dependent electronic wavefunction is calculated and the surface hopping probabilities are obtained by projecting the time-dependent wavefunction on the CASSCF states. Hops occur when the probabilities for the running state go below a threshold of 0.20 (see the Supporting Information of Ref. 279 for more details). The current implementation of this algorithm only allows for state-averaging between the two highest excited states. For this reason, after the hops from S_2 to S_1, the trajectories were stopped once the S_2-S_1 energy gap went over a threshold of 10 mhartree, and they were restarted on S_1 using the final geometry and velocity of the first run.

The length of the trajectories (approximately 500fs, with approximately 2 steps per fs) and the need to calculate accurate gradients by solving the SA-CPMCSCF equations, forced us to reduce the active space to 8 electrons in 6 orbitals. In the trajectories started on S_2 and those running on the n,π* state (S_1), the active space consisted of the 5 π orbitals approximately localized on the N_1-C_6-C_5-C_4-O_8 fragment and the in-plane O_8 oxygen lone pair. This active space remained stable during those trajectories, and no discontinuities were observed. In the S_1 trajectories on the π,π* state, the occupation of the oxygen lone pair active orbital was 2.0. This orbital was replaced by the N_3 π orbital at the beginning of these trajectories, and the corresponding active space also remained stable.

The initial conditions of the dynamics calculations were obtained by sampling the transition state zero-point energy, which generated a random set of $3N$-7 coordinates with the corresponding kinetic energy distribution.[271,280] To model the approach to a conical intersection and drive the trajectories in that direction, 1 kcal/mol of extra kinetic energy was added to the mode with the imaginary frequency. The trajectories were run with the analytical Hessian for the first few steps, after which a Hessian updating algorithm[239] was used. The updating algorithm failed to conserve the total energy in 11 out of 25 trajectories, which had to be discarded. The remaining 14 trajectories are the ones discussed here. Energy plots for these trajectories are provided in Appendix II.

4.3.3   *High level potential energy surface*

The paths sketched in Figure 28 have been characterized with CASSCF/6-311G* minimum energy path calculations from the FC structure on the $\pi.\pi^*$ state (see Figure 30, which includes the structures of the critical points).
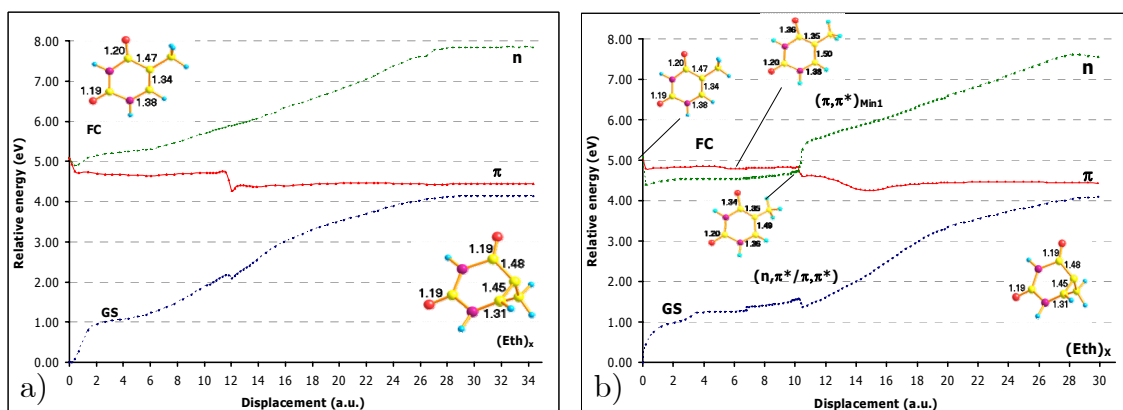
Figure 30. MS-CASPT2(12,9) energy profiles along the CASSCF/6-311G* minimum energy paths from the FC structure: a) Direct path (path 1 on Figure 28); b) Indirect path (path 2 on Figure 28).

Similar to what is described in Ref. 131, the resulting path depends on the CASSCF active space used in the calculation. The calculation with a (10,8) active space leads directly to $(Eth)_X$.[131] Along the first steps of this path, the molecule stays planar and the C_5-C_6 and C_4-O_8 bonds are expanded to 1.5 Å and 1.25 Å, approximately. Later on, the molecule loses its planarity and the C_4-O_8 bond is contracted again to its initial value. Thus, structure $(Eth)_X$ has a short C_4-O_8 and a long C_5-C_6 bond (1.19 Å and 1.45 Å, respectively), and the methyl group is twisted out of the plane of the ring. In contrast to this, the calculation with the (12,9) active space leads to a quasi-planar minimum on the π,π^* state, $(\pi,\pi^*)_{Min}$. The main difference with the first path is that the C_4-O_8 bond is expanded to approximately 1.36 Å, and the methyl group stays in the plane of the molecule. From $(\pi,\pi^*)_{Min}$ we have characterized an indirect path similar to the one reported in Ref. 138. This path continues along a transition structure, $(\pi,\pi^*)_{TS}$, and a conical intersection with the n,π^* state, $(\pi,\pi^*/n,\pi^*)_X$, and ends at $(Eth)_X$. To compare the energies of the two paths, the energy profiles have been recalculated at the MS-CASPT2 level, using a (12,9) active space that includes all π orbitals and the oxygen lone pair involved in the n,π^* excitation. The MS-CASPT2(12,9) energy profiles along the direct and indirect paths are shown in Figure 30a and Figure 30b, respectively. The MS-CASPT2 barriers along the two paths are small (<0.1 eV), and the discontinuities around 10 a.u.

in both profiles are due to degeneracies of the CASSCF reference wavefunction. The two MS-CASPT2 profiles are similar in energy, and from the static calculations it is not clear which of the two paths is preferred.

The possibility of populating the n,π* state along the indirect path and the fate of this state, have been studied by calculating the decay path from $(\pi,\pi^*/n,\pi^*)_X$ on the n,π* state, which leads to a minimum, $(n,\pi^*)_{Min}$. This minimum is characterized by a long C_4-O_8 bond (1.37 Å) and C_6 pyramidalization. From this minimum, we have optimized a decay path through a conical intersection with the ground state, $(n,\pi^*/GS)_X$. This intersection is similar to structure MXS4 from Ref. 267, with the C_4-O_8 bond stretched to 1.44 Å and a distorted ring. The MS-CASPT2 energy is 5.3 eV, which is higher than the vertical excitation of 5.1 eV. In addition to that, the $(n,\pi^*)_{Min}$ population can also decay through $(Eth)_X$ via a switch back to the π,π* state. This path will involve an avoided crossing near $(\pi,\pi^*/n,\pi^*)_X$, and its estimated barrier is the difference between $(n,\pi^*)_{Min}$ and $(\pi,\pi^*/n,\pi^*)_X$, 0.4 eV.
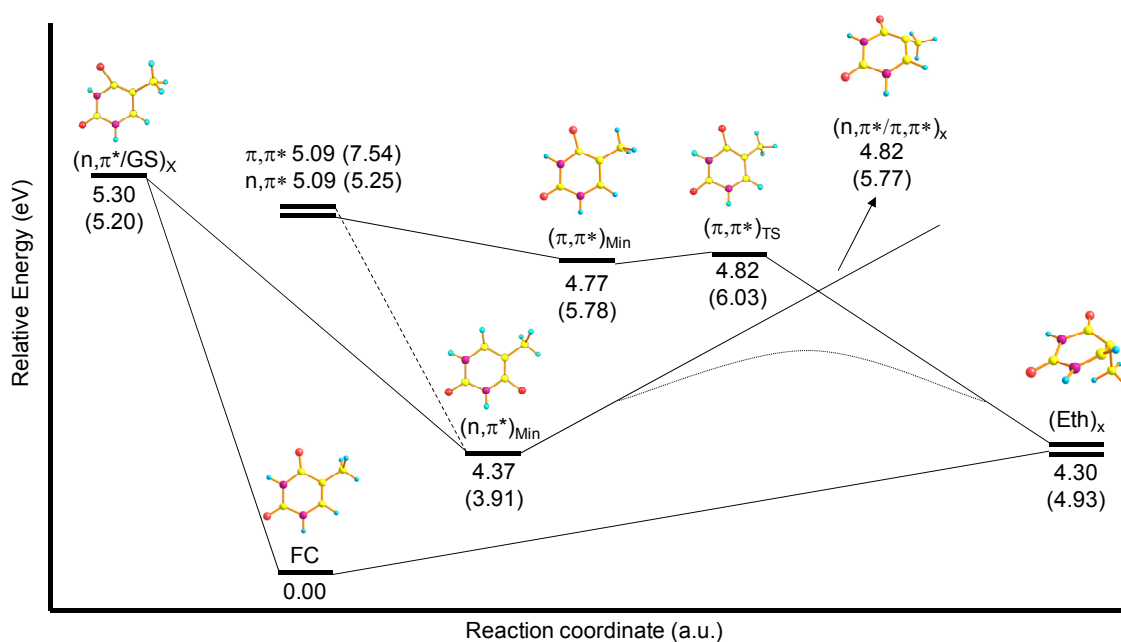
Figure 31. Energies of relevant critical points on the excited state surface of thymine at the MS-CASPT2(12,9)/6-311G* level of theory (CASSCF(8,6)/6-31G* optimized energies in brackets).

The MS-CASPT2 static picture is summarized in Figure 31, and the CASSCF(8,6) energies of the critical points (see Computational Details) are shown in brackets for comparison. There are several differences between the MS-CASPT2 and CASSCF energy profiles. The largest dynamic correlation effect is found along the initial part of the indirect path. The vertical π,π* and n,π* excitation energies are approximately 5.1 eV at the MS-CASPT2 level.

However, the π,π^* excitation energy is increased by more than 2 eV at the CASSCF level, and the CASSCF barrier between $(\pi,\pi^*)_{\text{Min}}$ and $(\text{Eth})_{\text{X}}$ is overestimated by 0.2 eV with respect to the MS-CASPT2 value. Another relevant difference regards the energy of $(n,\pi^*/\text{GS})_{\text{X}}$ relative to the vertical excitation. At the CASSCF level, the intersection lies below the vertical excitation and is accessible (see below), but at the MS-CASPT2 level it lies higher in energy than the FC structure.

4.3.4   Dynamics simulations

The CASSCF trajectories started at the FC structure in Ref. 138 follow the indirect path and get trapped at $(\pi,\pi^*)_{\text{Min}}$ for at least 500 fs. For this reason, the dynamics were started at $(\pi,\pi^*)_{\text{TS}}$ on S_2, following the transition vector direction toward $(n,\pi^*/\pi,\pi^*)_{\text{X}}$ and $(\text{Eth})_{\text{X}}$. $(\pi,\pi^*)_{\text{TS}}$ is characterized by a weak boat-like puckering of the ring (Figure 32) and a slightly pyramidalized methyl group (C_9-C_5-C_4-N_3 dihedral angle: -153.9 °).
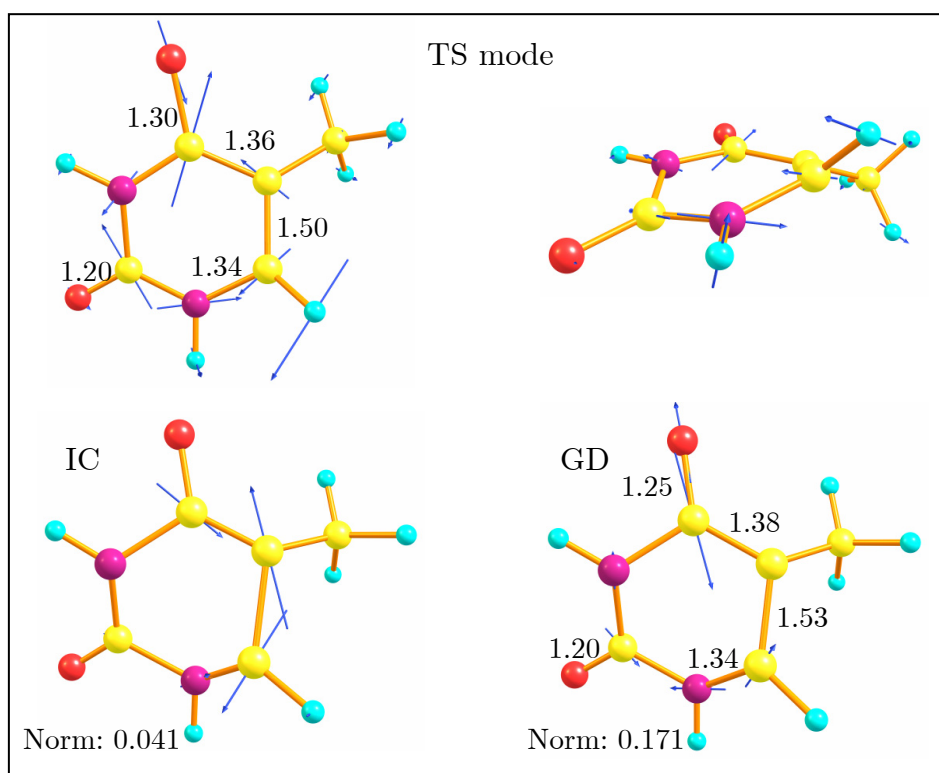
Figure 32. Transition vector at $(\boldsymbol{\pi},\boldsymbol{\pi}^*)_{\text{TS}}$ and branching space vectors (interstate coupling, IC, and gradient difference, GD) at $(n,\boldsymbol{\pi}/\boldsymbol{\pi},\boldsymbol{\pi}^*)_{\text{X}}$, calculated at the CASSCF(8,6)/6-31G* level.

The most relevant bond distances of $(\pi,\pi^*)_{TS}$ have a mixed single-double bond character (C_4-O_8: 1.29 Å, C_4-C_5: 1.36 Å, C_5-C_6: 1.49 Å). The transition vector is mainly compound of C_4-O_8 shortening, C_5-C_6 lengthening and pyramidalization of N_1 and C_6.

Figure 33 shows the evolution of the three lowest excited states of thymine together with the C_4-O_8 distance along time for a representative trajectory. The state followed by the trajectory is shown with a bold red line. We center on the approach to the CI seam along the reaction coordinate. Thus, the gradient difference vector at the seam (Figure 32) is similar to the transition vector (large C_4-O_8 stretch component), and the seam lies approximately perpendicular to the coordinate (see Figure 28). For this reason the seam is reached from the TS in a few fs. The trajectory oscillates back and forth around the seam, and the oscillations are approximately in phase with the C_4-O_8 stretching mode (see the evolution of the C_4-O_8 distance in Figure 33).
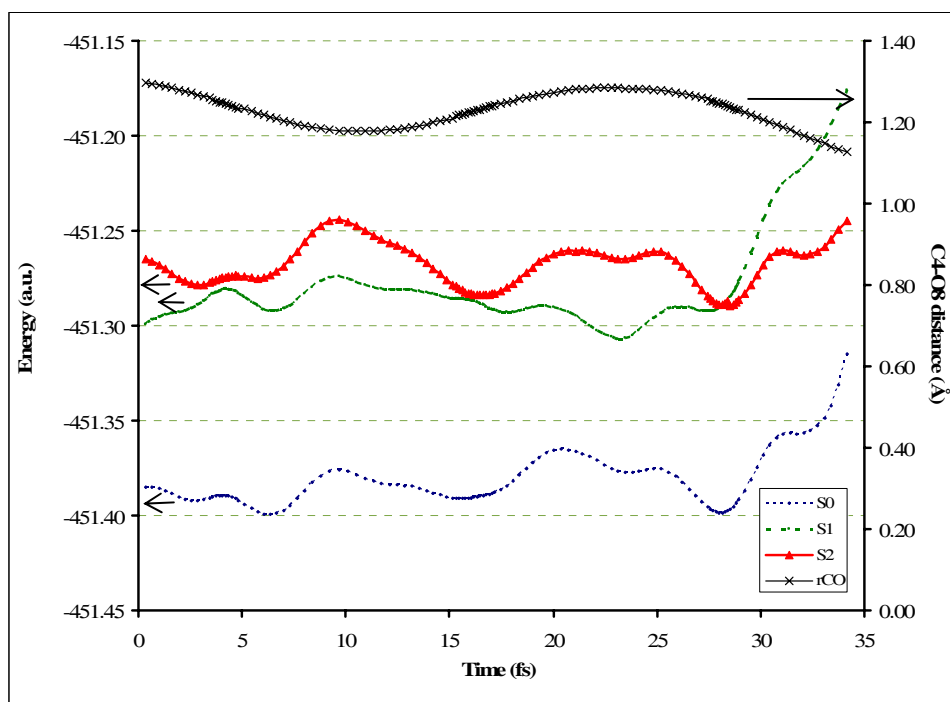
Figure 33. Time evolution of the CASSCF(8,6) energy of the S_0-S_2 states of thymine and the C_4-O_8 distance for a representative trajectory on S_2 from $(\boldsymbol{\pi},\boldsymbol{\pi}^*)_{TS}$. The labels of the states refer to the order at the beginning of the trajectory.

Each oscillation on S_2 is associated with an 'adiabatic switch' between the π,π^* and n,π^* states, until the trajectory hops to S_1. In the example shown in Figure 33, the hop to S_1 takes place at the third approach to the seam, and the trajectory continues on the π,π^* state. Once the trajectories decay to the π,π^*

state, the C_4-O_8 bond is shortened, while this bond is stretched for those trajectories that decay to the n,π^* state. After the hop to S_1, all trajectories have been continued to monitor their subsequent relaxation on the π,π^* or n,π^* state, reducing the number of roots to two (see section 4.3.2). Out of 14 valid trajectories (see section 4.3.2), 9 stay on the π,π^* state after the S_2/S_1 crossing and 5 populate the n,π^* state. A representative trajectory for the decay of the π,π^* state on S_1 is shown in

Figure 34. In this case, the C_4-O_8 bond remains short and oscillates around 1.2 Å, while the C_5-C_6 bond is stretched up to 2 Å. At the same time, the methyl starts to bend out of plane. The C_5-C_6 bending and methyl out of plane coordinates are contained in the branching space vectors at $(Eth)_X$ (see Figure 35) and take the trajectory to the S_1/S_0 seam of intersection centered around this structure. The decay to the ground state takes place 15 fs - 200 fs after the hop from S_2 to S_1 at geometries where the out-of-plane bending angle of the methyl group ranges from 50º to 100º and the C_2-C_3 bond length ranges from 1.2 Å to 2.0 Å. After the hop to S_0, the 'hot' molecule keeps vibrating on the ground state. In the condensed phase (solution or DNA environment) the vibrational excess energy will be dissipated to the environment.
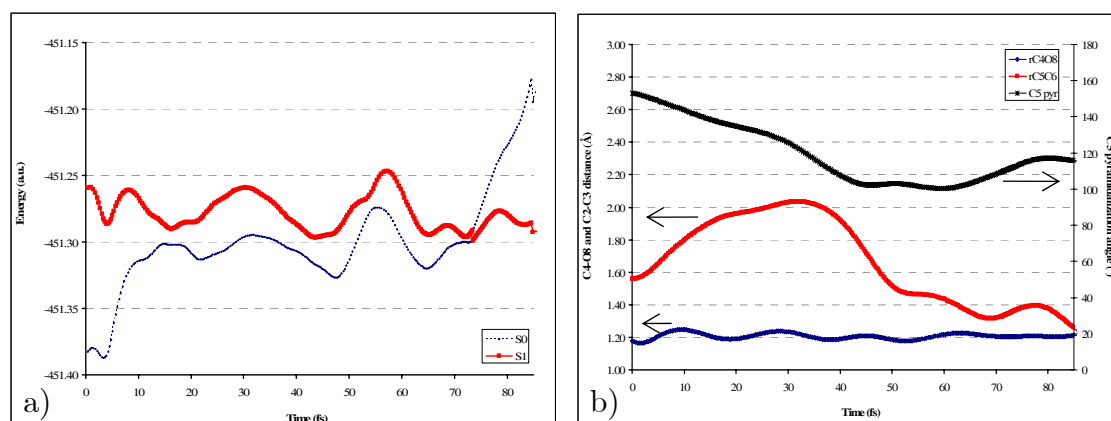
Figure 34. a) Time evolution of the CASSCF(8,6) energy of the S_0-S_1 states of thymine for a representative trajectory on S_1 (π,π^* state); (b) the same for the C_4-O_8 and C_5-C_6 distance and the C_5 pyramidalization (C_9-C_5-C_4-N_3 dihedral angle). The label of the states in (a) refers to the order at the beginning of the trajectory.
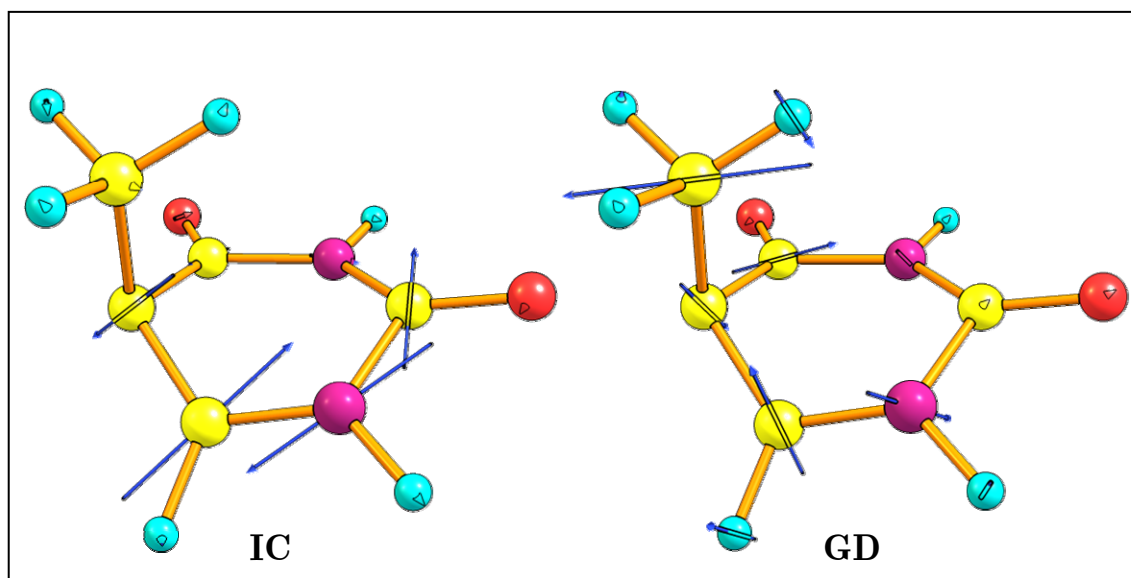
Figure 35. Branching space vectors (interstate coupling, IC, and gradient difference, GD) at $(\text{Eth})_X$, calculated at the CASSCF(8,6)/6-31G* level.

In contrast to the π,π^* trajectories, the trajectories on the n,π^* state are mainly driven by the C_4-O_8 stretch coordinate. This coordinate takes the molecule to a seam of intersection with the ground state which is reached in about 10 fs (see Figure 36 for a representative trajectory). At the seam region, the C_4-O_8 distance is larger than 1.5 Å. The trajectories oscillate around the seam, and three out of five trajectories decay to the ground state after 100 fs or less. The remaining two trajectories reached the computational time limit without decaying to S_0, but it can be assumed that they would decay shortly after that. However the analysis of the intersection and the comparison with the MS-CASPT2 energies show that the propensity for this decay is overestimated in the dynamics (see Figure 31). At the CASSCF level, the energy of the starting point, $(\pi,\pi^*)_{TS}$, is approximately 0.8 eV higher than $(n,\pi^*/GS)_X$, and the intersection is energetically accessible, but at the MS-CASPT2 level the minimum of the $(n,\pi^*/GS)_X$ seam lies slightly higher in energy than the FC point. The decay through this intersection is therefore less favored at the higher level of theory.
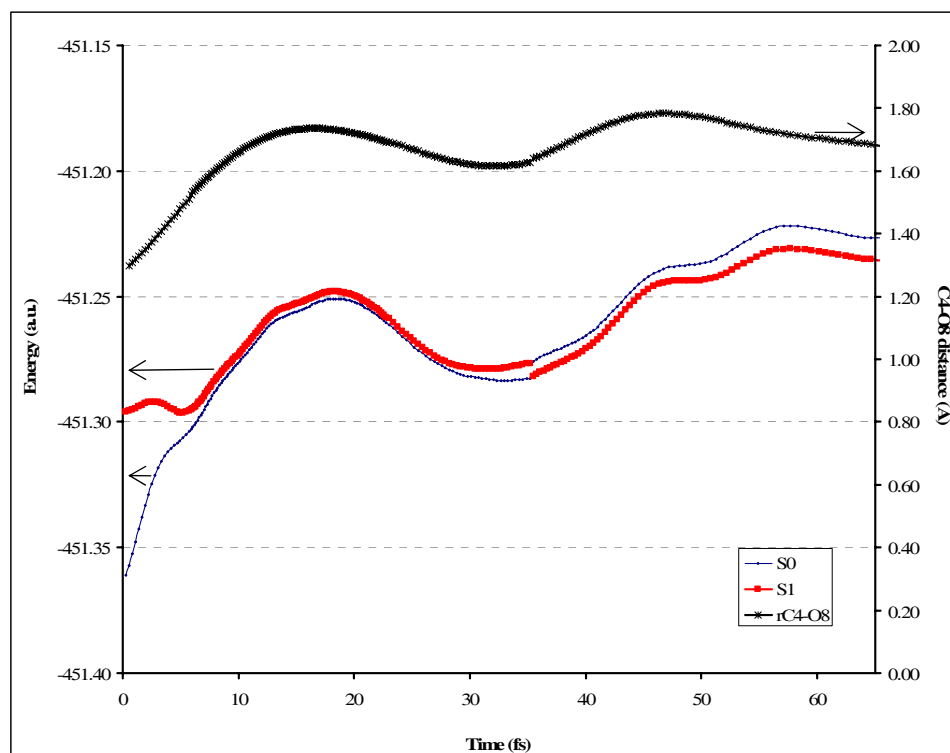
Figure 36. Time evolution of the CASSCF(8,6) energy of the S_0-S_1 states of thymine and the C_4-O_8 distance for a representative trajectory on S_1 (n,π^* state). The label of the states refers to the order at the beginning of the trajectory.

4.3.5   Discussion

The indirect excited-state decay path of thymine shown in Figure 28 has been studied with semiclassical dynamics started at the transition state that separates the minimum from the conical intersection with the ground state, $(Eth)_X$. Qualitatively, our results are similar to those found in a semiempirical OM2/MRCI dynamics study after excitation to the FC region.[266] Thus, the trajectories are funneled along the decay path towards an S_2/S_1 conical intersection seam between the π,π^* and n,π^* states which is perpendicular to the decay coordinate (C_4-O_8 bond stretch, see Figure 28). The quenching to the S_1 state is highly efficient and takes place in 5 - 60 fs, and a branching of the trajectories between the π,π^* and n,π^* states is observed. The trajectories that stay on the π,π^* state continue to the S_1/S_0 seam associated to $(Eth)_X$ and decay to the ground state in 200 fs or less. The trajectories that decay on the n,π^* state also reach a seam of conical intersection where they decay to the ground state on a similar time scale, but comparison with MS-CASPT2

calculations shows that the CASSCF trajectories overestimate the probability of this decay.

Experimentally, the short-time excited-state photodynamics of thymine are described with one or two components in the femtosecond range and one component in the picosecond range. From the present CASSCF dynamics and the previous study of the decay from the FC structure, this multiexponential decay could be associated to the indirect path shown on Figure 28, as already proposed.[138] In this case, the first step of the decay would correspond to the decay to $(\pi,\pi^*)_{Min}$ in the fs scale, and the second step to the access of $(Eth)_X$ through $(\pi,\pi*)_{TS}$, in the ps scale. However, comparison of the CASSCF decay paths with the MS-CASPT2 ones shows that this conclusion is not as straightforward as it may seem. At the higher level of theory there is a direct path (see Figure 28) that could also account for the fs decay component. This path could not be optimized at the CASSCF(8,6)/6-31G* level and is not observed in the short-time dynamics. However, the MS-CASPT2 results show that the direct and indirect paths are very similar from the energetic point of view. It is therefore likely that the CASSCF semiclassical dynamics only explain one side of the story (the indirect decay), while in reality the wavepackets created by the excitation may be distributed along the two paths shown in Figure 28. The MS-CASPT2 energy profile along the indirect path also suggests that the lifetime of $(\pi,\pi^*)_{Min}$ is overestimated by the CASSCF trajectories, as the barrier that separates it from $(Eth)_X$ is substantially lowered at the dynamically correlated level. Thus, the estimated MS-CASPT2 value is approximately 0.1 eV (800 cm^{-1}). In this case, the lifetime of $(\pi,\pi^*)_{Min}$ will mainly depend on intrastate vibrational redistribution from the in-plane modes activated in the first part of the decay to the TS mode, and this process can be expected to be quite efficient, as the TS mode also has a large in-plane component (see Figure 32). Therefore it is likely that the lifetime of $(\pi,\pi^*)_{Min}$ lies in the sub-picosecond range. This suggests that the experimental sub-picosecond decay component(s) reflects the quenching to the ground state along both direct and indirect paths.

Comparison between the CASSCF and MS-CASPT2 energies is also necessary to discuss the lifetime of the n,π^* state in the dynamics. The CASSCF trajectories on the n,π^* state decay to the ground state quickly because the energy of the $(n,\pi^*/GS)_X$ intersection seam is below the energy of $(\pi,\pi*)_{TS}$, the dynamics starting point. This is similar to the OM2/MRCI dynamics study, where the lifetime of the n,π^* state is less than 1 ps because the minimum energy intersection between the n,π^* and the ground state lies 1 eV below the FC point.[266] However our relative MS-CASPT2 energies show that the short n,π^* lifetime may be an artifact of the CASSCF and OM2 methods. Thus, the relative MS-CASPT2 energy of the $(n,\pi^*/GS)_X$ intersection minimum,

which marks the threshold for decay of the n,π^* state along this route, is 5.3 eV. This value is above the calculated vertical excitation of approximately 5.1 eV (see Figure 31). Other intersection minima of (n,π^*/GS) character, such as the intersection labeled MXS3 in Ref. 267, lie higher in energy at the MS-CASPT2 level. This suggests that the decay of the n,π^* state along this route will only take place at excitation wave lengths of 230 - 240 nm or less. At lower excitation energies there are other decay paths available for $(n,\pi^*)_{Min}$. The first one is the decay through $(Eth)_X$. This path goes through an avoided crossing between the π,π^* and n,π^* states associated to a MS-CASPT2 barrier of approximately 0.4 eV (see Figure 31) and is not observed in the dynamics because the trajectories are biased towards the $(n,\pi^*/GS)_X$ intersection. The second path is intersystem crossing to the triplet state. This process is possible due to the large spin-orbit coupling[137] of approximately 60 cm^{-1}, but it has been not observed in our dynamics because only singlet states are considered. However, the presence of these two paths suggests that the decay component of approximately 5 ps determined experimentally may correspond to the lifetime of the n,π^* state.

4.4   Thymine S_2/S_1 CI seam analysis and quantum dynamics

In section 4.3 we have seen that the S_2/S_1 seam of intersection represents a bottleneck for the decay of thymine along the indirect path. Thus, trajectories started on S_2 have to pass the seam to access S_1. Moreover, at the seam there is a branching of the trajectories between the degenerate states, and depending on the branching the trajectories remain in the dark (n,π^*) state or decay further to S_0 on the (π,π^*) state. Therefore, the outcome of the trajectories is determined by the passage through the S_2/S_1 seam.
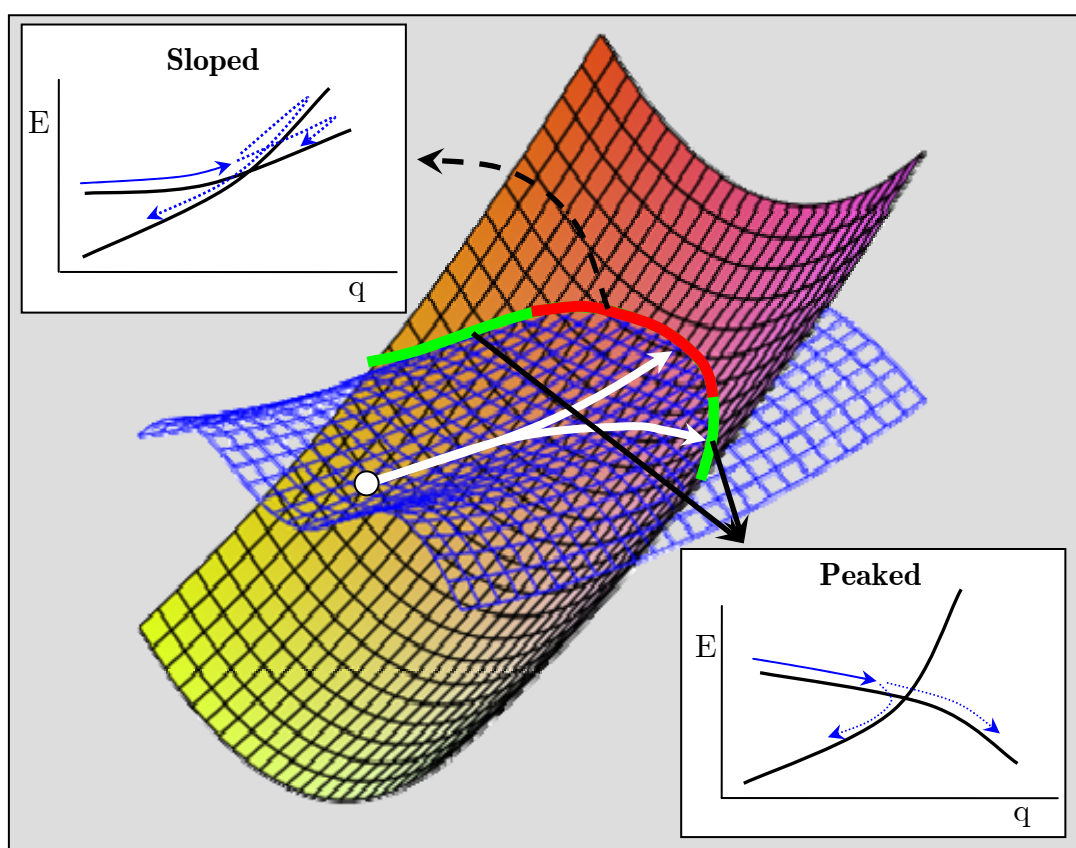
Figure 37. Sketch of a sloped-to-peaked CI intersection seam.

In this section we explore the S_2/S_1 seam in more detail by combining a topological analysis with quantum dynamics trajectories directed to different segments of the seam. The topological analysis shows that the determining feature of the energetically accessible region of the seam is the presence of segments with a peaked and sloped topology. This is shown schematically in Figure 37, together with the expected consequences for the reactivity. The central part of the seam has a sloped topology, where the decay to S_1 should lead to population of the n,π^* state and compete with return to $(\pi,\pi^*)_{Min}$. In contrast to this, at the peaked segments, the decay to S_1 should be more efficient and should be characterized by a branching between the two states. To explore the different segments of the seam and prove this picture, quantum dynamics propagations have been started at the transition state $(\pi,\pi^*)_{TS}$, adding momenta along different coordinates to drive the wavepackets towards the regions of interest. The quantum dynamics study has been carried out using the DD-vMCG method described in section 2.2. In principle, this method should allow for a better treatment of the non-adiabatic events compared to the trajectory surface hopping method. Previous applications of these method have considered up to 15 degrees of freedom (a reduced dimensionality study of benzene), and the present dynamics calculations provide a test of the performance of the method with a medium-sized system where 39 degrees of freedom are taken into account in the propagation.

4.4.1   Computational details

The calculations have been carried out at the CASSCF(8,6)/6-31G* level of theory, which is the level used in the previous trajectory surface hopping study. For consistency with the dynamics, the static study of the seam topology has been carried out at the same level of theory. A development version of the Gaussian 03[278] package which allows to run constrained IRC calculations and evaluate second derivatives at the CI has been used for the characterization of the CI seam.

The quantum dynamics simulations have been carried out with the direct dynamics variational Multi-Configurational Gaussian wavepacket method (DD-vMCG)[258,259] implemented in a development version of the Heidelberg MCTDH package.[281] With the actual implementation, this method, which is explained in section 2.2.3.2.a, propagates a wavepacket formed of a number of Gaussian functions on a maximum of two PES. The adiabatic PES are calculated on-the-fly at each step and are transformed to the diabatic representation,[261] which is more convenient to describe non-adiabatic events (see section 2.2.4). The corresponding PES are obtained with a development version of the Gaussian package[278] which is interfaced with the vMCG code. The DD-vMCG method

requires the calculation of the energy and first and second derivatives at each step, however it benefits from the use of a database in which these parameters are stored at every step. The database is also used by the dynamics code to avoid calculating the same point twice. Moreover, the CASSCF energies and gradients are obtained using state-averaged orbitals averaging over S_2 and S_1 with equal weights, as described for the semiclassical dynamics. Therefore, when the Gaussian functions approaches a region of S_1/S_0 degeneracy, it is not possible to converge the CASSCF calculation, and the trajectories are interrupted.

As in the semi-classical dynamics case, the dynamics calculations were started on $(\pi,\pi^*)_{TS}$. The normal modes at this structure provide the frequency-mass weighted coordinates for the propagation, and the diabatic states were chosen so that they coincide with the adiabatic ones at the starting point. They correspond to the π,π^* and (n,π^*) states (S_2 and S_1, respectively). The nuclear motion is described by a wavepacket composed of up to 8 pairs of Gaussian functions, with one function per state for each pair. The wavepacket is propagated along all degrees of freedom. An initial momentum of 0.1 eV was added to the wavepacket on the imaginary vibrational normal mode to drive it towards the CI in all calculations. Extra momenta of 0.2 eV were added to other normal modes to analyze the different parts of the CI.

4.4.2   Topological analysis of the S₂/S₁ CI seam

4.4.2.1   Intersection space characterization

To start the topological analysis of the seam, a conical intersection of *Cs* symmetry has been optimized on the seam and has been characterized with the intersection space Hessian analysis (Figure 39). This structure lies 1.07 eV above the minimum energy conical intersection $(\pi,\pi^*/n,\pi^*)_X$, and 0.3 eV below $(\pi,\pi^*)_{TS}$. The structure is characterized by a short C_4O_8 bond (1.18 Å) and a stretched C_5-C_6 (1.60 Å) bond, and the gradient difference vector is dominated by the C_4-O_8 stretch. The structure has three imaginary intersection space frequencies of *a''* symmetry, of $391i$ cm^{-1}, $218i$ cm^{-1} and $62i$ cm^{-1}. The first two modes are shown in Figure 39 and correspond to pyramidalization of C_5 and C_6. According to the second-order model for the analysis of conical intersection seams,[38] a seam of intersection lies along curved coordinates composed of displacements along any of the intersection space modes, combined with the branching space vectors. The imaginary character of the frequencies indicates

that the energy of the seam must decrease along the corresponding displacements. This suggests that the *Cs* conical intersection is a saddle point in the space of conical intersection. It connects two symmetric $(\pi,\pi^*/n,\pi^*)_X$ structures, along a continuous seam.

| A" = *391i* | A" = 218*i* | GD | IC |
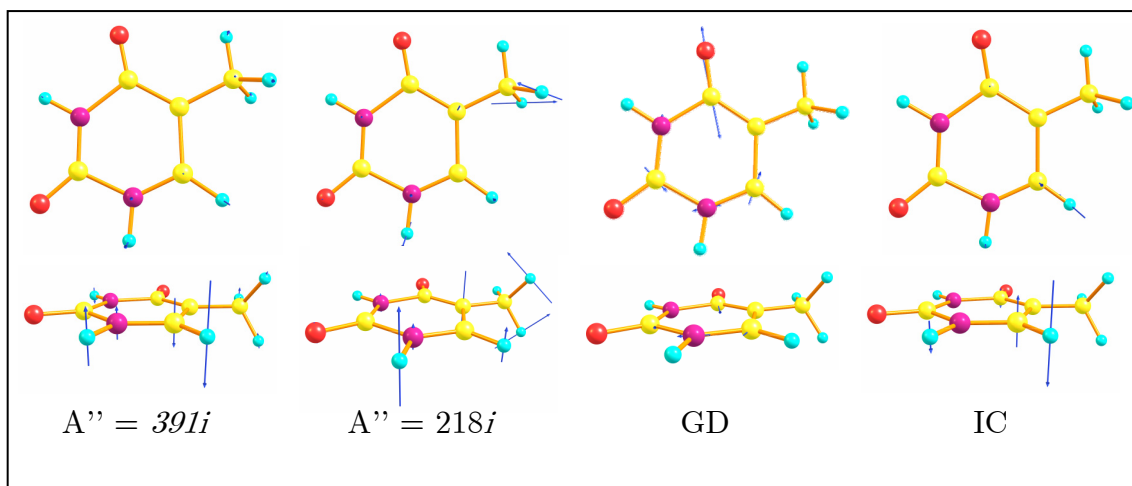
Figure 38. The two lowest vibrational modes and branching space vectors of the Cs structure of the S_2/S_1 CI.

The conical intersection seam has been characterized with a constrained IRC calculation in the intersection space started from the *Cs* conical intersection, in the direction of the pyramidalization. The results are shown in Figure 39.
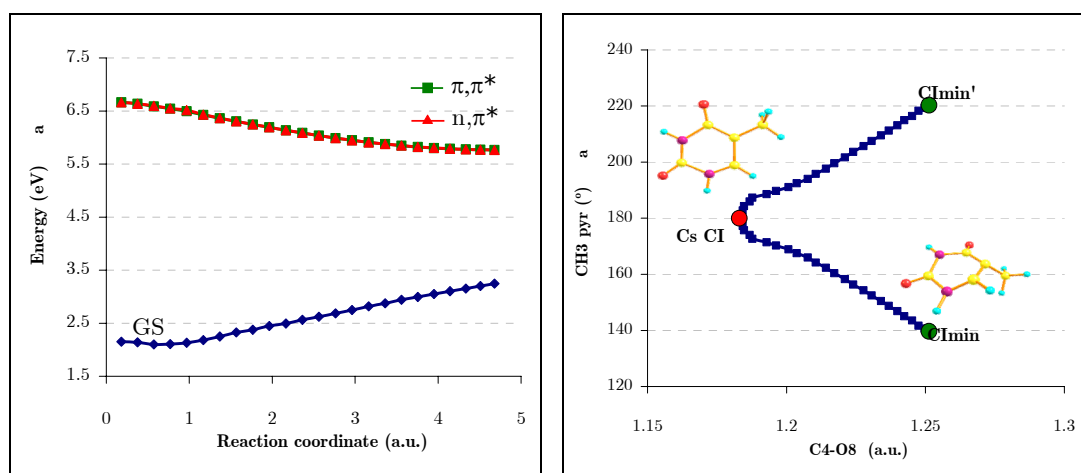
Figure 39. a) Energy profile of the three lowest states along the interstate vector. b) C_5 pyramidalization angle vs C_4-O_8 distance along the intersection space of the S_2/S_1 seam.

The IRC in the seam space ends at $(\pi,\pi^*/n,\pi^*)_X$, which is the global minimum of the seam. The intersection structures are found along a

combination of the C_4-O_8 stretch and methyl pyramidalization coordinates, as predicted by the second-order analysis at the *Cs* structure. Along the whole mapped segment of the seam, the energy lies below that of $(\pi,\pi^*)_{TS}$. This suggests that also the planar segments of the seam are dynamically accessible.

To complete this analysis, the topology of the seam has been examined at the structures that compose the IRC by comparing the signs of the gradients of the two states along the gradient difference coordinate. To carry out this comparison, it is necessary to ensure that the gradients are consistently defined along the seam, *i.e.* to avoid rotations of the configuration interaction eigenvectors of the degenerate CASSCF states. For this purpose, the gradients have been transformed as proposed by Yarkony[282] before examining the sign. The lengths of the gradients along the seam, including the sign, are shown in Figure 40. The CI seam has a sloped-to-peaked character. Thus, the length of the gradients decreases along the seam, and one of the gradients changes its sign. In the segment of the seam close to the planar intersection (left side of Figure 40), the gradients have the same sign, *i.e.* the two states are parallel and the CI is sloped. On the other hand, near the minimum energy CI structure (right side of Figure 40), the gradients have opposite sign and the topology is peaked. The peaked and sloped regions of the seam are separated by a segment with intermediate topology, where one of the gradients is close to zero.
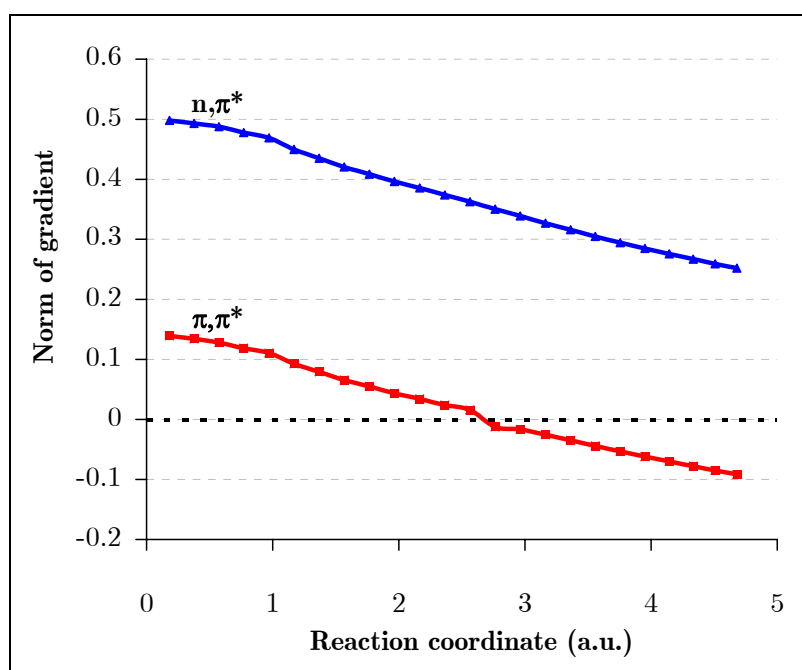
Figure 40. Norm of the two state gradient vectors along the constrained IRC on the seam.

4.4.2.2   Analysis of the normal modes at the dynamics starting point $(\pi,\pi^*)_{TS}$

In the present DD-vMCG application, the frequency-mass-weighted coordinate system is based on the normal modes at $(\pi,\pi^*)_{TS}$. This structure is the origin of the coordinate system, and the coordinates of every structure correspond to the distance to $(\pi,\pi^*)_{TS}$ along the different modes. These coordinates can be used to drive the dynamics toward a given structure, by adding extra momentum to the coordinate with the highest value for the desired structure.
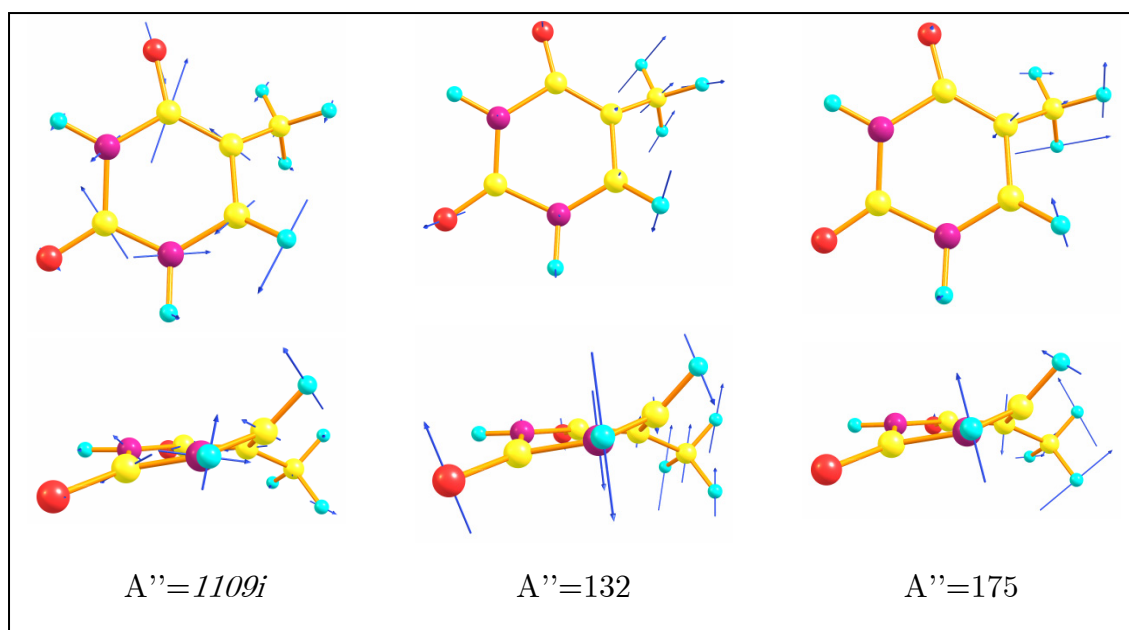
Figure 41. Vibrations of the 3 lowest normal modes at the dynamics starting point. Frequencies expressed in cm^{-1} (imaginary values in italics).

Out of the 39 frequency-mass-weighted coordinates, the three coordinates which change more significantly among the structures correspond to displacements along the imaginary frequency, which is dominated by C_4-O_8 bond stretching, and the two lowest-frequency modes at $(\pi,\pi^*)_{TS}$, which are different combinations of C_5 and C_6 pyramidalization. They are referred to as modes 1 - 3 (see Figure 41). The values of these coordinates for the relevant structures are shown in Table 9, together with the total distance of every structure to the origin.

Mode	$(\pi,\pi^*)_{Min}$	$(n,\pi^*/\pi,\pi^*)_X$	$(n,\pi^*/\pi,\pi^*)_{XCs}$	$(n,\pi^*)_{Min}$	$(Eth)_X$
1 (TS)	-4.400	3.067	-1.341	-5.246	4.488
2	3.230	-0.408	5.290	5.059	-6.955
3	3.481	-3.873	3.101	2.851	-6.028
Total dist.	7.636	6.125	24.515*	17.745*	20.075

Table 9. Critical points position with respect to $(\pi,\pi^*)_{TS}$ in Fmw coordinates. * The total distances of these points with respect to the $(\pi,\pi^*)_{TS}$ in Fmw coordinates are overestimated because of a rotation of the methyl group.

It is clear from Table 9 that the dynamics start close to the peaked segment of the seam, since $(\pi,\pi^*/n,\pi^*)_X$ lies closer to the origin than the other structures. Moreover, $(\pi,\pi^*,n,\pi^*)_X$ and the S_1/S_0 intersection $(Eth)_X$ are found approximately in the same direction from $(\pi,\pi^*)_{TS}$, while $(\pi,\pi^*)_{Min}$, $(n,\pi^*)_{Min}$ and the Cs intersection lie along the opposite direction. Overall, the coordinate analysis suggests that modes 2 and 3 can be used to guide the propagation to the different parts of the seam. Thus, momentum in the positive direction along modes 2 and 3 should drive the wavepacket to the region of $(\pi,\pi^*)_{Min}$ and the *Cs* intersection, whereas momentum in the negative direction of mode 3 should drive it toward $(\pi,\pi^*/n,\pi^*)_X$.

4.4.3   Quantum dynamics at the S_2/S_1 CI seam

The role of the different segments of the S_2/S_1 seam in the decay has been studied with quantum dynamics calculations, by driving the propagation along the vibrational normal modes shown in Figure 41. Thus, three sets of propagations with a different number of Gaussian functions to simulate the wavepacket have been carried out, where momentum has been added along some of the three modes. All the simulations described in this section are gathered on Table 10 with their corresponding features. The energy that has been added to each mode for each run is displayed in columns 2 and 3, whereas the pairs of Gaussian functions that have been used for the wavepacket are described in column 4. To facilitate the description of the results, the figures that describe each run are gathered on the last column of Table 10.

Run	Mode	Momentum (eV)	Pairs of Gaussians	Figure(s)
1a	1	-0.1	1	42a-c
1b	1	0.1	1	43a-c, 44a, AIII-1
1c	1	0.1	2	44b, AIII-2
1d	1	0.1	4	44c, AIII-3
1e	1	0.1	8	44d, AIII-4
2a	1	0.1	2	AIII-5-10
	2	-0.2		
2b	1	0.1	2	45, AIII-11-15
	2	0.2		
3a	1	0.1	4	46a-b, AIII-16-24
	3	-0.2		
3b	1	0.1	4	47a-b, AIII-25-33
	3	0.2		

Table 10. Simulations described in this chapter with their corresponding characteristics (modes to which momentum has been added, momentum (eV), functions per state, and figures that describe each run).

4.4.3.1   Propagation with additional momentum on mode 1

Mode 1 is associated to the imaginary frequency at $(\pi,\pi^*)_{TS}$. It can be expected that addition of momentum along that coordinate will drive the propagation to the seam or to $(\pi,\pi^*)_{Min}$, depending on the direction of the momentum. The details of the DD-vMCG method are illustrated for two simple cases, where a single pair of Gaussian centered functions is used to simulate the wavepacket and a momentum of 0.1 eV is added in either direction of mode 1 (runs 1a and 1b). There is one function running on each diabatic state, π,π^* and n,π^*, which correspond to the adiabatic S_2 and S_1 states at the beginning of the propagation, respectively. Initially, the population of the π,π^* function is 1.0, and that of the n,π^* function 0.0. The results for the backward propagation (run 1a) are shown in Figure 42a - Figure 42c.
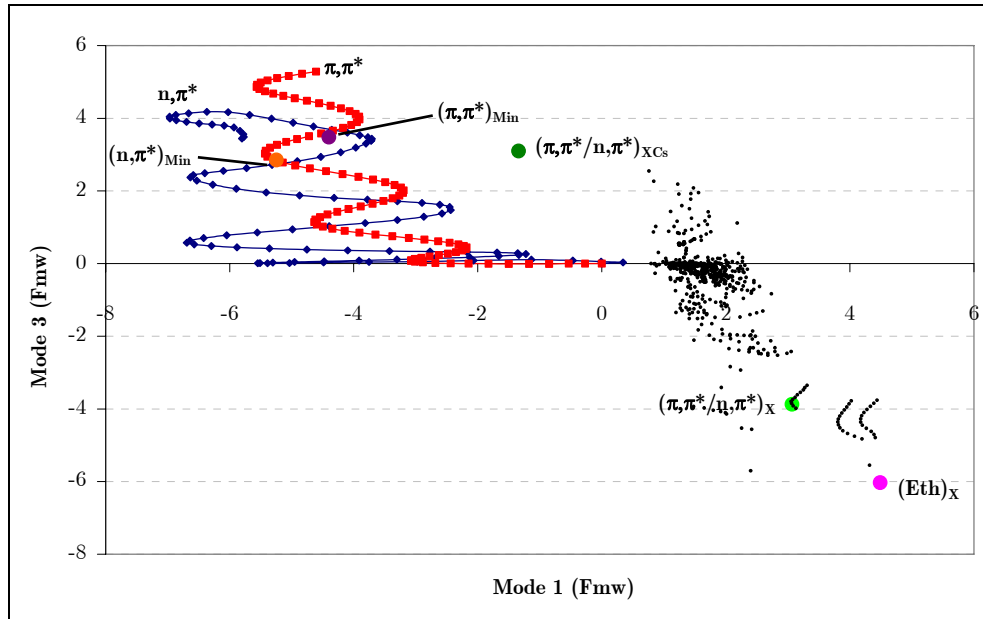
Figure 42a. Position of the center of the Gaussian functions of run 1a with respect to mode 1 and mode 3 (position relative to $(\pi,\pi^*)_{TS}$).

Figure 42a shows the evolution of the position of the two Gaussian functions along modes 1 and 3. The seam of intersection is represented by the black dots, which correspond to the data base structures (see section 4.4.1) where the energy gap is below a threshold of 0.35 eV. In the backward propagation, the function of the π,π^* state (red squares) evolves toward $(\pi,\pi^*)_{Min}$, and that of the n,π^* state (blue diamonds) toward $(n,\pi^*)_{Min}$.
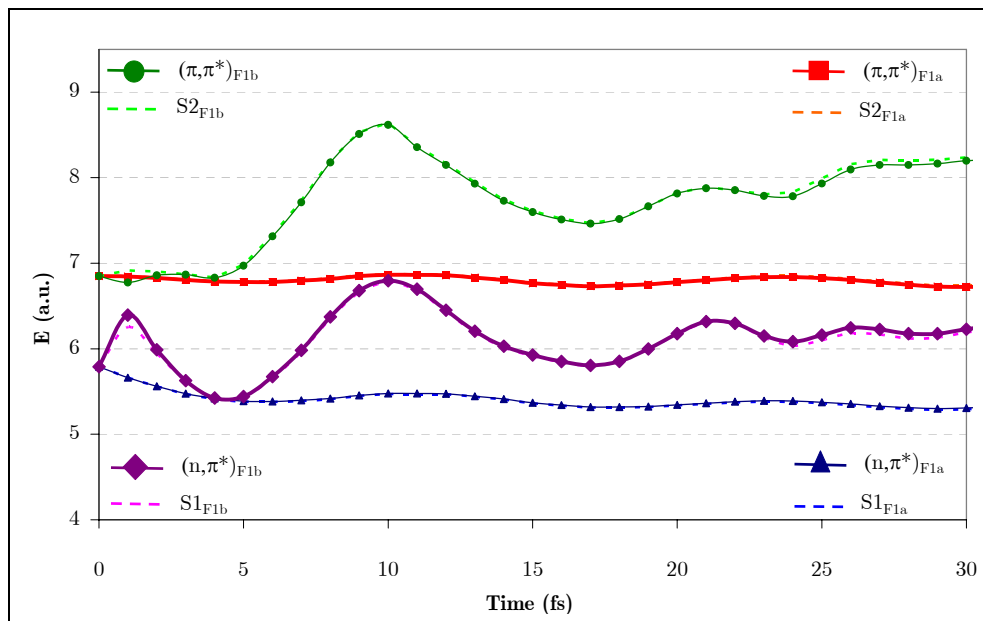
Figure 42b. Diabatic and adiabatic energies of the Gaussian functions of the π,π^* (F1a) and n,π^* states (F1b) of run 1a.

Figure 42b shows the energetics of the propagation. In the vMCG method, every Gaussian function is associated to four energetic parameters: the adiabatic S_2 and S_1 energies, and the diabatic energies obtained after the transformation described in section 2.2.4. Thus Figure 42b contains eight energy profiles, four for each function. However, the diabatic and adiabatic energies for each state are undistinguishable in the present propagation because the wavepacket is driven away from the seam and the S_2/S_1 energy gap is large. Moreover, only the energy of the diabatic state associated with each Gaussian function (π,π^* or n,π^*) is relevant for the propagation. The corresponding profiles are drawn with a bold line in Figure 42b. Finally, Figure 42c presents the diabatic populations of the two Gaussians. At the beginning of the propagation, there is a small population transfer to the n,π^* function because of the proximity of $(\pi,\pi^*)_{TS}$ to the seam, but the majority of the population stays on the π,π^* function as the wavepacket evolves toward $(\pi,\pi^*)_{Min}$.
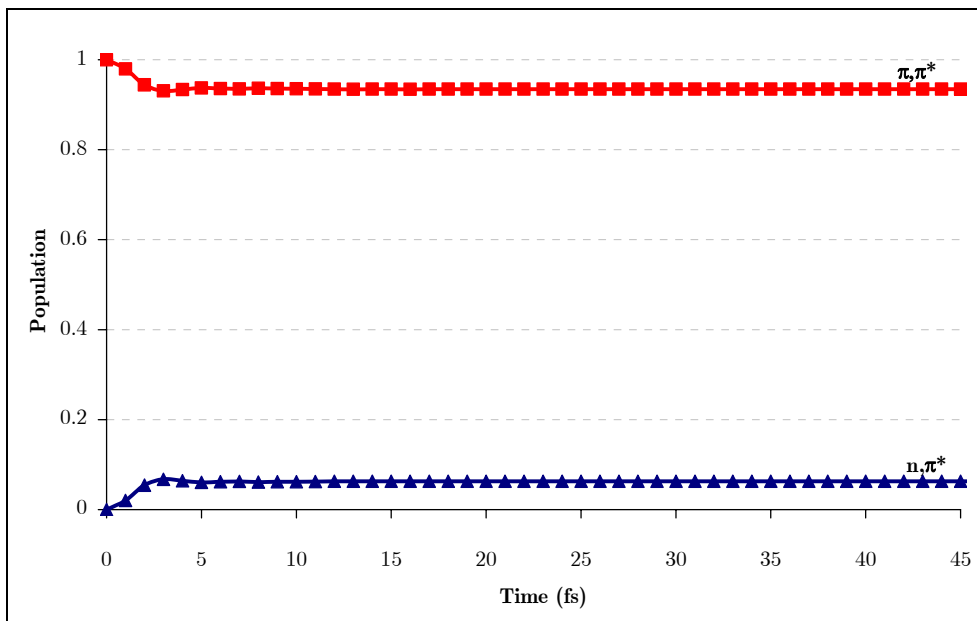
Figure 42c. Diabatic populations of the π,π^* and n,π^* states of run 1a.

The results for the propagation of one pair of Gaussian functions adding momentum to the forward direction of mode 1 are shown in Figure 43a-c (run 1b). In this case, the π,π^* function evolves towards $(\pi,\pi^*/n,\pi^*)_X$, while the n,π^* function oscillates along mode 1. The behavior of the wavepacket at the seam can be monitored with the energy profiles for the π,π^* function, which are shown in Figure 43a.
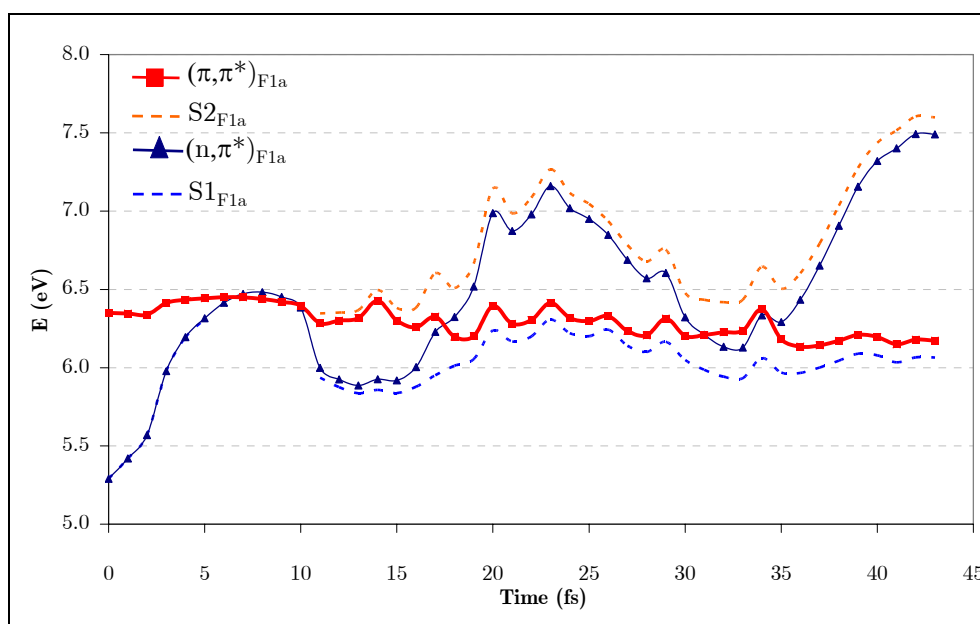
Figure 43a. Diabatic and adiabatic energies of the Gaussian function of the π,π^* state (F1a) of run 1b.

The energy profiles for the n,π^* function are shown in Appendix III. The π,π^* function approaches the seam three times, and the passages through the seam are in phase with the oscillation shown for that function along mode 1 in Figure 43b. After the third passage, the function, which runs on the diabatic π,π^* state, leaves the seam region as S_1. The population analysis (Figure 43c) shows that most of the population stays on the π,π^* state, except for a small amount of population transferred to the n,π^* function at the beginning of the trajectory. Thus, the propagation is similar to the semiclassical trajectories described above where the trajectories pass through the seam, stay on the π,π^* state and approach the S_1/S_0 intersection. However, in the present case, the propagation was interrupted when the function approached the S_1/S_0 degeneracy because it was not possible to converge the CASSCF calculation.
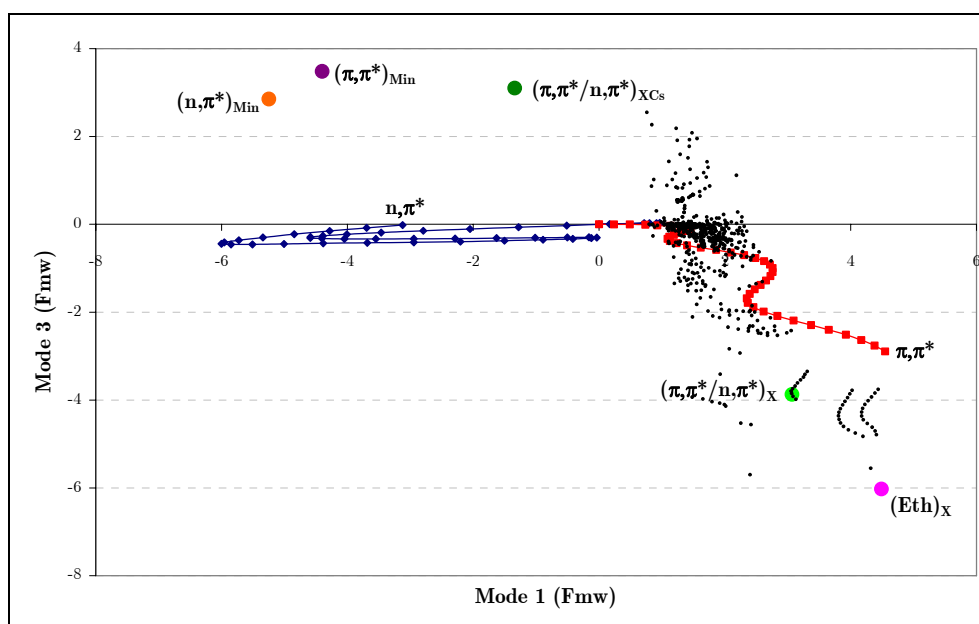
Figure 43b. Position of the center of the Gaussian functions of run 1b (position relative to $(\pi,\pi^*)_{TS}$).

Overall, the quantum dynamics with a single pair of Gaussian functions are in agreement with the expectation that the initial momentum in the backward direction leads the wavepacket to $(\pi,\pi^*)_{Min}$, while in the forward direction the wavepacket is driven towards the S_2/S_1 seam and continues toward the S_1/S_0 intersection.
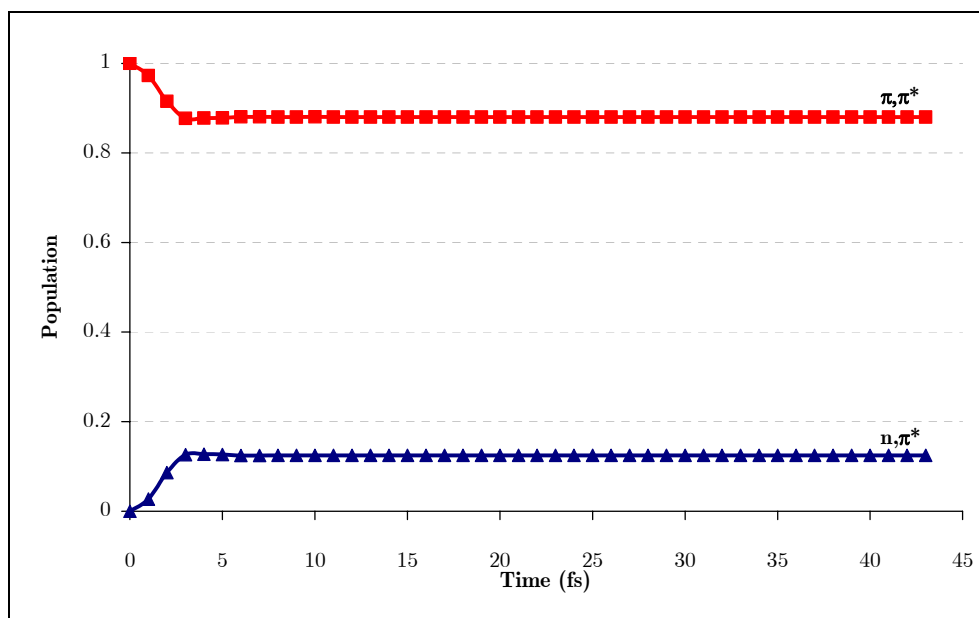
Figure 43c. Diabatic populations of the π,π^* and n,π^* states of run 1b.

The propagations with 0.1 eV extra momentum in the forward direction of mode 1 were also carried out with 2, 4, and 8 pairs of Gaussian wavefunctions (runs 1c - 1e, respectively). The Gaussian functions are situated at $(\pi,\pi^*)_{TS}$ at the beginning of the propagation, from where they can evolve along different directions. Ideally, the increase in the number of Gaussians should lead to convergence of the propagations. In runs 1c - 1e, all propagated functions follow the fate of the functions shown in Figure 42a and Figure 43a (see Appendix III). However, the difference between the propagations with different number of functions lies in the evolution of the diabatic populations (Figure 44a-d). For instance, with four pairs of functions more than 50% of population is transferred to the n,π^* state. However, the series of propagations with an increasing number of functions is not converged, as seen by comparison of the populations with four and eight pairs. Increasing the number of Gaussian function pairs beyond 8 was not possible, because in the propagations it is necessary to carry out one CASSCF calculation per Gaussian function center at every step, using the same active space for each point. With an increasing number of Gaussian functions, the regions of space sampled during the propagation increase, and it becomes impossible to converge the CASSCF calculations for all Gaussian functions with the same active space. However, although the propagations are not converged with regard to the amount of population transferred to the n,π^* state, the mechanistic picture provided for the passage through the peaked region of the seam is clear from the propagations.
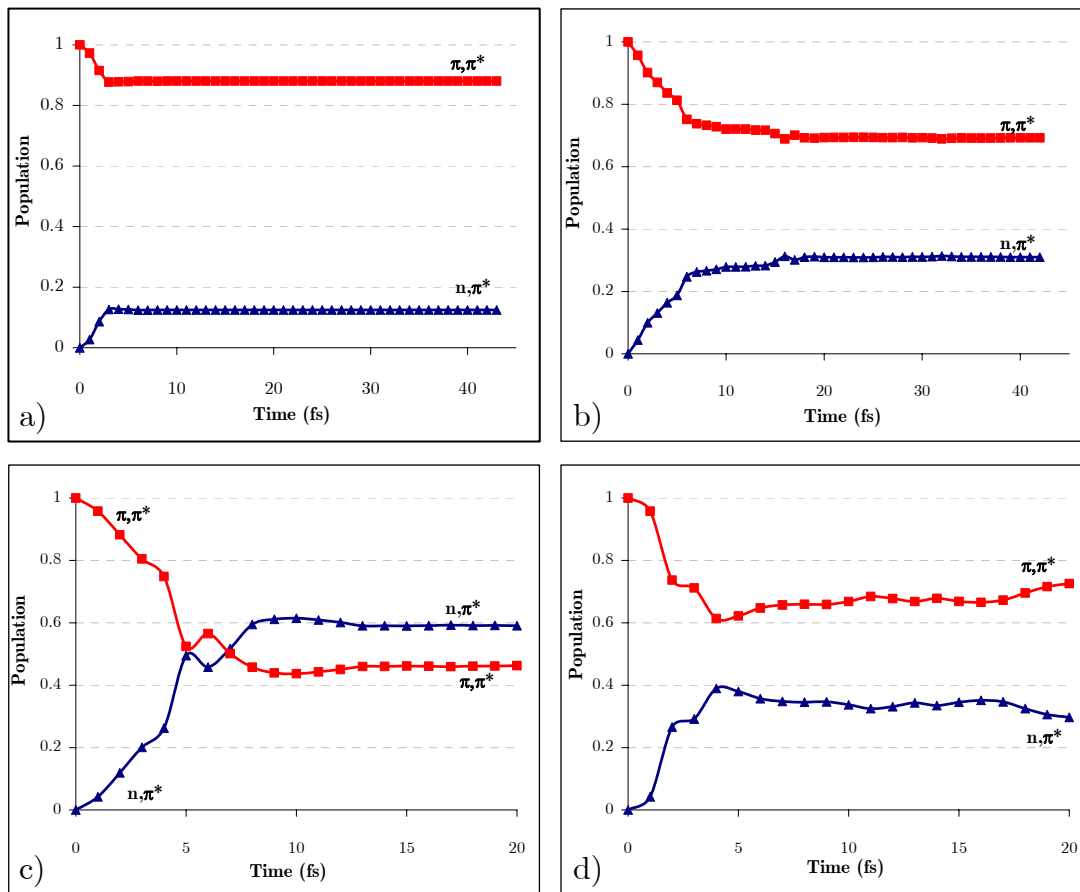
Figure 44. Diabatic populations of propagations with 0.1 eV of extra momentum on mode 1 with 1, 2, 4, and 8 functions per state, respectively (runs 1a, 1c, 1d, and 1e).

4.4.3.2 Propagation with additional momentum on mode 2

The effect of adding momentum along mode 2 is studied in the second set of propagations (runs 2a and 2b). This mode corresponds to an out-of-phase combination of the pyramidalizations at C_5 and C_6, and Table 9 suggests that addition of momentum along this coordinate in the positive direction may drive the wavepacket toward the planar region of the seam. To test this hypothesis, two propagation runs were carried out using two pairs of Gaussian functions, adding a momentum of 0.2 eV in the positive and negative directions of mode 2 (runs 2a and 2b, respectively). A momentum of 0.1 eV was also added in the positive direction of mode 1. The position of the two pairs of functions during run 2b are shown in Figure 45 (for energetics and populations see Appendix III).
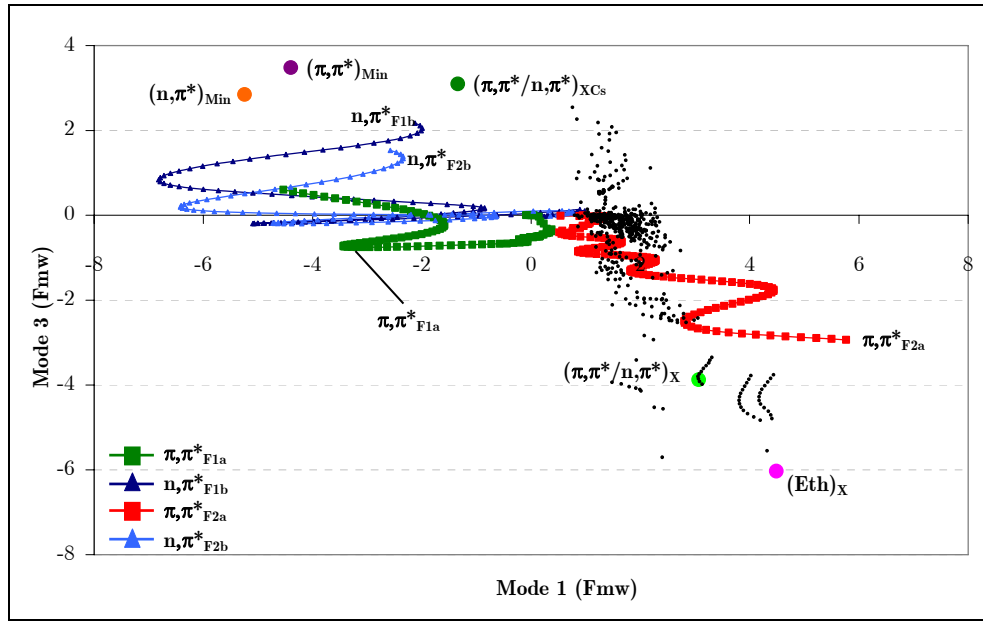
Figure 45. Position of the center of the Gaussian functions with respect to mode 1 and mode 3 along run 2b (position relative to $(\pi,\pi^*)_{TS}$).

Here most of the population stays on the two π,π^* functions. One of these functions crosses the seam near $(\pi,\pi^*/n,\pi^*)_X$ and continues on S_1 toward $(Eth)_X$, while the other function evolves towards $(\pi,\pi^*)_{Min}$. This behavior is similar to the one described for runs 1b-e, and to the one observed in run 2a (see Appendix III). This shows that mode 2 is not suitable to regulate the access of the wavepacket to the seam.

4.4.3.3   Propagation with additional momentum on mode 3

The third set of propagations is run with additional momentum along mode 3 (in-phase pyramidalization of C_5 and C_6). Additional momentum in the positive direction should lead the wavepacket toward the planar conical intersection, whereas momentum in the negative direction should lead it toward $(\pi,\pi^*/n,\pi^*)_X$ (see Table 9). Therefore, two propagations were run with 0.2 eV extra momentum in the negative and positive directions of this mode, together with 0.1 eV momentum in the forward direction of mode 1 (runs 3a and 3b). Four pairs of Gaussian functions were used for these runs. Figure 46a and Figure 46b show the results for run 3a (negative momentum along mode 3).
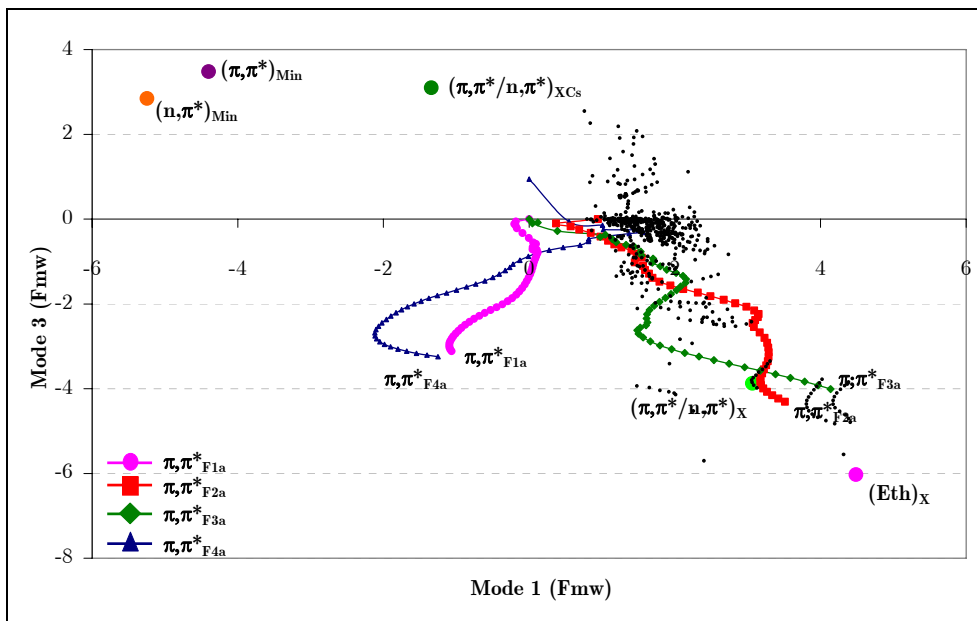
Figure 46a. Position of the center of the Gaussian functions of the π,π* state of run 3a (position relative to $(\pi,\pi^*)_{TS}$).

Out of the four π,π* Gaussian functions, two pass the seam near $(\pi,\pi^*/n,\pi^*)_X$ (F2a and F3a on Figure 46a). The diabatic populations (see Figure 46b) show that approximately 90% of the wavepacket stays on the π,π* state, and the diabatic energies for F2a and F3a (see figures AIII-17-18 and AIII-22-23 of Appendix III) show that the π,π* state is S_1 at the end of the propagation. Therefore, the two functions reproduce the part of the wavepacket that crosses the seam and continues to evolve on the π,π* state in the direction of $(Eth)_X$. In contrast to this, the other two functions (F1a and F4a) go to a flat region of the surface (see the energy profiles of figures AIII-16, AIII-19, AIII-21, and AIII-24 in Appendix III) which is only accessed under these initial conditions and has not been identified further. The overall amount of population transferred to the (n,π*) state in this run is small.
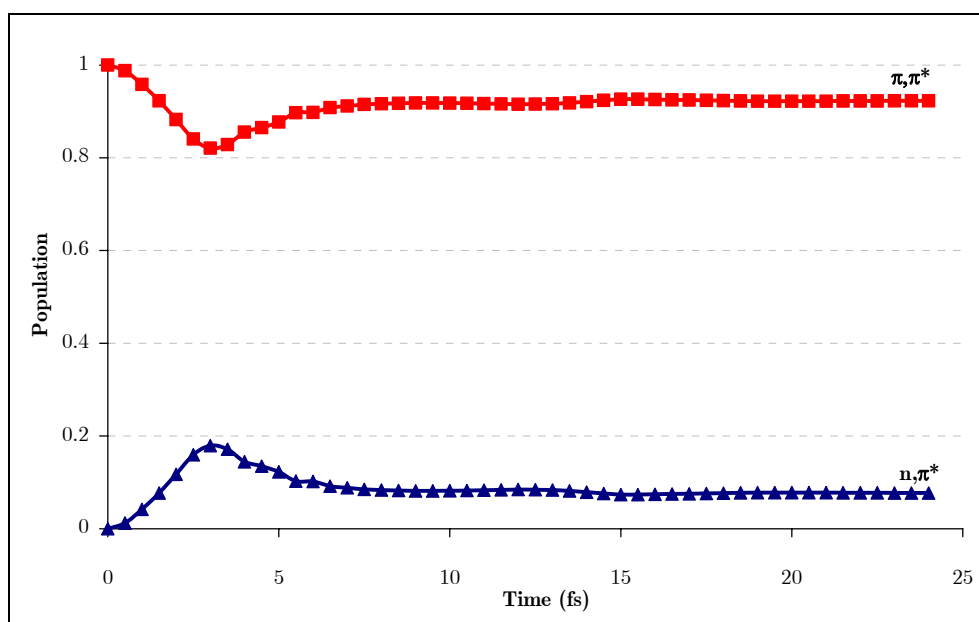
Figure 46b. Diabatic populations of the π,π* and n,π* states of run 3a.

Run 3b, where the momentum is added in the positive direction, shows a different behavior (see Figure 47a,b). Out of the four π,π* Gaussian functions, function F1a evolves directly toward $(\pi,\pi^*)_{\text{Min}}$, whereas the remaining three functions are directed toward the seam (see Appendix III). Compared with the previous run, the functions reach the seam closer to the *Cs* geometry. Function F2a oscillates there and then evolves further to $(\pi,\pi^*)_{\text{Min}}$, while function F4a keeps oscillating around the seam until the end of the trajectory.
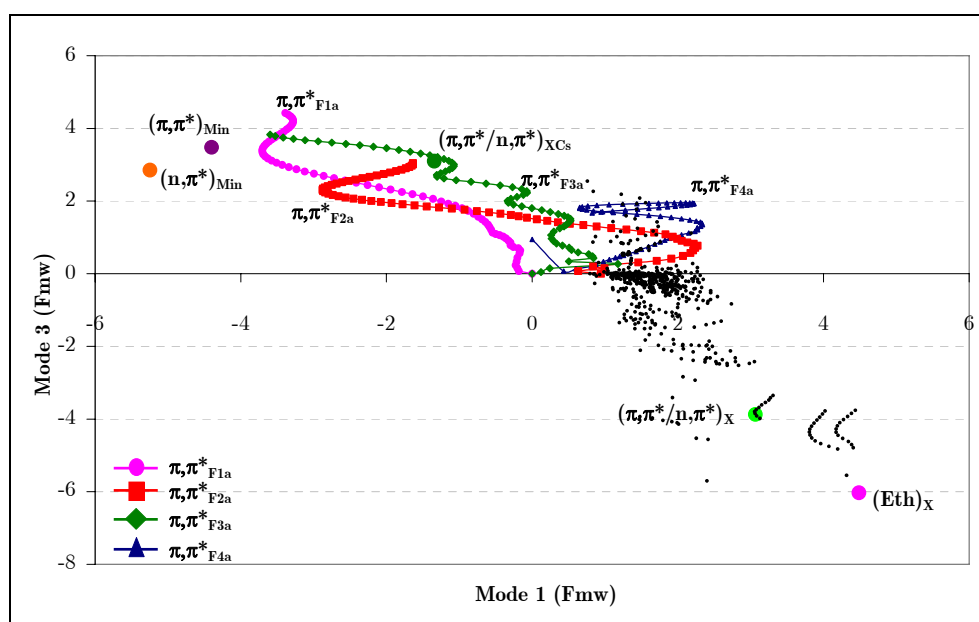
Figure 47a. Position of the center of the Gaussian functions of the π,π* state of run 3b (position relative to $(\pi,\pi^*)_{\text{TS}}$).

Turning to F3a, the propagation of this Gaussian function is affected by a failure of the diabatic transformation. Thus, the function leaves the seam with the diabatic π,π^* state being S_1, but this occurs at a region where the π,π^* state is the adiabatic S_2 state. This failure of the transformation is presumably caused by intruder states in the electronic wavefunction, *i.e.* states with different configurations from the ones of the initial π,π^* and n,π^* states. Inclusion of more states in the propagation would be necessary to correct this behavior.
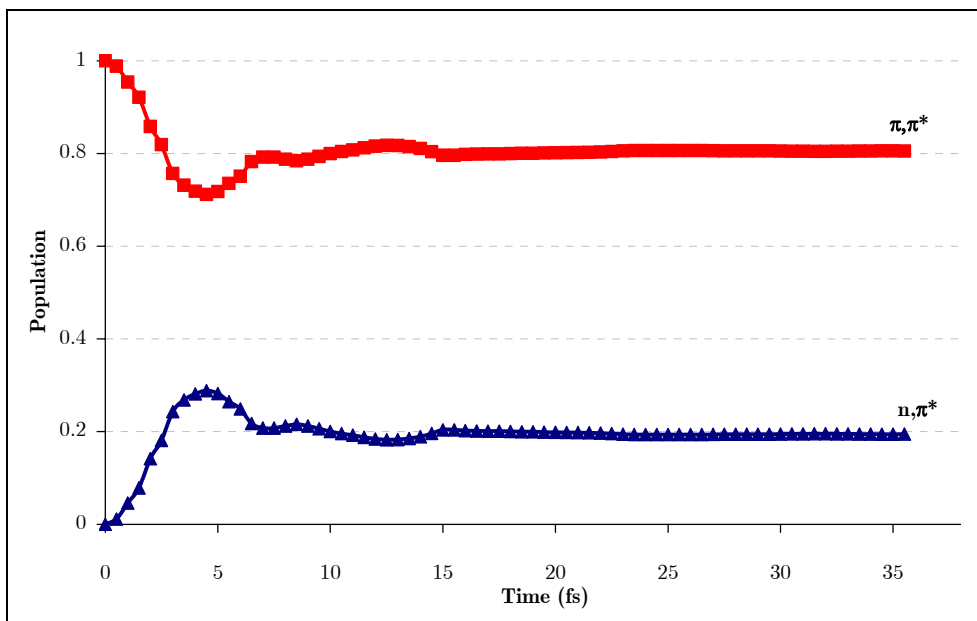
Figure 47b. Diabatic populations of the π,π^* and n,π^* states of run 3b.

Turning to the n,π^* state, overall there is a population transfer of 20% to that state (see Figure 47), and all the corresponding functions evolve towards $(n,\pi^*)_{Min}$ (see Appendix III). In summary, under the conditions of this run, the evolution of a part of the wavepacket toward $(Eth)_X$ described for run 3a is not observed. Instead, the wavepacket touches the seam and evolves towards $(\pi,\pi^*)_{Min}$, staying on S_2, or crosses to S_1 to populate the n,π^* state. This behavior corresponds, approximately, to the one schematized for the sloped region of the seam in Figure 37. Clearly, this conclusion has to be considered with care because of the problems described for function F3a. However, other simulations run with the same initial momentum and less Gaussian functions show a similar behavior to the one described for the remaining functions of run 3b, and it can be assumed that the results for the present run are qualitatively correct.

4.4.4   Discussion

The present quantum dynamics study, where 39 degrees of freedom are considered and the propagation is carried out on-the-fly, is a challenge for the DD-vMCG method. This is reflected in some of the problems found in our study. Thus, the number of Gaussian centered functions cannot be increased, as it would be necessary to converge the propagation, because it leads to convergence problems for the electronic structure. Similar problems have been found in runs with a higher amount of extra momentum. Moreover, because of the proximity of other excited states, intruder states appear that lead to a failure of the two-state diabatic transformation required for the propagation. However, in spite of these limitations, the propagations, combined with the topological analysis of the seam, have provided valuable insight into the behavior of the wavepacket at the S_2/S_1 seam of thymine. The resulting picture can be completed with the results of the previous study (section 4.3) and is illustrated best with the help of Figure 48.
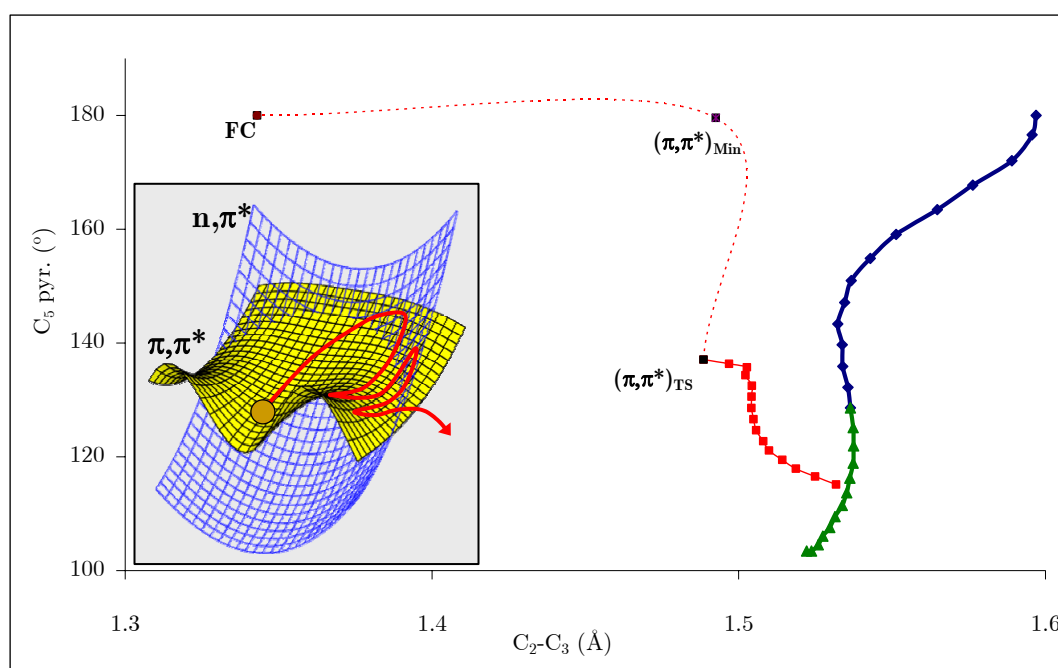
Figure 48. Representation of the S_2/S_1 CI seam and MEP from the dynamics starting point. Inset represents the typical behavior of a wavepacket accessing the peaked region of the seam.

Figure 48 represents the topology of the surface along the C_2-C_3 stretch and C_5 pyramidalization coordinates. It contains the indirect path from the Franck-Condon structure to $(\pi,\pi^*/n,\pi^*)_X$ and the S_2/S_1 seam. The red squares

represent the minimum energy path from $(\pi,\pi^*)_{TS}$ to the S_2/S_1 seam. In the seam, the peaked segment is displayed by green triangles, and the sloped segment by blue diamonds. The minimum energy path from $(\pi,\pi^*)_{TS}$ encounters the seam in the peaked region, near $(\pi,\pi^*/n,\pi^*)_X$. The semiclassical dynamics and the quantum dynamics runs 1a-1e are carried out with only 0.1 eV extra momentum along the imaginary frequency mode at $(\pi,\pi^*)_{TS}$ and therefore follow the minimum energy path quite closely. This explains why the trajectories encounter the peaked region of the seam. The typical behavior of a Gaussian function in the quantum dynamics propagations is shown in the inset of Figure 48. Thus, the function oscillates along the seam in the direction of mode 1, which is orthogonal to the seam, until it reaches the part of the surface with sufficient gradient along mode 3 to escape the seam. Overall, a substantial part of the wavepacket that reaches this region of the seam continues on the S1 π,π^* state, reaches $(Eth)_X$ and decays to the ground state. This behavior is not changed by adding energy along mode 2, but addition of energy along mode 3 in the direction that induces the planarization of the ring (positive direction in Table 9) leads to a different behavior. Thus, when the seam is encountered at planar structures, the route that leads ultimately to $(Eth)_X$ is not populated, at least in first instance, and the trajectories are bounced back to $(\pi,\pi^*)_{Min}$ or populate the n,π^* state. The different behavior found for the two seam segments in the quantum dynamics runs may explain the difference between the results of our trajectory surface hopping and quantum dynamics studies with a similar trajectory surface hopping dynamics study by Szymczak *et al.*[283] In that study, the trajectories are started at the FC structure and are propagated for more than 2 ps, until they escape $(\pi,\pi^*)_{Min}$ and reach the S_2/S_1 seam. In contrast to our results, all trajectories in that study populate the n,π^* state and do not proceed along the route that leads to the ground state through $(Eth)_X$. This may be due to the different initial conditions of that study, which may have led the trajectories to different regions of the S_2/S_1 seam, resulting in a different outcome. Therefore, the implication of the present results for the interpretation of the dynamics is that, depending on the initial conditions, one may sample different regions of the same seam, and this can have a substantial effect on the results.

The characterization of the S_2/S_1 seam of thymine with static and dynamics calculations is also interesting from a general point of view. The seam is characterized by the sloped-to-peaked topology. This type of topology arises when one of the intersecting states has a relatively large gradient, while the other state is more or less flat and has a negative curvature along a given coordinate. In the case of thymine, the n,π^* has a large gradient and the π,π^* state is flat in the region of the seam, as can be seen from the energy profiles of our propagations. The important point is that thymine is one in a series of molecules where a seam with sloped-to-peaked topology has been characterized.

Other examples are fulvene,[284] benzene,[285] and 5-bromo uracil.[286] In these examples, the intersection is associated with a photochemical reaction in the peaked region, namely double bond isomerization in fulvene, intramolecular cyclization in benzene, and carbon bromine bond cleavage in 5-bromo uracil, while the sloped region leads to recovery of the reactant. Similar to the recent benzene study, the present results for thymine illustrate the different behavior in the two segments at the dynamics level and suggest a way to control the reactivity in coherent control simulations, by controlling the access of the wavepacket to the seam.

5   CONCLUSIONS

In the first part of the thesis, we have shown that the CASSCF method and 6-311G* and 6-31G* basis sets can properly describe the planar structure of thymine. Thus, we have found the level of theory that will be used for the description of the photophysics of thymine. We have also found that some Pople's basis sets used with the MP2 method present problems in describing the planarity of nucleobases. This suggests such problems are common to correlated methods describing planar aromatic rings with Pople's basis sets. We have investigated further the origin of this phenomenon, and concluded that the loss of planarity can be explained in terms of intramolecular BSSE. We have shown that the application of conventional BSSE-correction techniques, such as the Counterpoise method, provide proper assessment and correction whenever spurious results occur and do not produce meaningful effects in those cases already correctly described. We have used a code[287] that allows for flexible definitions of the Counterpoise function. With this code we have been able to address charged systems and determine that BSSE has a delocalized nature. The correction of BSSE for a single fragment can fix the imaginary frequencies, however, only if all fragments are included in the Counterpoise function the frequency values of the correct descriptions are recovered. The fragments' definition and fragments' multiplicity assignment does not affect the performance of the method as long as isolated fragment and associated ghost orbital calculations correspond to the same state with the same orientation of singly-occupied degenerate orbitals.

The photophysics of thymine have been studied with MS-CASPT2 static calculations and CASSCF on-the-fly dynamics. The MS-CASPT2 calculations show there exist two paths that lead from the FC structure to $(Eth)_X$, a conical intersection with the ground state. The indirect path goes through a shallow minimum of the π,π^* state, $(\pi,\pi^*)_{Min}$, which is separated by a barrier from a S_2/S_1 CI, $(n,\pi^*/\pi,\pi^*)_X$. Decay through this CI can lead to either the n,π^* state or $(Eth)_X$, which further leads to the ground state. The direct path is a barrierless path from the FC to $(Eth)_X$. We propose that both paths are populated during the decay since they have similar slopes at FC region. Therefore we propose that the experimental subpicosecond decay component(s) come from the two groups of trajectories (direct decay and indirect decay staying on the π,π^* state). In addition to that, we assign the picosecond component to the n,π^* state which is populated via $(n,\pi^*/\pi,\pi^*)_X$.

We have simulated the decay along the indirect path with semiclassical dynamics. The results confirm the role proposed for this path at the MS-

CASPT2 level. That is, part of the population of the indirect path is responsible for the picosecond since it is funneled to the n,π^* state at $(n,\pi^*/\pi,\pi^*)_X$, whereas the rest of the population decays in the subpicosecond range.

Due to $(n,\pi^*/\pi,\pi^*)_X$ is a key point in the photophysics of thymine, a topological analysis of this seam has been carried out. This study shows it forms a sloped-to-peaked topology and that all parts are accessible. The quantum dynamics simulations carried out with the DD-vMCG method show that the segment of the seam that is reached during the decay has a large influence on the photophysics. In general it is observed that the peaked region of the seam favors the regeneration of the ground state, whereas the sloped one delays the deactivation as this region is responsible for trappings at $(\pi.\pi^*)_{Min}$ and the n,π^* state.

The DD-vMCG method has been applied to the present study of thymine and all 39 degrees of freedom have been taken into account. Problems associated with the use of a small active space, which limits the number of Gaussian centered wavefunctions that form the wavepacket, and the appearance of intruder states that invalidate the diabatic transformation at some points, have been encountered. However, the method still can be used to get a qualitative, mechanistic insight into the photophysics of thymine.

6   BIBLIOGRAPHY

(1)   Miescher, F. Medizinisch-chemische Untersuchungen **1871**, 4, 502-509.

(2)   Watson, J. D.; Crick, F. H. C. Nature **1953**, 171, 737-738.

(3)   Hall, S. S. Science **1993**, 259, 1532-1533.

(4)   Sayre, A. Rosalind Franklin and DNA; W.W. Norton & Co. Inc.: New York, 1975.

(5)   "Tribute to a Forgotten Pioneer"; In New Scientist; **1987**; Vol. 115, p. 71-72

(6)   DNA   Molecule: Two   views, http://www.accessexcellence.org/AE/AEC/CC/images/DNA_views.gif

(7)   Sinha, R. P.; Hader, D. P. Photochem. Photobiol. Sci. **2002**, 1, 225-236.

(8)   Blum, H. Carcinogenesis by Ultraviolet light. ; Princeton University Press: Princeton, , 1959.

(9)   Becker, D.; Sevilla, M. D. "The Chemical Consequences Of Radiation-Damage To Dna"; In Advances In Radiation Biology **1993**; Vol. 17, pag. 121-180.

(10)   Brash, D. E. Photochem. & Photobiol. **1988**, 48, 59-66.

(11)   Forbes, P. 2nd Annual Meeting of American Chemical Society for Photobiology, Book of Abstracts **1974**, 102.

(12)   Setlow, R. B.; Carrier, W. L.; Bollum, F. J. Proc. Natl. Acad. Sci. U. S. A. **1965**, 53, 1111.

(13)   Smith, K.; Hanawalt, D. C. Molecular Photobiology; Academic: New York 1969.

(14)   Beukers, R.; Berends, W. Biochimica Et Biophysica Acta **1960**, 41, 550-551.

(15)   Hearst, J. E. Annu. Rev. Biophys. Biochem. **1981**, 10, 69-86.

(16)   Thymine   UV   induced   photoproducts http://www.medicalecology.org/atmosphere/a_app_strat.htm

(17)   Absorption   and   Emission   Spectrum, http://people.whitman.edu/~sofiauj/classes/astro110.html

(18)   Kasha, M.

Faraday Discussions **1950**, 14-19.

(19)   Condon, E. Phys. Rev. **1926**, 28, 1182-1201.

(20)   Franck, J. Transactions of the Faraday Society **1926**, 21, 536-542.

(21)   Molecular   Expressions   website.   Physics   of   Light   and   Color, http://www.microscopy.fsu.edu/primer/java/jablonski/lightandcolor/

(22)   Daniels, M.; Hauswirt.W Science **1971**, 171, 675.

(23)   Kang, H.; Lee, K. T.; Jung, B.; Ko, Y. J.; Kim, S. K. J. Amer. Chem. Soc. **2002**, 124, 12958-12959.

(24)   He, Y. G.; Wu, C. Y.; Kong, W. J. Phys. Chem. A **2003**, 107, 5145-5148.

(25)   Michl, J. Mol. Photochem. **1972**, 4, 243, 257, 287.

(26) Van der Lugt, W. T. A. M.; Oosterhoff, L. J. J. Amer. Chem. Soc. **2002**, 91, 6042-6049.

(27) Dirac, P. A. M. Proceedings of the Royal Society of London Series A-Mathematical and Physical Sciences **1927**, 114, 243-265

(28) Fermi, E. Nuclear Physics; University of Chicago Press: Chicago, 1950.

(29) Landau, L. Physik. Z. **1932**, 2, 46-51.

(30) Zener, C. Proceedings of the Royal Society of London Series A-Mathematical and Physical Sciences **1932**, 137, 696-702.

(31) Hund, F. Z. Physik. **1927**, 40, 742.

(32) Teller, E. J. Phys. Chem. **1936**, 41, 109-116.

(33) von Neumann, I.; Wigner, E. Physik. Z. **1929**, 30, 467.

(34) Atchity, G. J.; Xantheas, S. S.; Ruedenberg, K. J. Chem. Phys. **1991**, 95, 1862-1876.

(35) Köppel, H.; Barentzen, H.; Yarkony, D. "The Jahn-Teller effect - Advances and Perspectives"; In Springer Series in Chemical Physics; Springer in press (expected year of publication: 2009); Vol. 97.

(36) Paterson, M. J.; Bearpark, M. J.; Robb, M. A.; Blancafort, L. J. Chem. Phys. **2004**, 121, 11562-11571.

(37) Paterson, M. J.; Bearpark, M. J.; Robb, M. A.; Blancafort, L.; Worth, G. Phys. Chem. Chem. Phys. **2005**, 121, 2100-2115.

(38) Sicilia, F.; Blancafort, L.; Bearpark, M. J.; Robb, M. A. J. Phys. Chem. A **2007**, 111, 2182-2192.

(39) Meyer, H.-D.; Gatti, F.; Worth, G. A. Multidimensional Quantum Dynamics. MCTDH Theory and Applications; 1 ed.; Wiley-VCH: Weinheim, 2009.

(40) Sicilia, F.; Blancafort, L.; Bearpark, M. J.; Robb, M. A. J. Chem. Theor. Comput. **2008**, 4, 257-266.

(41) Bearpark, M. J.; Robb, M. A.; Schlegel, H. B. Chem. Phys. Lett. **1994**, 223, 269-274.

(42) Koga, N.; Morokuma, K. Chem. Phys. Lett. **1985**, 119, 371-374.

(43) Manaa, M. R.; David, R. Y. J. Chem. Phys. **1993**, 99, 5251-5256.

(44) Ragazos, I. N.; Robb, M. A.; Bernardi, F.; Olivucci, M. Chem. Phys. Lett. **1992**, 197, 217-223.

(45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B. et al. Gaussian 03, Revision B.02; Gaussian, Inc.: Pittsburgh PA, 2003.

(46) Fukui, K. Acc. Chem. Res. **1981**, 14, 363-368.

(47) Gonzalez, C.; Schlegel, H. B. J. Chem. Phys. **1989**, 90, 2154-2161.

(48) Gonzalez, C.; Schlegel, H. B. J. Phys. Chem. **1990**, 94, 5523-5527.

(49) Kutateladze, A. G. Computational Methods in Photochemistry; CRC Press: New York, 2005.

(50) Andersson, K.; Malmqvist, P. A.; Roos, B. O. J. Chem. Phys. **1992**, 96, 1218-1226.

(51)   Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K.
        J. Phys. Chem. **1990**, 94, 5483-5488.

(52)   Cimiraglia, R.; Persico, M. J. Comput. Chem. **1987**, 8, 39-47.

(53)   Foresman, J. B.; Headgordon, M.; Pople, J. A.; Frisch, M. J. J. Phys.
        Chem. **1992**, 96, 135-149.

(54)   Grimme, S.; Waletzke, M. J. Chem. Phys. **1999**, 111, 5645-5655.

(55)   Huron, B.; Malrieu, J. P.; Rancurel, P. J. Chem. Phys. **1973**, 58, 5745-
        5759.

(56)   Cizek, J. J. Chem. Phys. **1966**, 45, 4256-4266.

(57)   Roos, B. O. Acc. Chem. Res. **1999**, 32, 137-144.

(58)   Cederbaum, L. S.; Yarkony, D. R.; Köppel, H. et al. Conical
        Intersections: Electronic Structutre, Dynamics & Spectroscopy; World
        Scientific Publ Co Pte Ltd: Singapore, 2004; Vol. 15.

(59)   Klessinger, M.; Michl, J. Excited states and photochemistry of organic
        molecules; VCH Publishing: New York 1994.

(60)   Bolton, K.; Hase, W. L.; Peslherbe, G. H. "Direct dynamics simulations
        of reactive systems"; In Modern methods for multi-dimensional dynamics
        computations in chemistry; Thompson, D. L., Ed.; World Scientic:
        Singapore, **1998**, pag. 143-149.

(61)   Truhlar, D. G. "Direct dynamics method for the calculation of reaction
        rates"; In The reaction path in chemistry; Heidrich, D., Ed.; Kluwer
        Academic Publishers: Dordrecht, **1995**, pag. 229-255.

(62)   Worth, G. A.; Robb, M. A. Adv. Chem. Phys. **2002**, 124, 355-431.

(63)   Donoso, A.; Kohen, D.; Martens, C. C. J. Chem. Phys. **2000**, 112, 7345-
        7354.

(64)   Domcke, W.; Yarkony, D. R.; Köppel, H. Conical Intersections:
        Electronic Structure, Dynamics & Spectroscopy; World Scientific:
        Singapore, 2004; Vol. 15.

(65)   Brauns, E. B.; Madaras, M. L.; Coleman, R. S.; Murphy, C. J.; Berg, M.
        A. J. Amer. Chem. Soc. **1999**, 121, 11644-11649.

(66)   Brauns, E. B.; Madaras, M. L.; Coleman, R. S.; Murphy, C. J.; Berg, M.
        A. Phys. Rev. Lett. **2002**, 88.

(67)   Brauns, E. B.; Murphy, C. J.; Berg, M. A. J. Amer. Chem. Soc. **1998**,
        120, 2449-2456.

(68)   Hess, S.; Davis, W. B.; Voityuk, A. A.; Rosch, N.; Michel-Beyerle, M. E.;
        Ernsting, N. P.; Kovalenko, S. A.; Lustres, J. L. P. Chem. Phys. Chem
        2002, 3, 452.

(69)   Nikogosyan, D. N.; Letokhov, V. S. Riv. Nuovo Cimento **1983**, 6, 1-72.

(70)   Ballini, J. P.; Daniels, M.; Vigny, P. J. Lum. **1982**, 27, 389-400.

(71)   Kobayashi, S.; Yamashita, M.; Sato, T.; Muramatsu, S. Ieee Journal of
        Quantum Electronics **1984**, 20, 1383-1385.

(72)   Nikogosyan, D. N.; Angelov, D.; Soep, B.; Lindqvist, L. Chem. Phys.
        Lett. **1996**, 252, 322-326.

(73)   Reuther, A.; Nikogosyan, D. N.; Laubereau, A. J. Phys. Chem. **1996**, 100, 5570-5577.

(74)   Haupl, T.; Windolph, C.; Jochum, T.; Brede, O.; Hermann, R. Chem. Phys. Lett. **1997**, 280, 520-524.

(75)   Pecourt, J. M. L.; Peon, J.; Kohler, B. J. Amer. Chem. Soc. **2001**, 123, 10370-10378.

(76)   Fujiwara, T.; Kamoshida, Y.; Morita, R.; Yamashita, M. J. Photochem. Photobiol **1997**, 41, 114-121.

(77)   Pecourt, J. M. L.; Peon, J.; Kohler, B. J. Amer. Chem. Soc. **2000**, 122, 9348-9349.

(78)   Gustavsson, T.; Sharonov, A.; Markovitsi, D. Chem. Phys. Lett. **2002**, 351, 195-200.

(79)   Gustavsson, T.; Sharonov, A.; Onidas, D.; Markovitsi, D. Chem. Phys. Lett. **2002**, 356, 49-54.

(80)   Onidas, D.; Markovitsi, D.; Marguet, S.; Sharonov, A.; Gustavsson, T. J. Phys. Chem. B **2002**, 106, 11367-11374.

(81)   Hare, P. M.; Crespo-Hernandez, C. E.; Kohler, B. Proc. Natl. Acad. Sci. U. S. A. **2007**, 104, 435-440.

(82)   Brady, B. B.; Peteanu, L. A.; Levy, D. H. Chem. Phys. Lett. **1988**, 147, 538-543.

(83)   Fujii, M.; Tamura, T.; Mikami, N.; Ito, M. Chem. Phys. Lett. **1986**, 126, 583-587.

(84)   Kang, H.; Lee, K. T.; Kim, S. K. Chem. Phys. Lett. **2002**, 359, 213-219.

(85)   He, Y. G.; Wu, C. Y.; Kong, W. J. Phys. Chem. A **2004**, 108, 943-949.

(86)   Ullrich, S.; Schultz, T.; Zgierski, M. Z.; Stolow, A. J. Amer. Chem. Soc. **2004**, 126, 2262-2263.

(87)   Ullrich, S.; Schultz, T.; Zgierski, M. Z.; Stolow, A. Phys. Chem. Chem. Phys. **2004**, 6, 2796-2801.

(88)   Canuel, C.; Mons, M.; Piuzzi, F.; Tardivel, B.; Dimicoli, I.; Elhanine, M. J. Chem. Phys. **2005**, 122, 074316.

(89)   Oraevsky, A. A.; Sharkov, A. V.; Nikogosyan, D. N. Chem. Phys. Lett. **1981**, 83, 276-280.

(90)   Satzger, H.; Townsend, D.; Zgierski, M. Z.; Patchkovskii, S.; Ullrich, S.; Stolow, A. Proc. Natl. Acad. Sci. U. S. A. **2006**, 103, 10196-10201.

(91)   Crespo-Hernandez, C. E.; Cohen, B.; Hare, P. M.; Kohler, B. Chem. Rev. **2004**, 104, 1977-2019.

(92)   Stevens, B. Nature **1961**, 192, 725.

(93)   Bircks, J. B. Photophysics of aromatic molecules New York, 1970.

(94)   Vigny, P. Comptes Rendus Hebdomadaires Des Seances De L'Academie Des Sciences Serie D **1973**, 277, 1941-1944.

(95)   Rhodes, W. J. Amer. Chem. Soc. **1961**, 83, 3609.

(96)   Tinoco, I. J. Amer. Chem. Soc. **1960**, 82, 4785-4790.

(97) Gador, N.; Samoylova, E.; Smith, V. R.; Stolow, A.; Rayner, D. M.; Radloff, W. G.; Hertel, I. V.; Schultz, T. J. Phys. Chem. A **2007**, 111, 11743-11749.

(98) Samoylova, E.; Lippert, H.; Ullrich, S.; Hertel, I. V.; Radloff, W.; Schultz, T. J. Amer. Chem. Soc. **2005**, 127, 1782-1786.

(99) Samoylova, E.; Schultz, T.; Hertel, I. V.; Radloff, W. Chem. Phys. **2008**, 347, 376-382.

(100) Crespo-Hernandez, C. E.; Cohen, B.; Kohler, B. Nature **2005**, 436, 1141-1144.

(101) Löwdin, P. O. Rev. Mod. Phys. **1963**, 35, 724.

(102) Abo-Riziq, A.; Grace, L.; Nir, E.; Kabelac, M.; Hobza, P.; de Vries, M. S. Proc. Natl. Acad. Sci. U. S. A. **2005**, 102, 20-23.

(103) Schultz, T.; Samoylova, E.; Radloff, W.; Hertel, I. V.; Sobolewski, A. L.; Domcke, W. Science **2004**, 306, 1765-1768.

(104) Crespo-Hernandez, C. E.; Cohen, B.; Kohler, B. Nature **2006**, 441, E8-E8.

(105) Crespo-Hernandez, C. E.; de La Harpe, K.; Kohler, B. J. Amer. Chem. Soc. **2008**, 130, 10844.

(106) Markovitsi, D.; Sharonov, A.; Onidas, D.; Gustavsson, T. Chem. Phys. Chem **2003**, 4, 303.

(107) Markovitsi, D.; Talbot, F.; Gustavsson, T.; Onidas, D.; Lazzarotto, E.; Marguet, S. Nature **2006**, 441, E7-E7.

(108) Plessow, R.; Brockhinke, A.; Eimer, W.; Kohse-Hoinghaus, K. J. Phys. Chem. B **2000**, 104, 3695-3704.

(109) Schwalb, N. K.; Temps, F. Science **2008**, 322, 243-245.

(110) Eisinger, J.; Gueron, M.; Shulman, R. G.; Yamane, T. Proc. Natl. Acad. Sci. U. S. A. **1966**, 55, 1015-&.

(111) Lamola, A. A. Photochem. & Photobiol. **1968**, 7, 619.

(112) Pechenaya, V. I.; Danilov, V. I.; Slyusarchuk, O. N.; Alderfer, J. L. Photochem. & Photobiol. **1995**, 61, 435-441.

(113) Vigny, P.; Ballini, J. P. Excited states in organic chemistry and biochemistry, ; D. Reidel:Dordrecht: The Netherlands 1977.

(114) Crespo-Hernandez, C. E.; Kohler, B. J. Phys. Chem. B **2004**, 108, 11182-11188.

(115) Schreier, W. J.; Schrader, T. E.; Koller, F. O.; Gilch, P.; Crespo-Hernandez, C. E.; Swaminathan, V. N.; Carell, T.; Zinth, W.; Kohler, B. Science **2007**, 315, 625-629.

(116) Buchvarov, I.; Wang, Q.; Raytchev, M.; Trifonov, A.; Fiebig, T. Proc. Natl. Acad. Sci. U. S. A. **2007**, 104, 4794-4797.

(117) Takaya, T.; Su, C.; de La Harpe, K.; Crespo-Hernandez, C. E.; Kohler, B. Proc. Natl. Acad. Sci. U. S. A. **2008**, 105, 10285-10290.

(118) Kwok, W. M.; Ma, C. S.; Phillips, D. L. J. Amer. Chem. Soc. **2006**, 128, 11894-11905.

(119) Markovitsi, D.; Gustavsson, T.; Talbot, F. Photochem. Photobiol. Sci. **2007**, 6, 717-724.

(120) Frenkel, J. Physical Reviews **1931**, 37, 17-24.

(121) Georghiou, S.; Nordlund, T. M.; Saim, A. M. Photochem. & Photobiol. **1985**, 41, 209-212.

(122) Georghiou, S.; Zhu, S.; Weidner, R.; Huang, C. R.; Ge, G. J. Biomol. Struct. Dyn. **1990**, 8, 657-674.

(123) Lochbrunner, S.; Larsen, J. J.; Shaffer, J. P.; Schmitt, M.; Schultz, T.; Underwood, J. G.; Stolow, A. Journal of Electron Spectroscopy and Related Phenomena **2000**, 112, 183-198.

(124) Mestdagh, J. M.; Visticot, J. P.; Elhanine, M.; Soep, B. J. Chem. Phys. **2000**, 113, 237-248.

(125) Leszczynski, J. Int. J. Quantum Chem. **1992**, 19, 43-55.

(126) Sponer, J.; Hobza, P. J. Phys. Chem. **1994**, 98, 3161-3164.

(127) Stewart, E. L.; Foley, C. K.; Allinger, N. L.; Bowen, J. P. J. Amer. Chem. Soc. **1994**, 116, 7282-7286.

(128) Shukla, M. Chem. Phys. **1999**, 240, 319-329.

(129) Matsika, S. J. Phys. Chem. A **2004**, 108, 7584-7590.

(130) Perun, S.; Sobolewski, A. L.; Domcke, W. J. Phys. Chem. A **2006**, 110, 13238-13244.

(131) Merchan, M.; Gonzalez-Luque, R.; Climent, T.; Serrano-Andres, L.; Rodriuguez, E.; Reguero, M.; Pelaez, D. J. Phys. Chem. B **2006**, 110, 26471-26476.

(132) Zgierski, M. Z.; Patchkovskii, S.; Fujiwara, T.; Lim, E. C. J. Phys. Chem. A **2005**, 109, 9384-9387.

(133) Bonačić-Koutecký, V.; Koutecký, J.; Michl, J. Angw. Chem. Int. Ed. Engl. **1987**, 26, 170-189.

(134) Freund, L.; Klessinger, M. Int. J. Quantum Chem. **1998**, 70, 1023-1028.

(135) Ben-Nun, M.; Martínez, T. J. J. Chem. Phys. **2000**, 259, 237-248.

(136) Barbatti, M.; Paier, J.; Lischka, H. J. Chem. Phys. **2004**, 121, 11614-11624.

(137) Serrano-Perez, J. J.; Gonzalez-Luque, R.; Merchan, M.; Serrano-Andres, L. J. Phys. Chem. B **2007**, 111, 11880-11883.

(138) Hudock, H. R.; Levine, B. G.; Thompson, A. L.; Satzger, H.; Townsend, D.; Gador, N.; Ullrich, S.; Stolow, A.; Martinez, T. J. J. Phys. Chem. A **2007**, 111, 8500-8508.

(139) Lan, Z.; Fabiano, E.; Thiel, W. J. Phys. Chem. B **2009**, 113, 3548-3555.

(140) Moran, D.; Simmonett, A. C.; Leach, F. E.; Allen, W. D.; Schleyer, P. V.; Schaefer, H. F. J. Amer. Chem. Soc. **2006**, 128, 9342-9343.

(141) Moller, C.; Plesset, M. S. Phys. Rev. **1934**, 46, 618.

(142) Cizek, J. Adv. Chem. Phys. **1969**, 35-89.

(143) Cizek, J.; Paldus, J. Int. J. Quantum Chem. **1971**, 5, 359.

(144) Hartree, D. Proceedings of the Cambridge Phylosophical Society **1928**, 24, 111.

(145) Becke, A. D. Phys. Rev. A **1988**, 38, 3098-3100.

(146) Lee, C. T.; Yang, W. T.; Parr, R. G. Phys. Rev. B **1988**, 37, 785-789.

(147) Becke, A. D. J. Chem. Phys. **1993**, 98, 5648-5652.

(148) Goodman, L.; Ozkabak, A. G.; Thakur, S. N. J. Phys. Chem. **1991**, 95, 9044-9058.

(149) Martin, J. M. L.; Taylor, P. R.; Lee, T. J. Chem. Phys. Lett. **1997**, 275, 414-422.

(150) Dkhissi, A.; Adamowicz, L.; Maes, G. J. Phys. Chem. A **2000**, 104, 2112-2119.

(151) Lampert, H.; Mikenda, W.; Karpfen, A. J. Phys. Chem. A **1997**, 101, 2254-2263.

(152) Michalska, D.; Zierkiewicz, W.; Bienko, D. C.; Wojciechowski, W.; Zeegers-Huyskens, T. J. Phys. Chem. A **2001**, 105, 8734-8739.

(153) Saeki, M.; Akagi, H.; Fujii, M. J. Chem. Theor. Comput. **2006**, 2, 1176-1183.

(154) Torii, H.; Ishikawa, A.; Takashima, R.; Tasumi, M. J. Mol. Struct. (TheoChem) **2000**, 500, 311-321.

(155) Simandiras, E. D.; Rice, J. E.; Lee, T. J.; Amos, R. D.; Handy, N. C. J. Chem. Phys. **1988**, 88, 3187-3195.

(156) Shahbazian, S. Chem. Phys. Lett. **2007**, 443, 147-151.

(157) Sellers, H.; Almlof, J. J. Phys. Chem. **1989**, 93, 5136-5139.

(158) Jensen, F. J. Chem. Phys. **2003**, 118, 2459-2463.

(159) Jansen, H. B.; Ros, P. Chem. Phys. Lett. **1969**, 3, 140-143.

(160) Vanduijneveldt, F. B.; Vanduijneveldtvanderijdt, J.; Vanlenthe, J. H. Chem. Rev. **1994**, 94, 1873-1885.

(161) Halasz, G. J.; Vibok, A.; Suhai, S.; Mayer, I. Int. J. Quantum Chem. **2002**, 89, 190-197.

(162) Jensen, F. Chem. Phys. Lett. **1996**, 261, 633-636.

(163) Kobko, N.; Dannenberg, J. J. J. Phys. Chem. A **2001**, 105, 1944-1950.

(164) Salvador, P.; Duran, M.; Dannenberg, J. J. J. Phys. Chem. A **2002**, 106, 6883-6889.

(165) Csontos, J.; Palermo, N. Y.; Murphy, R. F.; Lovas, S. J. Comput. Chem. **2008**, 29.

(166) Holroyd, L. F.; van Mourik, T. Chem. Phys. Lett. **2007**, 442, 42-46.

(167) Palermo, N. Y.; Csontos, J.; Owen, M. C.; Murphy, R. F.; Lovas, S. J. Comput. Chem. **2007**, 28, 1208-1214.

(168) van Mourik, T. J. Comput. Chem. **2008**, 29, 1-3.

(169) Torrent-Sucarrat, M.; Geerlings, P.; Luis, J. M. Chem. Phys. Chem **2007**, 8, 1065-1070.

(170) Hobza, P.; Havlas, Z. Theor. Chem. Acc. **1998**, 99, 372-377.

(171) Salvador, P.; Paizs, B.; Duran, M.; Suhai, S. J. Comput. Chem. **2001**, 22, 765-786.

(172) Salvador, P.; Simon, S.; Duran, M.; Dannenberg, J. J. J. Chem. Phys. **2000**, 113, 5666-5674.

(173) Kobayashi, R. J. Phys. Chem. A **1998**, 102, 10813-10817.

(174) Broo, A.; Holmen, A. J. Phys. Chem. A **1997**, 101, 3589-3600.

(175) Cai, Z. L.; Sha, G. H.; Zhang, C. H.; Huang, M. B. Chem. Phys. Lett. **1991**, 178, 273-278.

(176) Colominas, C.; Luque, F. J.; Orozco, M. J. Amer. Chem. Soc. **1996**, 118, 6811-6821.

(177) Szabo A., O. N. S. Modern Quantum Chemistry. Introduction to Advanced Electronic Structure Theory; Dover, 1996.

(178) Helgaker, T.; Jorgensen, P.; Olsen, J. Molecular Electronic-Structure Theory; Wiley, 2000.

(179) Roos, B. O.; Widmark, P. O. European Summerschol in Quantum Chemistry; Chemical Centre Printshop: Lund, 2000; Vol. II.

(180) Hehre, W. J.; Radom, L.; Schleyer, P. V.; Pople, J. A. Ab Initio Molecular Orbital Theory; John Wiley & Sons Inc.: New York, 1986.

(181) Yarkony, D. R. Modern Electronic Structure World Scientific: Singapore, 1995; Vol. I.

(182) Schrödinger, E. Ann. Phys. **1926**, 79, 361.

(183) Born, M., Oppenheimer, J. R. Ann. Phys. **1927**, 84, 457-484.

(184) Lichten, W. Phys. Rev. **1963**, 131, 229.

(185) Smith, F. T. Phys. Rev. **1969**, 179, 111.

(186) Pauli, W. Z. Physik. **1925**, 31, 765.

(187) Slater, J. C. Phys. Rev. **1929**, 34, 1293-1322.

(188) Slater, J. C. Phys. Rev. **1930**, 35, 210-211.

(189) Rothenberg. S; Schaefer, H. F. J. Chem. Phys. **1971**, 54, 2764-&.

(190) Huzinaga, S. J. Chem. Phys. **1965**, 42, 1293-&.

(191) Huzinaga, S.; Arnau, C. J. Chem. Phys. **1970**, 53, 348-&.

(192) Huzinaga, S.; Arnau, C. J. Chem. Phys. **1970**, 52, 2224-&.

(193) Huzinaga, S.; Sakai, Y. J. Chem. Phys. **1969**, 50, 1371-&.

(194) Ditchfie.R; Hehre, W. J.; Pople, J. A. J. Chem. Phys. **1971**, 54, 724-&.

(195) Hehre, W. J.; Ditchfie.R; Pople, J. A. J. Chem. Phys. **1972**, 56, 2257-&.

(196) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. J. Chem. Phys. **1980**, 72, 650-654.

(197) Binkley, J. S.; Pople, J. A.; Hehre, W. J. J. Amer. Chem. Soc. **1980**, 102, 939-947.

(198) Frisch, M. J.; Pople, J. A.; Binkley, J. S. J. Chem. Phys. **1984**, 80, 3265-3269.

(199) Hehre, W. J.; Stewart, R. F.; Pople, J. A. J. Chem. Phys. **1969**, 51, 2657-&.

(200) Harihara, P.; Pople, J. A. Theor. Chim. Acta **1973**, 28, 213-222.

(201) Dunning, T. H. J. Chem. Phys. **1970**, 53, 2823-&.

(202) Dunning, T. H. Chem. Phys. Lett. **1970**, 7, 423-427.

(203) Dunning, T. H. J. Chem. Phys. **1971**, 55, 3958-&.

(204) Dunning, T. H. J. Chem. Phys. **1971**, 55, 716-&.

(205) Almlof, J.; Taylor, P. R. J. Chem. Phys. **1987**, 86, 4070-4077.

(206) Gutowski, M.; Chalasinski, G. J. Chem. Phys. **1993**, 98, 5540-5554.

(207) Gutowski, M.; Vanduijneveldtvanderijdt, G. C.; Vanlenthe, J. H.; Vanduijneveldt, F. B. J. Chem. Phys. **1993**, 98, 4728-4738.

(208) Liu, B.; McLean, A. D. J. Chem. Phys. **1973**, 59, 4557-4558.

(209) Masamura, M. Theor. Chem. Acc. **2001**, 106, 301-313.

(210) van Mourik, T.; Wilson, A. K.; Peterson, K. A.; Woon, D. E.; Dunning, T. H. "The effect of basis set superposition error (BSSE) on the convergence of molecular properties calculated with the correlation consistent basis sets"; In Adv. Quantum Chem. **1999**; Vol. 31, pag. 105-135.

(211) Boys, S. F.; Bernardi, F. Mol. Phys. **1970**, 19, 553-566.

(212) Iwata, S.; Nagata, T. Theor. Chem. Acc. **2007**, 117, 137-144.

(213) Mayer, I. Int. J. Quantum Chem. **1983**, 23, 341-363.

(214) Nagata, T.; Iwata, S. J. Chem. Phys. **2004**, 120, 3555-3562.

(215) Simon, S.; Duran, M.; Dannenberg, J. J. J. Chem. Phys. **1996**, 105, 11024-11031.

(216) Hartree, D. R. Physik. Z. **1929**, 30, 517.

(217) Fock, V. Z. Physik. **1930**, 61, 126.

(218) Roothaan, C. C. J. Rev. Mod. Phys. **1951**, 23, 69-89.

(219) Pople, J. A.; Binkley, J. S.; Seeger, R. Int. J. Quantum Chem. **1976**, 1-19.

(220) Boys, S. F. Proceedings of the Royal Society of London Series A-Mathematical and Physical Sciences **1950**, 201, 125-137.

(221) Boys, S. F. Proceedings of the Royal Society of London Series A-Mathematical and Physical Sciences **1950**, 200, 542-554.

(222) Brillouin, L. Actualities Sci. Ind. **1933**, 71.

(223) Brillouin, L. Actualities Sci. Ind. **1934**, 159.

(224) Bartlett, R. J. Annu. Rev. Phys. Chem. **1981**, 32, 359.

(225) Malmqvist, P. A.; Rendell, A.; Roos, B. O. J. Phys. Chem. **1990**, 94, 5477-5482.

(226) Olsen, J.; Roos, B. O.; Jorgensen, P.; Jensen, H. J. A. J. Chem. Phys. **1988**, 89, 2185-2192.

(227) Golab, J. T.; Yeager, D. L. J. Chem. Phys. **1987**, 87, 2925-2944.

(228) Davidson, E. R. J. Comput. Phys. **1975**, 17, 87-94.

(229) Roos, B. O.; Andersson, K. Chem. Phys. Lett. **1995**, 245, 215-223.

(230) Forsberg, N.; Malmqvist, P. A. Chem. Phys. Lett. **1997**, 274, 196-204.

(231)   Ghigo, G.; Roos, B. O.; Malmqvist, P. A. Chem. Phys. Lett. **2004**, 396,
        142-149.

(232)   Malrieu, J. P.; Heully, J. L.; Zaitsevskii, A. Theor. Chim. Acta **1995**, 90,
        167-187.

(233)   Serrano-Andres, L.; Merchan, M.; Nebotgil, I.; Lindh, R.; Roos, B. O. J.
        Chem. Phys. **1993**, 98, 3151-3162.

(234)   Tully, J. C.; Preston, R. K. J. Chem. Phys. **1971**, 55, 562-572.

(235)   Gear, C. W. Numerical Initial Value Problems in Ordinary Dierential
        Equations; Prentice-Hall: Englewood Clis, U.S.A., 1971.

(236)   Verlet, L. Phys. Rev. **1967**, 159, 98-103.

(237)   Chen, W.; Hase, W. L.; Schlegel, H. B. Chem. Phys. Lett. **1994**, 228, 436-
        442.

(238)   Helgaker, T.; Uggerud, E.; Jensen, H. J. A. Chem. Phys. Lett. **1990**, 173,
        145-150.

(239)   Bofill, J. M. J. Comput. Chem. **1994**, 15, 1-11.

(240)   Bofill, J. M. Chem. Phys. Lett. **1996**, 260, 359-364.

(241)   Bofill, J. M.; Comajuan, M. J. Comput. Chem. **1995**, 16, 1326-1338.

(242)   Bjerre, A.; Nikitin, E. E. Chem. Phys. Lett. **1967**, 1, 179-181.

(243)   Tully, J. C. J. Chem. Phys. **1990**, 93, 1061.

(244)   Herman, M. F. Annu. Rev. Phys. Chem. **1994**, 45, 83-111.

(245)   Tully, J. C. J. Chem. Phys. **1990**, 93, 1061.

(246)   Heller, E. J. J. Chem. Phys. **1975**, 62, 1544.

(247)   Feit, M. D.; Fleck Jr., J. A. J. Chem. Phys. **1983**, 78, 301.

(248)   Feit, M. D.; Fleck Jr., J. A. J. Chem. Phys. **1984**, 80, 2578.

(249)   Feit, M. D.; Fleck Jr., J. A.; Steiger, A. J. Comput. Phys. **1982**, 47, 412.

(250)   Park, T. J.; Light, J. C. J. Chem. Phys. **1986**, 85, 5870.

(251)   Tal-Ezer, H.; Kosloff, R. J. Chem. Phys. **984**, 81, 3967.

(252)   Dirac, P. A. M. Proc. Cambridge Philos. Soc. **1930**, 26, 376.

(253)   Frenkel, J. Wave Mechanics; Claredon Press: Oxford, 1934.

(254)   McLachlan, A. D. Mol. Phys. **1964**, 8, 39.

(255)   Meyer, H. D.; Manthe, U.; Cederbaum, L. S. Chem. Phys. Lett. **1990**,
        165, 73.

(256)   Beck, M. H.; Jackle, A.; Worth, G. A.; Meyer, H. D. Phys. Rep.-Rev.
        Sec. Phys. Lett. **2000**, 324, 1-105.

(257)   Worth, G.; Meyer, H. D.; Cederbaum, L. S. J. Chem. Phys. **1996**, 105,
        4412.

(258)   Worth, G.; Robb, M. A.; Burghardt, I. Faraday Discussions **2004**, 127,
        307.

(259)   Burghardt, I.; Meyer, H. D.; Cederbaum, L. S. J. Chem. Phys. **1999**, 111,
        2927.

(260)   Worth, G.; Burghardt, I. Chem. Phys. Lett. **2003**, 368, 502-508.

(261) Köppel, H.; Gronki, J.; Mahapatra, S. J. Chem. Phys. **2001**, 115, 2377-2388.

(262) Köppel, H.; Schubert Mol. Phys. **2006**, 104, 1069.

(263) Thiel, A.; Koppel, H. J. Chem. Phys. **1999**, 110, 9371-9383.

(264) Handy, N. C.; Murray, C. W.; Amos, R. D. J. Phys. Chem. **1993**, 97, 4392-4396.

(265) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B. et al. Gaussian 03, Revision C.01; Gaussian, Inc.: Pittsburgh, PA, 2003.

(266) Mercier, Y.; Santoro, F.; Reguero, M.; Improta, R. J. Phys. Chem. B **2008**, 112, 10769-10772.

(267) Zechmann, G.; Barbatti, M. J. Phys. Chem. A **2008**, 112, 8273-8279.

(268) Hunt, P. A.; Robb, M. A. J. Amer. Chem. Soc. **2005**, 127, 5720-5726.

(269) Paterson, M. J.; Hunt, P.; Robb, M. A.; Takahashi, O. J. Phys. Chem. A **2002**, 106, 10494-10504.

(270) Weingart, O.; Migani, A.; Olivucci, M.; Robb, M. A.; Buss, V.; Hunt, P. J. Phys. Chem. A **2004**, 108, 4685-4693.

(271) Worth, G. A.; Hunt, P.; Robb, M. A. J. Phys. Chem. A **2003**, 107, 621-631.

(272) Barbatti, M.; Lischka, H. J. Amer. Chem. Soc. **2008**, 130, 6831-6839.

(273) Celani, P.; Robb, M. A.; Garavelli, M.; Bernardi, F.; Olivucci, M. Chem. Phys. Lett. **1995**, 243, 1-8.

(274) Yamamoto, N.; Vreven, T.; Robb, M. A.; Frisch, M. J.; Schlegel, H. B. Chem. Phys. Lett. **1996**, 250, 373-378.

(275) Karlstrom, G.; Lindh, R.; Malmqvist, P. A. et al. Comput. Mater. Sci. **2003**, 28, 222-239.

(276) Lorentzon, J.; Fulscher, M. P.; Roos, B. O. J. Amer. Chem. Soc. **1995**, 117, 9265-9273.

(277) Warshel, A. Chem. Phys. Lett. **1975**, 32, 11-17.

(278) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B. et al. Gaussian 03, Developement version (Revision B.07); Gaussian, Inc.: Pittsburgh, PA, 2003.

(279) Blancafort, L.; Hunt, P.; Robb, M. A. J. Amer. Chem. Soc. **2005**, 127, 3391-3399.

(280) Tannor, D. J. Chem. Phys. **1986**, 85, 5805-5820.

(281) Worth, G. A.; Beck, M. H.; Jäckle, A.; Meyer, H. D. The MCTDH Package; Development version 9.0; University of Heidelberg: Heidelberg, 2006.

(282) Yarkony, D. R. Faraday Discussions **2004**, 127, 325-336.

(283) Szymczak, J. J.; Barbatti, M.; Soo Hoo, J. T.; Adkins, J. A.; Windus, T. L.; Nachtigallová, D.; Lischka, H. J. Phys. Chem. A **2009**, ASAP.

(284) Bearpark, M. J.; Blancafort, L.; Paterson, M. J. Mol. Phys. **2006**, 104, 1033-1038.

(285) Lasorne, B.; Bearpark, M. J.; Robb, M. A.; Worth, G. A. J. Phys. Chem.
       A **2008**, 112, 13017-13027.

(286) Kobylecka, M.; Migani, A.; Asturiol, D.; Rak, J.; Blancafort, L. J. Phys.
       Chem. A **2009**, 113, 5489-5495.

(287) Salvador, P. CP-opt driver for Gaussian98 Girona, 2001.

Appendix I

CP-opt driver: The CP-opt program automatically calls a slightly modified version (for higher accuracy of print out reasons) of Gaussian98 in order to perform either energy, gradient optimizations or frequencies corrected for Basis Set Superposition Error using Counterpoise-type methods. Moreover, the user can define which fragment calculations are necesary in each case in terms of the fragment symmetry of the supermolecule and build any Counterpoise-type function. The package includes several UNIX scripts (drive files) and FORTRAN 77 programs. From a conventional GAUSSIAN input file, 2N+1 input files for each calculation are generated and computed sequentially. Then CP-corrected, energy, gradient or hessian, depending on the calculation requested, is determined by the corresponding linear combination, either authomaticaly or as defined by the user. In case of geometry optimisations, the new point in the CP-corrected PES is calculated externally using a generalized DIIS combined with a variable metric optimizer and the next set of 2N+1 calculations are carried out again and untill the desired convergence.

P. Salvador implemented automatic Counterpoise correction to energy, gradients, second and third derivatives for up to 10 fragments into Gaussian, being readily available since versions Gaussian 98 rev A11. However, no handling of the fragment's symmetry was implemented and generally the NOSYMM keyword is necessary.

References to program packages used in this study:

Gaussian 98, Revision A.7: M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998.

Gaussian 03 Revision B.02: M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K.

N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, 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, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople; Gaussian, Inc., Wallingford CT, 2004.

Full frequencies of the calculations described in this study:

Table AI-1. Uncorrected and Counterpoise-corrected harmonic vibrational frequencies (cm^{-1}) for benzene (D_{6h}) at the Hartree-Fock level of theory. The average difference between uncorrected and CP-corrected frequencies is 1.6% and 1.3% for Hartree-Fock and B3LYP, respectively. All data in the table were computed with Gaussian 03.

Hartree-Fock											
6-31+G(d) CP-corrected		Unc.	% diff	6-311G CP-corrected		Unc.	% diff	6-311++G CP-corrected		Unc.	% diff
E2U	448	451	0.7	E2U	459	463	0.9	E2U	449	462	3.1
E2G	663	663	0.0	E2G	688	683	0.7	E2G	680	682	1.3
A2U	745	760	2.0	A2U	747	771	3.2	A2U	723	772	1.0
B2G	765	775	1.3	B2G	784	794	1.3	B2G	788	803	9.9
E1G	940	960	2.1	E1G	949	978	3.1	E1G	939	983	2.0
A1G	1073	1076	0.3	A1G	1067	1070	0.3	A1G	1091	1140	5.2
E2U	1073	1108	3.3	E2U	1081	1126	4.2	E1U	1064	1185	19.4
B2G	1092	1141	4.5	B2G	1097	1152	5.0	B1U	1129	1067	0.2
B1U	1101	1097	0.4	E1U	1135	1134	0.1	E2U	1130	1123	2.3
E1U	1133	1137	0.4	B1U	1147	1124	2.0	B2U	1122	1132	0.3
B2U	1195	1201	0.5	B2U	1237	1242	0.4	E2G	1230	1243	0.2
E2G	1288	1291	0.2	E2G	1312	1304	0.6	B2U	1285	1302	1.3
B2U	1349	1353	0.3	B2U	1372	1362	0.7	B2G	1337	1361	1.4
A2G	1505	1506	0.1	A2G	1551	1524	1.7	A2G	1506	1521	2.2
E1U	1642	1644	0.1	E1U	1662	1644	1.1	E1U	1627	1640	1.6
E2G	1782	1786	0.2	E2G	1775	1772	0.2	E2G	1755	1766	0.3
B1U	3365	3351	0.4	B1U	3314	3323	0.3	B1U	3341	3313	1.1
E2G	3376	3362	0.4	E2G	3324	3334	0.3	E2G	3351	3326	1.0
E1U	3393	3380	0.4	E1U	3343	3355	0.4	E1U	3396	3347	0.8
A1G	3403	3391	0.4	A1G	3358	3371	0.4	A1G	3383	3361	0.7

Table AI-1 Cont.

B3LYP											
6-31+G(d) CP-corrected		Unc.	% diff	6-311G CP-corrected		Unc.	% diff	6-311++G CP-corrected		Unc.	% diff
E2U	408	412	1.0	E2U	419	421	0.5	E2U	409	418	2.2
E2G	619	621	0.3	E2G	644	640	0.5	E2G	642	639	0.5
A2U	678	689	1.6	A2U	683	704	3.1	A2U	671	703	4.8
B2G	708	712	0.6	B2G	727	736	1.2	B2G	687	740	7.7
E1G	851	863	1.4	E1G	860	884	2.9	E1G	843	887	5.2
E2U	964	981	1.8	E2U	971	1006	3.7	E2U	932	1006	7.9
B2G	994	1014	2.0	B2G	1001	1009	0.8	B2G	940	1046	11.3
A1G	1014	1014	0.0	A1G	1006	1044	3.8	A1G	1004	1006	0.2
B1U	1022	1019	0.3	B1U	1061	1048	1.2	B1U	1058	1054	0.4
E1U	1062	1064	0.2	E1U	1064	1064	0.0	E1U	1062	1062	0.0
B2U	1177	1184	0.6	B2U	1204	1207	0.2	B2U	1208	1206	0.2
E2G	1199	1205	0.5	E2G	1219	1221	0.2	E2G	1220	1219	0.1
B2U	1354	1355	0.1	B2U	1337	1338	0.1	B2U	1337	1337	0.0
A2G	1385	1386	0.1	A2G	1432	1418	1.0	A2G	1428	1415	0.9
E1U	1520	1523	0.2	E1U	1538	1530	0.5	E1U	1534	1526	0.5
E2G	1644	1646	0.1	E2G	1636	1634	0.1	E2G	1631	1630	0.1
B1U	3179	3174	0.2	B1U	3116	3155	1.3	B1U	3148	3149	0.0
E2G	3188	3183	0.2	E2G	3134	3166	1.0	E2G	3159	3161	0.1
E1U	3203	3199	0.1	E1U	3153	3185	1.0	E1U	3177	3180	0.1
A1G	3213	3209	0.1	A1G	3168	3200	1.0	A1G	3190	3194	0.1

Table AI-2. Uncorrected and Counterpoise-corrected harmonic vibrational frequencies (cm^{-1}) for benzene (D$_{6h}$) at the MP2 and CISD levels of theory. The average difference between uncorrected and CP-corrected frequencies is 17.8% and 9.8% for MP2 and CISD, respectively. Excluding the problematic out-of-plane vibrational modes the differences are 4.4% and 4.2%, respectively. All data in the table were computed with Gaussian 03 and Gaussian98.

MP2											
6-31+G(d) CP-corrected		Unc.	% diff	6-311G CP-corrected		Unc.	% diff	6-311++G CP-corrected		Unc.	% diff
E2U	384	379	1.3	E2U	382	333	12.8	E2U	409	470	53.4
B2G	473	182	61.5	B2G	442	722i	263.3	E2G	631	627	0.7
E2G	613	618	0.8	E2G	625	630	0.8	A2U	703	573	18.4
A2U	659	672	2.0	A2U	650	620	4.6	B2G	730	721	1.3
E1G	824	829	0.6	E1G	812	736	9.4	E1G	890	620	30.4
E2U	894	877	1.9	E2U	861	648	24.7	A1G	961	984	2.3
B2G	896	859	4.1	B2G	877	779	11.2	E2U	1015	468i	146.1
A1G	1004	1020	1.6	A1G	966	988	2.3	E1U	1031	1050	1.9
B1U	1026	1018	0.8	E1U	1034	1055	2.0	B2U	1054	1208	14.6
E1U	1061	1077	1.5	B1U	1050	1029	2.0	B2G	1093	1852i	269.4

Table AI-2 Cont.

B2U	1181	1206	2.1	B2U	1183	1210	2.3	B1U	1192	1021	14.3
E2G	1204	1225	1.7	E2G	1203	1225	1.8	E2G	1209	1223	1.1
A2G	1388	1389	0.1	B2U	1330	1362	2.4	B2U	1342	1364	1.6
B2U	1439	1464	1.7	A2G	1423	1410	0.9	A2G	1426	1402	1.7
E1U	1516	1530	0.9	E1U	1515	1516	0.1	E1U	1521	1511	0.7
E2G	1640	1660	1.2	E2G	1593	1615	1.4	E2G	1594	1609	0.9
B1U	3226	3208	0.6	B1U	3129	3134	0.2	B1U	3144	3110	1.1
E2G	3235	3217	0.6	E2G	3139	3145	0.2	E2G	3154	3127	0.9
E1U	3249	3232	0.5	E1U	3156	3165	0.3	E1U	3171	3150	0.6
A1G	3257	3242	0.5	A1G	3170	3180	0.3	A1G	3183	3166	0.5

CISD

6-31+G(d) CP-corrected		Unc.	% diff	6-311G CP-corrected		Unc.	% diff	6-311++G CP-corrected		Unc.	% diff
E2U	426	418	1.9	E2U	382	397	3.9	E2U	455	281	38.2
E2G	639	643	0.6	B2G	442	189i	142.8	E2G	653	650	0.5
A2U	713	720	1.0	E2G	625	652	4.3	A2U	761	652	14.3
B2G	716	591	17.5	A2U	650	678	4.3	B2G	875	801	8.5
E1G	904	899	0.6	E1G	812	830	2.2	E2U	991	666	32.8
E2U	1028	999	2.8	E2U	861	839	2.6	A1G	1011	1031	2.0
A1G	1047	1061	1.3	B2G	877	847	3.4	B1U	1060	1057	0.3
B2G	1049	971	7.4	A1G	966	1034	7.0	E1G	1089	768	29.5
B1U	1063	1055	0.8	E1U	1034	1093	5.7	B2U	1176	1215	3.3
E1U	1099	1115	1.5	B1U	1050	1063	1.2	E1U	1216	1090	10.4
B2U	1190	1218	2.4	B2U	1183	1215	2.7	E2G	1250	1255	0.4
E2G	1241	1261	1.6	E2G	1203	1257	4.5	B2U	1320	1321	0.1
B2U	1335	1358	1.7	B2U	1330	1320	0.8	A2G	1465	1448	1.2
A2G	1446	1445	0.1	A2G	1423	1454	2.2	E1U	1583	1570	0.8
E1U	1582	1594	0.8	E1U	1515	1574	3.9	B2G	1591	1384i	187.0
E2G	1724	1743	1.1	E2G	1593	1697	6.5	E2G	1678	1691	0.8
B1U	3302	3288	0.4	B1U	3129	3205	2.4	B1U	3206	3187	0.6
E2G	3311	3299	0.4	E2G	3139	3216	2.5	E2G	3224	3202	0.7
E1U	3327	3316	0.3	E1U	3156	3237	2.6	E1U	3252	3225	0.8
A1G	3337	3327	0.3	A1G	3170	3254	2.6	A1G	3270	3243	0.8

Table AI-3. Uncorrected and Counterpoise-corrected harmonic vibrational frequencies (cm^{-1}) for naphtalene (D$_{2h}$) with the 6-31G basis set. All data in the table were computed with Gaussian 03.

HF CP-corrected		HF	% diff	B3LYP CP-corrected		B3LYP	% diff	MP2 CP-corrected		MP2	% diff
B3U	188	193	2.9	B3U	176	180	2.2	B3U	159	159	0.0
AU	209	209	0.1	AU	192	192	0.2	AU	181	177	2.3
B1U	397	400	0.7	B1U	373	373	0.1	B1U	363	369	1.6
B1G	436	441	1.0	B1G	401	403	0.7	B1G	368	359	2.7
B2G	533	538	1.0	B2G	490	490	0.0	B3U	446	419	5.9
B3U	542	551	1.5	B3U	495	500	0.9	B2G	450	426	5.3
AG	564	566	0.2	AG	531	531	0.0	AG	516	523	1.3
B3G	572	573	0.1	B3G	530	531	0.2	B3G	522	523	0.2
B2U	700	700	0.0	AU	647	649	0.4	AU	565	450	20.4
AU	701	709	1.1	B2U	654	654	0.1	B2G	570	402i	170.5
B1G	795	831	4.5	B1G	734	748	1.9	B2U	638	642	0.6
AG	837	839	0.2	AG	782	782	0.0	B1G	691	680	1.7
B3U	873	915	4.8	B2G	793	800	0.9	B3U	744	716	3.8
B2G	879	891	1.4	B3U	806	821	2.0	AG	764	769	0.7
B1U	888	885	0.4	B1U	828	826	0.2	AU	804	776	3.5
AU	927	976	5.2	AU	852	873	2.5	B1U	814	819	0.5
B2G	970	1039	7.2	B2G	892	922	3.4	B2G	816	783	4.0
B1G	1055	1122	6.4	B3G	977	973	0.4	B1G	883	838	5.1
B3G	1062	1054	0.7	B1G	950	977	2.8	B3U	886	839	5.3
B3U	1071	1148	7.1	B3U	966	997	3.2	AU	897	821	8.5
AU	1081	1172	8.4	AU	990	1021	3.1	B2G	916	855	6.6
B2G	1086	1183	8.9	B2G	1002	1030	2.8	B3G	967	958	0.9
B2U	1091	1099	0.7	B2U	1046	1052	0.5	B2U	1011	1034	2.3
AG	1122	1127	0.4	AG	1058	1062	0.4	AG	1028	1048	1.9
B2U	1236	1242	0.5	B1U	1172	1175	0.2	B1U	1154	1168	1.2
B1U	1266	1265	0.1	B3G	1194	1200	0.5	B3G	1172	1195	2.0
B3G	1290	1292	0.1	B2U	1206	1216	0.9	B2U	1187	1220	2.7
AG	1313	1313	0.0	AG	1211	1218	0.6	AG	1197	1224	2.3
B2U	1330	1333	0.2	B2U	1255	1260	0.4	B2U	1227	1257	2.4
B3G	1409	1403	0.5	B3G	1306	1303	0.2	B3G	1285	1292	0.5
B1U	1414	1415	0.0	B1U	1316	1320	0.3	B1U	1295	1315	1.5
B2U	1501	1497	0.3	B2U	1421	1419	0.2	B2U	1409	1429	1.4
AG	1513	1519	0.4	AG	1430	1431	0.0	AG	1415	1425	0.7
B1U	1582	1565	1.1	B1U	1464	1453	0.8	B1U	1447	1441	0.4
B3G	1645	1636	0.5	B3G	1525	1520	0.3	B3G	1495	1501	0.3
AG	1652	1638	0.9	AG	1527	1521	0.5	AG	1502	1505	0.2
B2U	1694	1689	0.3	B2U	1574	1572	0.2	B2U	1537	1550	0.8
AG	1781	1781	0.0	AG	1632	1631	0.0	AG	1593	1603	0.6
B1U	1806	1804	0.1	B1U	1661	1660	0.1	B1U	1608	1625	1.0

Table AI-3 Cont.

B3G	1845	1846	0.1	B3G	1690	1690	0.0	B3G	1644	1669	1.5
B3G	3349	3347	0.1	B3G	3154	3185	1.0	B3G	3164	3158	0.2
B1U	3351	3349	0.1	B1U	3156	3186	1.0	B1U	3167	3159	0.2
B2U	3353	3352	0.0	B2U	3158	3189	1.0	B2U	3167	3160	0.2
AG	3358	3357	0.1	AG	3161	3192	1.0	AG	3170	3163	0.2
B3G	3368	3370	0.1	B3G	3172	3205	1.0	B3G	3184	3184	0.0
B1U	3370	3372	0.1	B1U	3174	3206	1.0	B1U	3184	3185	0.0
B2U	3380	3386	0.2	B2U	3186	3221	1.1	B2U	3196	3202	0.2
AG	3382	3388	0.2	AG	3188	3222	1.1	AG	3198	3203	0.2

Table AI-4. Uncorrected and Counterpoise-corrected harmonic vibrational frequencies (cm^{-1}) for indenyl anion (C$_{2v}$) at the MP2/6-311G level of theory. All data in the table were computed with the CP-opt driver and Gaussian 98.

	MP2 CP-corrected	MP2	% diff
B1	197	199	1.0
A2	221	305	38.2
A2	331	1072i	424.3
B1	363	444	22.3
B2	390	391	0.3
A1	550	544	1.2
B2	604	594	1.6
A2	604	458i	175.8
B1	620	609	1.7
B1	660	570	13.6
A2	705	523	25.9
A1	733	730	0.4
B1	781	504	35.5
B1	823	200i	124.3
A2	848	125i	114.7
A2	869	677	22.1
A1	874	874	0.0
B2	891	875	1.8
A1	986	979	0.8
A1	1034	1029	0.4
B2	1046	1049	0.3
B1	1058	555i	152.4
B2	1111	1104	0.6
A1	1167	1169	0.1
B2	1208	1195	1.1
B2	1250	1228	1.7
A1	1259	1242	1.3
A1	1349	1337	0.9

Table AI-4 Cont.

B2	1383	1350	2.4
B2	1424	1405	1.3
A1	1451	1422	2.0
B2	1484	1463	1.4
A1	1493	1468	1.7
A1	1543	1518	1.6
B2	1590	1561	1.9
B2	3064	3061	0.1
A1	3072	3069	0.1
A1	3072	3076	0.1
B2	3089	3088	0.0
B2	3108	3110	0.1
A1	3109	3110	0.0
A1	3123	3124	0.0

Table AI-5. Uncorrected and Counterpoise-corrected harmonic vibrational frequencies (cm^{-1}) for cyclopentadienyl anion (D_{5h}) at the MP2/6-311G level of theory. All data in the table were computed with the CP-opt driver and Gaussian 98.

MP2 CP-corrected		MP2	% diff
E2"	614	401i	165.3
E1"	640	284	55.7
A2"	680	523	23.0
E2"	848	632	25.4
E2'	854	860	0.6
E1'	995	1020	2.5
E2'	1065	1089	2.3
A1'	1089	1115	2.4
A2'	1311	1298	1.0
E2'	1383	1402	1.4
E1'	1449	1454	0.4
E2'	3060	3081	0.7
E1'	3087	3111	0.8
A1'	3119	3145	0.8

Appendix II

Full references used in this study:

Gaussian 03 Revision C.01: Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M.A.; Cheeseman, J. R.; Montgomery Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.;Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.;Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota,K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai,H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo,C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.;Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.;Dannenberg, J. J.; Zakrzewski, G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas,O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui,Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.;Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng,C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; Gaussian 03, Revison C.01 ed.; Gaussian,Inc.: Pittsburgh, PA, 2003.

Molcas 5.4: Karlstrom, G.; Lindh, R.; Malmqvist, P. A.; Roos, B. O.; Ryde, U.;Veryazov, V.; Widmark, P. O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L.Comput. Mater. Sci. 2003, 28, 222-239.

Gaussian 03 Revision B.07: Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M.A.; Cheeseman, J. R.; Montgomery Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.;Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.;Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota,K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai,H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo,C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.;Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.;Dannenberg, J. J.; Zakrzewski, G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas,O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui,Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.;Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng,- 2 -C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.;Wong, M. W.;

Gonzalez, C.; Pople, J. A.; Gaussian 03, Developement version (Revision B.07) ed.; Gaussian, Inc.: Pittsburgh, PA, 2003.
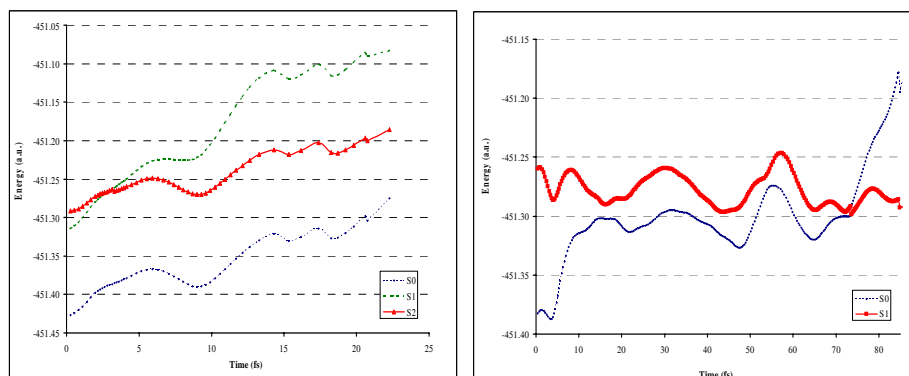
Figure AII-1. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
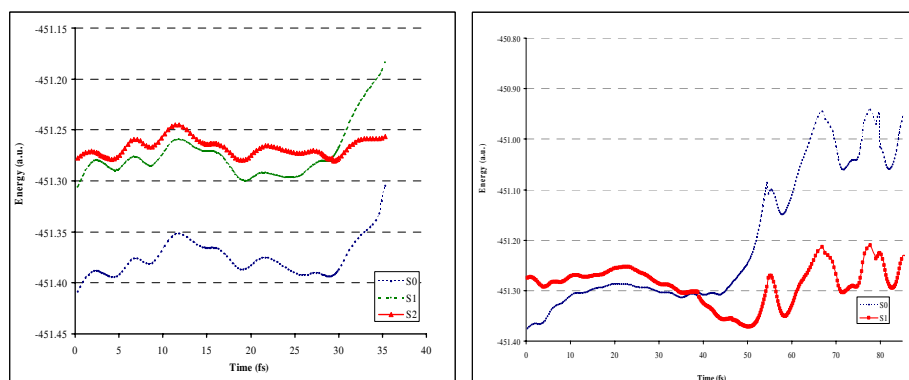
Figure AII-2. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
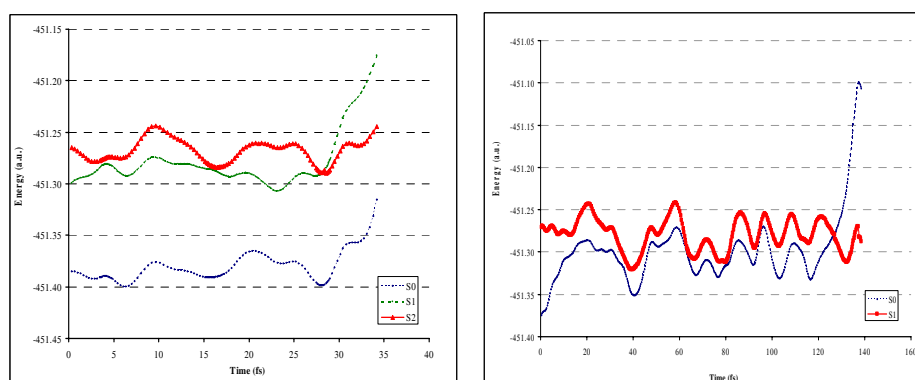
Figure AII-3. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
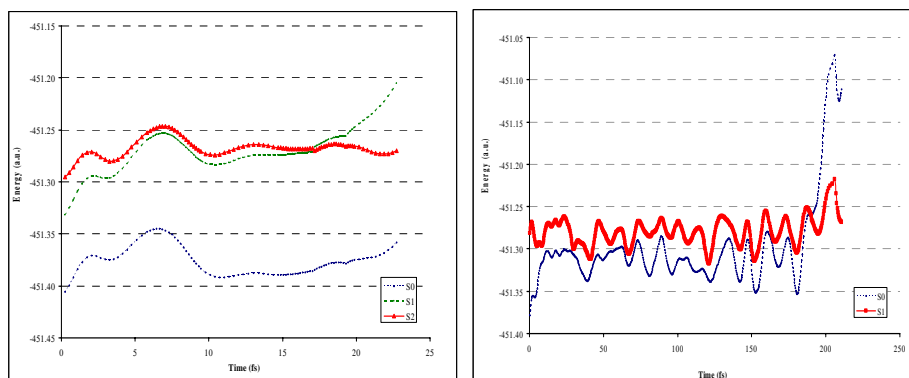
Figure AII-4. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
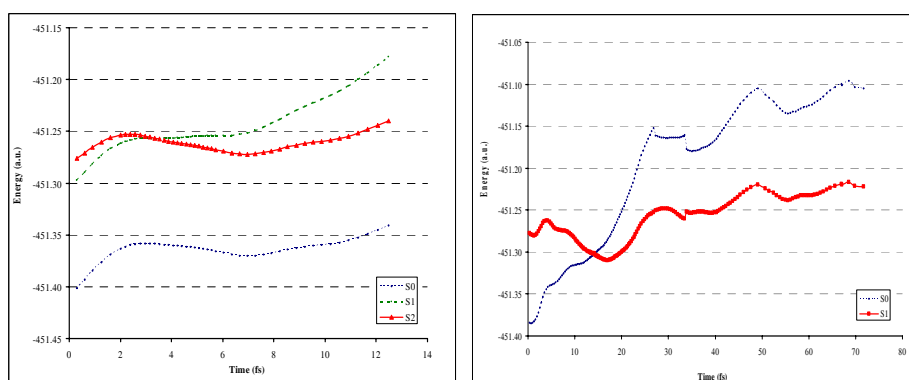
Figure AII-5. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
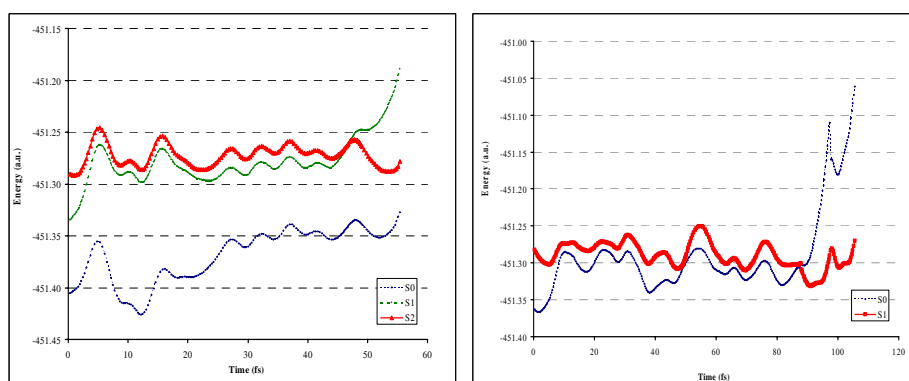
Figure AII-6. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
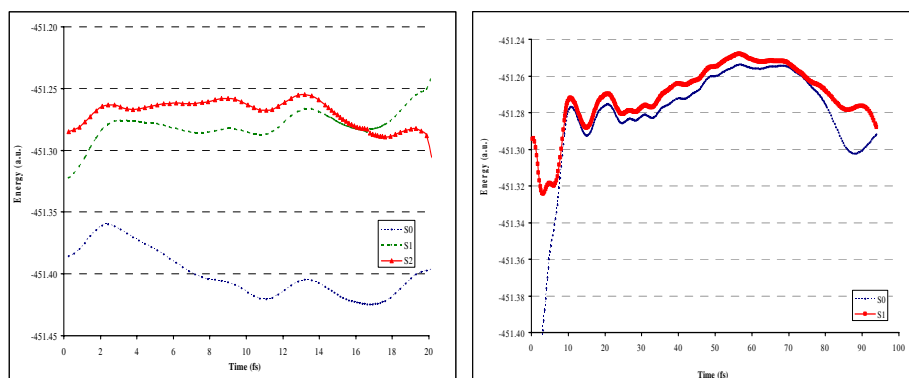
Figure AII-7. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
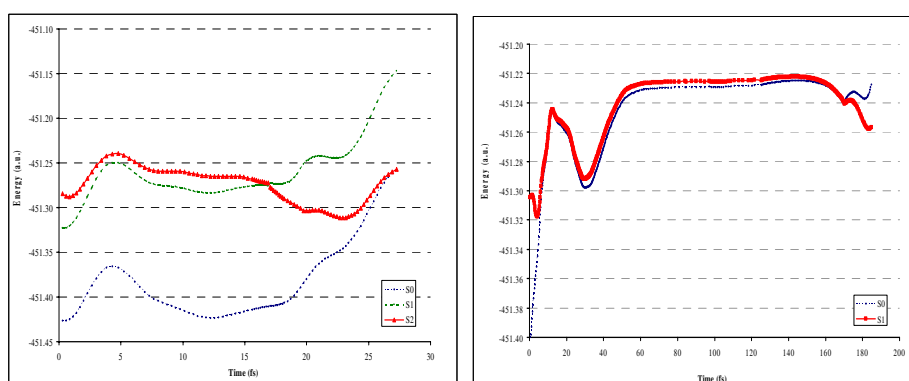
Figure AII-8. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
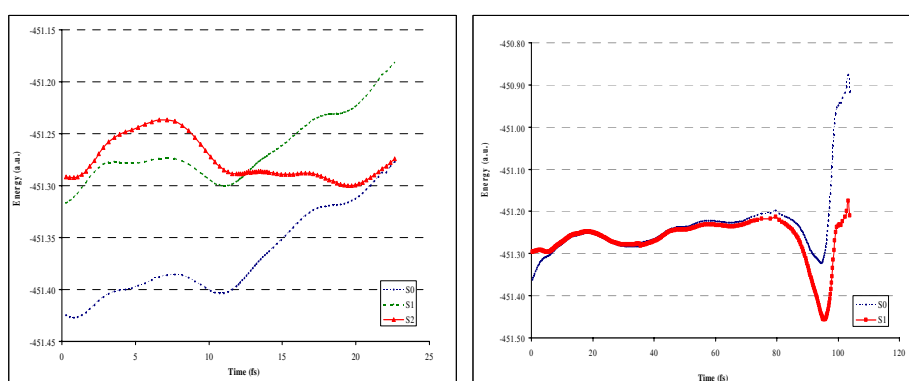
Figure AII-9. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
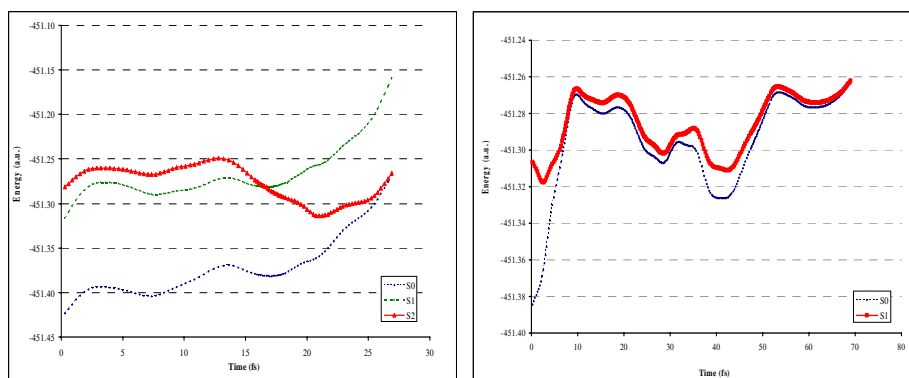
Figure AII-10. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
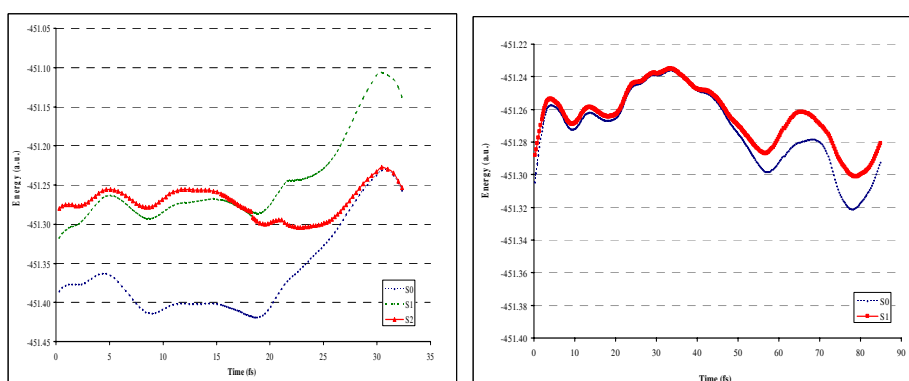
Figure AII-11. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
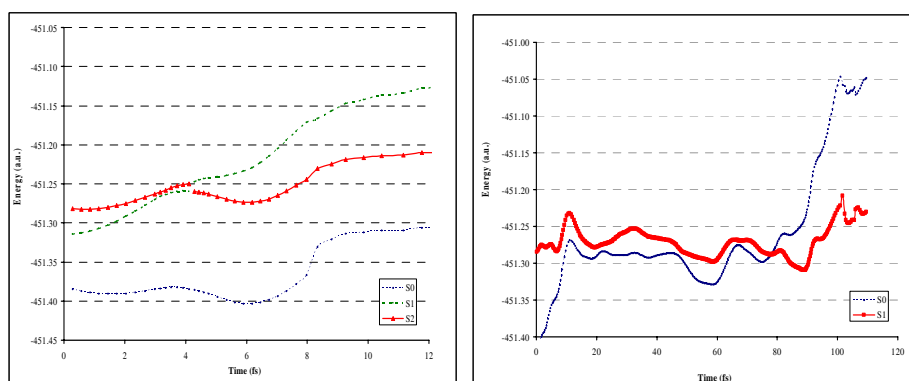
Figure AII-12. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
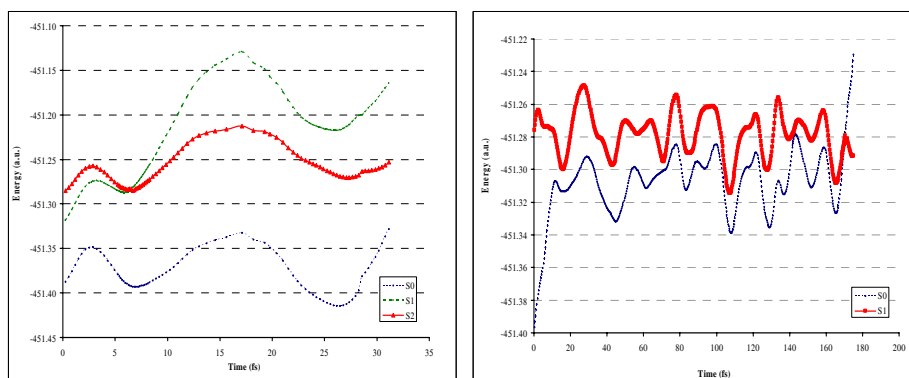
Figure AII-13. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)
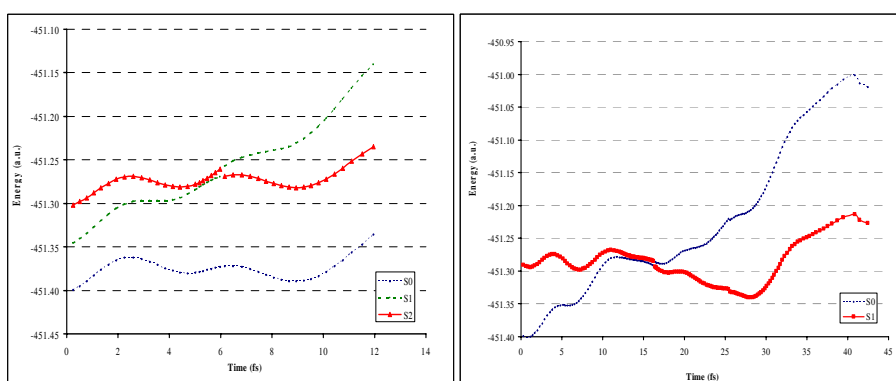
Figure AII-14. Trajectory 1. CASSCF(8,6,nr=3)/6-31G* + subsequent trajectory on S1 (nr=2)

Appendix III

Run 1b
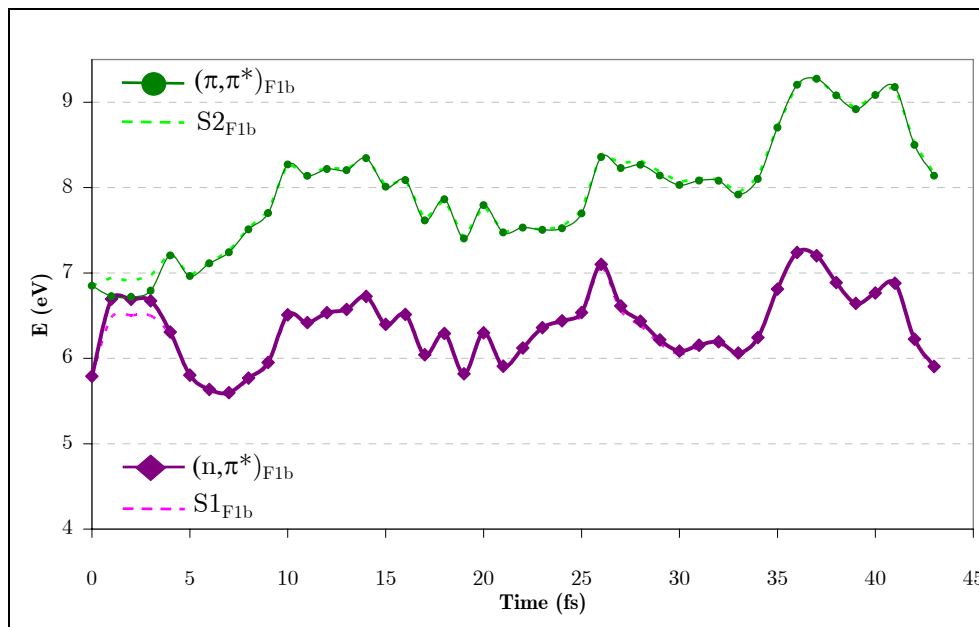
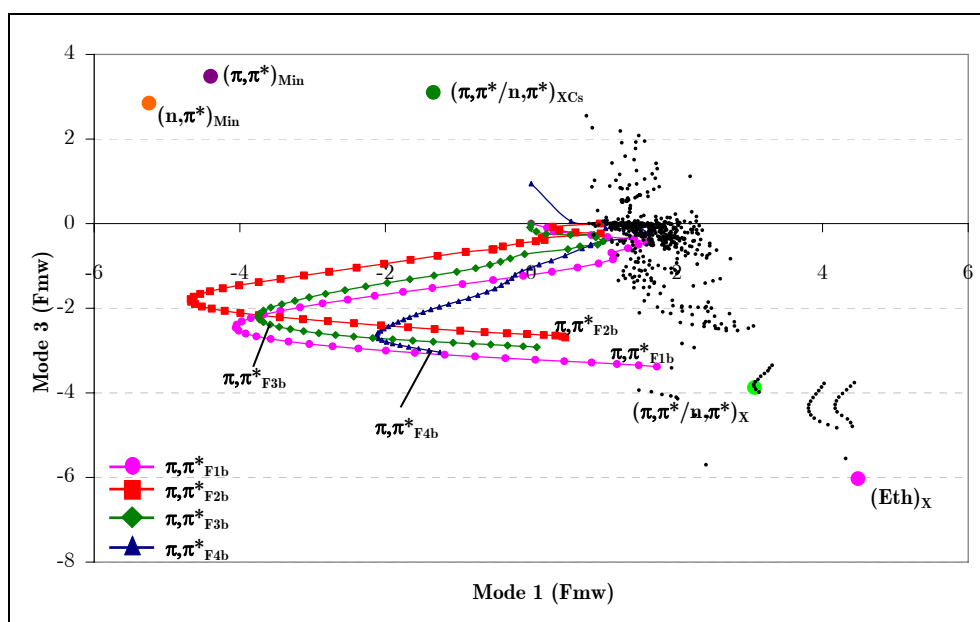**Figure AIII-1.** Diabatic and adiabatic energies of the F1b function of run 1b.

Run 1c
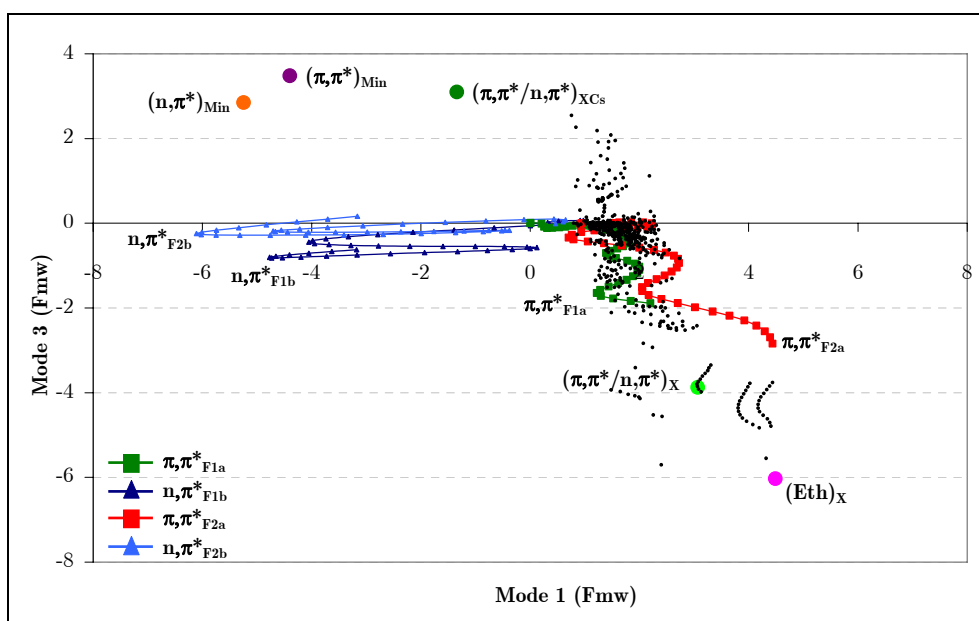
Figure AIII-2. Position of the center of the Gaussian functions of run 1c (position relative to $(\pi,\pi^*)_{TS}$).

Run 1d
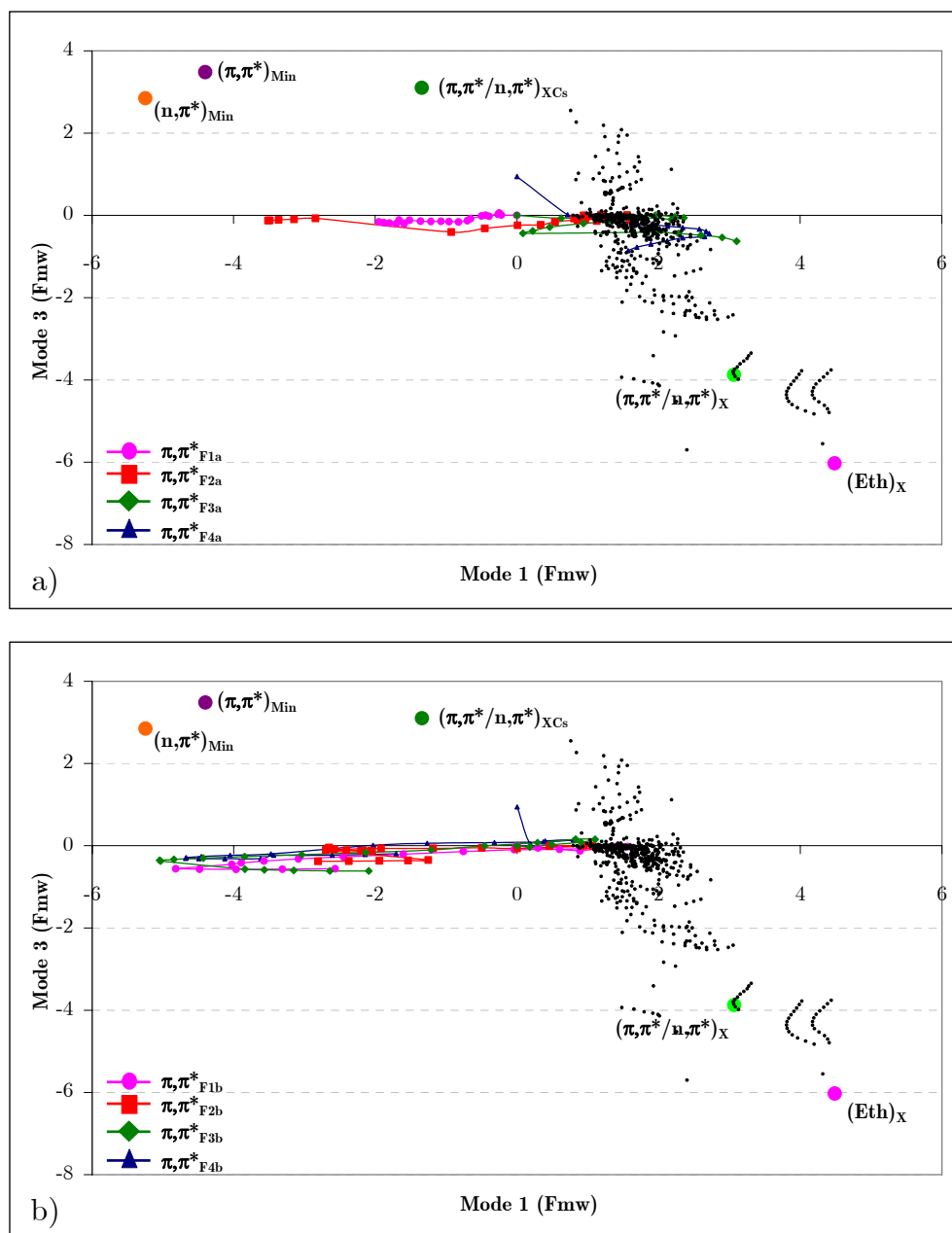
Figure AIII-3. a) Position of the center of the Gaussian functions of the π,π^* state of run 1d. b) Position of the center of the Gaussian functions of the n,π^* state of run 1d (positions relative to $(\pi,\pi^*)_{TS}$).

Run 1e
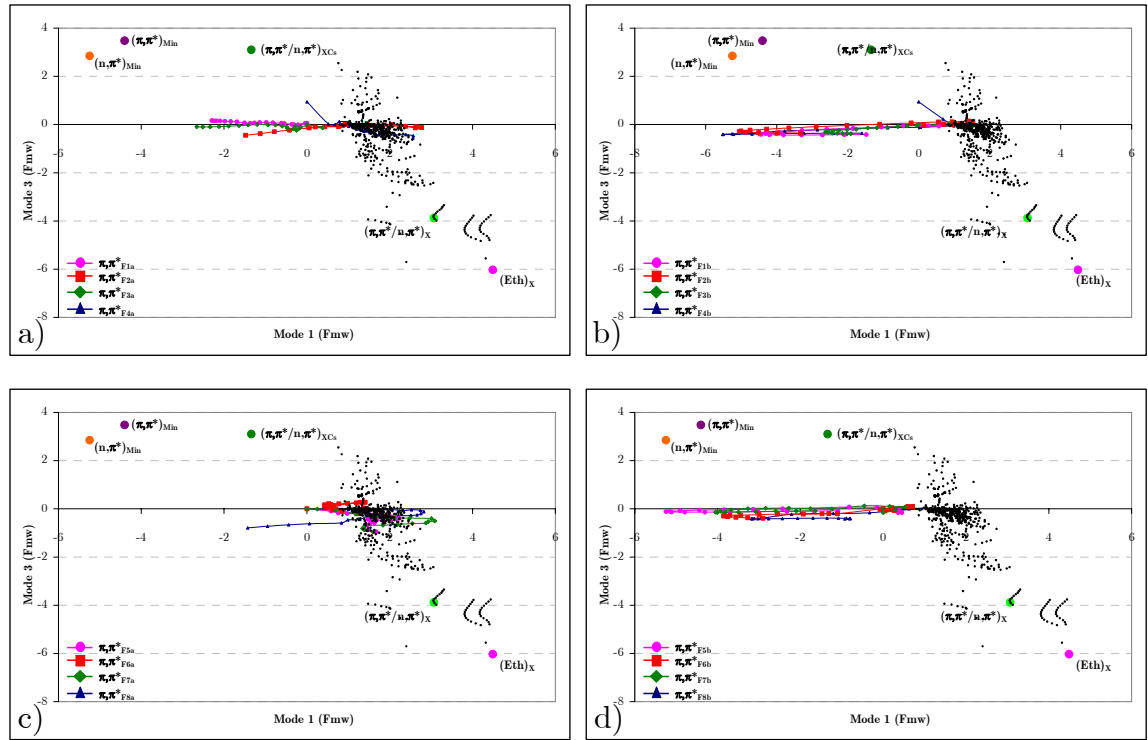

Figure AIII-4. a) Position of the center of the F1a-F4a Gaussian functions of run 1e. b) Position of the center of the F5a-F8a Gaussian functions of run 1e. c) Position of the center of the F1b-F4b Gaussian functions of run 1e. d) Position of the center of the F5b-F8b Gaussian functions of run 1e (positions relative to $(\pi,\pi^*)_{TS}$).

Run 2a
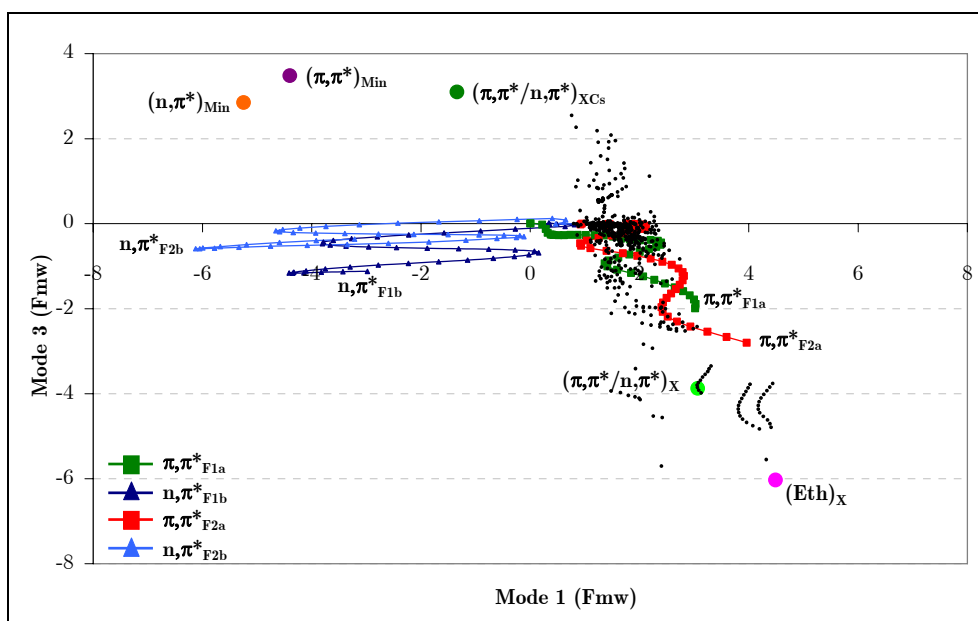
Figure AIII-5. Position of the center of the Gaussian functions of run 2a (position relative to $(\pi,\pi^*)_{TS}$).
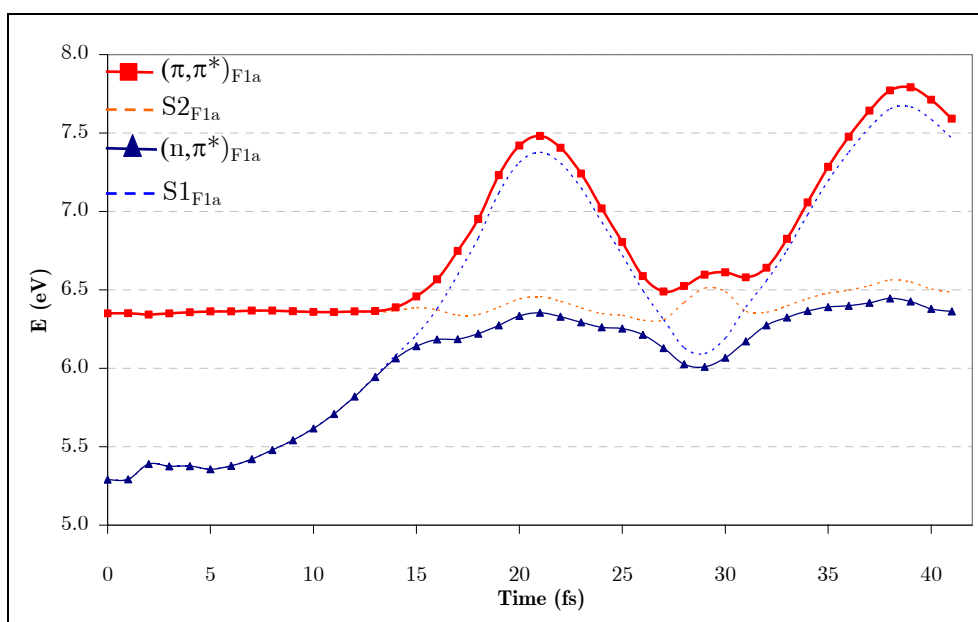
Figure AIII-6. Diabatic and adiabatic energies at the center of function F1a of run 2a.
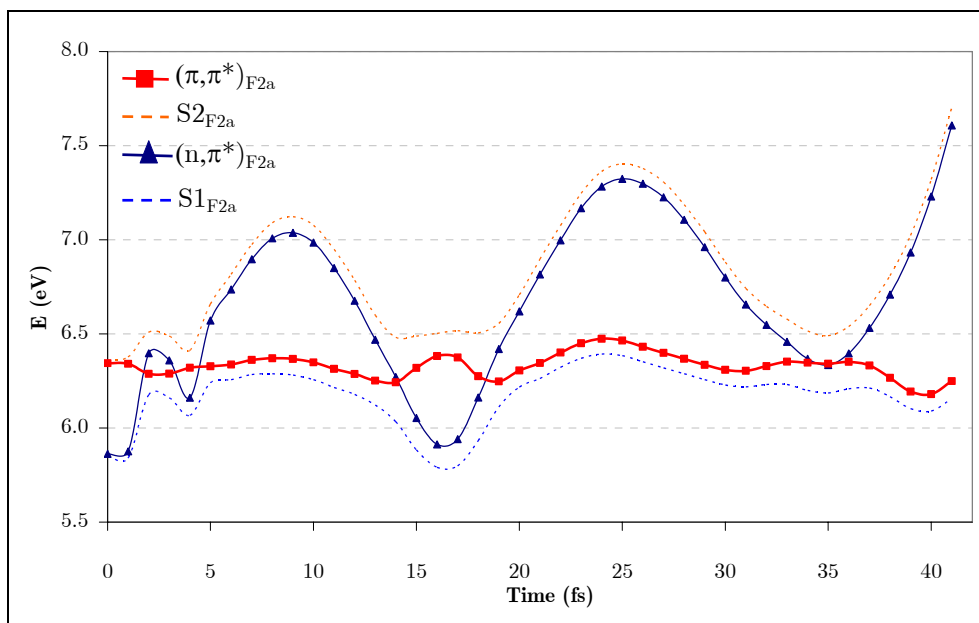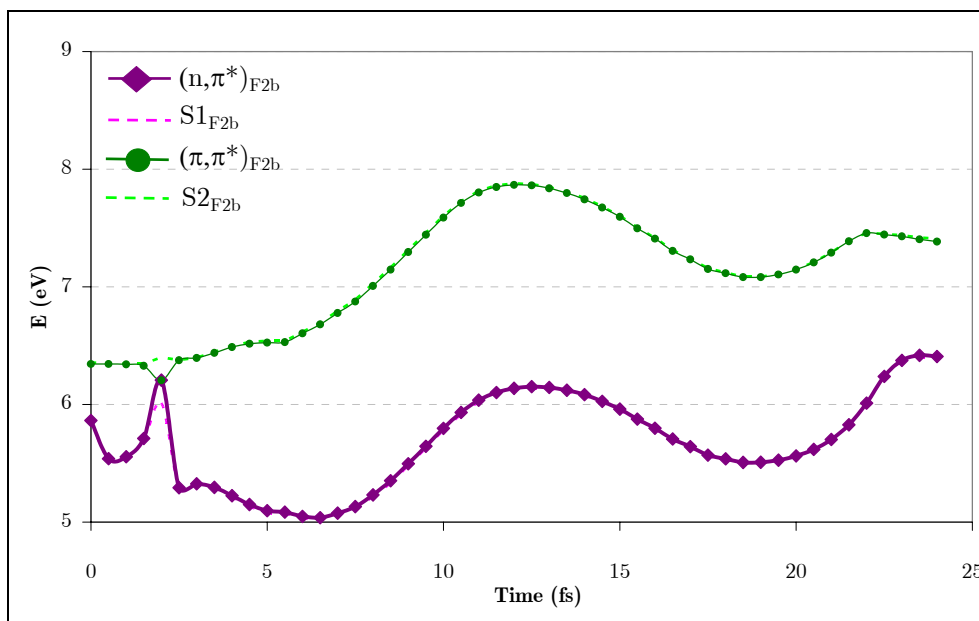
Figure AIII-7. Diabatic and adiabatic energies at the center of function F2a of run 2a.
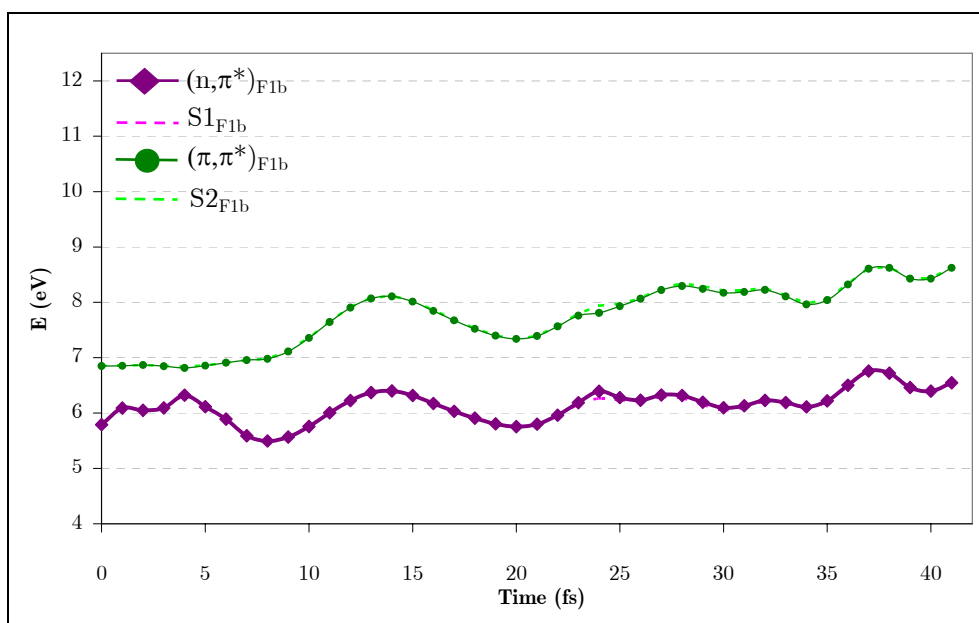
Figure AIII-8. Diabatic and adiabatic energies at the center of function F1b of run 2a.
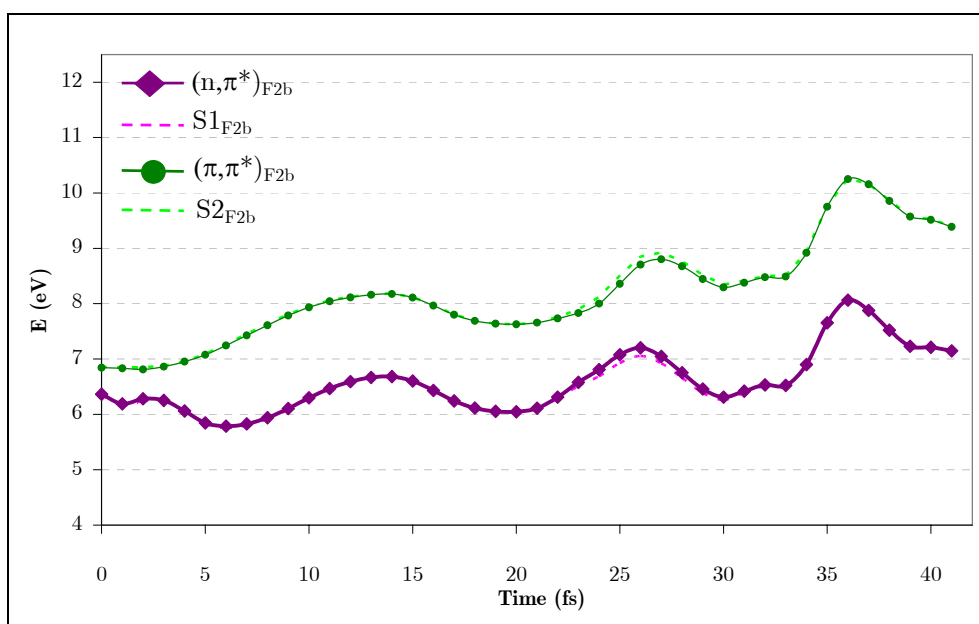
Figure AIII-9. Diabatic and adiabatic energies at the center of function F2b of run 2a.
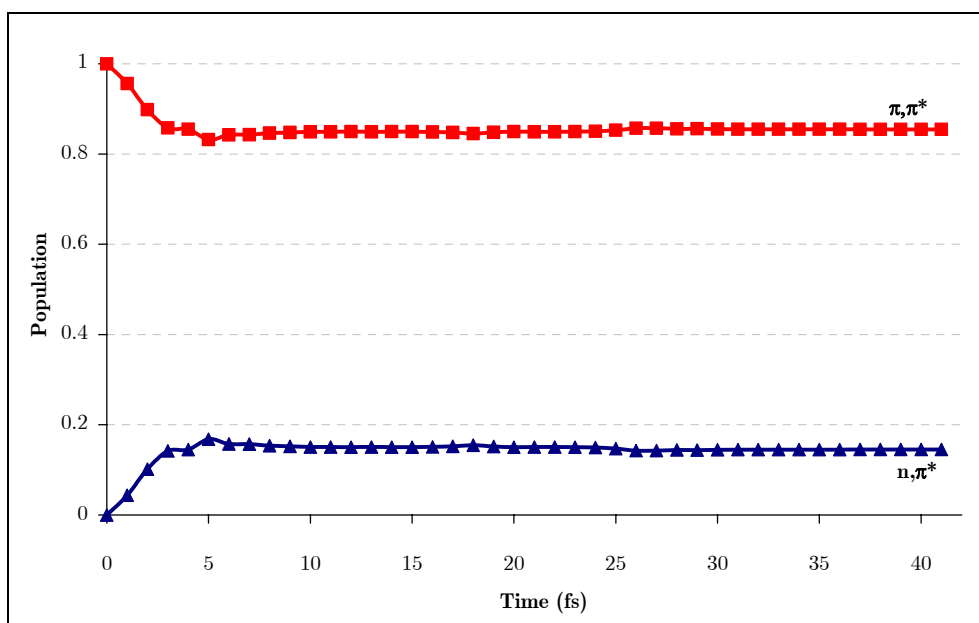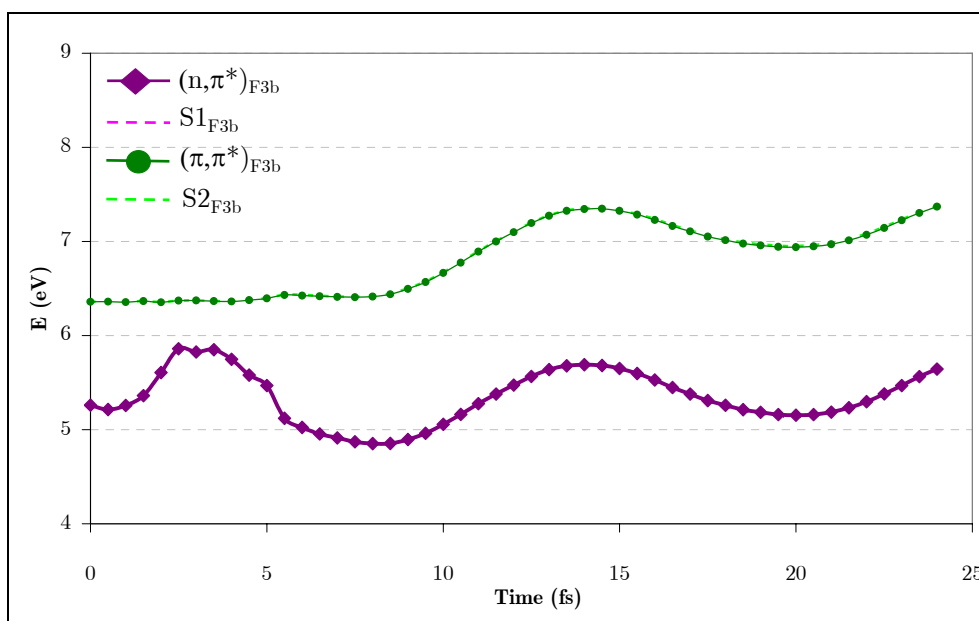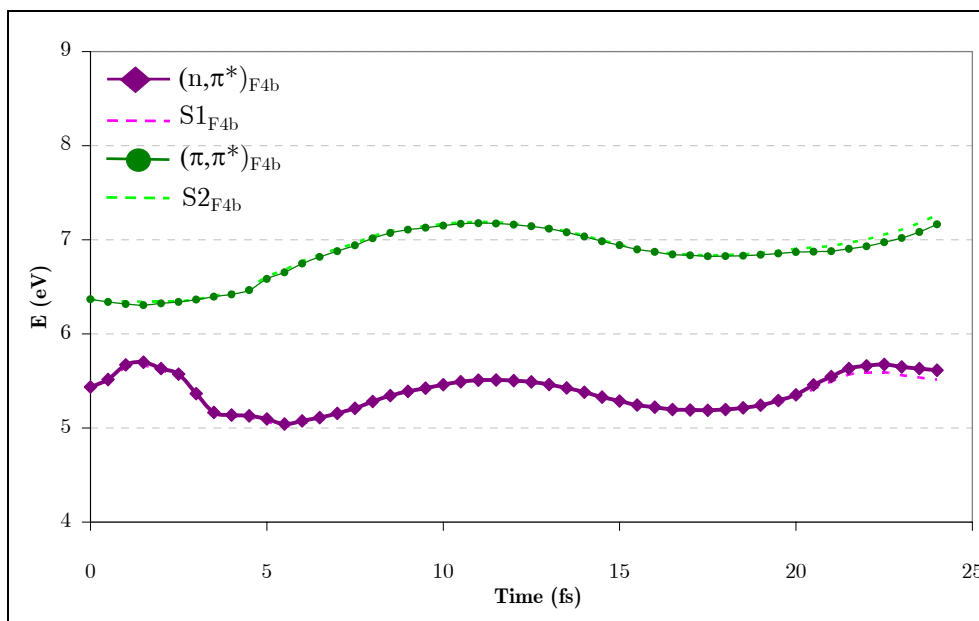

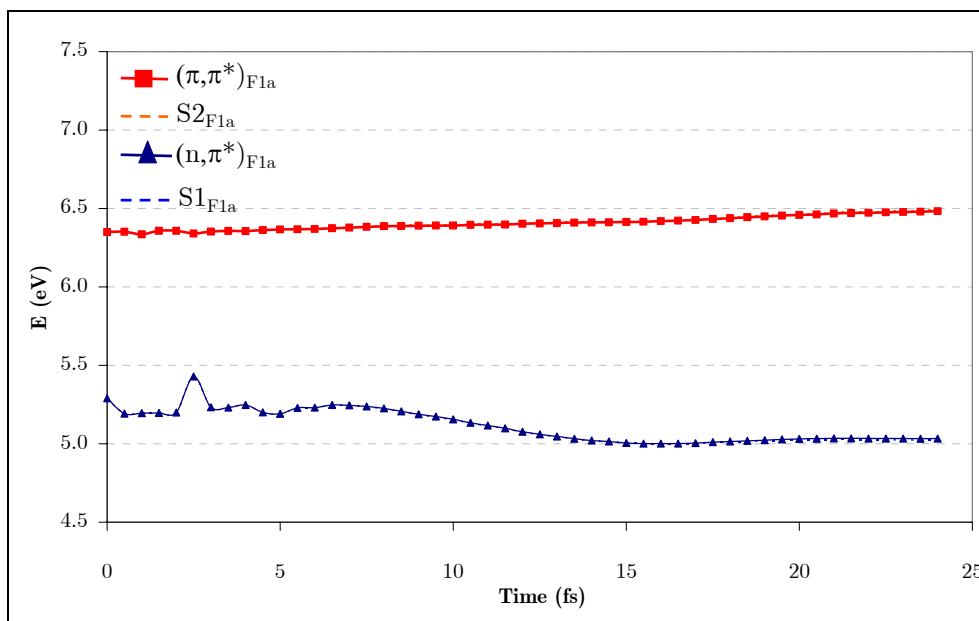**Figure AIII-10.** Population of the π,π* and n,π* state of run 2a.

Run 2b
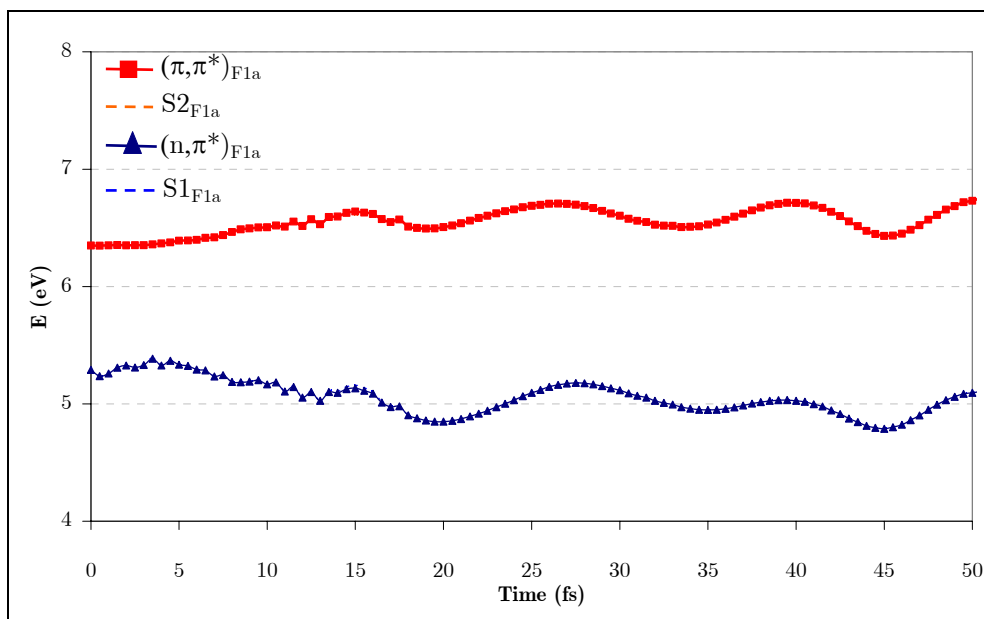
Figure AIII-11. Diabatic and adiabatic energies at the center of function F1a of run 2b.
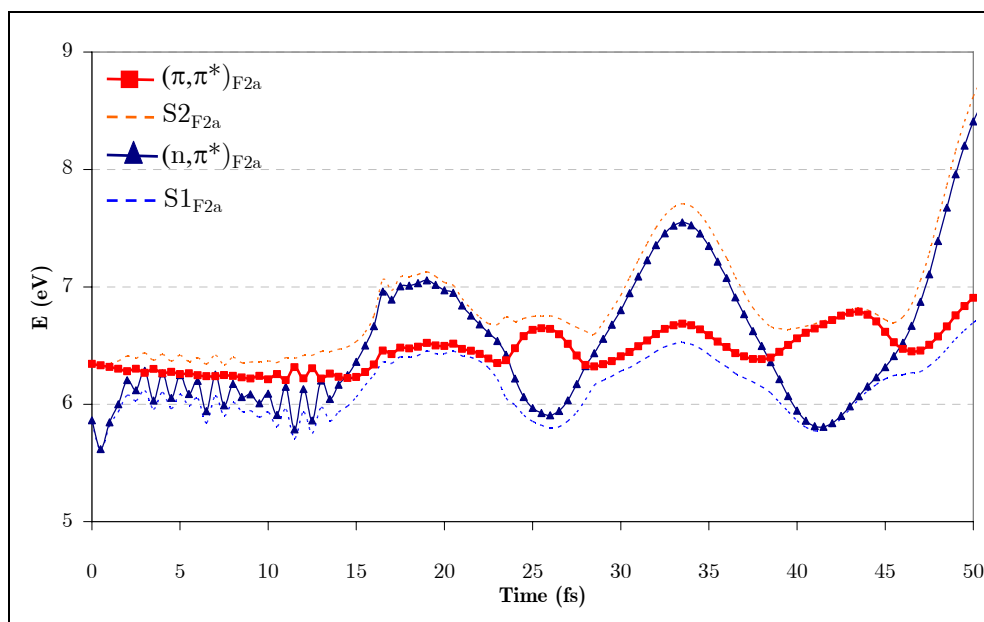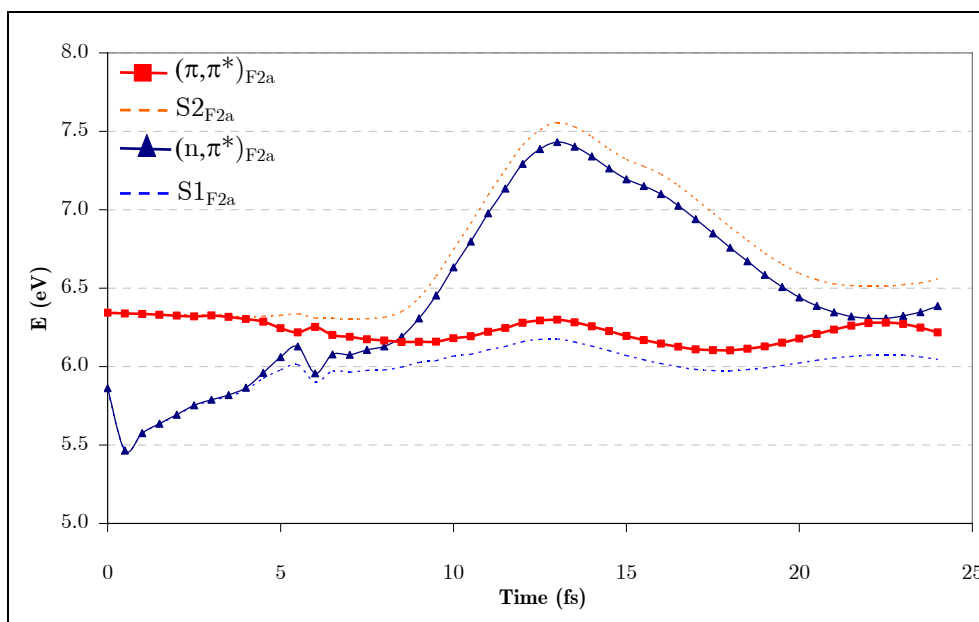

Figure AIII-12. Diabatic and adiabatic energies at the center of function F2a of run 2b.
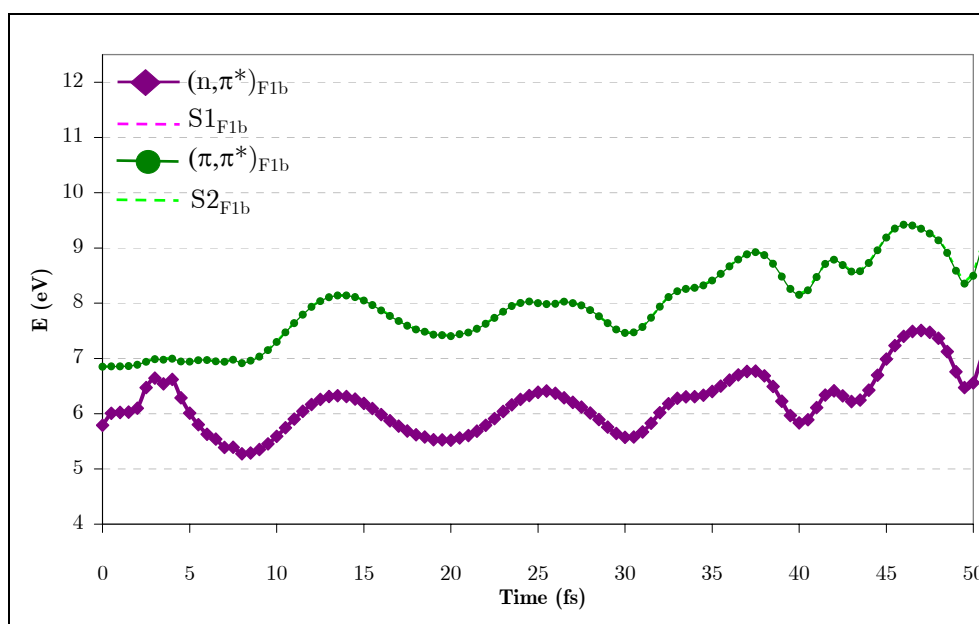
Figure AIII-13. Diabatic and adiabatic energies at the center of function F1b of run 2b.
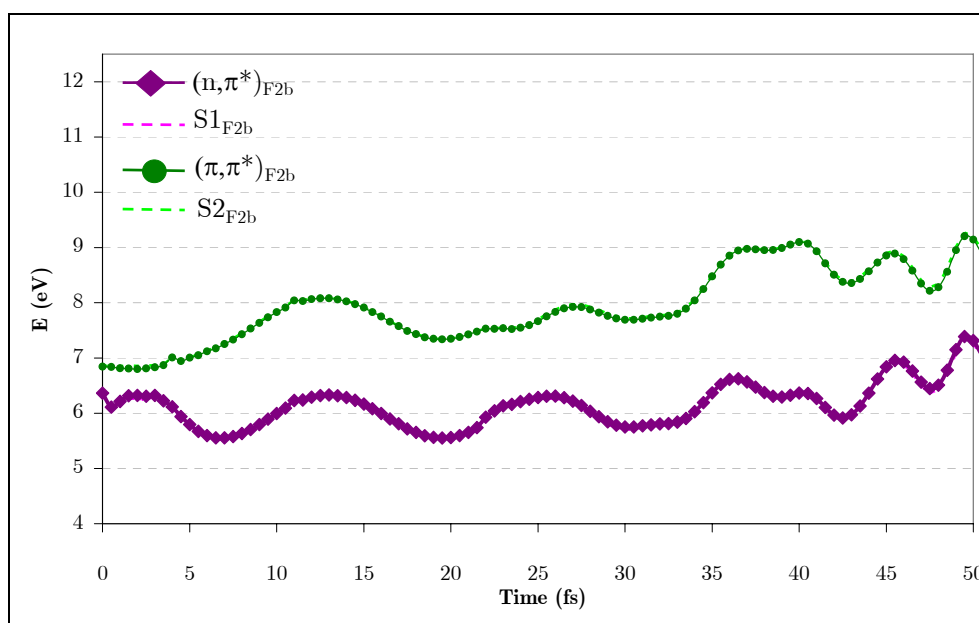
Figure AIII-14. Diabatic and adiabatic energies at the center of function F2b of run 2b.
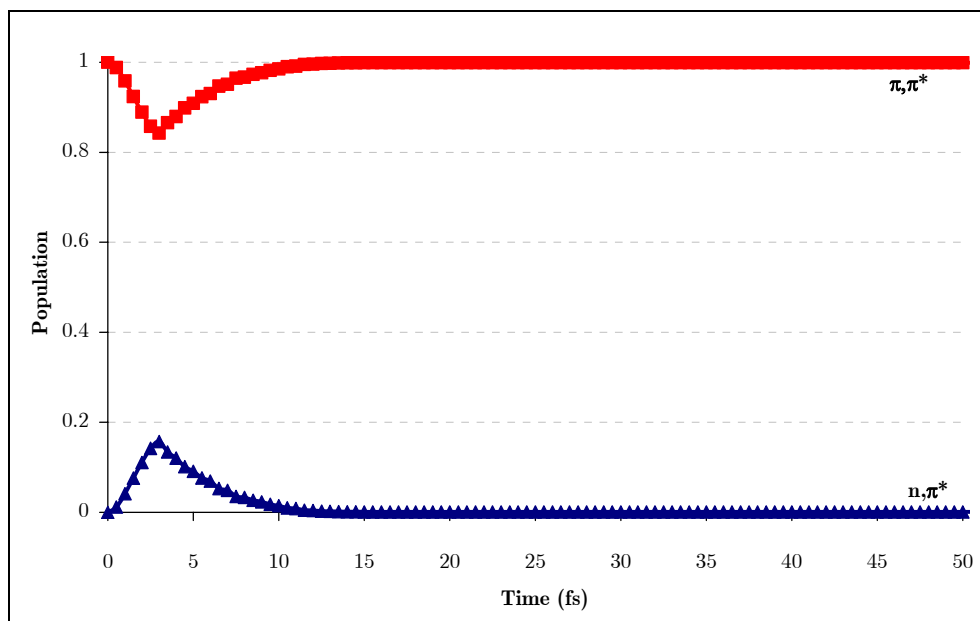
Figure AIII-15. Population of the π,π^* and n,π^* state of run 2b.

Run 3a
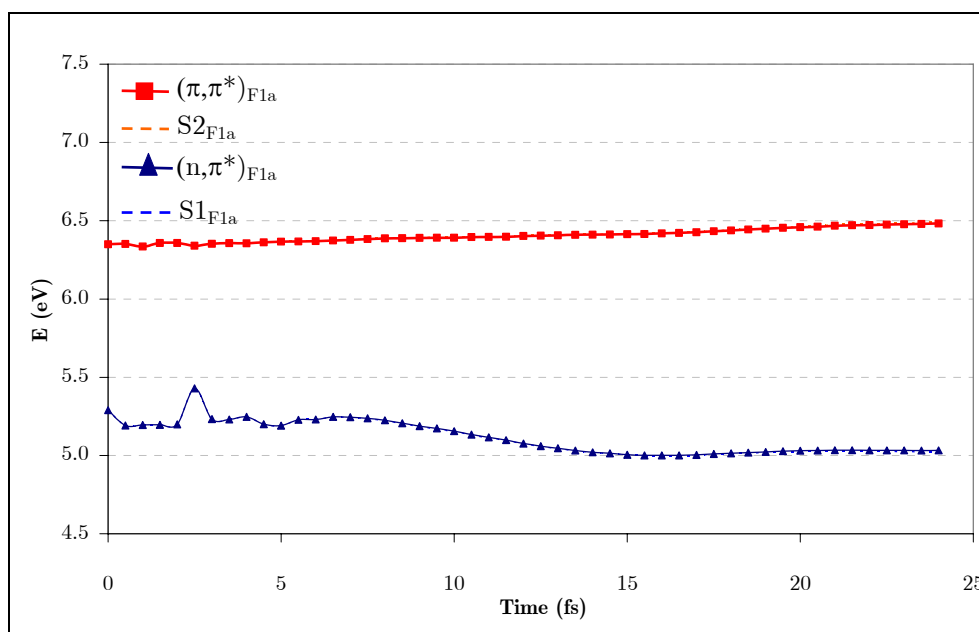
Figure AIII-16. Diabatic and adiabatic energies at the center of function F1a of run 3a.
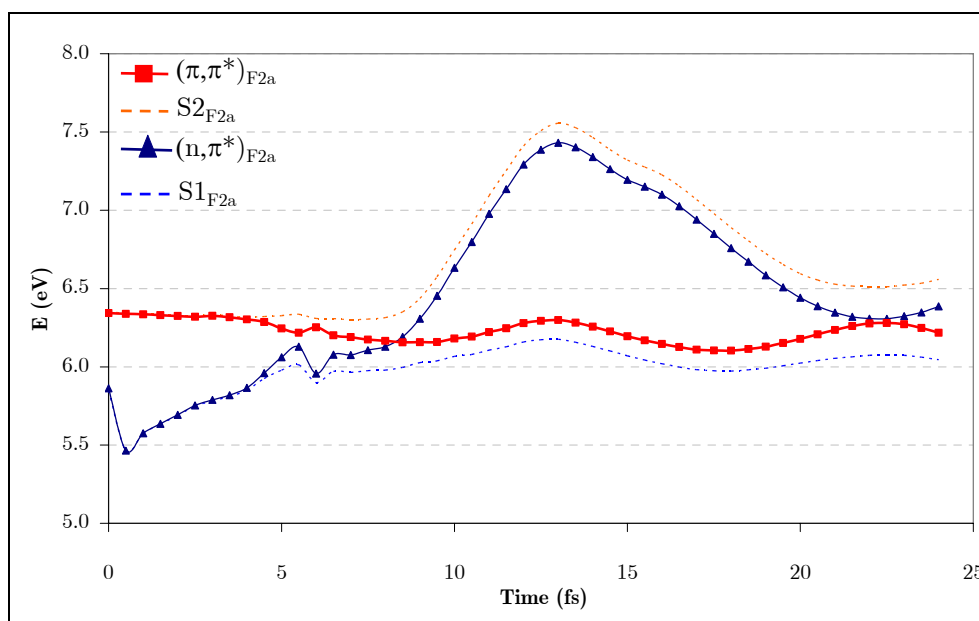
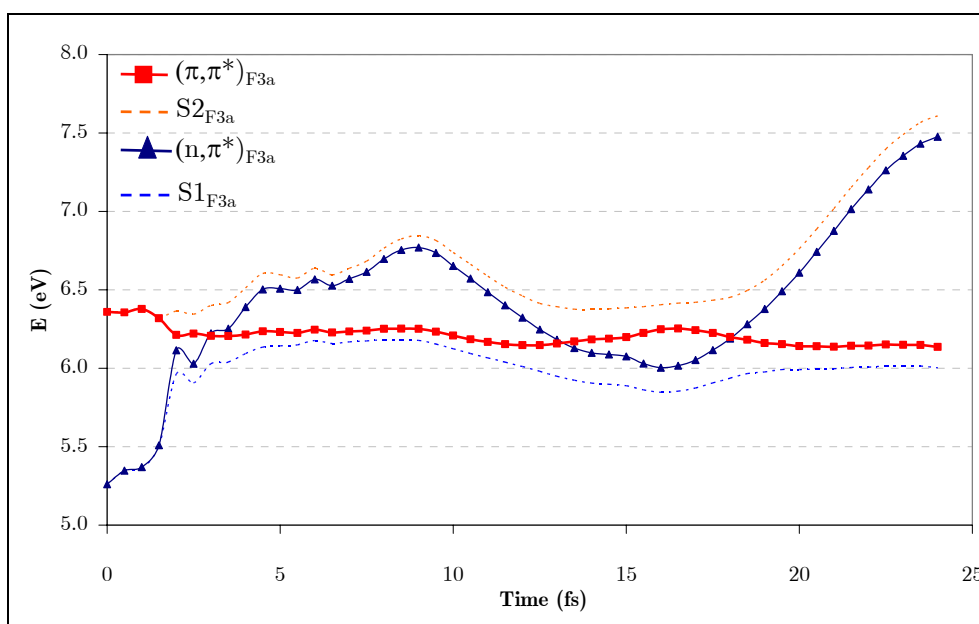**Figure AIII-17.** Diabatic and adiabatic energies at the center of function F2a of run 3a.
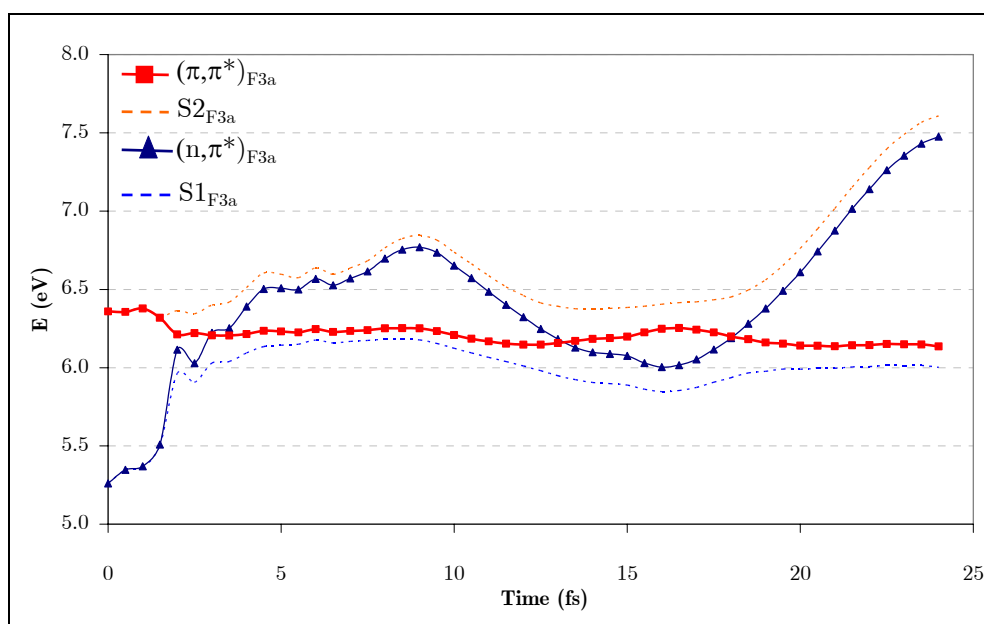
Figure AIII-18. Diabatic and adiabatic energies at the center of function F3a of run 3a.
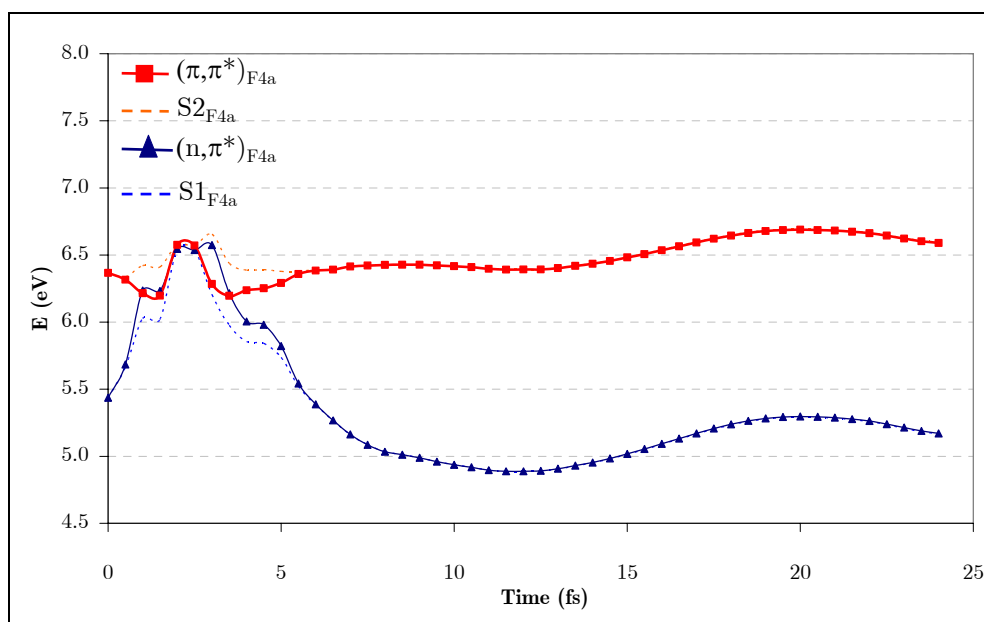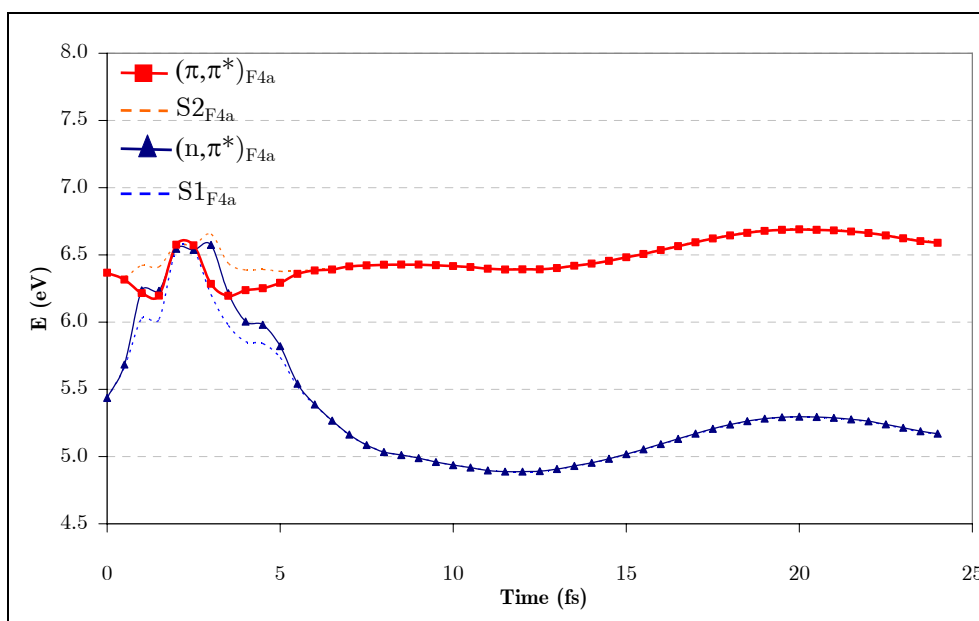

Figure AIII-19. Diabatic and adiabatic energies at the center of function F4a of run 3a.
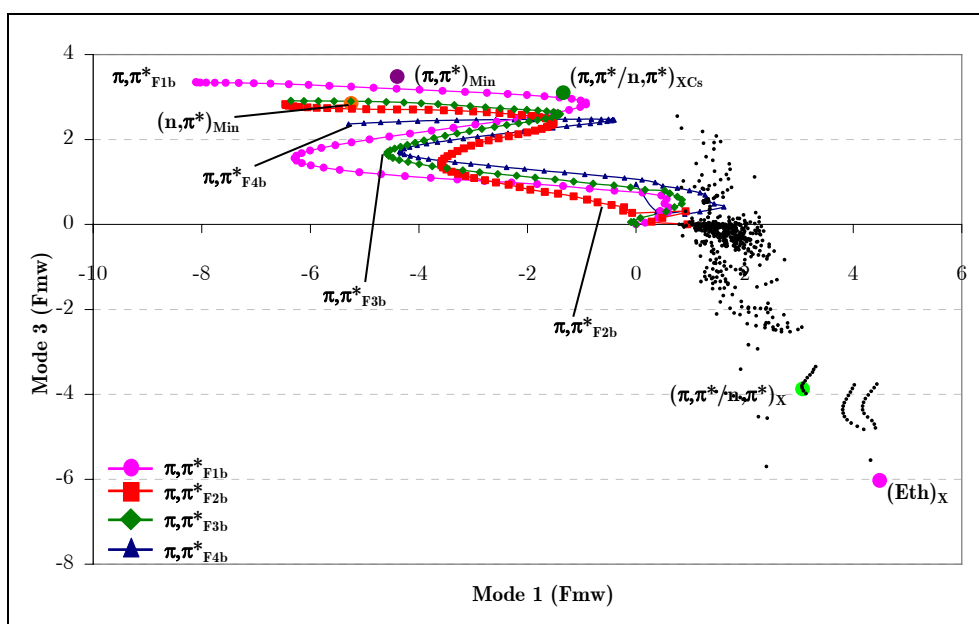
Figure AIII-20. Position of the center of the Gaussian functions of the n,π* state of run 3a (position relative to $(\pi,\pi^*)_{TS}$).
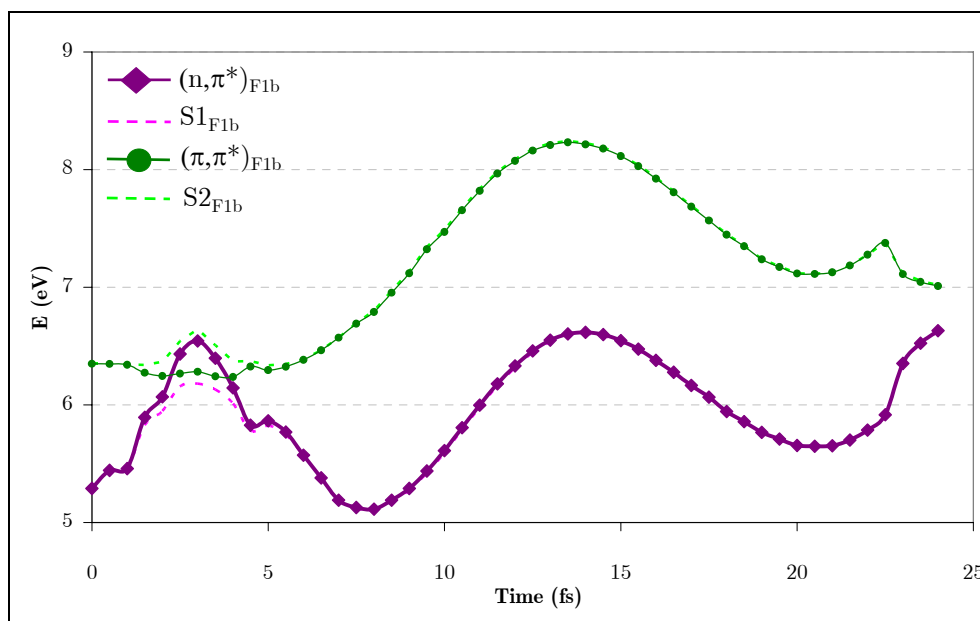
Figure AIII-21. Diabatic and adiabatic energies at the center of function F1b of run 3a.

Figure AIII-22. Diabatic and adiabatic energies at the center of function F2b of run 3a.

Figure AIII-23. Diabatic and adiabatic energies at the center of function F3b of run 3a.

Figure AIII-24. Diabatic and adiabatic energies at the center of function F4b of run 3a.
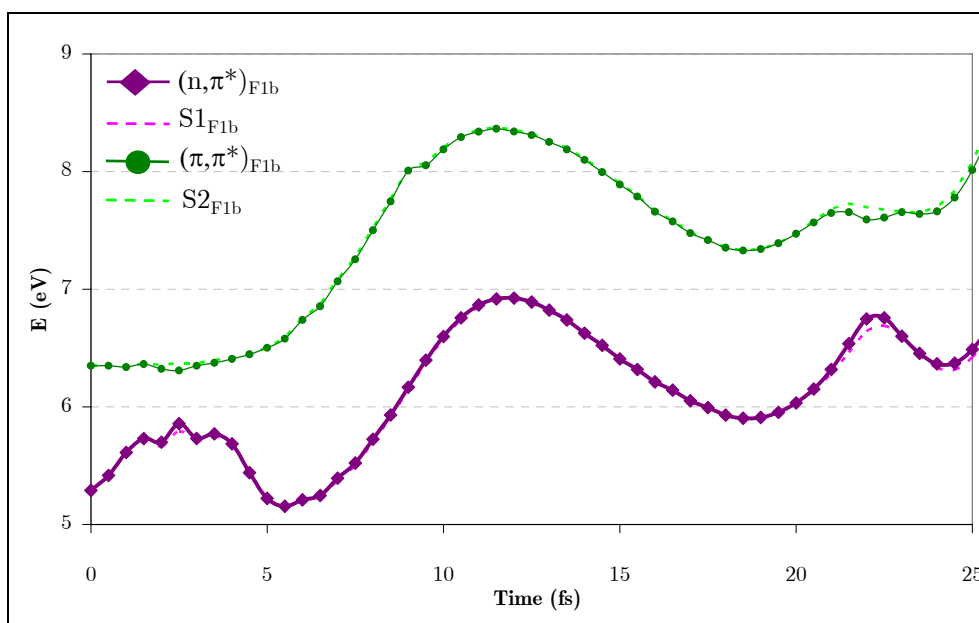
Run 3b

Figure AIII-25. Diabatic and adiabatic energies at the center of function F1a of run 3b.

Figure AIII-26. Diabatic and adiabatic energies at the center of function F2a of run 3b.

Figure AIII-27. Diabatic and adiabatic energies at the center of function F3a of run 3b.

Figure AIII-28. Diabatic and adiabatic energies at the center of function F4a of run 3b.

Figure AIII-29. Position of the center of the Gaussian functions of the n,π* state of run 3b (position relative to $(\pi,\pi^*)_{TS}$).

Figure AIII-30. Diabatic and adiabatic energies at the center of function F1b of run 3b.
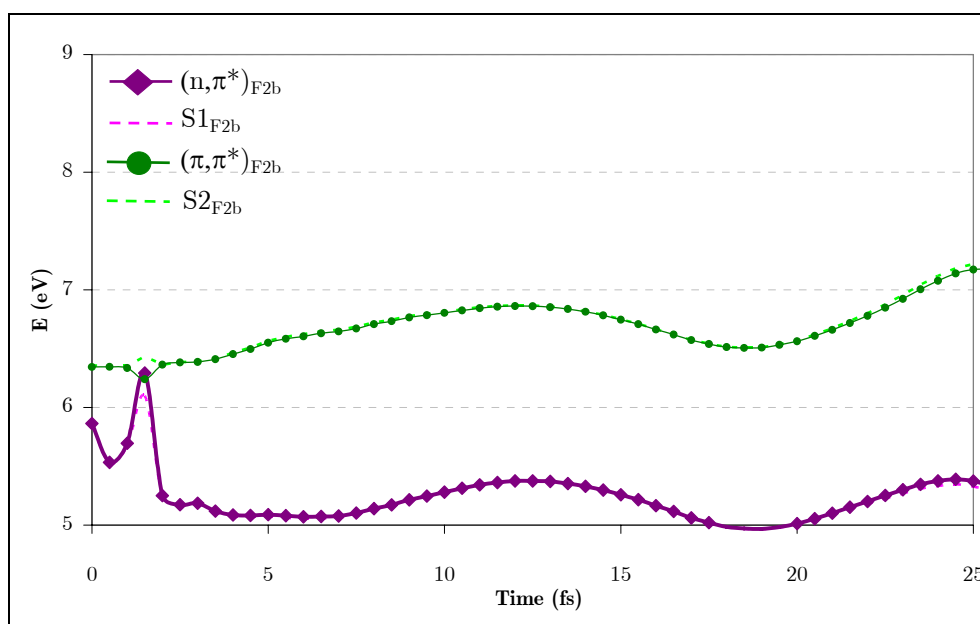
Figure AIII-31. Diabatic and adiabatic energies at the center of function F2b of run 3b.
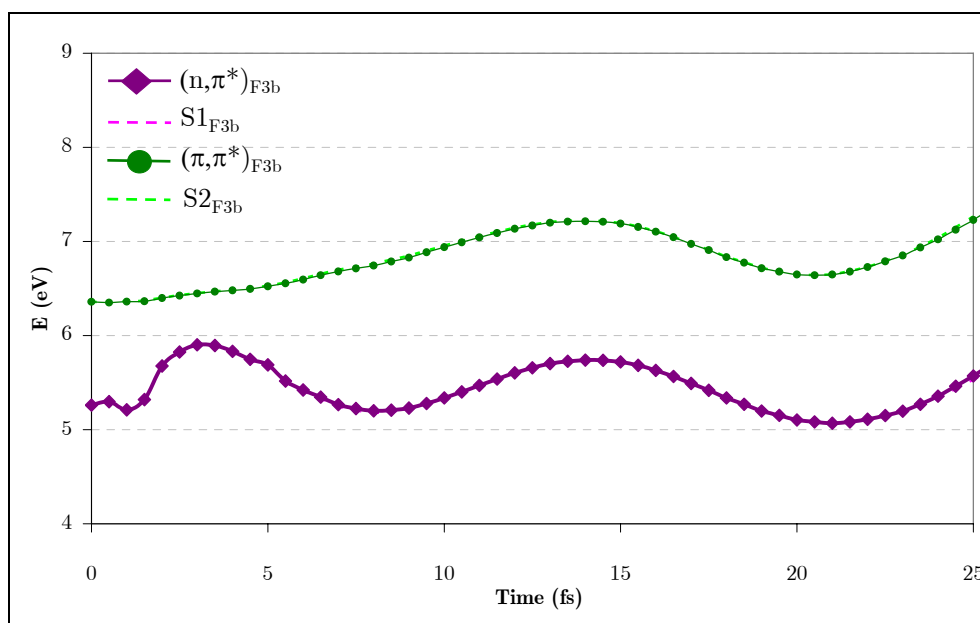

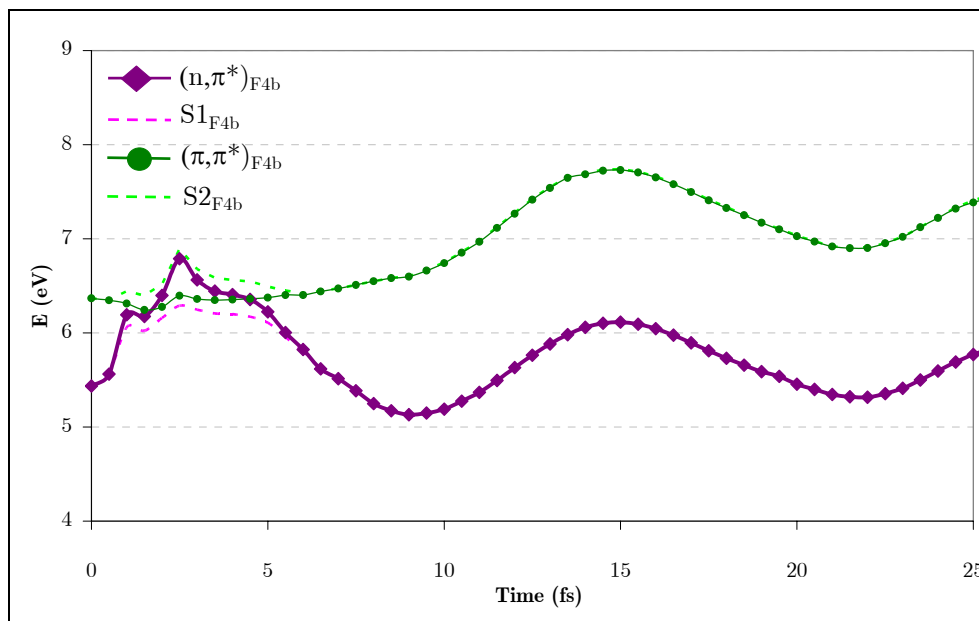**Figure AIII-32.** Diabatic and adiabatic energies at the center of function F3b of run 3b.

Figure AIII-33. Diabatic and adiabatic energies at the center of function F4b of run 3b.