



UNIVERSITAT DE
BARCELONA

Modelització multiescala del paper de l'entorn en processos de transferència de càrrega i d'energia en biomolècules

Marina Corbella Morató

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (deposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (deposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (deposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ
DEPARTAMENT DE FARMÀCIA I TECNOLOGIA FARMACÈUTICA I
FISICOQUÍMICA

**MODELITZACIÓ MULTIESCALA DEL PAPER DE L'ENTORN EN
PROCESSOS DE TRANSFERÈNCIA DE CÀRREGA I D'ENERGIA EN
BIOMOLÈCULES**

MARINA CORBELL MORATÓ

BARCELONA, 2018



UNIVERSITAT DE BARCELONA
FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ
DEPARTAMENT DE FARMÀCIA I TECNOLOGIA FARMACÈUTICA I
FISICOQUÍMICA

PROGRAMA DE DOCTORAT EN RECERCA, DESENVOLUPAMENT I
CONTROL DE MEDICAMENTS

**MODELITZACIÓ MULTIESCALA DEL PAPER DE L'ENTORN EN
PROCESSOS DE TRANSFERÈNCIA DE CÀRREGA I D'ENERGIA EN
BIOMOLÈCULES**

Memòria presentada per Marina Corbella Morató per optar al títol de doctora per la
Universitat de Barcelona

Dr. Carles Eduard Curutchet Barat
Director

Dr. Francisco Javier Luque Garriga
Director

Marina Corbella Morató
Doctoranda

Dr. Francisco Javier Luque Garriga
Tutor

MARINA CORBELL MORATÓ
BARCELONA, 2018

Al Roc,
per haver-me ensenyat
un nou ordre de prioritats

Agraïments

Això s'acaba... i ho fa amb una mescla de sensacions estranyes, per un cantó alegria i unes ganes boges de tancar aquesta etapa, però d'altra banda, amb un xic de pena i de nostàlgia. Encara que en començar saps que és un projecte temporal, cinc anys han estat una temporalitat molt llarga! Encara recordo el primer dia que vaig posar un peu a "dalt", jo ni tant sols sabia que hi havia una altra planta al damunt de la tercera planta, recordo aquesta sensació (que encara tinc en moltes ocasions) de no saber res i alhora la fascinació de veure que hi ha un horitzó nou per descobrir. En aquests cinc anys he descobert un nou camp que m'ha impressionat, que m'ha permès aprendre i redescobrir les meves possibilitats i els meus límits, ha passat molta gent pel departament durant aquests cinc anys i vull pensar que de tothom n'he extret alguna cosa positiva.

En primer lloc voldria donar les gràcies als meus directors, al Javi per la seva amabilitat i la seva paciència, i per ser un "jefe" tant accessible i pròxim tot i no parar mai ni cinc minuts. Però vull donar les gràcies especialment al Curu, en primer lloc per donar-me la oportunitat de treballar amb ell, quan recordo la neòfita que era quan vaig començar imagino lo exasperant que deuria arribar a ser explicar-me les coses una vegada i una altra, i la paciència que ha tingut sempre! Pel que més haig de donar les gràcies per això és pel seu optimisme, un optimisme que s'encomana i que en cinc anys crec que no he vist decaure mai. Aquest positivisme és una de les grans coses que també m'ha aportat embarcar-me en aquesta tesi.

En segon lloc vull donar les gràcies també a la gent de casa que sempre han estat allà. Especialment al Sergi, que és el que ha hagut de pagar més plats trencats els dies que els càculs no "corrien" o els dies que aquell "script" no funcionava, o simplement els dies que tornava a casa feta una boleta ben petita i preguntant-me on m'havia ficat. Gràcies per creure en mi sempre un xic més del que ho faig jo. A la mare, al pare i a l'Helena gràcies per donar-me suport quan ho he necessitat, sé que al principi, quan vaig deixar la feina per embarcar-me a fer el màster i

finalment el doctorat, no ho vau veure gens clar, però que mica en mica heu anat veient que no era una idea tant desbaratada.

Vull donar les gràcies també a tota la gent del departament, no m'agradaria deixar-me ningú però segur que ho faig. A la Salo i el Jordi, en primer lloc perquè, a part del Curu, han estat les persones que més m'han ensenyat, van començar sent companys de feina però han acabat sent bons amics. A la Ornella perquè els dies que no estava estressada per alguna reacció que no funcionava feia que el departament agafés un altre aire. Ja no és el mateix des que no hi ets. A la Silvana, perquè hem fet tot aquest camí juntes i ha estat una gran companya de congressos, i una gran estilista! Als altres companys de grup, que tot i haver estat més de pas no per això heu tingut menys pes en aquesta experiència. A la Sònia per ser tant "fan" i haver aportat aquest punt d'innocència que tantes estones de riure ha generat. A l'Andreana perquè ens vas donar un exemple del que era anar a un altre país i semblar més local que molts dels que hi viuen, i al Teo per aportar aquest toc francès al grup.

També vull donar les gràcies als recentment batejats com a "Torriberos" però amb els que vam passar moltes estones divertides els primers anys, i quan vau marxar a Torribera em vau deixar molt sola! Gràcies Toni, Javi, Will i.... a sí! Constantí, t'haig de fer una menció especial perquè quan hi eres no hi havia qui treballés però quan no hi ets se't troba a faltar. Gràcies també a l'Àxel i a la Carol, amb vosaltres les pràctiques a Torribera es fan més agradables. Tampoc em puc oblidar del Ramón, un persona peculiar, però amb la que desconnectar i parlar sobre art o teatre, especialment, sempre ha estat molt agradable.

Al grup dels Barrils, a la "vella" generació: Dani, Aïda, Kevin, Montse i Sergi, amb els que vaig coincidir poc temps però dels que vaig aprendre molt al començament. I també a la nova generació: Serena, Mírian, Moira, Maciej, Guillermo, Morena, Marina i Salvo, que van arribar com una plaga un bon dia però han acabat sent el meu grup adoptiu, especialment a la hora del tupper, gràcies.

També vull agrair a les diverses persones amb les qui he tingut el plaer de col·laborar durant aquests anys. Especialment al Professor Scholes i al seu grup per haver-me acollit durant quatre mesos, una estada que em va servir per entendre moltes coses, puc assegurar que hi ha hagut un abans i un després d'aquells mesos a Amèrica! També al Prof. Voityuk, essencialment per la seva increïble rapidesa, i al Lorenzo, que tot i tenir maneres de fer tant diferents, m'ha ensenyat i ajudat molt.

Per acabar vull donar les gràcies també a totes les persones que d'una manera indirecta m'heu ajudat a acabar aquesta tesi, de vegades sembla que els amics no influeixen en el terreny laboral, però puc dir amb certesa que han sigut claus a l'hora de fer aquest camí més lleuger. Gràcies a tots per ser al meu costat!

Índex

Resum	13
Summary.....	14
Abreviatures.....	15
1. Introducció.....	19
1.1. Mecanismes d'EET i CT en biosistemes	23
1.1.1. Límit d'acoblament dèbil	24
1.1.1.1. Teoria de Förster per EET	25
1.1.1.2. Teoria de Marcus per CT	27
1.1.2. Límit d'acoblament fort	28
1.1.3. Règim d'acoblament intermedi	30
1.2. Modelització d'EET i CT en biosistemes	31
1.2.1. Efectes de l'entorn: Mètodes híbrids QM/Clàssics.....	31
1.2.1.1. Models continus de solvatació	32
1.2.1.2. Mètodes QM/MM	33
1.2.1.2.1. Mètodes d'inmersió electrostàtica QM/MM	34
1.2.1.2.2. Mètodes d'immersió polaritzable QM/MMPol	35
1.2.1.3. Models multiescala QM/MM/CD.....	37
1.2.1.4. Fluctuacions estructurals: Estratègia MD-QM/MM	37
1.2.2. Paràmetres EET	38
1.2.2.1 Energies d'excitació	38
1.2.2.2 Acoblaments electrònics	40
1.2.2.2.1. El model EET QM/MMPol.....	43
1.2.2.3 Densitat espectral d'acoblament vibrònic	44
1.2.2.3.1. Mètodes basats en simulacions MD	45
1.2.2.3.2. Mètodes basats en NMA	46
1.2.3. Paràmetres CT	49
1.2.3.1. Energies d'estats radical catió: Aproximació de Koopmans	49
1.2.3.2. Acoblaments electrònics: El mètode FCD.....	50
1.3. Modelització de processos EET en complexes fotosintètics: ficobiliproteïnes	52
1.3.1. Estructura i funció de les ficobiliproteïnes	53
1.3.1.1. Criptòfits.....	56
1.4. Modelització de processos CT en ADN	59
1.4.1. Estructura de l'ADN	59
1.4.2. CT a través de l'ADN.....	61
1.4.3. CT en sistemes ADN-proteïna	62
2. Objectius	67
2.1. Efectes de l'entorn en processos EET en ficobiliproteïnes.....	67
2.2. Efectes de l'entorn en processos CT en l'ADN.....	68
3. Resultats.....	75
3.1. Efectes de l'entorn en processos EET en ficobiliproteïnes.....	75
3.1.1. Molecular basis of exciton-phonon interactions in the PE545 light-harvesting complex.....	79

3.1.2. Determination of the protonation preferences of bilin pigments in cryptophyte antena complexes.....	93
3.1.3. Spectral variability in phycocyanin cryptophyte antenna complexes is controlled by changes in the α polypeptide chains	109
3.2. Efectes de l'entorn en processos CT en l'ADN.....	133
3.2.1. Single Amino Acid Mutation Controls Hole Transfer Dynamics in DNA-Methyltransferase <i>HhaI</i> Complexes.....	137
3.2.1. How Abasic Sites Impact Hole Transfer Dynamics in GC-rich DNA Sequences	147
4. Discussió	161
4.1. Efectes de l'entorn en processos EET en ficobiliproteïnes.....	162
4.1.1. Anàlisi de les funcions de densitat espectral en el complex PE545	163
4.1.2. Determinació de l'estat de protonació de les ficobilines presents en els complexes PE545, PC577, PC612, PC630 i PC645	167
4.1.3. Propietats de captura de llum dels complexes PC577, PC612, PC630 i PC645.....	170
4.2. Efectes de l'entorn en processos CT en l'ADN.....	173
4.2.1. Sistemes d'ADN en els que s'ha escindit una base de la seqüència.	174
4.2.1.1. Paràmetres estructurals	175
4.2.1.2. Paràmetres HT	177
4.2.2. Sistemes ADN-proteïna.....	179
4.2.2.1. Paràmetres estructurals	180
4.2.2.2. Paràmetres de HT	181
5. Conclusions.....	188
Bibliografia	196
APÈNDIXS.....	208
Apèndix I. Informació suplementària: <i>Molecular basis of the exciton-phonon interactions in the PE545 light-harvesting complex.</i>	210
Apèndix II. Informació suplementària: <i>Determination of the protonation preferences of bilin pigments in cryptophyte antena complexes.</i>	224
Apèndix III. Informació suplementària: <i>Spectral variability in phycocyanin cryptophyte antenna complexes is controlled by changes in the α polypeptide chains.</i>	232
Apèndix IV. Informació suplementària: <i>Single Amino Acid Mutation Controls Hole Transfer Dynamics in DNA-Methyltransferase <i>HhaI</i> Complexes.</i>	238

Resum

Aquesta tesi doctoral es centra en l'aplicació d'eines de la química teòrica i computacional a fi de comprendre l'impacte de l'entorn en i) processos de transferència d'energia electrònica (EET) en complexes pigment-proteïna fotosintètics i ii) processos de transferència de càrrega (CT) en l'ADN. Comprendre la complexa interrelació entre l'estructura de l'entorn biològic (proteïna, àcid nucleic, solvent) i la dinàmica ultrarràpida d'aquests processos és de gran interès, amb aplicacions en el desenvolupament de materials orgànics fotovoltaics o de sensors per monitoritzar la reparació de l'ADN. La multitud d'escales de temps i d'espai que caracteritzen aquests processos en biomolècules constitueixen però un repte important per les tècniques computacionals actuals. En aquesta tesi es combinen simulacions de dinàmica molecular (MD) amb càlculs multiescala de mecànica quàntica/mecànica molecular (QM/MM) a fi d'entendre diversos aspectes de la interrelació entre estructura i dinàmica d'EET/CT tant en complexes fotosintètics d'algues criptofícties com en l'ADN. En primer lloc, els resultats mostren com l'entorn local modula de forma significativa la densitat espectral d'acoblament vibrònic – i per tant la dinàmica d'EET – en la biliproteïna antena PE545 d'algues criptofícties. També resolen la incertesa relacionada amb l'estat de protonació dels pigments bilina que caracteritzen els complexes fotosintètics en criptòfits. Finalment, es presenta un model de les energies relatives dels pigments en les biliproteïnes PC577, PC612, PC630 i PC645, aspecte de gran importància ja que determina les rutes d'EET i per tant la seva funció biològica. Respecte a processos CT en ADN, la tesi estableix el mecanisme mitjançant el qual interaccions ADN/proteïna poden tant inhibir com accelerar el procés de transferència, depenent de la natura de l'aminoàcid inserit en la doble hèlix. A més, també estableix l'impacte de llocs abàsics, lesions molt comuns caracteritzades per l'absència d'una nucleobase, en la dinàmica de CT, depenent si la base desaparellada adopta una conformació a l'interior o exterior de la doble hèlix.

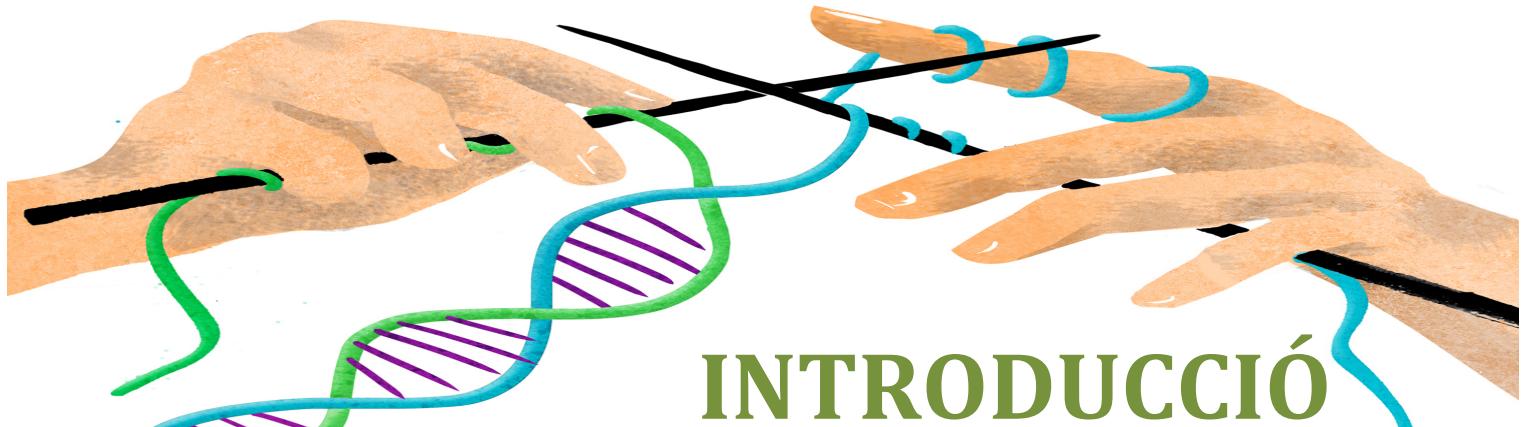
Summary

This thesis is focused on the application of theoretical and computational chemistry tools to understand the impact of the environment on i) electronic energy transfer (EET) processes in photosynthetic pigment-protein complexes and ii) charge transfer (CT) processes in DNA. Understanding the complex interrelation between the structure of the biological environment (protein, nucleic acid, solvent) and the ultrafast dynamics of these processes is of current interest, with applications in the development of photovoltaic organic materials or sensors to monitor DNA repair. The multiple time and space scales that characterize these processes in biomolecules represent however an important challenge for current computational techniques. In this thesis we combine molecular dynamics (MD) simulations with multiscale quantum mechanics/molecular mechanics (QM/MM) calculations in order to understand several aspects of the interrelation between structure and EET/CT dynamics in both photosynthetic complexes of cryptophyte algae and in DNA. First, our results show how the environment significantly modulates the spectral density of vibronic coupling – and thus EET dynamics – in the PE545 biliprotein of cryptophyte algae. We also solve the uncertainty related to the protonation state of bilin pigments that characterize cryptophyte antenna complexes. Finally, we present a model of the relative energies for the pigments in biliproteins PC577, PC612, PC630 and PC645, an important issue that determines EET pathways and therefore biological function. Regarding CT processes in DNA, the thesis establishes the mechanism by which DNA/protein interactions can both inhibit or boost the transfer process, depending on the nature of the amino acid inserted in the double helix. In addition, it establishes the impact of abasic sites, common lesions characterized by the absence of a nucleobase, in the CT dynamics, depending on whether the unpaired base adopts an intra or extrahelical conformation.

Abreviatures

A	Acceptor / <i>Acceptor</i>
A	Adenina / <i>Adenine</i>
ADN	Àcid Desoxiribonuceic / <i>Deoxyribonucleic Acid</i>
ASC	Càrregues Superficials Aparents / <i>Apparent Surface Charges</i>
ATP	Trifosfat d'Adenosina / <i>Adenosine Triphosphate</i>
BER	Reparació per Excisió de Bases / <i>Base Excision Repair</i>
C	Citosina / <i>Cytosine</i>
CBS	Conjunt Complet de funcions de Base / <i>Complete Basis Set</i>
CCSD	- / <i>Coupled Cluster method with Single and Double excitations</i>
CD	Dielèctric Continu / <i>Continuum Dielectric</i>
COSMO	- / <i>Conductor-like Screening Model</i>
CT	Transferència de Càrrega / <i>Charge Transfer</i>
D	Donador / <i>Donor</i>
DBV	15,16-Dihidrobiliverdina / <i>15,16-Dihydrobiliverdin</i>
ddCOSMO	- / <i>domain decomposition Conductor-like Screening Model</i>
DFT	Teoria del Funcional de la Densitat / <i>Density Functional Theory</i>
DHO	Oscil·lador Harmònic Desplaçat / <i>Displaced Harmonic Oscillator</i>
EET	Transferència d'Energia Electrònica / <i>Electronic energy transfer</i>
ET	Transferència d'Electrons / <i>Electron Transfer</i>
eV	Electronvolts / <i>Electronvolts</i>
FCD	- / <i>Fragment Charge Difference</i>
G	Guanina / <i>Guanine</i>
GMH	- / <i>Generalized Mulliken-Hush</i>
HOMO	Orbital Molecular Ocupat de més Energia / <i>Highest Occupied Molecular Orbital</i>
HT	Transferència de Forats / <i>Hole Transfer</i>
KMC	- / <i>Kinetic Monte Carlo</i>
MBV	Mesobiliverdina / <i>Mesobiliverdin</i>
MD	Dinàmica Molecular / <i>Molecular Dynamics</i>
MM	Mecànica Molecular / <i>Molecular Mechanics</i>
MMPol	Mecànica Molecular Polaritzable / <i>Polarizable Molecular Mechanics</i>

NADPH	Nicotinamida Adenina Dinucleòtid Fosfat reduïda / <i>Reduced Nicotinamide Adenine Dinucleotide Phosphate</i>
NMA	Anàlisi de Modes Normals / <i>Normal Mode Analysis</i>
PBP	Ficobiliproteïna / <i>Phycobiliprotein</i>
PBS	Ficobilisoma / <i>Phycobilisome</i>
PCB	Ficocianobilina / <i>Phycocyanobilin</i>
PCM	Model Continu Polaritzable / <i>Polarizable Continuum Model</i>
PDA	Aproximació de Dipols Puntuals / <i>Point Dipole Approximation</i>
PEB	Ficoeritrobilina / <i>Phycoerythrobilin</i>
PES	Superfície d'Energia Potencial / <i>Potential Energy Surface</i>
QM	Mecànica Quàntica / <i>Quantum Mechanics</i>
QMD	Dinàmica Molecular Quàntica / <i>Quantum Molecular Dynamics</i>
RMSD	Desviació de l'Arrel Quadràtica Mitjana / <i>Root Mean Square Deviation</i>
SCS-MP2	Teoria perturbativa de Møller-Plesset de segon ordre amb Components d'Espín Escalats / <i>Spin-Component-Scaled Møller Plesset second order perturbation theory</i>
SD	Densitat Espectral / <i>Spectral Density</i>
SS(V)PE	- / <i>Surface and Simulation of Volume Polarization and Electrostatics</i>
T	Timina / <i>Thymine</i>
TDC	Cub de Densitats de Transició / <i>Transition Density Cube</i>
TDDFT	Teoria del Funcional de la Densitat Dependent del Temps / <i>Time Dependent Density Functional Theory</i>
UV	Ultraviolat / <i>Ultraviolet</i>

A photograph of two hands, one light-skinned and one dark-skinned, holding a physical model of a DNA double helix. The model is made of plastic and features a green and blue color scheme with purple diagonal stripes on the blue segments. The hands are positioned as if presenting or examining the model.

INTRODUCCIÓ

1. Introducció

Comprendre com l'estructura d'un sistema biomolecular controla processos de transferència d'energia electrònica (EET) o de càrrega (CT) en sistemes biològics és clau per tal d'entendre importants processos biològics tals com la fotosíntesi o la protecció de l'ADN en front del dany oxidatiu. La mecànica quàntica representa el punt de partida per a la descripció teòrica d'aquest tipus de processos a nivell atòmic. No obstant, un "sistema biomolecular" compren des de petites molècules orgàniques fins a grans macromolècules, com ara proteïnes, àcids nucleics o inclús macrocomplexes proteïcs de membrana. La enorme complexitat d'aquests sistemes fa que un tractament exclusivament basat en la mecànica quàntica resulti intractable des d'un punt de vista computacional, mentre que un tractament aïllat de la regió d'interès on té lloc el procés estaria obviant, en essència, l'important paper que juga l'entorn i que sovint determina la funció característica i única d'un biosistema dins de la complexa maquinària biològica.

Determinar la interrelació entre la dinàmica ultrarràpida de transferència d'aquests processos, típicament de l'ordre de fs o ps i localitzada en una certa regió de l'espai, amb les múltiples escales temporals i espacials que caracteritzen l'estructura i les fluctuacions d'un sistema biomacromolecular és, per tant, un repte important per les eines de la química teòrica. Tot i que alguns grups de recerca han presentat recentment les primeres simulacions explícites de la dinàmica quàntica d'aquests processos en biosistemes,¹ aquesta estratègia té evidents limitacions en la mida dels sistemes que es poden abordar i, sobretot, en la possibilitat d'explorar l'impacte de les fluctuacions estructurals d'una biomolècula en el procés. Així, l'estratègia habitual en la descripció de fenòmens de EET o CT en biosistemes es basa en l'ús d'expressions per a la velocitat de transferència basades, per exemple, en les populars teories de Förster per EET o de Marcus per CT, i per tant en la simulació acurada dels ingredients necessaris en tals teories per a predir la velocitat: les energies dels estats electrònics involucrats, l'acoblament electrònic entre aquests estats i, l'acoblament entre tals estats i les vibracions nuclears del sistema, l'anomenat acoblament vibrònic.^{2,3} L'entorn molecular juga un paper clau en la modulació dels paràmetres EET/CT necessaris

per simular aquests processos (energies, acoblament electrònic i acoblament vibrònic). No obstant, la descripció acurada de l'efecte de l'entorn en fenòmens electrònics com l'EET o CT representa un repte important per als mètodes computacionals actuals.

En aquesta tesi s'han emprat una estratègia computacional adoptada per diversos laboratoris al món a fi d'estudiar processos d'EET i CT en biomolècules, la qual consisteix en simular les fluctuacions estructurals mitjançant simulacions de MD clàssica i posteriorment avaluar els paràmetres d'EET/CT per un conjunt d'estructures suficientment gran a partir de càlculs QM/MM multiescala. Aquesta estratègia s'ha usat per aprofundir en l'impacte de l'entorn en diversos aspectes relacionats amb fenòmens CT en l'ADN i EET en complexes fotosintètics de recaptació de llum.

Els fenòmens CT a través de la cadena d'ADN són actualment d'interès, no només per les seves implicacions en la comprensió del dany oxidatiu y les mutacions que tenen lloc en l'ADN, sinó també per les aplicacions d'aquest fenomen en el desenvolupament de biosensors capaços de monitoritzar la maquinària de reparació de l'ADN, clau en l'estudi de les bases moleculars del càncer.⁴⁻⁷ A nivell biològic, la transferència de forats (HT, migració de càrregues positives) és la més habitual i estudiada actualment. Aquesta transferència es dóna bàsicament entre guanines, ja que és la base nitrogenada que posseeix el potencial d'ionització menor.⁸⁻¹⁰ No obstant, les altres bases juguen un paper important a l'hora de modular l'acoblament electrònic entre guanines o les energies de lloc (*site energies*), aspectes claus que determinen la velocitat de HT. Diversos estudis, per exemple, han mostrat que els ponts d'hidrogen en la doble hèlix tenen un efecte important en quan a l'estabilització dels radicals catió formats en el procés de HT.¹¹⁻¹³ També es conegeuda la extremada sensibilitat de l'acoblament electrònic entre bases a la disposició relativa entre elles, de manera que, degut a la gran flexibilitat de la doble hèlix, és essencial tenir en compte una correcta descripció de l'entorn i les seves fluctuacions.⁷

D'altra banda, s'ha observat que quan l'ADN interacciona amb proteïnes es poden donar processos de CT a aminoàcids aromàtics de la proteïna, protegint així el material genètic de possibles mutacions.¹⁴⁻¹⁶ També s'ha observat que la presència de llocs abàsics, és a dir de lesions caracteritzades per l'absència d'una nucleobase,

generalment redueixen el rendiment de CT en l'ADN, fenomen que s'ha usat en biosensors per tal de monitoritzar la reparació de l'ADN, ja que el mecanisme de reparació per excisió de bases (BER) comença amb la formació d'un lloc abàsic per l'excisió d'una base danyada o desaparellada per acció de la ADN glicosilasa.^{17,18} En aquesta tesi s'ha estudiat el mecanisme mitjançant el qual una interacció ADN-proteïna o la presència de llocs abàsics modulen les propietats d'HT a l'ADN, mostrant com tal impacte depèn de la naturalesa del aminoàcid inserit a la doble hèlix o de la conformació adoptada per la base desaparellada en presència de llocs abàsics.

Pel que fa als mecanismes fotosintètics de recaptació de llum, les anomenades proteïnes antena, que contenen nombrosos pigments tals com clorofil·les, bilines o carotenoides, són les encarregades d'absorir la llum solar i transferir l'energia d'excitació resultant a través de múltiples passos d'EET fins arribar al centre de reacció, on es produeix la separació de càrregues.¹⁹ L'elevada eficiència quàntica d'aquests processos EET ha motivat una creixent interès en entendre la relació entre la funció d'aquestes proteïnes i la seva estructura des de que es va resoldre la primera estructura cristall d'una antena a alta resolució el 1975, la corresponent al complex Fenna-Matthews-Olson de les bactèries verdes del sofre.²⁰ Existeixen una gran varietat de proteïnes antena, i aquesta tesi es centra en l'estudi de les anomenades ficobiliproteïnes, que contenen pigments anomenats ficobilines, i que presenten una elevada activitat fotosintètica a intensitats molt baixes de llum.²¹ Als últims anys, la implementació de la tècnica de espectroscòpia electrònica bidimensional, a més, han posat de manifest efectes de coherència en els processos EET en complexes, inicialment interpretats com coherències electròniques, fet que ha creat una gran expectació en el camp.²²⁻²⁵ En efecte, fins al moment es creia que aquests fenòmens de coherència electrònica decauen en pocs fs degut a l'acoblament dels pigments amb els modes vibracionals desordenats de l'entorn, però l'observació d'aquests fenòmens en sistemes biològics capaços de persistir fins a centenars de femtosegons ha creat una gran debat respecte l'origen molecular i l'impacte que aquests fenòmens poden tenir sobre l'eficiència en el procés de transferència.

L'origen molecular i la naturalesa electrònica, vibracional o vibrònica de tals coherències son font de constant debat a la comunitat científica. Inicialment, es va

postular que aquests efectes podrien ser fruit de fluctuacions correlacionades de les energies d'excitació dels pigments causades pels modes vibracionals de l'entorn.^{23,24,26} No obstant, varis estudis teòrics semblen desestimar aquesta idea.²⁷⁻²⁹ Una altra hipòtesis més recent proposa que l'acoblament a vibracions amb energies capaces de completar la diferència d'energia entre diferents estats excitònics podria ser el responsable de les coherències observades.^{30,31} En ambdues hipòtesis queda clar, en tot cas, el paper clau jugat per l'entorn – i l'acoblament vibrònic – en el procés.

En aquest context, cal destacar que aquest fenomen va ser observat per primera vegada a temperatura ambient en la família de ficobiliproteïnes estudiades en aquesta tesis,²² fet que dóna una especial rellevància al seu estudi. A més, la recent obtenció de les estructures cristall de dues noves ficobiliproteïnes, en les que a diferència de la estructures obtingudes fins al moment es presenta una estructura quaternària diferent (oberta), i en les que els efectes de coherència abans mencionats desapareixen,^{32,33} ha reafirmat la importància específica de l'entorn en modular les energies i els acoblaments entre els diferents pigments presents, així com l'interès en aquests sistemes.

En aquesta tesis, doncs, s'estudien diversos aspectes de rellevància relacionats amb l'impacte de l'entorn en els mecanismes fotosintètics en criptòfits. En primer lloc, s'ha estudiat com l'entorn modula de forma diferencial la densitat espectral d'acoblament vibrònic en l'antena ficoeritrina PE545, típicament assumida igual per a tots els pigments en un complex donat. En segon lloc, s'ha estudiat l'impacte de l'entorn en els estats de protonació d'aquesta proteïna PE545 així com a les ficocianines PC577, PC612, PC630 i PC645, fet clau a fi d'obtenir simulacions fiables d'aquests sistemes. La incertesa en aquest aspecte, per exemple, ha donat lloc a diferents estudis teòrics en els quals els pigments bilina s'assumeixen tots en un estat neutre o bé protonat. Recentment, però, un estudi ha posat de manifest que l'estat de protonació d'aquests pigments té un impacte important en les energies d'excitació dels pigments, i per tant en els espectres d'absorció, fluorescència i dicroisme circular.³⁴ Finalment, s'ha estudiat la modulació de les energies dels pigments en les antenes PC577, PC612, PC630 i PC645, aspecte més difícil donades les petites diferències d'energia entre bilines, però de gran

rellevància ja que tal ordenament dona lloc als camins de EET i per tant determina la funció d'aquests complexes.

1.1. Mecanismes d'EET i CT en biosistemes

La dinàmica quàntica d'un procés EET o CT en un sistema quàntic aïllat es pot descriure resolent l'equació de Schrödinger dependent del temps. Tot i que recentment s'han presentat les primeres simulacions explícites de la dinàmica no-adiabàtica d'un sistema multicromofòric de mida considerable,^{1,35} tals simulacions són encara massa costoses computacionalment, fet que limita les escales de temps i el mida dels sistemes que es poden simular. Així, l'estratègia més comuna es basa en l'ús d'equacions perturbatives per a la velocitat. En aquest context, el Hamiltonià total del sistema es representa com la suma de les diferents parts involucrades:

$$H = H_{el} + H_{vib} + H_{el-vib} \quad (11.1)$$

on H_{el} és el Hamiltonià electrònic del sistema d'interès (energia dels estats del donador i de l'acceptor i el seu acoblament electrònic), H_{vib} és el hamiltonià del "bany", que inclou les vibracions de l'entorn (normalment representades com un conjunt d'oscil·ladors harmònics), mentre que H_{el-vib} descriu la interacció del sistema d'interès amb el bany. A partir d'aquesta expressió, i aplicant l'aproximació de Born-Oppenheimer, es pot introduir de manera perturbativa un dels elements per tal d'obtenir una expressió per a la probabilitat (velocitat) de transferència. Per tant, dependent de la força relativa de l'acoblament electrònic i l'acoblament vibrònic (amb les vibracions de l'entorn), ens trobem davant de tres escenaris diferents.

En aquest capítol discutim els tres règims que s'originen de la interacció entre l'acoblament electrònic i l'acoblament vibrònic. En primer lloc, el límit de l'acoblament dèbil, en què l'acoblament electrònic entre donador i acceptor, V , és menor que l'acoblament de cada un amb les vibracions del seu entorn, el qual es pot quantificar mitjançant l'energia de reorganització, λ . En situacions on $V \ll \lambda$,

doncs, podem descriure l'impacte de V pertorbativament, aproximació que dona lloc a les teories de Förster³⁶ i Marcus³⁷ per EET i CT, respectivament, que deriven de l'anomenada *Fermi Golden Rule* (Regla d'or de Fermi). Ambdues teories descriuen un mecanisme de *incoherent hopping* (salts incoherents) entre estats completament localitzats en una sola molècula, ja sigui en el donador o en l'acceptor, mediat per l'acoblament electrònic entre ells.

Per contra, quan $V \gg \lambda$ ens trobem en el límit d'acoblament fort. En aquest cas, l'estat electrònic està compartit de forma coherent entre el donador i l'acceptor. Mentre que en CT aquesta situació és poc habitual, en EET es dona sovint, donant lloc a estats deslocalitzats anomenats excitons. En aquest cas, el procés EET és més aviat un procés de relaxació entre estats mediat per la interacció amb el bany, ja que no hi ha transferència real entre molècules.

Més enllà d'aquests dos límits, però, en complexos fotosintètics sovint ens trobem en un règim intermedi ja que $V \sim \lambda$. En aquesta situació intermèdia, doncs, el tractament pertorbatiu d'una de les dues contribucions (acoblament electrònic entre estats o acoblament dels estats amb el bany) no és vàlid i, per tant, descriure correctament aquests processos, representa un dels reptes actuals en el camp.

1.1.1. Límit d'acoblament débil

En el límit de l'acoblament débil la velocitat de EET o CT ve donada per l'anomenada *Fermi Golden Rule*:

$$k_{if} = \frac{2\pi}{\hbar} |\langle i | \hat{V} | f \rangle|^2 \delta(E_i - E_f) \quad (11.2)$$

on E_i i E_f són les energies de l'estat inicial $|i\rangle$ i final $|f\rangle$, respectivament, i \hat{V} descriu l'acoblament electrònic entre els dos estats. A partir d'aquesta expressió, les teories de Förster per EET i de Marcus per CT introduceixen diferents aproximacions per tal d'arribar a una expressió útil per descriure la velocitat, en ambdós casos assumint que la contribució nuclear es pot separar de l'acoblament electrònic.

1.1.1.1. Teoria de Förster per EET

La transferència d'energia electrònica de manera no radiativa, o sigui, sense emissió/absorció de fotons, descriu la des-excitació d'una molècula (donador, D) que es troba en el seu estat electrònic excitat $|A\rangle$ acoblada a l'excitació d'una altra molècula (acceptor, A) produint-se per tant la promoció a un estat electrònic excitat $|DA^*\rangle$:



Per als processos EET més habituals que involucren estats singlet, l'acoblament entre D i A que fa possible la transició és bàsicament una interacció coulòmbica entre anomenades “densitats de transició”.³⁸

El 1948, Theodor Förster va proporcionar una elegant teoria³⁶ que relaciona la velocitat d'EET amb senzilles propietats espectroscòpiques del donador i de l'acceptor: el temps de vida de D (τ_D), el rendiment quàntic de fluorescència de D (ϕ_D), la distància (R) i la orientació relativa entre D i A (κ)³⁹, i la integral de solapament (J_F). L'expressió de la velocitat ve donada per:

$$k_{EET} = \frac{1}{\tau_D} \frac{9(\ln 10)\kappa^2 \phi_D J_F}{128\pi^5 N_A n^4} \frac{1}{R^6} \quad (11.4)$$

on

$$J_F = \int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda \quad (11.5)$$

on n és l'índex de refracció del medi, N_A és el nombre d'Avogadro, F_D l'espectre d'emissió de D normalitzat per àrea, i ε_A l'espectre del coeficient d'extinció molar d'A. J_F per tant s'obté a partir del solapament entre l'espectre d'emissió de D i l'espectre d'absorció d'A (Fig. 1) i assegura la condició de conservació de l'energia il·lustrada per la funció δ a la equació 11.2. Com més gran és la integral de solapament, doncs, més gran és la velocitat de transferència d'energia.

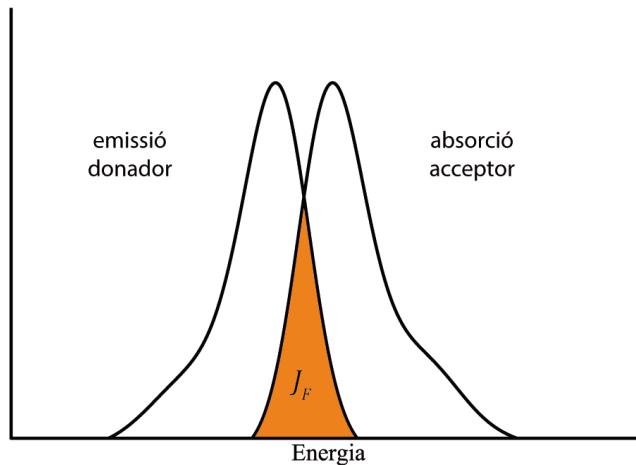


Figura 1. Integra de solapament de Förster entre l'espectre d'emissió del donador i l'espectre d'absorció de l'acceptor.

Des d'un punt de vista teòric, però, es convenient reescriure l'expressió de Förster seguint la Regla d'or de Fermi, on s'aprecia la dependència amb l'acoblament electrònic al quadrat:

$$k_{EET} = \frac{2\pi}{\hbar} |sV_{dd}|^2 J_F \quad (11.6)$$

on s és l'apantallament de l'acoblament electrònic degut al solvent (que en el model de Förster equival a $1/n^2$), i V_{dd} és l'acoblament electrònic en l'aproximació dipol-dipol que dóna lloc a la dependència R^{-6} de la velocitat mostrada a l'equació 11.4.

La teoria de Förster és àmpliament usada en nombroses aplicacions basades en EET. Tot i això, la seva aplicabilitat resta subjecta a certes suposicions²⁶, com per exemple l'adopció de l'aproximació dipol-dipol, que resulta inapropiada a distàncies curtes entre D i A; o l'adopció d'un factor d'apantallament $1/n^2$ contant. A més, en escurçar la distància D/A l'acoblament electrònic augmenta, fet que ens pot portar un règim intermedi d'EET, situació que es tracta en apartats següents (Secció 1.1.3.)

1.1.1.2. Teoria de Marcus per CT

Un procés de transferència de càrrega, com bé diu el seu nom, suposa el desplaçament d'una càrrega d'una molècula donadora a una d'acceptora. Es pot parlar de *Electron Transfer* (ET, transferència d'electrons) o de *Hole Transfer* (HT, transferència de forats) dependent de si es transporta una càrrega negativa o positiva, respectivament:



Rudolph A. Marcus va desenvolupar una teoria per a les reaccions de CT que li va suposar el Premi Nobel el 1992³⁷. Aquesta teoria va ser inicialment formulada per a reaccions CT entre molècules acoblades dèbilment. Per tant, permet treballar novament en el marc de la Regla d'Or de Fermi, considerant que els estats inicial i final estan completament localitzats en D i A.

$$k_{CT} = \frac{2\pi}{\hbar} |V_{DA}|^2 F \quad (11.9)$$

on F representa el factor de Frank-Condon que inclou l'acoblament vibrònic. No obstant, a través de certes simplificacions⁴⁰, es pot arribar a l'expressió clàssica de l'equació (11.8), que representa l'equació principal de la teoria de Marcus:

$$k_{CT} = \frac{2\pi}{\hbar} |V_{DA}|^2 \sqrt{\frac{1}{4\pi k_B T \lambda}} e^{-\frac{(\lambda + \Delta G^0)^2}{4\lambda k_B T}} \quad (11.10)$$

en què V_{DA} representa l'acoblament electrònic entre els estats inicial i final, k_B és la constant de Boltzmann, T la temperatura, ΔG^0 és l'energia lliure de Gibbs de la reacció de CT i λ és l'energia de reorganització.

L'energia lliure de Gibbs es pot determinar com la diferència entre els potencials d'oxidació del donador i l'acceptor, mentre que l'energia de reorganització descriu el canvi en l'energia lliure corresponent al pas de la configuració de reactius a

productes un cop ja s'ha donat la transferència electrònica. Aquest terme de reorganització nuclear conté dues contribucions, $\lambda = \lambda_i + \lambda_s$, una component

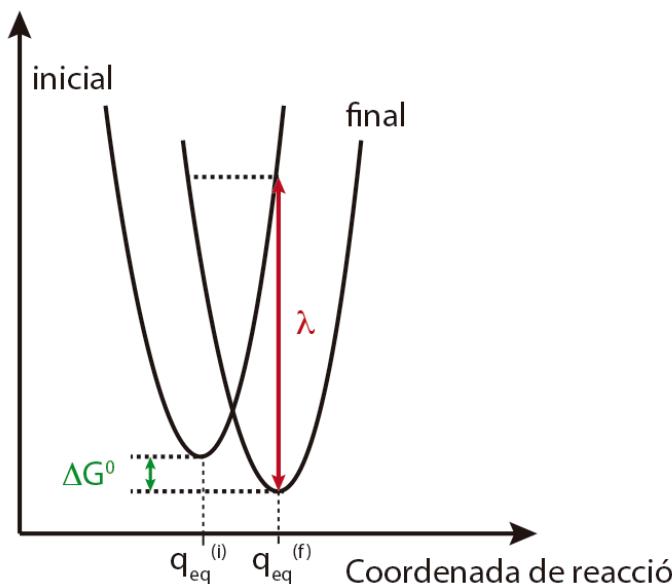


Figura 2. Representació de les superfícies d'energia potencial dels estats inicial i final en un procés de transferència de càrrega.

internal, associada a canvis en la geometria de D i A, i una component del solvent associada a canvis en la polarització de l'entorn.

Des d'un punt de vista gràfic, si representem els estats inicial i final com a paràboles en la coordenada de reacció (Fig. 2), la diferència d'energia entre el mínim de l'estat inicial i el mínim de l'estat final és la ΔG^0 . D'altra banda l'energia de reorganització és la diferència entre l'energia de l'estat inicial i final, calculat sobre la paràbola de l'estat final.

$$\lambda = E_f\left(q_{eq}^{(i)}\right) - E_f\left(q_{eq}^{(f)}\right) \quad (12.11)$$

1.1.2. Límit d'acoblament fort

Com s'ha mencionat anteriorment, quan $V \gg \lambda$ ens trobem en el límit d'acoblament fort. Tal situació és poc habitual en CT, però es dona sovint en casos d'EET. En aquest cas el tractament és més complicat, ja que degut al fort acoblament els estats excitats no estan localitzats en D i A, sinó que formen un

estat deslocalitzat que s'anomena excitó. Si prenem com a base els estats individuals de cada molècula i només considerem un estat excitat a cadascuna, aquests excitons es poden descriure mitjançant el següent Hamiltonià excitònic de Frenkel:

$$H = \sum_{i=1}^N \varepsilon_i |i\rangle\langle i| + \sum_{i \neq j} V_{ij} |i\rangle\langle j| \quad (11.12)$$

en què cada molècula i esta representada per una única energia de transició electrònica, ε_i , a la que ens referirem com a *site energy* i V_{ij} és novament l'acoblament entre les molècules i i j . Per tant, per un sistema amb N molècules acoblades tindrem una matriu amb les *site energies* a la diagonal i els acoblaments fora de la diagonal. La diagonalització d'aquest Hamiltonià permet obtenir els corresponents estats excitònics:

$$H = \sum_{k=1}^N \omega_k |k\rangle\langle k|; |k\rangle = \sum_{i=1}^N c_i^k |i\rangle \quad (11.13)$$

on $|k\rangle$ són els estat excitònics descrits com a combinació lineal dels estats de cada molècula, ω_k són les energies d'aquests estats, mentre que c_i^k són les contribucions de cada molècula i a l'excitó $|k\rangle$.

El 1957 A. G. Redfield va proposar una teoria per descriure el procés de relaxació, induït per les vibracions de l'entorn entre aquests estats excitònics deslocalitzats⁴¹. Partint de la base de que l'acoblament vibrònic és introduït de manera perturbativa, l'expressió per a la velocitat de relaxació entre dos estats excitònics deslocalitzats $|k\rangle$ i $|m\rangle$ pren la forma següent⁴²:

$$k_{k \rightarrow m} = 2\gamma_{km} \tilde{C}^{(Re)}(\omega_{km}) \quad (11.14)$$

on γ_{km} és el factor de solapament, format per pels coeficients excitònics, el radi de correlació entre les vibracions de l'entorn (R_c) i la distància centre a centre entre les molècules involucrades (R_{ij})

$$\gamma_{km} = \sum_{i,j} c_i^{(k)} c_i^{(m)} c_j^{(m)} c_j^{(k)} e^{-R_{ij}/R_c} \quad (11.15)$$

mentre que $\tilde{C}^{(Re)}(\omega)$ és la part real de la transformada de Fourier de la funció de correlació $C_{ij}(t) = \langle \delta\varepsilon_i(t)\delta\varepsilon_j(t) \rangle = C(t)e^{-R_{ij}/R_c}$, que descriu com de favorable és l'intercanvi d'energia entre l'entorn i les molècules involucrades, i que pren la forma:

$$\tilde{C}^{(Re)}(\omega) = \pi\omega^2 \{ (1 + n(\omega))J(\omega) + n(-\omega)J(-\omega) \} \quad (11.16)$$

on $J(\omega)$ és la funció de densitat espectral (de la que es parlarà en detall en la secció 1.2.2.3.) i $n(\omega)$ és el nombre mitjà d'estats vibracionals que són excitats a una determinada temperatura:

$$J(\omega) = \sum_{\xi} g_{\xi}^2 \delta(\omega - \omega_{\xi}); \quad n(\omega) = \frac{1}{e^{\hbar\omega/kT}-1} \quad (11.17)$$

La teoria de Redfield ha estat àmpliament emprada per descriure processos de transferència d'energia en sistemes fotosintètics de recaptació de llum amb pigments fortament acoblats.⁴²⁻⁴⁴ Tanmateix, falla en descriure processos en els que és necessària la transferència d'energia a múltiples estats vibracionals del bany, com és el cas de sistemes en que la diferència d'energia entre excitons és gran, o quan es treballa a baixes temperatures.⁴⁵

1.1.3. Règim d'acoblament intermedi

A la natura no tot és blanc o negre, i és per això que en nombrosos biosistemes (igual que en cristalls moleculars) l'acoblament electrònic i l'acoblament vibrònic presenten magnituds comparables, de manera que el tractament perturbatiu falla i el desenvolupament de teories per a la velocitat d'EET representa encara un repte considerable en el camp. Una primera aproximació en aquesta direcció, va ser l'anomenada *modified Redfield theory*^{46,47} (teoria de Redfield modificada). La idea, és que enllloc de tractar de manera perturbativa tot l'acoblament vibrònic, aquest es divideix en dos, els elements de la diagonal s'inclouen en el Hamiltonià de

manera no perturbativa, però els termes de fora de la diagonal no. Per tant, tot i presentar millors respecte la teoria de Redfield incial, no és capaç de descriure coherències dinàmiques.

Posteriorment s'han desenvolupat aproximacions més rigoroses, com les anomenades *Hierarchical Equations of Motion* (HEOMs, Equacions Jeràrquiques del Moviment)^{48,49}, que tot i donar formalment una solució exacta, no estan exemptes d'aproximacions.

1.2. Modelització d'EET i CT en biosistemes

Independentment de la teoria i/o les aproximacions que s'utilitzin per descriure la velocitat de CT i EET, les seves prediccions depenen de tres elements clau, a partir dels quals es podrà decidir si ens trobem en un règim d'acoblament dèbil, fort o intermedi. Aquests elements són: les anomenades *site energies* de les diferents molècules involucrades; l'acoblament electrònic entre elles i la funció de densitat espectral de l'acoblament vibrònic, que descriu la modulació dinàmica de les *site energies* degut a les vibracions de l'entorn.

La correcta caracterització d'aquests tres paràmetres mitjançant mètodes computacionals és un dels objectius bàsics d'aquesta tesi. Per tant, aquest capítol es centra en explicar les metodologies que s'han usat per a la seva descripció. En primer lloc, s'expliquen les principals estratègies que permeten incorporar l'efecte de l'entorn a nivell clàssic en càlculs QM, mentre que a continuació es descriuen les estratègies usades per a calcular els paràmetres d'EET i CT en aquesta tesi.

1.2.1. Efectes de l'entorn: Mètodes híbrids QM/Clàssics

En l'actualitat, els avenços tant a nivell de metodologies com de potència de les arquitectures dels ordinadors actuals permeten tractar a nivell QM grans sistemes, inclús formats per milers d'àtoms. Tot i això, la necessitat d'explorar l'espai conformacional en sistemes biològics de gran tamany fa que l'estudi de processos electrònics en biosistemes generalment es basi en la combinació d'una descripció a nivell QM de la regió d'interès amb una descripció clàssica més aproximada de

l'entorn. A continuació, es descriuen les característiques principals d'aquests mètodes, típicament basats en una descripció continua polaritzable del l'entorn (models QM/DC) o bé mitjançant una descripció atomística basada en camps de forces clàssics (models QM/MM).

1.2.1.1. Models continus de solvatació

Els models continus de solvatació són àmpliament usats en el context de càlculs QM per tal d'introduir de forma eficient, però tanmateix força acurada, l'efecte del solvent. En aquest models, el solvent es descriu com un medi dielèctric continu, caracteritzat per la seva constant dielèctrica macroscòpica, mentre que el solut, descrit a QM, es col·loca en una cavitat dins del dielèctric amb dimensions apropiades, típicament construïda a partir d'esferes centrades en els àtoms. Existeixen diferents models continus de solvatació, tals com el mètode Possion-Boltzmann, el model generalitzat de Born, o els models basats en ASCs, tals com el model PCM o COMSO.⁵⁰⁻⁵² Tals models difereixen principalment en la manera en que es resol la equació electrostàtica de Poisson, la qual permet calcular la polarització del medi dielèctric en resposta a la distribució de càrrega del solut, l'anomenat cap de reacció (*reaction field*). Així, es possible incorporar els efectes de l'entorn resolent l'equació de Schrödinger a partir d'un Hamiltonià efectiu format pel terme habitual del solut més el potencial creat pel camp de reacció:

$$\hat{H}_{eff}|\Psi\rangle = (\hat{H}_0 + V_{RF})|\Psi\rangle = E|\Psi\rangle \quad (12.1)$$

Donat que el terme V_{RF} , la polarització de l'entorn, depèn de la densitat electrònica del solut, aquesta expressió es resol iterativament, fet que fa que sovint tals mètodes s'anomenin mètodes SCRF (self-consistent-reaction-field), en analogia amb el procediment SCF habitual en càlculs QM. Els mètodes QM/DC, per tant, tenen en compte els efectes de mútua polarització entre el solut i el solvent. La descripció de la interacció electrostàtica solut-solvent es sol complementar amb termes empírics no-electrostàtics a fi, per exemple, de calcular energies lliures de solvatació. En aquesta tesi s'ha usat dues parametritzacions de models continus

basades en el mètode IEFPCM, el model MST desenvolupat a Barcelona^{53,54} i el model SMD desenvolupat per Cramer i Truhlar,⁵⁵ a fi de calcular els *pKas* en aigua dels pigments bilina presents en complexos fotosintètics criptòfits.

1.2.1.2. Mètodes QM/MM

Els models continus representen una estratègia eficient per tal d'incorporar l'efecte del solvent en càlculs QM. En sistemes biològics, però, l'entorn té una naturalesa heterogènia, i sovint les interaccions específiques amb aquest entorn són claus a fi d'entendre la seva funcionalitat. En aquests casos és preferible combinar la descripció QM de la regió d'interès amb una descripció atomística clàssica de l'entorn mitjançant un camp de forces (*force field*) basat en mecànica molecular (MM), donant lloc als coneguts com a models mixtes o híbrids QM/MM. Aquesta secció es centra en aquest tipus de models, que s'han emprat extensament durant aquesta tesi per tal de descriure l'impacte de l'entorn en processos EET i CT.

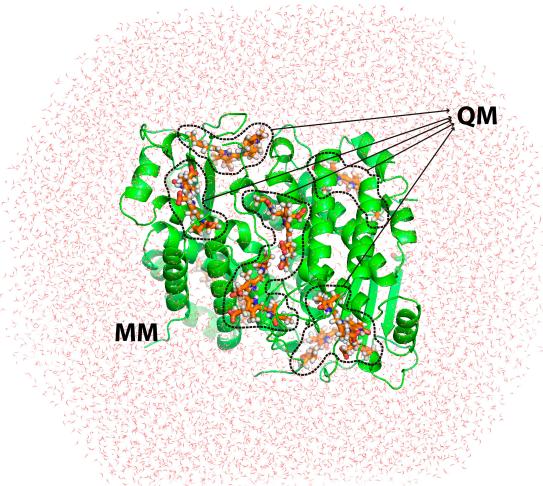


Figura 3. Representació esquemàtica de les particions clàssica (MM) i quàntica (QM) d'un sistema proteic amb un model de solvatació explícit.

Els mètodes QM/MM van ser introduïts per primer cop el 1976 per Warshel i Levitt.⁵⁶ En els coneguts com a models QM/MM additius, l'energia total del sistema s'expressa com la suma d'una contribució deguda a la molècula o molècules QM

d'interès, un terme MM degut a l'entorn, i un terme QM/MM que descriu la interacció entre les dues regions. De forma similar al descrit per models continus, això permet definir un Hamiltonià efectiu, H_{eff} , que permet resoldre l'equació de Schrödinger a partir de la següent expressió:

$$\hat{H}_{eff}|\Psi\rangle = (\hat{H}_0 + \hat{H}_{MM} + \hat{H}_{QM/MM})|\Psi\rangle = E|\Psi\rangle \quad (12.2)$$

on \hat{H}_0 és el Hamiltonià en fase gas del subsistema d'interès (QM), mentre que \hat{H}_{MM} representa l'energia de l'entorn i $\hat{H}_{QM/MM}$ introduceix l'acoblament entre la regió QM i la MM.

1.2.1.2.1. Mètodes d'inmersió electrostàtica QM/MM

Als mètodes QM/MM estàndard la descripció electrostàtica de l'entorn es sol limitar a un conjunt de càrregues puntuals extretes d'un *force field* clàssic, estratègia coneguda amb el nom de immersió electrostàtica (*electrostatic embedding*). En aquest cas, els operadors MM i QM/MM venen donar per les següents expressions:

$$\hat{H}_{MM} = \sum_m \sum_{n>m} \frac{q_m q_n}{r_{mn}} \quad (12.3)$$

$$\hat{H}_{QM/MM} = \sum_m q_m \hat{V}(r_m) \quad (12.4)$$

on $\hat{V}(r_m)$ correspon a l'operador del potencial electrostàtic degut als electrons i als nuclis de la regió QM sobre els àtoms de la regió MM, mentre que els índex m (n) corresponen a les càrregues q de la regió MM.

Així, en aquests models la regió MM és capaç de polaritzar la funció d'ona de la regió QM, però no es té en compte la polarització que la regió QM exerceix sobre la regió MM, ja que en aquesta última sols es considera el terme electrostàtic. Per tant, tot i representar un salt de qualitat a l'hora de descriure processos que tenen lloc a l'interior de grans biomolècules, són incapços de descriure, per exemple, la

reorganització electrònica ultrarràpida que té lloc a la regió MM quan es dóna una transició electrònica en la regió QM, per exemple després de l'absorció d'un fotó.

1.2.1.2.2. Mètodes d'immersió polaritzable QM/MMPol

Els models QM/MM poden ser estesos per tenir en compte la mútua polarització entre els subsistemes quàntic i clàssic mitjançant un cap de forces polaritzable. En aquests casos, la descripció electrostàtica basada en càrregues puntuals es complementa amb un model explícit dels efectes de polarització basat en el model de dipols induïts⁵⁷, el model de càrregues fluctuants⁵⁸, o el model basat en oscil·ladors clàssics de Drude⁵⁹. En aquesta tesi s'ha utilitzat bàsicament el model QM/MM polaritzable (QM/MMPol) basat en dipols induïts desenvolupat al nostre grup, en col·laboració amb el grup de la Prof. Mennucci de la Universitat de Pisa,⁵⁷ i és per això que aquest apartat es centra en la descripció d'aquest model en particular.

En el model de dipols induïts, cada àtom dins el subsistema MM està descrit per una càrrega puntual i una polaritzabilitat isotòpica⁶⁰, de manera que tant \hat{H}_{MM} com $\hat{H}_{QM/MM}$ es desglossen en un terme electrostàtic i un corresponent a la polarització, $\hat{H} = \hat{H}^{el} + \hat{H}^{pol}$. Per tant, les expressions anteriors prenen la forma:

$$\hat{H}_{MM} = \sum_m \sum_{n>m} \frac{q_m q_n}{r_{mn}} - \frac{1}{2} \sum_a \mu_a^{ind} \sum_m \frac{q_m (r_a - r_m)}{|r_a - r_m|^3} \quad (12.5)$$

$$\hat{H}_{QM/MM} = \sum_m q_m \hat{V}(r_m) - \frac{1}{2} \sum_a \mu_a^{ind} \hat{E}_a^{solut}(r_a) \quad (12.6)$$

on $\hat{E}_a^{solut}(r_a)$ correspon a l'operador del camp elèctric degut als electrons i als nuclis de la regió QM sobre els àtoms de la regió MM, mentre que l'índex a correspon als àtoms polaritzables i μ_a^{ind} als dipols induïts situats en la regió MM. Aquests dipols, situats en cada àtom polaritzable de la regió MM, s'obtenen a través de l'expressió següent:

$$\mu_a^{ind} = \alpha_a (E_a^{solut} + E_a^{entorn}\{q; \mu_a^{ind}\}) \quad (12.7)$$

en la que s'assumeix una aproximació lineal, s'obvia qualsevol contribució de caràcter magnètic i s'utilitza una polaritzabilitat isotòpica (α_a)⁶⁰. Com que E_a^{entorn} correspon al camp elèctric de l'entorn calculat a cada punt de la regió MM i al mateix temps conté la suma de contribucions de les càrregues puntuals i dipols induïts de cada àtom d'aquesta, això fa que cada dipol induït depengui de tots els altres dipols induïts de l'entorn, i per tant, aquesta expressió hagi de ser resolta iterativament o mitjançant tècniques d'inversió de matriu.

Cal remarcar que, en l'expressió de \hat{H}_{MM} , mentre que el primer terme representa l'energia electrostàtica de les càrregues de la regió MM, el segon terme descriu la interacció de polarització entre aquestes càrregues i els dipols induïts. Per tant, el primer terme representa només una constant dins el Hamiltonià efectiu, mentre que el segon es considera de forma explícita en l'operador de Fock degut a la dependència dels dipols induïts en la funció d'ona de la regió QM. Cal destacar també que en la pràctica només s'inclou la polarització dels àtoms de la regió MM que es troben dins d'una certa distància límit respecte la regió QM, l'anomenat *cutoff* de polarització, ja que l'efecte de la polarització mútua entre àtoms que es troben a distàncies majors de 15-20 Å té un efecte negligible en l'energia i les propietats del sistema QM, reduint d'aquesta forma considerablement el cost computacional.

El desenvolupament d'un model híbrid QM/MM polaritzable ha representat un salt de qualitat a l'hora de descriure processos EET en entorns heterogenis complexes, en els que una descripció continua de l'entorn representa un limitació important. L'impacte més important es veu reflectit a l'hora de calcular els acoblaments electrònics (secció 1.2.2.2.1.), ja que la inclusió d'un entorn polaritzable heterogeni permet simular els efectes d'apantallament deguts a l'entorn de forma atomística.⁶¹ Tot i això, els efectes de polarització també poden canviar qualitativament la direcció del desplaçament solvatocròmic en cromòfors biològics, tal i com s'ha mostrat recentment, per exemple, per la proteïna verda fluorescent.⁶²

1.2.1.3. Models multiescala QM/MM/CD

En aplicar un model híbrid QM/MM o QM/MM polaritzable, tant la proteïna com les aigües de solvatació estan representades de forma explícita, i per tal de capturar-ne l'efecte dinàmic és precís realitzar una simulació de MD per tal de mostrejar vàries estructures i realitzar els càlculs QM/MM en cada una d'elles (secció 1.2.1.4), amb el cost computacional que això comporta. Si es vol realitzar un càlcul sobre una estructura cristall incloent l'efecte del solvent, això directament no és possible més enllà de les aigües cristal·logràfiques presents en l'estructura. Una alternativa interessant en aquest cas és utilitzar un model híbrid a tres nivells QM/MM/CD, en el qual la regió d'interès segueix rebent un tractament QM, l'entorn proteic es descriu mitjançant un model MM o MMPol, i el solvent es descriu mitjançant un model dielèctric continu. D'aquesta forma es defineix un Hamiltonià efectiu similar al descrit en l'equació (12.2) que incorpora la interacció dels subsistemes QM i MM amb el continu dielèctric (CD):

$$\hat{H}_{eff} = \hat{H}_0 + \hat{H}_{MM} + \hat{H}_{QM/MM} + \hat{H}_{QM/CD} + \hat{H}_{MM/CD} \quad (12.8)$$

Tals models en principi es poden desenvolupar a partir de qualsevol model dielèctric continu de solvatació, com ara el PCM⁶³ desenvolupat a Pisa, el SS(V)PE⁶⁴, el COSMO⁶⁵ o el recentment desenvolupat ddCOSMO.^{66,67} En aquesta tesi s'ha usat la recent implementació QM/MMpol/ddCOSMO⁶⁸ per calcular els energies de transició de pigments en complexes fotosintètics. El model ddCOSMO usa un nou algoritme de descomposició per dominis que permet reduir el cost computacional associat a la resolució de les equacions COSMO dràsticament, permeten ser aplicat d'aquesta forma a biomolècules de grans dimensions.

1.2.1.4. Fluctuacions estructurals: Estratègia MD-QM/MM

Les fluctuacions estructurals en sistemes com l'ADN modulen de forma molt important tant les energies com sobretot els acoblaments electrònics necessaris per tal de descriure processos CT.^{69,70} En menor grau, tals efectes també són

importants a fi d'obtenir paràmetres EET realistes en complexes fotosintètics. Tals fluctuacions es poden descriure de forma en principi acurada mitjançant la realització de simulacions MD a nivell QM/MM, o inclús a nivell QM, estratègies comunes en estudis, per exemple, de catàlisi enzimàtica. Això, no obstant, limita les escales temporals accessibles a les simulacions. En aquesta tesi, s'usa una estratègia adoptada per diferents grups d'investigació en l'estudi de processos EET i CT basada en combinar simulacions de MD clàssiques amb un processat *a posteriori* d'un conjunt representatiu d'estructures a fi d'avaluar els paràmetres CT o EET amb càlculs QM/MM o QM/MMPol. Avui en dia, en efecte, les simulacions de dinàmica molecular han evolucionat de tal manera que els *force fields* actuals permeten descriure de manera força acurada les fluctuacions de proteïnes i de l'ADN en l'escala de ms.⁷¹⁻⁷³ L'estratègia que s'ha seguit en aquesta tesi, doncs, consisteix en realitzar simulacions de MD per tal de mostrejar el sistema en el seu l'estat fonamental. A partir d'aquesta trajectòria s'extreuen diverses estructures sobre les quals es realitzen els càlculs QM/MM d'estats excitats, permeten així obtenir valors mitjanats estadísticament per als paràmetres CT o EET calculats.

1.2.2. Paràmetres EET

En aquesta secció es descriuen els mètodes usats per tal d'estimar els paràmetres EET en complexes pigment-proteïna fotosintètics: les energies d'excitació, els acoblaments electrònics, i la densitat espectral d'acoblament vibrònic amb els graus de llibertat nuclears del sistema.

1.2.2.1 Energies d'excitació

Les conegudes com a *site energies* són les energies d'excitació electrònica de les diferents molècules que intervenen en el procés de EET. En aquesta tesi, les energies dels pigments bilina en complexes fotosintètics s'han calculat usant la metodologia QM/MMPol descrita anteriorment, en la seva extensió al context de la teoria del funcional de la densitat dependent del temps (TD-DFT).⁷⁴ Tot i que no està exclosa de problemes, aquesta metodologia permet calcular de forma eficient i acurada excitacions electròniques en pigments de mida considerable, tals com els

pigments bilina estudiants en aquesta tesi. El càlcul d'energies de transició a partir de mètodes QM/MM a nivell TD-DFT s'ha discutit en nombrosos treballs. Les energies d'excitació es poden obtenir a través de la següent expressió:

$$\begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix} \begin{pmatrix} X_n \\ Y_n \end{pmatrix} = \omega_n \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \quad (12.9)$$

On les matrius A i B constitueixen la matriu Hessiana de l'energia electrònica, i els vectors $(X_n \ Y_n)$ representen modes normals col·lectius de la matriu densitat amb freqüències ω_n . Per tal d'incloure els efectes de l'entorn polaritzable en les matrius A i B es consideren els dipols induïts en la regió MM per la matriu densitat, associats als vectors de transició $(X_n \ Y_n)$, de manera que squestes prenen la forma:⁵⁷

$$\begin{aligned} A_{ai,bj} &= \delta_{ab} \delta_{ij} (\varepsilon_a - \varepsilon_i) + K_{ai,bj} + C_{ai,bj}^{pol} \\ B_{ai,bj} &= K_{ai,bj} + C_{ai,bj}^{pol} \end{aligned} \quad (12.10)$$

on ε_r són les energies dels orbitals, on (i, j, \dots) representen els orbitals moleculars ocupats i (a, b, \dots) els virtuals, i $K_{ai,bj}$ i $C_{ai,bj}^{pol}$ són la matriu d'acoblament i la matriu polaritzable MM, respectivament:

$$K_{ai,bj} = \int dr \int dr' \phi_i(r') \phi_a^*(r') \left(\frac{1}{|r'-r|} + g_{xc}(r', r) \right) \phi_j(r) \phi_b^*(r) \quad (12.11)$$

$$C_{ai,bj}^{pol} = - \sum_k \left(\int dr \phi_i(r) \phi_a^*(r) \frac{(r_k - r)}{|r_k - r|^3} \right) \mu_k^{ind}(\phi_j \phi_b^*) \quad (12.12)$$

on l'índex k en l'equació 12.12 representa el nombre total d'àtoms polaritzables a la regió MM. A continuació ens centrem en els aspectes més específics relacionats amb processos EET, en concret en el càlcul dels acoblaments electrònics i de la densitat espectral.

1.2.2.2 Acoblaments electrònics

L'acoblament electrònic és el segon paràmetre clau a l'hora simular processos EET en sistemes biològics. Com s'ha mencionat anteriorment, la magnitud d'aquest acoblament comparada amb la de l'acoblament amb les vibracions de l'entorn, determinarà si els estats del sistema estan localitzats o deslocalitzats formant excitons. Formalment, l'acoblament electrònic descriu l'element de matriu entre l'estat inicial i final del procés, $V_{DA} = \langle D^* A | \hat{H}_{el} | DA^* \rangle$, on \hat{H}_{el} és el hamiltonià electrònic del sistema, mentre que $\langle D^* A |$ i $| DA^* \rangle$ són els estats excitats localitzats en D i A. Aquest terme es pot escriure mitjançant una expressió perturbativa⁷⁵ com la suma de tres contribucions diferents, $V_{DA} = V_{Coul} + V_{exch} + V_{ovlp}$ on V_{Coul} és la interacció coulòmbica entre les densitats de transició, V_{exch} és la interacció d'intercanvi de Dexter,⁷⁶ i V_{ovlp} és un terme proporcional al solapament entre les densitats de transició:

$$V_{DA}^{Coul} = \int dr \int dr' \rho_D^{T*}(r') \frac{1}{|r-r'|} \rho_A^T(r) \quad (12.13)$$

$$V_{DA}^{exch} = - \iint dr dr' \gamma_D^{T*}(r, r') \frac{1}{|r-r'|} \gamma_A^T(r, r') \quad (12.14)$$

$$V_{DA}^{ovlp} = - \int dr \int dr' \rho_D^{T*}(r') \rho_A^T(r) \quad (12.15)$$

on la densitat de transició (ρ_A^T) correspon a la part diagonal de la matriu densitat d'una partícula construïda a partir de les funcions d'ona de l'estat fonamental (*ef*) i l'estat excitat (*eex*):

$$\gamma_{eex-ef}^T(r, r') = N \int \dots \int \Psi_{eex}^*(r, r_2 \dots r_N) \Psi_{ef}(r', r_2 \dots r_N) dr_2 dr_3 \dots dr_N \quad (12.16)$$

$$\rho_{eex-ef}^T(r) = \gamma_{eex-ef}^T(r, r) \quad (12.17)$$

Quan considerem transferències d'energia permeses entre estats singlet, l'acoblament electrònic està dominat per la seva component Coulòmbica, mentre que si considerem excitacions prohibides o que involucrin estats triplet, la

component és nul·la i entren en joc altres mecanismes d'interacció electrònica de curt abast. En aquests casos, una forma més precisa d'estimar l'acoblament és a partir de càlculs d'estats excitats realitzats sobre la supermolècola D+A. Si el sistema és simètric, l'acoblament correspon a la meitat del desdoblament excitònic entre els dos estats ressonants involucrats, $|V| = |E_1 - E_2|/2$.⁷⁷ Altres mètodes més generals es basen en la diabatització dels estats involucrats i el càlcul directe del seu acoblament.²

Aquesta tesi es centra només en l'acoblament electrònic per EETs entre estats singlet, els quals caracteritzen els processos de captura de llum en els complexes fotosintètics criptofites estudiats. En aquests casos el terme Coulòmbic és dominant i els termes de curt abast es poden negligir. En aquesta tesi s'ha calculat aquest terme a partir de les densitats de transició derivades de càlculs TD-DFT QM/MMPol, seguint la equació 12.13. Aquesta estratègia, doncs, permet evitar l'aproximació en dipols puntuals (PDA) habitualment usada en estudis d'EET, la qual dona lloc a la coneguda dependència R^{-6} de la velocitat d'EET en la teoria de Förster:

$$V_{DA}^{PDA} = |\mu_D| |\mu_A| \frac{\kappa}{R_{DA}^3} \quad (12.18)$$

$$\kappa = \hat{\mu}_D \cdot \hat{\mu}_A - 3(\hat{\mu}_D \cdot \hat{R}_{DA})(\hat{\mu}_A \cdot \hat{R}_{DA}) \quad (12.19)$$

on $|\mu_D|$ i $|\mu_A|$ són els mòduls dels dipols de transició de D i A, \hat{R}_{DA} és la distància entre dipols i κ representa la orientació relativa entre ells. L'avantatge d'aquesta aproximació rau bàsicament en la seva simplicitat, ja que sols requereix conèixer els dipols de transició i posició relativa dels cromòfors, informació que sovint es pot extreure de mesures experimentals. D'altra banda, el seu límit d'aplicabilitat està acotat a distàncies intermoleculars R_{DA} majors que les dimensions de la molècula en si, obtenint errors significatius per distàncies curtes.⁷⁸

Una alternativa eficient al càlcul explícit del terme Coulòmbic, però que a diferència de l'aproximació PDA també té en compte els detalls de la forma dels cromòfors, es basa en expressar les densitats de transició a partir d'un conjunt de càrregues atòmiques de transició (TrCh). En aquest cas l'acoblament Coulòmbic ve donar per:

$$V_{AD}^{TrCh} = \sum_{K,L} \frac{q_A(K)q_D(L)}{|R_K - R_L|} \quad (12.20)$$

on K i L iteren sobre els diferents àtoms de D i A, respectivament, i q_X i R_X són les càrregues de transició i les seves posicions a l'espai. Aquestes es poden estimar a partir d'anàlisis de població com el de Mulliken. Una estratègia més precisa, però, consisteix en derivar-les per ajust del potencial electrostàtic al voltant de la molècula, tal i com es fa habitualment per derivar models de càrregues en camps de forces clàssics.⁷⁹ Tot i obtenir l'acoblament de manera aproximada, aquest mètode representa una alternativa ràpida i plausible quan s'ha de calcular un gran nombre d'acoblaments, per exemple, quan es parteix d'una simulació MD.

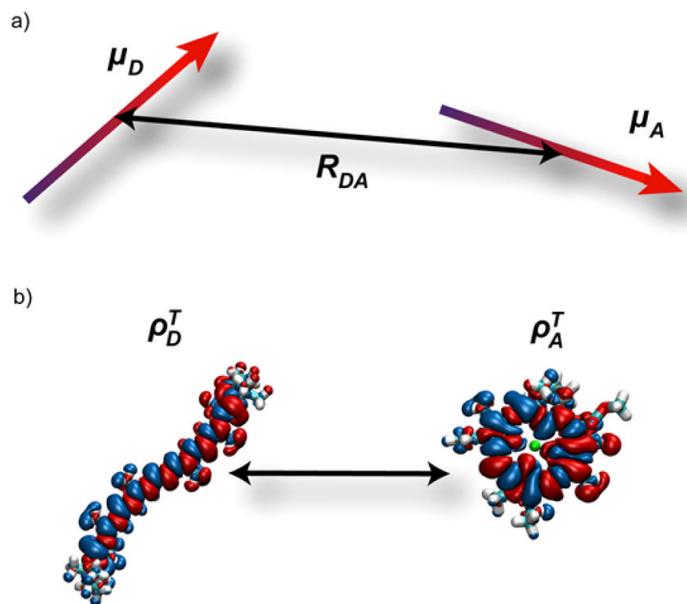


Figura 4. Representació gràfica de les densitats de transició de dues molècules utilitzant diferents aproximacions per calcular-ne la interacció. (a) PDA , equació (13.18). (b) Interacció completa entre les densitats de transició, equació (13.13).

D'altra banda, l'entorn pot introduir grans variacions en l'acoblament electrònic. En primer lloc, l'entorn pot modificar l'estrucció tant geomètrica com electrònica, impactant tant les energies d'excitació com les densitats i els dipols de transició de les molècules involucrades. Aquest efecte indirecte és menor però comparat amb l'impacte explícit de l'entorn en la interacció Coulòmbica entre D i A, que induceix un fort apantallament en l'acoblament. La teoria de Förster descriu aquest efecte mitjançant el factor d'apantallament $s = 1/\epsilon_{opt}$, on la constant dièlectrica del medi

s'expressa a partir del l'índex del seu índex de refracció $\varepsilon_{opt} = n^2$. En biomolècules és comú adoptar un valor $\varepsilon_{opt} = 2$, el qual redueix la velocitat d'EET per 4, fet que il·lustra la importància d'aquest efecte. En estudis teòrics d'EET es sol adoptar aquest factor d'apantallament, o altres també constants basats en considerar dos dipols situats a cavitats esfèriques dins del dielèctric, $s = 3/(2\varepsilon_{opt} + 1)$.⁸⁰ Tanmateix, el desenvolupament de mètodes híbrids QM/Clàssics ha revelat que el factors d'apantallament depèn de la forma i orientació de les molècules acoblades, així com de la distància entre elles.^{81,82} En aquesta tesi aquest efecte s'ha descrit a nivell atòmic mitjançant el model EET QM/MMPol descrit a continuació.

1.2.2.2.1. El model EET QM/MMPol

Tal i com s'ha mencionat anteriorment (secció 1.2.1.2.2.) el model QM/MMPol adopta una descripció atomística polaritzable de l'entorn MM clàssic. En aquest context, l'acoblament es pot expressar com la suma del terme Coulòmbic i un terme explícit mediat per l'entorn MM:⁵⁷

$$V_{DA} = V_{DA}^{Coul} + V_{DA}^{MMPol} \quad (12.21)$$

$$V_{DA}^{Coul} = \int dr \int dr' \rho_D^T(r') \frac{1}{|r-r'|} \rho_A^T(r) \quad (12.22)$$

$$V_{DA}^{MMPol} = - \sum_k \left(\int dr \rho_D^T(r') \frac{(r'-r_k)}{|r'-r_k|^3} \right) \mu_k^{ind}(\rho_A^T) \quad (12.23)$$

A partir d'aquesta expressió, es pot definir el factor d'apantallament com:

$$s = \frac{V_{DA}}{V_{DA}^{Coul}} = \frac{V_{DA}^{Coul} + V_{DA}^{MMPol}}{V_{DA}^{Coul}} \quad (12.24)$$

El model MMPol ha permès avaluar per primera vegada l'impacte que té un entorn polaritzable heterogeni en processos EET, suposant un salt qualitatiu a l'hora de descriure l'impacte de l'entorn en els acoblaments electrònics en biosistemes. En efecte, la seva aplicació a diferents complexes fotosintètics,^{61,83-85} a àcids nucleics^{86,87} i a complexes proteïna-lligand⁷⁴ indiquen que de mitja s'obté un efecte

similar al descrit per un model continu a partir d'un valor $\varepsilon_{opt} = 2$, tal i com es sol assumir en biosistemes. Tanmateix, s'ha observat que aquests valors depenen fortament de l'entorn local i la posició relativa dels cromòfors involucrats, donen

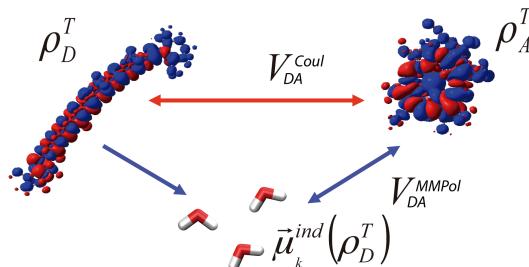


Figura 5. Representació gràfica de les densitats de transició de dues molècules i els acoblaments directe i mediats per l'entorn.

lloc a valors d'apantallament corresponents a constants dielèctriques efectives entre ~ 1 i 3.

1.2.2.3 Densitat espectral d'acoblament vibrònic

Les propietats dinàmiques de l'entorn juguen un paper important a l'hora de descriure processos EET. És fonamental descriure la modulació dinàmica, tant de les energies d'excitació com dels acoblaments electrònics, degut a l'acoblament entre els estats electrònics i les vibracions nuclears de l'entorn (proteïna i solvent). Aquest acoblament dóna lloc a fluctuacions en els nivells d'energia dels diferents elements involucrats, que pot conduir, per exemple, a una localització excitònica. Aquestes propietats dinàmiques derivades de l'acoblament amb l'entorn es concentren en el que s'anomena la funció de densitat espectral $J_i(\omega)$, que descriu la força de l'acoblament entre el sistema i les vibracions de l'entorn, en funció de la freqüència, per a cada excitació i . La funció de densitat espectral es divideix generalment en dos termes:

$$J_i(\omega) = J_{0,i}(\omega) + J_{vib,i}(\omega) \quad (12.25)$$

on $J_{0,i}(\omega)$ descriu l'acoblament de les transicions electròniques a un «bany» continu de modes vibracionals de baixa freqüència deguts a la proteïna i el solvent.

D'altra banda, $J_{vib,i}(\omega)$ descriu l'acoblament a modes vibracionals discrets d'alta freqüència, deguts bàsicament a vibracions intramoleculars.

Experimentalment, la funció de densitat espectral es pot mesurar mitjançant diferents tècniques espectroscòpiques, com les anomenades *spectral hole-burning* o *fluorescence line-narrowing*. No obstant, per un sistema amb múltiples estats excitònics, només és possible obtenir la funció de densitat espectral per l'excító de més baixa energia, o una mitjana entre diferents estats poblats de baixa energia. En aquest context, les simulacions teòriques permeten explorar si tals densitats espectrals poden ser modulades per diferents entorns locals, i impactar d'aquesta forma la dinàmica d'EET i els possibles efectes de coherència observats experimentalment.

A rel de la importància de la densitat espectral, als últims anys ha crescut l'interès per desenvolupar tècniques computacionals precises per a la estimació d'aquesta quantitat. Les diferents tècniques a l'abast es divideixen bàsicament en dos grans grups, les basades en l'exploració de les fluctuacions estructurals del sistema a través de simulacions MD, i les basades en anàlisis de modes normals de vibració (NMA).

1.2.2.3.1. Mètodes basats en simulacions MD

Aquestes tècniques es basen en descriure l'evolució temporal de les energies d'excitació, combinant simulacions de dinàmica molecular clàssiques en l'estat fonamental amb càlculs QM/MM dels estats excitats al llarg de la trajectòria. És el que es podria anomenar novament MD-QM/MM (secció 1.2.1.4.). A diferència del cas anterior però, en el que només es pretén mitjanar certes propietats sobre les fluctuacions estructurals, en aquest cas l'interès es centra en l'evolució temporal d'aquestes. Per tant, aquesta tècnica requereix la realització de càlculs QM/MM d'estats excitats al llarg de la trajectòria cada 2-5 fs, per tal de capturar les oscil·lacions més ràpides del sistema que s'acoblen a les energies d'excitació, resultants de les vibracions dels enllaços entre àtoms pesats. Això significa una quantitat massiva de càlculs QM/MM, fet que resulta un dels factors limitants d'aquesta tècnica, restringint la seva aplicació a pocs centenars de picosegons, de

manera que els moviments més lents queden inexplorats, al mateix temps que limita el nivell de càlcul emprat en la regió QM.

Aquesta tècnica permet la determinació simultània de les dues contribucions mencionades $J_{0,i}(\omega)$ i $J_{vib,i}(\omega)$, ja que la funció de densitat espectral s'obté a partir de la transformada de Fourier de la funció d'autocorrelació de les fluctuacions de les energies d'excitació:

$$J_i(\omega) = \frac{\beta\omega}{\pi} \int_0^\infty C_i^{cl}(t) \cos(\omega t) dt \quad (12.26)$$

on $\beta = 1/K_B T$ i $C_i^{cl}(t)$ és la funció d'autocorrelació clàssica de les fluctuacions de les energies d'excitació que ve donada per l'expressió:

$$C_i^{cl}(t_j) = \frac{1}{N-j} \sum_{k=1}^{N-j} \Delta E_i(t_j + t_k) \Delta E_i(t_k) \quad (12.27)$$

El principal inconvenient d'aquesta tècnica rau en la utilització de simulacions MD basades en camps de forces clàssics aproximats per descriure els moviments nuclears, fet que porta a una sobreestimació de l'acoblament als modes vibracionals intramoleculars d'alta freqüència. Una possible solució a aquestes limitacions és el desenvolupament de camps de forces especialitzats capaços de descriure de manera acurada la superfície d'energia potencial (PES) quàntica de les molècules involucrades, o l'ús de simulacions QM/MM MD, en que les molècules d'interès són descrites a nivell QM al llarg de tota la dinàmica.⁸⁸

1.2.2.3.2. Mètodes basats en NMA

Una alternativa al mètodes basat en MD descrit en l'apartat anterior consisteix en calcular l'acoblament vibrònic a partir de tècniques QM d'anàlisi dels modes normals de vibració. L'essència d'un anàlisi de modes normals és la diagonalització de la matriu Hessiana, la inversa de la qual proporciona la matriu de covariància, que conté una descripció detallada dels desplaçaments de tots els àtoms sobre la posició d'equilibri (mínim d'energia), incloent les fluctuacions individuals i les

correspondents correlacions. La idea principal d'aquesta tècnica, és que tot sistema en equilibri fluctua al voltant d'una sola conformació definida, de manera que aquestes fluctuacions es poden calcular assumint una forma simple i harmònica del potencial.

Un model que permet relacionar aquest anàlisi NMA amb la densitat espectral és l'anomenat *displaced harmonic oscillator model* (model de l'oscil·lador harmònic desplaçat) o *vertical gradient model* (model del gradient vertical), que descriu l'acoblament vibracional en l'estat fonamental i l'estat excitat. Aquest model es basa en certes aproximacions, la primera, igual que en el cas dels modes normals, és que la PES de l'estat fonamental té una forma harmònica. La segona, rau en que les segones derivades tant de la PES de l'estat fonamental com de l'estat excitat són iguals, de manera que la única diferència entre aquests dues PES és la posició del mínim, que es troba desplaçat.

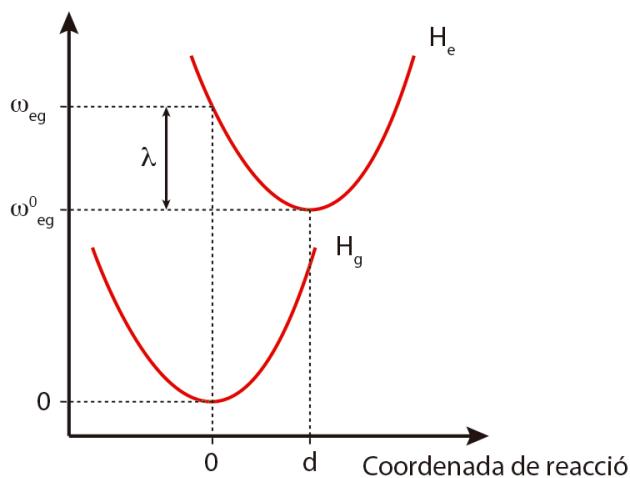


Figura 6. Representació del model DHO. On H_g i H_e són les superfícies d'energia potencial dels estats fonamental i excitat, respectivament.

La idea principal és que en la posició d'equilibri de l'estat fonamental, el gradient de la PES de l'estat excitat (*vertical gradient*) és igual que el gradient de la diferència d'energia (*energy gap*) U , que ve donat per la diferència entre les dues PES:

$$H_g = \sum_{i=1}^M \left(\frac{p_i^2}{2m_i} + \frac{1}{2} m_i \omega_i^2 q_i^2 \right) \quad (12.28)$$

$$H_e = \omega_{eg}^0 + \sum_{i=1}^M \left(\frac{p_i^2}{2m_i} + \frac{1}{2} m_i \omega_i^2 (q_i - d_i)^2 \right) \quad (12.29)$$

$$U(\tilde{q}) = H_e - H_g = \omega_{eg}^0 - \sum_i \left(\omega_i^2 \tilde{d}_i \tilde{q}_i + \frac{1}{2} \omega_i^2 \tilde{d}_i^2 \right) \quad (12.30)$$

on H_g i H_e són les PES dels estats fonamental i excitat, d és el desplaçament entre les posicions d'equilibri de l'estat fonamental i l'excitat, ω_{eg}^0 és l'energia d'excitació adiabàtica i ω_i és la freqüència del mode vibracional i .

El desplaçament d_i del mode i es pot obtenir mitjançant $d_i = -f_i/\omega_i^2$, on f és el gradient vertical en coordenades normals, que es pot obtenir a partir del gradient vertical en coordenades cartesianes, la matriu de masses nuclears i els modes normals de vibració expressats en coordenades cartesianes mitjanades per massa. D'aquesta manera s'obtenen finalment els factors Huang-Rhys de cada mode vibracional, a partir dels quals s'obté la funció de densitat espectral:

$$S_i = \frac{\omega_i \tilde{d}_i^2}{2\hbar} = \frac{\tilde{f}_i^2}{2\hbar \omega_i^3} = \frac{\lambda_i}{\hbar \omega_i} \quad (12.31)$$

$$\tilde{C}''(\omega) = \pi \omega \sum_{i=0} S_i \omega_i \quad (12.32)$$

Aquesta tècnica ha estat emprada per al càlcul de les contribucions intramoleculars de la funció de densitat espectral,⁸⁹ $J_{vib,i}(\omega)$, mentre que la contribució contínua de la funció, $J_{0,i}(\omega)$, s'introduceix mitjançant un oscil·lador Brownià sobreatenuat (*overdamped Brownian oscillator*):

$$\tilde{C}_{OBO}''(\omega) = \pi \lambda \omega \frac{\gamma}{\omega^2 + \gamma^2} \quad (12.33)$$

on λ és l'energia de reorganització total i γ és la constant d'atenuació.

Paral·lelament, la tècnica NMA també ha sigut aplicada per Renger et al.⁹⁰ per obtenir la part de baixa freqüència de la densitat espectral, a partir d'una minimització i anàlisis NMA d'un complex fotosintètic sencer basada en camps de forces clàssics, mantenint els modes vibracionals intramoleculars congelats,

permeten estimar d'aquesta forma la modulació de les energies de transició al llarg de les coordenades de cada mode.

Un dels problemes fonamentals d'aquests mètodes és l'aproximació a un sol mínim en l'anàlisi de modes normals, ja que especialment les biomolècules es caracteritzen per tenir PES amb múltiples mínims.

1.2.3. Paràmetres CT

En aquesta secció es descriuen els mètodes usats per tal d'estimar els paràmetres CT en l'ADN: les energies dels estats radical catió i els acoblaments electrònics. L'energia de reorganització no s'ha estimat, sino que hem assumit en tots els casos un valor de 0.7 eV.⁷ En processos CT en biomolècules, es solen assumir valors de l entre 0.5 i 1.5 eV. El valor 0.7 eV representa un compromís entre un valor de 0.6 ± 0.1 eV estimat experimentalment per a processos HT en oligonucleòtids,⁹¹ i valors $\sim 0.8\text{--}1.5$ eV obtinguts en diferents estudis computacionals per HT entre llocs adjacents.

1.2.3.1. Energies d'estats radical catió: Aproximació de Koopmans

Per a un procés de transferència de càrrega entre dues molècules, l'energia o la diferència d'energia entre aquestes dues es pot estimar de manera raonable aplicant l'aproximació de Koopmans.⁹² Aquesta postula que en un sistema de capa tancada, aplicant la teoria de Hartree-Fock, el potencial d'ionització correspon al negatiu de l'energia de l'HOMO per al sistema neutre. En un sistema DA, per tant, podem estimar els potencials d'ionització corresponents a A^+D i AD^+ a partir de les energies dels orbitals ocupats de més alta energia localitzats sobre D i A. Així, la diferència d'energia entre aquests dos estats es pot estimar a partir de la diferència entre les energies de l'HOMO i l'HOMO-1 calculades pel sistema neutre AD.

$$\Delta E_{AD} = \varepsilon_{HOMO} - \varepsilon_{HOMO-1} \quad (12.34)$$

Aquesta estratègia és pot estendre a sistemes neutres amb múltiples molècules. En aquesta tesi s'ha seguit aquesta estratègia per a calcular les energies de múltiples llocs mitjançant càlculs QM/MM semiempírics INDO/S en sistemes neutres formats per múltiples parells de nucleobases. Estudis previs han mostrat que el mètode INDO/S és capaç d'estimar energies i acoblaments electrònics per a processos HT entre nucleobases apilades a partir de l'aproximació de Koopmans per comparació amb càlculs d'alt nivell CASPT2, especialment en el càlcul d'acoblaments electrònics en comparació amb altres mètodes semiempírics.⁹³

1.2.3.2. Acoblaments electrònics: El mètode FCD

L'acoblament electrònic és un paràmetre clau en el control de la velocitat de CT, i la seva definició pot variar depenent de si es consideren processos de migració d'electrons (ET) o de forats (HT). Igual que en un procés EET, si D i A tenen estats ressonants, l'acoblament electrònic correspon a la meitat de la diferència d'energia entre els estats desdoblats en el sistema DA, $V_{DA} = \frac{1}{2}E_{12}$. Malgrat això en sistemes reals asimètrics, aquest procediment resulta impracticable. Així, s'han desenvolupat diverses estratègies per a la obtenció directe de l'acoblament mitjançant tècniques de diabatització dels estats electrònics involucrats, tals com els mètodes *generalized Mulliken-Hush* (GMH)⁹⁴, *fragment charge difference* (FCD)⁹⁵ i *Boys localization*⁹⁶.

En aquesta tesi s'ha utilitzat el mètode FCD per a l'obtenció de l'acoblament electrònic en processos HT a través de la doble hèlix de l'ADN. Aquest mètode es basa en aplicar una transformació lineal d'estats adiabàtics $|\psi_i\rangle$ i $|\psi_f\rangle$ a estats diabàtics $|\psi_D\rangle$ i $|\psi_A\rangle$. Els estats diabàtics són aquells en que la càrrega està completament localitzada en una de les dues molècules, mentre que els estats adiabàtics corresponen als estats propis del sistema DA, on l'acoblament dona lloc a una certa deslocalització de la càrrega entre D i A.

Pels estats deslocalitzats $|\psi_i\rangle$ i $|\psi_f\rangle$, la càrrega $q_i(F)$ està localitzada en el fragment F = D, A (ja que F és una mescla de D i A). Si apliquem una transformació ortogonal als estats adiabàtics obtenim els estats diabàtics següents:

$$\begin{aligned} |\psi_D\rangle &= \cos \theta \psi_i + \sin \theta \psi_f, \\ |\psi_A\rangle &= -\sin \theta \psi_i + \cos \theta \psi_f \end{aligned} \quad (12.35)$$

de tal manera que la càrrega ja no està localitzada en el fragment format per la mescla entre D i A, sinó que està localitzada en D o en A:

$$\begin{aligned} q(D) &= (c_{Di} \cos \theta + c_{Df} \sin \theta)^2 = q_i(D) \cos^2 \theta + q_f(D) \sin^2 \theta + q_{if} \sin 2\theta, \\ q(A) &= (c_{Af} \cos \theta - c_{Ai} \sin \theta)^2 = q_f(A) \cos^2 \theta + q_i(A) \sin^2 \theta - q_{if} \sin 2\theta \end{aligned} \quad (12.36)$$

on $q_{if}(D) = c_{Di}c_{Df}$. Per tant, cal determinar l'angle de rotació θ en què els nous estats $|\psi_D\rangle$ i $|\psi_A\rangle$ representin una transferència de càrrega màxima del donador a l'acceptor, de manera que es maximitzi $q(D) + q(A)$:

$$\tan 2\theta = \frac{2\Delta q_{if}}{\Delta q_i - \Delta q_f} \quad (12.37)$$

Així, s'obté que l'acoblament electrònic entre els estats localitzats $|\psi_D\rangle$ i $|\psi_A\rangle$, $V_{DA} = \langle \psi_D | H | \psi_A \rangle = \frac{1}{2}(E_f - E_i) \sin 2\theta$, a partir de l'expressió següent:

$$V_{DA} = \frac{(E_f - E_i)|q_{if}|}{\sqrt{(\Delta q_f - \Delta q_i)^2 + 4q_{if}^2}} \quad (12.38)$$

1.3. Modelització de processos EET en complexes fotosintètics: ficobiliproteïnes

Un dels processos més importants d'EET en sistemes biològics que tenen lloc a la terra és la fotosíntesi. La fotosíntesi és un procés biològic en el que l'energia provinent del sol en forma de llum, és recol·lectada i emmagatzemada a través d'una sèrie de passos capaços de convertir aquesta energia en l'energia bioquímica necessària per sostener la vida a la terra. Per tant, directa o indirectament, la fotosíntesi és el motor de totes les cèl·lules活的. ^{97,98}

Es coneixen dos tipus de fotosíntesi, la oxigènica i l'anoxigènica. Les plantes, les algues i els cianobacteris porten a terme la fotosíntesi oxigènica, mentre que la resta de bacteris porten a terme la fotosíntesi anoxigènica. En la fotosíntesi oxigènica s'hi veuen implicades un seguit de reaccions coordinades que es poden dividir en tres fases diferenciades. La primera fase comença a la regió de la membrana del tilacoide (Figura 7) amb el procés d'absorció de llum a través de les anomenades proteïnes antena. Aquestes són les encarregades de transferir aquesta excitació provinent de la llum, amb la màxima conservació possible, fins al centre de reacció, on té lloc la segona fase. En aquesta, l'energia d'excitació es converteix en l'energia redox necessària per portar a terme el transport d'electrons i la producció d'ATP i NADPH, per finalment utilitzar aquests compostos en la tercera fase per a la reducció del CO₂ i fixació del carboni, entre d'altres reaccions metabòliques.

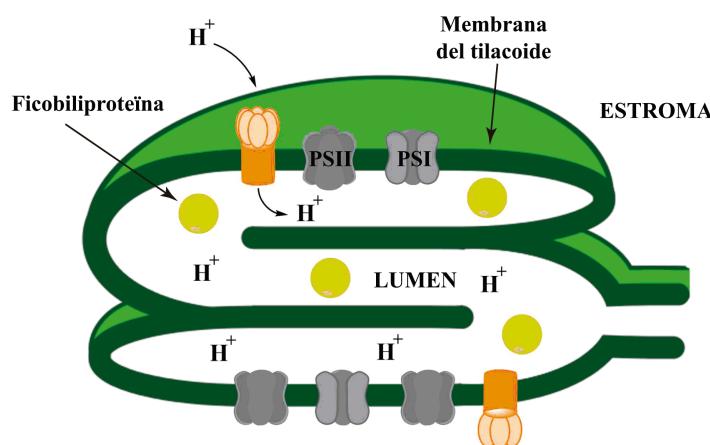


Figura 7. Estructura i parts del gránulo dins del cloroplast. En groc estan representades les ficobiliproteïnes que es troben suspesos en el lumen, a l'interior de la membrana del tilacoide.

Gran part d'aquesta tesi es centra en l'estudi i la modelització dels processos que tenen lloc en aquesta primera fase de la fotosíntesi, a través dels quals les proteïnes antena recapten la llum i en transfereixen l'energia al centre de reacció. És interessant remarcar que a diferència dels complexes presents en els centres de reacció, que presenten una arquitectura, composició i funció molt conservada entre tots els organismes fotosintètics⁹⁹, la gran diversitat estructural present en les proteïnes antena les ha fet especialment atractives a l'hora de buscar inspiració per a incrementar l'eficiència de les actuals cel·les solars. És el que s'anomena bioinspiració.

És fascinant com, al llarg dels últims 3 bilions d'anys, l'estructura d'aquestes proteïnes antena ha evolucionat en gran varietat d'organismes, no sols optimitzant-se, sinó també adaptant-se a nous habitats i situacions canviants, com ara a diferents condicions de disponibilitat de llum en diferents moments del dia, o en l'aprofitament de la llum incident disponible en organismes subaquàtics. És evident doncs, que el que fa interessant l'estudi d'aquestes proteïnes no és només la possibilitat de reproduir la seva eficiència en sistemes artificials, sinó també intentar entendre les característiques clau que les fan adaptables i funcionals sota diferents circumstàncies.

1.3.1. Estructura i funció de les ficobiliproteïnes

Com s'ha mencionat anteriorment, les proteïnes antena presenten una gran diversitat estructural entre organismes. En totes elles, però, el procés de recaptació de llum s'inicia gràcies a diferents pigments especialitzats, que a grans trets es poden dividir en tres grups: clorofil·les, carotenoides i ficobilines. Com es pot veure en la Figura 8, aquests pigments cobreixen pràcticament tot l'espectre de la llum visible, de manera que la distribució individual o combinada dels diferents pigments dins d'una antena fa possible una cobertura més o menys específica de la llum incident segons l'hàbitat on es trobi.

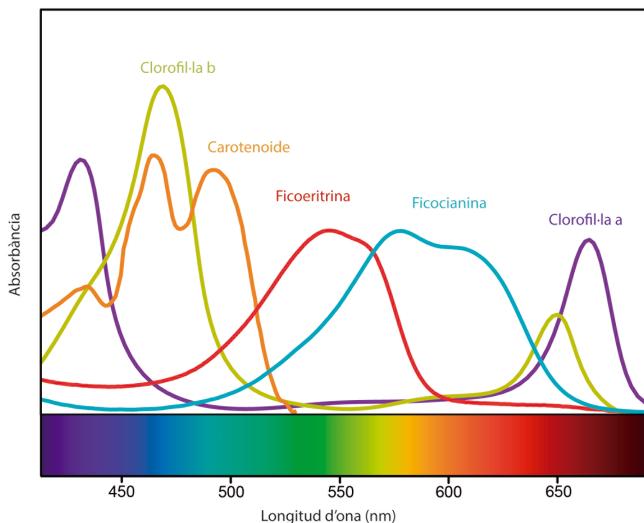


Figura 8. Espectres d'absorció de diferents pigments fotosintètics.

En el marc d'aquesta tesi s'han estudiat diferents tipus de proteïnes antena que tenen com a element en comú l'ús de ficobilines com a únic tipus de pigment: són les anomenades ficobiliproteïnes (PBP).

Les ficobiliproteïnes són les proteïnes antena presents en els cianobacteris i certs tipus d'algues (rhodophytes, cryptomonads, glaucocystophytes). Estan formades únicament per ficobilines, que presenten una estructura tertapirròlica lineal, i que a diferència de les clorofil·les, estan unides de manera covalent a l'estructura proteica a través de residus de cisteïna específics. És precisament gràcies a la presència d'aquests pigments que els organismes que posseeixen aquest tipus de proteïnes antena són especialment actius a la regió blava-verda de l'espectre, on les clorofil·les i carotenoides no presenten pràcticament activitat (Figura 8). És per això que els organismes que han adoptat aquest tipus d'antena són bàsicament organismes aquàtics, ja que la llum verda és la que és capaç de penetrar a grans profunditats d'aigua, de manera que la capacitat d'absorbir la llum a aquestes longituds d'ona confereix a aquests organismes la possibilitat de viure a més profunditat que els organismes que sols contenen clorofil·les en les seves antenes. Aquestes proteïnes antena però, són simplement les encarregades de recollir la llum i transferir-la, de manera eficient, al centre de reacció, constituït bàsicament per clorofil·les.

Una de les principals característiques que fa atractives les ficobiliproteïnes és la gran diversitat de ficobilines que es coneixen i que poden coexistir de manera

combinada en una mateixa antena. És precisament aquesta riquesa en la composició la que és capaç de modular l'absorció d'aquestes proteïnes antena, conferint-los la capacitat d'adaptar-se a diferents hàbitats amb diferents disponibilitats de llum.

Es coneixen nou tipus diferents de ficobilines (Figura 9) que poden variar en estat d'oxidació, nombre de dobles enllaços, grau de conjugació i nombre de punts d'unió covalents a la proteïna. D'aquestes nou, només dues, PCB i PEB, són presents en cianobacteris i algues rodofícies, mentre que tota la resta només s'ha trobat en criptòfits.

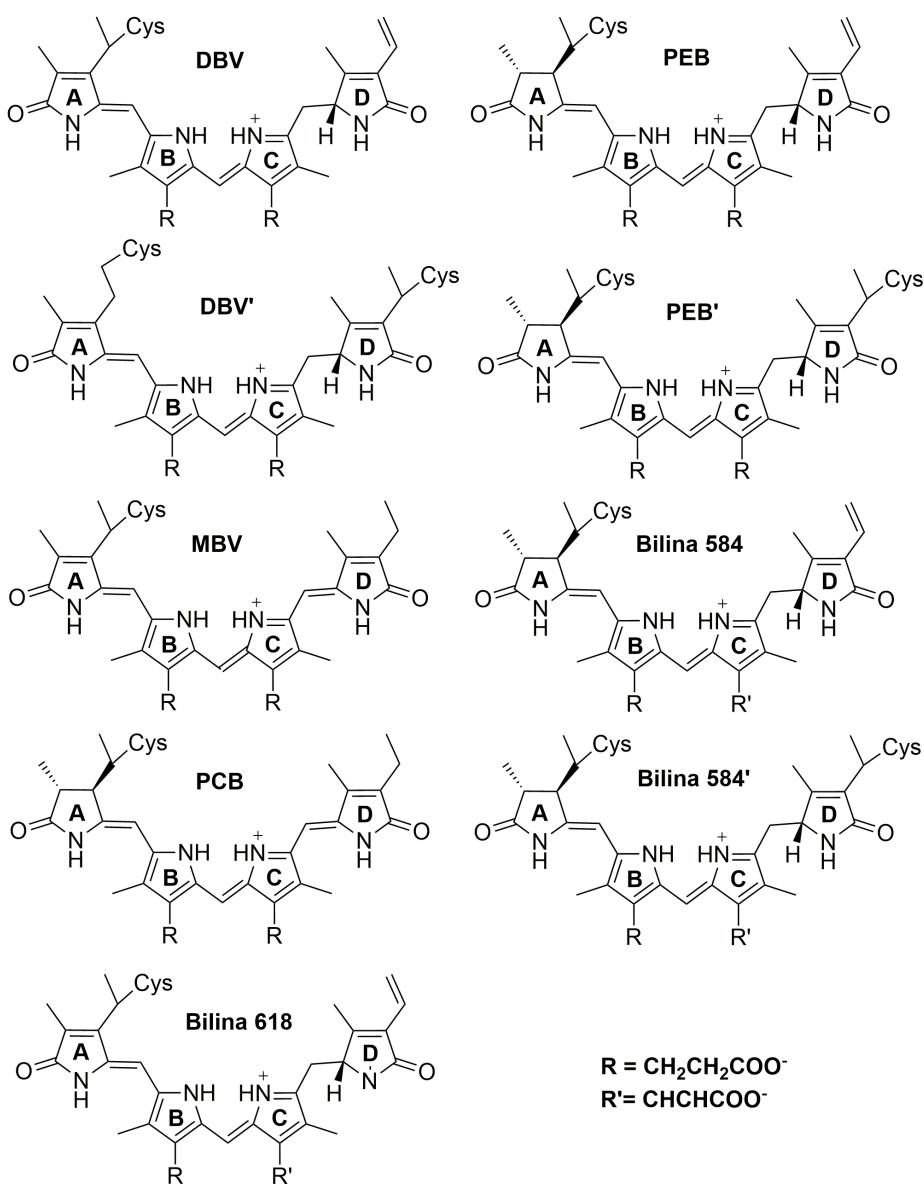


Figura 9. Estructura dels nou tipus diferents de ficobilines que es coneixen.

La seva estructura lineal conjugada, amb gran llibertat conformacional entre els quatre anells de pirrole, i la possibilitat de modular l'estat de protonació dels dos anells centrals, fa que la modelització d'aquest tipus de pigments sigui especialment complicada. Les petites diferències entre els diferents tipus de ficobilines i el gran paper que juga l'entorn local proteic fa que sigui un desafiament per als mètodes computacionals actuals obtenir de forma precisa les energies d'excitació relatives en complexes formats per múltiples ficobilines. En aquesta tesi s'ha treballat en aquesta direcció, desenvolupant i contrastant diferents metodologies per tal d'intentar obtenir un conjunt de valors fiable i amb sentit físic, capaç de reproduir algunes de les dades experimentals que s'han pogut adquirir i definir per tant models per als camins d'EET que tenen lloc en aquests sistemes.

1.3.1.1. *Criptòfits*

Les algues del fílum dominat per criptòfits són un grup d'algues unicel·lulars que es poden trobar tant en entorns d'aigua dolça com marina, generalment a grans profunditats. Se'n coneixen aproximadament unes 200 espècies¹⁰⁰, no obstant, cada espècie conté un sol tipus de ficobiliproteïna en la seva antena.

Els diferents tipus de PBPs presents en els criptòfits presenten una estructura similar basada en un heterodímer $\alpha\alpha'\beta_2$, no obstant, s'han observat casos en els que es presenta una simetria total $(\alpha\beta)_2$. Tot i que les subunitats β d'aquestes PBPs comparteixen un alt grau de identitat amb la seqüència de les subunitats α i β presents en les PBPs dels cianobacteris o les algues vermelles¹⁰¹, les subunitats α són més curtes, i no presenten similitud amb cap altra proteïna coneguda. Cada subunitat α conté solament un pigment, mentre que cada subunitat β en conté tres, un dels quals esta unit a la proteïna per ambdós extrems de l'estructura tetrapirròlica. És curiós observar que, tot i variar en la composició dels pigments presents en cada proteïna, les posicions en les que s'uneixen estan completament conservades: α -Cys 18 (o 19), β -Cys 50-61, β -Cys 82 i β -Cys 158.

Els dos pigments que es troben doblement units a l'estructura proteica són en tots els casos els dos pigments centrals. Fins fa uns anys s'havia observat en totes les

estructures que aquesta parella de pigments es trobaven a distàncies de contacte l'un de l'altre, fet que provocava un acoblament fort entre els dos pigments (acoblament excitònic). Aquesta disposició es creu que pot ser d'especial rellevància per explicar els fenòmens de coherència electrònica observats en els últims anys en aquestes PBPs.^{22,102} Tanmateix, però, recentment s'ha descrit una conformació quaternària nova per aquest tipus de PBPs.³² Aquesta nova conformació és fruit dels efectes estèrics generats degut a la inserció d'un

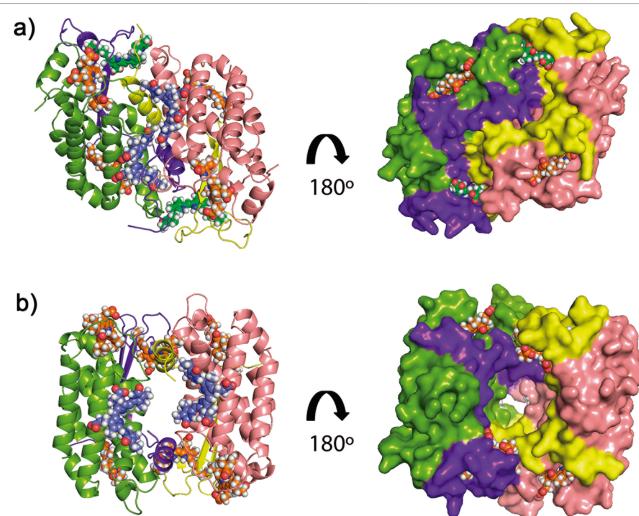


Figura 11. Estructures de les conformacions tancada (a) i oberta (b) de les fusicobiliproteïnes presents en els criptòfits.

aminoàcid en una regió conservada de la seqüència, just després del punt d'unió del pigment a la subunitat α . Aquesta inserció provoca una rotació d'uns 70° d'un dels dos monòmers generant una conformació "oberta", en la que els dos pigments centrals queden separats per un canal (Figura 11). El descobriment d'aquesta "nova" conformació, i en especial, el fet que els fenòmens de coherència mencionats anteriorment desapareixen en aquesta estructura "oberta", ha obert la porta al debat de si aquests fenòmens observats eren o no significatius per la transferència eficient de l'energia, o si es podria tractar d'una adaptació de certes espècies a noves circumstàncies lumíniques.

A diferència de les PBPs presents en els cianobacteris o en les algues vermelles, que s'organitzen formant el PBS, en aquestes algues es troben suspeses en el lumen, a l'altra banda de la membrana tilacoide. Aquesta característica, les fa diferents no només dels cianobacteris sinó de la resta d'organismes

fotosintètics¹⁰³, que tenen les seves proteïnes antena dins de la membrana tilacoide o adherides a la cara externa d'aquesta (estroma). Com s'ha mencionat anteriorment, quan la llum incident és excessiva, molts organismes han desenvolupat diferents mecanismes de fotoprotecció. Molts d'aquests mecanismes són regulats mitjançant un canvi en el gradient de pH a través de la membrana tilacoide, que pot fer variar el pH del lumen entre 5 i 7. Això fa que aquests PBPs es trobin immerses en un entorn canviant que pot tenir un efecte directe, modificant l'estat de protonació dels pigments o dels aminoàcids de la proteïna, o indirecte, ja que aquests canvis en l'estat de protonació poden fer variar les diferents energies d'excitació i modificar-ne el procés global de transferència d'energia.

D'altra banda, poc es coneix encara de l'agregació biològica d'aquestes proteïnes dins el lumen, i de com transfereixen l'energia recol·lectada al centre de reacció. Fins al moment, no s'han trobat evidències de si es troben en forma d'heterodímers com s'ha vist en les estructures cristall, "lligades" a la membrana o si s'organitzen formant algun tipus de complex, com en el cas del PBS.

1.4. Modelització de processos CT en ADN

L'ADN (àcid desoxiribonucleic) és àmpliament conegut, ja que és el responsable de la informació genètica de pràcticament tots els organismes vius. La seva estructura molecular característica en forma de doble hèlix, punt clau per a la seva modelització, va ser descoberta per F. Crick i J. D. Watson el 1953.¹⁰⁴ No va ser fins el 1962 quan es va descobrir que aquesta estructura particular presentava propietats electròniques,¹⁰⁵ que permetien a l'ADN actuar com a conductor orgànic d'electrons gràcies a les interaccions π - π entre les diferents bases apilades.

En els últims anys s'han fet grans avenços, tant des d'un punt de vista teòric com experimental, de cara a entendre els mecanismes CT a través de l'ADN. Avui en dia aquests processos són ben coneguts i se sap que, en funció de la seqüència, una càrrega pot migrar a través de la doble hèlix fins a 200 nanòmetres. Aquesta habilitat és d'especial interès en àrees tant diverses com la carcinogènesis o l'electrònica molecular.¹⁰⁶

Altrament, s'ha observat que aquests processos CT poden tenir lloc també entre l'ADN i certs aminoàcids, principalment el triptòfan i la tirosina, gràcies a la seva cadena lateral aromàtica.^{14,15} Per tant, l'estudi en profunditat dels paràmetres que dirigeixen la transferència de càrrega entre l'ADN i certs aminoàcids aromàtics d'una proteïna adjacent és fonamental, ja pot tenir implicacions importants en la reparació i protecció del material genètic.

Per a la modelització d'aquests processos de CT s'ha de tenir en compte que l'ADN es troba en un ambient aquós i, que tot i estar empaquetat, està subjecte a fluctuacions estructurals que influeixen en el procés CT. Per tant, és important tenir en compte els aspectes dinàmics, tant de l'ADN com de l'entorn.

1.4.1. Estructura de l'ADN

Com s'ha comentat anteriorment, l'ADN presenta una estructura característica en forma de doble hèlix . Aquesta estructura està formada per unes molècules que s'anomenen nucleòtids, constituïdes per un grup fosfat, que és el que fa de nexe entre els diferents nucleòtids, un sucre i una base nitrogenada.

En l'ADN es troben quatre tipus diferents de bases nitrogenades que es poden dividir en dos subgrups; de manera que en funció de si la seva estructura deriva de la purina o la pirimidina tindrem les bases púriques o pirimidínicas. Les bases púriques estan constituïdes per l'adenina (A) i la guanina (G), mentre que les bases pirimidínicas per la timina (T) i la citosina (C). Aquestes bases no s'aparellen de manera aleatòria per unir les dues cadenes, ja que perquè l'estructura sigui estable s'han de formar els ponts d'hidrogen necessaris que l'hi confereixin una màxima estabilitat i complementarietat. Per tant, les parelles de ball són sempre A-T i G-C, així que si imaginem l'estructura de la doble hèlix com una escala de cargol, les molècules de fosfat i els sucres serien els passamans mentre que les bases nitrogenades serien els graons.

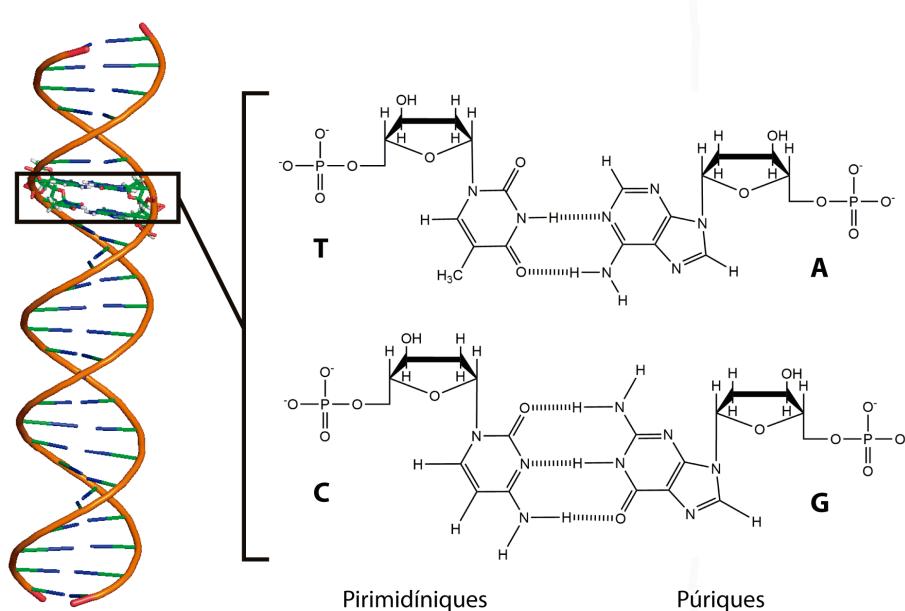


Figura 12. Estructura de la doble hèlix i estructura i interacció de les bases nitrogenades pirimidínicas (T, C) i púriques (A, G).

D'altra banda, l'ordre d'aquestes bases dins d'una mateixa cadena és el que determina el codi genètic, al mateix temps que l'estabilitat de la doble hèlix o l'eficiència en el procés CT.

1.4.2. CT a través de l'ADN

Quan es parla de transferència de càrrega a través de l'ADN, com s'ha mencionat anteriorment (secció 1.2.3.), es pot parlar tant de HT com d'ET. En cadenes oxidades, el que migra a través de la doble hèlix és un radical catió (un forat), mentre que en cadenes reduïdes és un radical anió (un electró). No obstant, com que l'estrés oxidatiu i la radiació UV són les principals causants de danys en l'ADN, la majoria d'estudis tan experimentals com teòrics s'han portat a terme per processos de HT, en cadenes oxidades.¹⁰⁷

Donat que els potencials d'oxidació de la guanina i l'adenina, i en especial el de la guanina, són menors que els de la citosina o la timina, el procés HT s'inicia mitjançant la formació d'un radical catió de la guanina ($G^{+ \cdot}$). Posteriorment, una guanina pròxima a la primera, l'hi transfereix un electró, de manera que el "forat" es transfereix de la primera a la segona guanina. D'aquesta manera, el procés HT se succeeix fins que el radical catió reacciona de manera irreversible amb una molècula d'aigua.¹⁰⁸ Per tant, el procés de transferència a través de l'ADN a llargues distàncies és possible sempre i quan la transferència de l'electró d'una guanina a l'altra sigui més ràpida que la velocitat de reacció amb l'aigua.

L'eficàcia del procés de transferència de càrrega depèn bàsicament de la seqüència de bases en la doble hèlix. Com s'ha mencionat anteriorment (secció 1.2.3.) un dels elements que dictaminen la velocitat de la transferència és l'acoblament electrònic entre les dues guanines, i aquest acoblament electrònic com ja s'ha discutit (secció 1.2.3.2.) depèn en gran part de la distància entre les dues molècules involucrades. Un fenomen important, és el que s'anomena *electron tunneling*, que es produeix entre guanines i està mediad per les parelles A-T. Tot i que aquestes dues bases tenen un potencial d'oxidació massa alt per ser oxidades, la interacció dels seus estats virtuals amb les guanines implicades, facilita significativament el procés de HT. S'ha observat però, que tot i que la velocitat de transferència es redueix al voltant d'un ordre de magnitud per cada parell A-T entre guanines, quan el "pont" està compost per quatre o més parelles A-T, la formació d'un catió radical de l'adenina entra en joc.¹⁰⁹

D'altra banda, l'impacte en el procés HT de defectes, lesions o conformacions inusuales en l'estructura, així com d'interaccions amb proteïnes, representa encara

un camp en estudi. En el marc d'aquesta tesi s'ha estudiat com es veu afectat el procés HT davant l'absència de bases nitrogenades en la seqüència, el que s'ha anomenat com a posicions abàsiques (*abasic sites*). Aquestes posicions abàsiques són lesions molt freqüents que es donen, en molts casos, de manera espontània degut al trencament de l'enllaç glicosídic i la pèrdua de la corresponent base nitrogenada. També s'ha estudiat l'efecte de la interacció amb certs aminoàcids de proteïnes que interaccionen amb la doble hèlix, com araenzims, i la possibilitat de que aquests aminoàcids intervinguin en el procés de transferència de càrrega.

1.4.3. CT en sistemes ADN-proteïna

Gran part de les funcions que desenvolupa l'ADN, com ara la transcripció, l'empaquetament, la replicació o la reparació de danys, depenen de la interacció amb diferents proteïnes oenzims. Aquestes interaccions poden ser específiques, on la proteïna s'uneix a una seqüència en particular de l'ADN, o no específiques.

S'ha observat que la unió de proteïnes a la doble hèlix pot afectar considerablement el processos CT a través de l'ADN. Per exemple, aquests processos són els precursores dels mecanismes de reparació que certes proteïnes específiques porten a terme per reparar el dany oxidatiu en l'ADN.

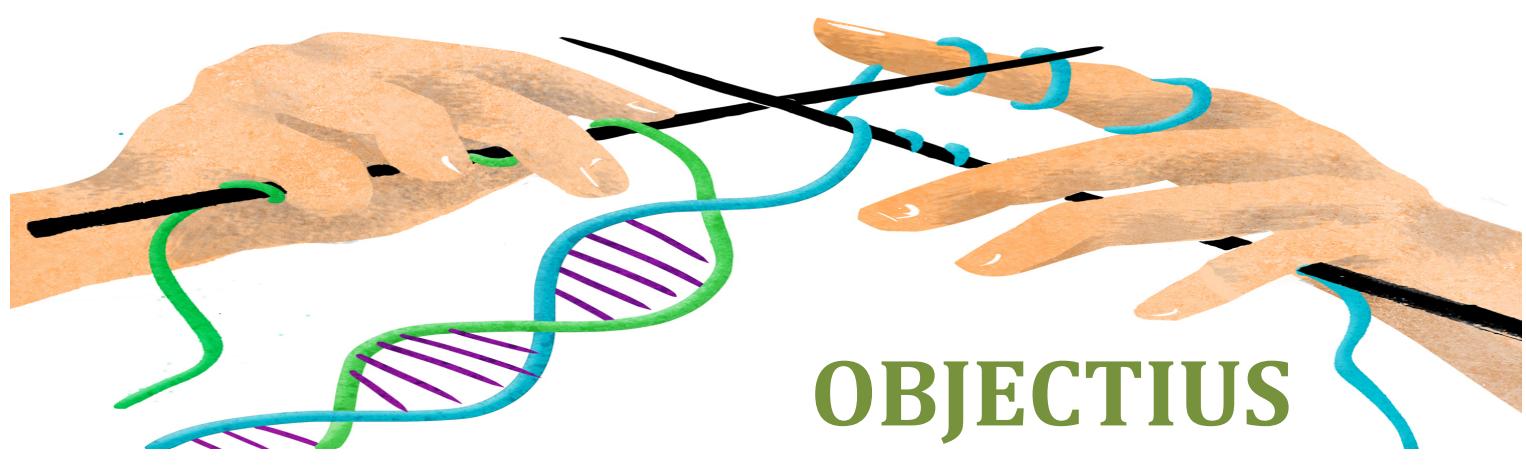
La gran majoria d'estudis que s'han portat a terme fins al moment han estat centrats en els processos de transferència a través de l'ADN aïllat. No obstant, també s'han estudiat aquests processos en complexos ADN-proteïna com ara la unió de l'enzim metiltrasferassa *Hhal*. S'ha observat experimentalment que la unió d'aquest enzim a la doble hèlix atenua dràsticament el procés de transferència de càrrega mediat per l'ADN, mentre que la unió d'un mutant de triptòfan en restitueix, i inclús n'accelera, el procés de transferència.^{14,15} També s'han observat processos de transferència de càrrega de histones a l'ADN¹¹⁰ igual que de peptids (Lys-Trp-Lys i Lys-Tyr-Lys) units a la doble hèlix.^{111,112}

És d'esperar que la distorsió de l'estructura estàndard de l'ADN degut a la interacció amb una proteïna, tingui cert impacte en l'acoblament entre les diferents bases nitrogenades i se'n vegi alterat el procés de transferència de càrrega. D'altra banda, la implicació de certs aminoàcids en el procés de

transferència de càrrega pot tenir conseqüències biològiques importants, com ara la protecció del material genètic contra possibles mutacions.

El triptòfan és el candidat ideal per intervenir en processos de transferència de càrrega amb l'ADN, ja que posseeix una cadena lateral aromàtica d'indole amb un potencial de ionització (7.90 eV) molt semblant al de la guanina (8.02 eV), i s'ha demostrat la seva implicació com a acceptor de "forats" en diferents processos de HT.^{113,114}

S'ha observat en diverses ocasions que les interaccions ADN-proteïna poden afectar dràsticament en el procés de HT. A més a més, certs aminoàcids, com el triptòfan, se sap que són capaços de reduir els cations radicals de la guanina reparant el dany oxidatiu en l'ADN. No obstant, encara hi ha molt camí a fer per entendre quins són els mecanismes que segueixen aquests processos de transferència de càrrega *in vivo*.



OBJECTIUS

2. Objectius

L'objectiu principal de la present tesi doctoral és aprofundir en diversos aspectes relacionats amb l'impacte de l'entorn en processos EET i CT en sistemes biològics. Al mateix temps, la tesi té com a objectiu avaluar i millorar les tècniques computacionals actuals en aquest àmbit. En particular, s'ha estudiat l'efecte de l'entorn en processos HT en l'ADN i en processos EET en proteïnes antena recol·lectores de llum.

2.1. Efectes de l'entorn en processos EET en ficobiliproteïnes

Les ficobiliproteïnes són un tipus de proteïnes fotosintètiques presents en alguns tipus d'algues i cianobacteris, on s'encarreguen de recol·lectar la llum i transferir-ne l'energia resultant al centre de reacció. Són proteïnes multicromofòriques, ja que contenen diferents pigments d'igual o diferent estructura. La seva naturalesa multicromofòrica i l'impacte que exerceix l'entorn en les propietats dels pigments que contenen fa que la descripció precisa dels elements necessaris per a la modelització del procés d'EET sigui encara un repte en l'actualitat. Per aquest motiu, en el marc d'aquesta línia d'investigació s'han utilitzat diferents metodologies per tal de descriure de manera acurada els elements clau per a la modelització de processos EET en aquests sistemes. Concretament s'han considerat els següents aspectes:

- Obtenir i comparar les funcions de densitat espectral global i individual per a cada pigment dins de la ficobiliproteïna PE545.
- Determinar l'origen molecular dels pics principals en les funcions de densitat espectral obtingudes i racionalitzar-ne les diferències en funció de la naturalesa de cada pigment.
- Comparar la dinàmica d'EET en la PE545 obtinguda usant funcions de densitat espectrals individuals per a cada pigment amb la obtinguda mitjançant una funció comuna.

- Determinar les propietats de protonació en aigua dels diferents tipus de pigments (bilines) presents en les cinc ficobiliproteïnes estudiades en aquesta tesi (PE545, PC577, PC612, PC630, PC645).
- Obtenir i avaluar els canvis en les propietats de protonació dels diferents pigments estudiats en ser transferits des d'un medi aquós fins a l'interior de cada proteïna.
- Determinar les energies d'excitació relatives i els acoblaments electrònics dels diferents pigments presents en les PBPs PC577, PC612, PC630 i PC645.
- Simular diferents espectres d'estat estacionari, com l'absorció, fluorescència i dicroisme circular, per a les PBPs PC577, PC612, PC630 i PC645, i validar l'ordre energètic dels pigments a cada complex per comparació amb els espectres experimentals.

2.2. Efectes de l'entorn en processos CT en l'ADN

L'oxidació de certes bases de l'ADN a través de la formació d'un estat radical catió és la precursora de diferents efectes mutagènics que poden tenir conseqüències nefastes per als organismes vius. Es coneix que aquests estats radical catió poden migrar a través de la doble hèlix, procés que depèn de la seqüència de bases. Tot i que els factors energètics i estructurals que governen aquests fenòmens CT a l'ADN han estat àmpliament estudiats, entendre com es veuen afectats aquests processos degut a la interacció amb proteïnes, o degut a la introducció de defectes estructurals tant comuns com la excisió d'una base de la seqüència, és encara un camp en desenvolupament. És per això que aquesta segona línia d'investigació es basa en la simulació del procés CT en seqüències d'ADN alterades estructuralment, ja sigui degut a la interacció amb una proteïna o a la excisió d'una base. Concretament, els objectius perseguits han sigut:

- Simular el procés CT en el sistema ADN-proteïna *HhaI* i el seu mutant Q237W.
- Validar la metodologia utilitzada mitjançant la comparació amb els resultats experimentals obtinguts en el grup de la Dra. Barton.

- Determinar el rol del triptòfan en el sistema mutant Q237W i la seva participació en el procés CT.
- Simular el procés CT en diferents sistemes model poli-G i poli-GC en els que s'ha escindit una base de la seqüència.
- Racionalitzar l'impacte estructural i l'impacte en el procés CT degut a l'excisió d'una base de la doble hèlix.

2. Objectives

The main goal of the present doctoral thesis is to deepen our understanding of different aspects related to the impact of the environment in EET and CT processes in biological systems. Simultaneously, the thesis aims to evaluate and improve the actual computational methods in this field. Particularly, we studied the effect of the environment in HT processes in DNA and EET processes in light-harvesting antenna proteins.

2.1. Effects of the environment in EET processes in phycobiliproteins

Phycobiliproteins are a kind of photosynthetic proteins found in some algae and cyanobacteria, where they harvest light and transfer the resultant excitation energy to the reaction center. They are multichromophoric proteins since they contain different pigments with equal or different structure. Their multichromophoric nature together with the impact of the environment on the properties of the pigments they contain make the precise description of the elements needed for the simulation of EET an important challenge. In the present research line we used different methodologies to get an accurate description of the key elements needed for the simulation of EET in these systems. We considered the following aspects:

- Obtain and compare the global and individual spectral densities for each pigment inside the phycobiliprotein PE545.
- Determine the molecular origin of the main peaks in the simulated spectral densities and rationalize their differences based on the nature of each pigment.
- Compare EET dynamics in PE545 described using individual spectral densities for each pigment or a common averaged one.
- Determine the protonation properties in water for the different kind of pigments (bilins) found in five phycobiliproteins studied along the present thesis (PE545, PC577, PC612, PC630, PC645).

- Obtain and evaluate the changes in the protonation properties of the pigments when transferred from water to their protein environment.
- Determine the relative excitation energies and the electronic couplings of the different pigments found in the PBPs PC577, PC612, PC630 and PC645.
- Simulate different steady-state spectra, such as absorption, fluorescence and circular dichroism, for the PBPs PC577, PC612, PC630 and PC645, and validate the energetic ladder of the pigments inside each complex by comparison with experimental spectra.

2.2. Environmental effects in CT processes in DNA

The oxidation of certain DNA bases due to the formation of a radical cation state is the precursor of different mutagenic effects with vast consequences for living organisms. It is well known that these radical cation states can migrate through the double helix depending on the DNA sequence. Although energetic and structural factors governing CT processes through DNA strands have been widely studied, how the interaction with proteins or the introduction of structural defects like the common excision of a base can affect CT properties is still poorly understood. This second research line is based on the simulation of CT processes through DNA structural altered sequences. The main objectives pursued here were:

- Simulate the CT process in the DNA-protein *HhaI* complex and in the mutant Q237W.
- Validate the methodology used in our simulations by comparison with experimental results from Dra. Barton's group.
- Determine the role of the tryptophan in the mutant system Q237W and its participation in the CT process.
- Simulate CT processes in different poli-G and poli-GC model systems, where a base has been excised from the DNA sequence.
- Rationalize the structural impact on the CT process due to the excision of a base from the double helix.



RESULTS

3. Resultats

Durant el transcurs d'aquesta tesis doctoral, el treball d'investigació desenvolupat a donat lloc a diverses publicacions científiques. A continuació se'n detallen les referències dividides en dos blocs.

3.1. Efectes de l'entorn en processos EET en ficobiliproteïnes

Lucas Viani, Marina Corbella, Carles Curutchet, Edward J. O'Reilly, Alexandra Olaya-Castro, and Benedetta Mennucci, *Molecular basis of exciton-phonon interactions in the PE545 light-harvesting complex*, Physical Chemistry Chemical Physics, **2014**, 16, 16302-16311.

Marina Corbella, Zi S. D. Toa, Gregory D. Scholes, F. Javier Luque, and Carles Curutchet, *Determination of the protonation preferences of bilin pigments in cryptophyte antenna complexes*, Physical Chemistry Chemical Physics, **2018** (submitted, minor revisions).

Marina Corbella, Lorenzo Cupellini, Benedetta Mennucci, Gregory D. Scholes and Carles Curutchet, *Spectral variability in phycocyanin cryptophyte antenna complexes is controlled by changes in the α polypeptide chains* (to be submitted).

Molecular basis of exciton-phonon interactions in the PE545 light-harvesting complex

Physical Chemistry Chemical Physics, **2014**, 16, 16302-16311

Lucas Viani,¹ Marina Corbella,² Carles Curutchet,² Edward J. O'Reilly,³ Alexandra Olaya-Castro,³ and Benedetta Mennucci¹

¹ Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Pisa, Italy.

² Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica i Institut de Química Teòrica i Computacional (IQTCUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona, Espanya.

³ Department of Physics and Astronomy, University College London, London, UK.

3.1.1. Molecular basis of exciton-phonon interactions in the PE545 light-harvesting complex

Resum

La fotosíntesi és un dels processos biològics més importants que tenen lloc a la Terra. Les anomenades proteïnes antena són les encarregades de captar la llum i transferir-ne l'energia resultant als centres de reacció dels diferents organismes fotosintètics. Aquestes proteïnes antena han despertat l'interès dins de la comunitat científica degut a la seva gran eficiència i per la possibilitat d'aplicar-ne les propietats a sistemes artificials de captura de llum. La recent observació d'efectes de coherència quàntica que poden perdurar fins a varis picosegons, inclús a temperatura ambient, en algunes proteïnes antena com la FMO o la PE545, ha generat un ampli debat, ja que en primera instància seria d'esperar que les interaccions amb un entorn aparentment desordenat, com són els múltiples modes vibracionals de la proteïna i les aigües de solvatació, dissipessin ràpidament qualsevol estat electrònic coherent format. Diversos grups han proposat que l'acoblament de les energies d'excitació als diferents modes vibracionals de l'entorn en rangs que concorden amb les diferències d'energia entre estats excitònics podria ser la causa de les coherències observades. Per tant, l'estudi en profunditat dels diferents modes vibracionals, igual que la seva interacció amb les energies d'excitació dels diferents pigments, també anomenades funcions de densitat espectral, és crucial per interpretar els efectes de coherència observats. Les funcions de densitat espectral, que quantifiquen la interacció fonó-excitó, poden ser obtingudes experimentalment, no obstant, en sistemes multicromofòrics només és possible obtenir-la per el cromòfor de més baixa energia o els cromòfors que contribueixen a l'estat de més baixa energia, perdent així la informació local de cada cromòfor. L'objectiu principal d'aquest treball és simular les funcions de densitat espectral per a cada cromòfor en particular dins del complex PE545, procedent de les algues del filum criptòfit, i realitzar un anàlisi dels graus de llibertat nuclears, tant dels cromòfors com del "bany" (proteïna i aigües de solvatació), per tal d'interpretar l'origen d'alguns dels pics més rellevants de les funcions de densitat espectral obtingudes. Per tal d'obtenir les diverses funcions de densitat espectral s'han realitzat càculs QM/MMPol de les energies d'excitació

de cada cromòfor al llarg d'una simulació de dinàmica molecular a intervals de temps de 5 femtosegons. Degut al gran nombre de càlculs que això implica s'ha utilitzat un nivell de teoria semiempíric ZINDO per als cromòfors, mentre que la proteïna i les aigües de solvatació s'han descrit mitjançant un camp de forces polaritzable. Per tal d'identificar l'origen molecular d'alguns dels pics característics que s'observen a les diferents funcions de densitat espectral, s'ha realitzat un ànalisi dels modes normal (NMA) dels cromòfors, que s'ha contrastat amb un ànalisi dels modes quasiharmònics (QHA) explorats al llarg d'una dinàmica molecular. Mentre que mitjançant un NMA s'obtenen els modes vibracionals dels cromòfors aïllats al voltant d'un mínim energètic, a través d'un QHA s'obtenen els modes al voltant una configuració mitjanada al llarg de la simulació de MD; per tant la comparació d'ambdós ànalisis posa de manifest les diferències entre modes intrínsecs de cada cromòfor i els que estan promoguts per l'entorn. En comparar les funcions de densitat espectrals obtingudes individualment, amb les funcions de densitat espectral global i experimental s'observa la pèrdua d'algunes característiques rellevants en la simulació del procés de transferència d'energia. D'altra banda, a través dels ànalisis dels diferents modes vibracionals, s'ha observat que, mentre que els modes al voltant de $1000\text{-}1500\text{ cm}^{-1}$ disminueixen degut a l'acoblament als modes vibracionals de la proteïna, el solvent es capaç de d'incrementar l'acoblament entre els modes vibracionals de la proteïna i els estats electrònics excitats de més baixa energia.

Molecular basis of the exciton–phonon interactions in the PE545 light-harvesting complex†

Cite this: *Phys. Chem. Chem. Phys.*,
2014, 16, 16302

Lucas Viani,^{*a} Marina Corbella,^b Carles Curutchet,^{*b} Edward J. O'Reilly,^c Alexandra Olaya-Castro^c and Benedetta Mennucci^{*a}

Long-lived quantum coherences observed in several photosynthetic pigment–protein complexes at low and at room temperatures have generated a heated debate over the impact that the coupling of electronic excitations to molecular vibrations of the relevant actors (pigments, proteins and solvents) has on the excitation energy transfer process. In this work, we use a combined MD and QM/MMPol strategy to investigate the exciton–phonon interactions in the PE545 light-harvesting complex by computing the spectral densities for each pigment and analyzing their consequences in the exciton dynamics. Insights into the origin of relevant peaks, as well as their differences among individual pigments, are provided by correlating them with normal modes obtained from a quasi-harmonic analysis of the motions sampled by the pigments in the complex. Our results indicate that both the protein and the solvent significantly modulate the intramolecular vibrations of the pigments thus playing an important role in promoting or suppressing certain exciton–phonon interactions. We also find that these low-frequency features are largely smoothed out when the spectral density is averaged over the complex, something difficult to avoid in experiments that underscores the need to combine theory and experiment to understand the origin of quantum coherence in photosynthetic light-harvesting.

Received 4th April 2014,
Accepted 18th June 2014

DOI: 10.1039/c4cp01477d

www.rsc.org/pccp

Introduction

Photosynthesis is one of the most important processes for sustaining life on earth. The understanding of its molecular mechanisms has always been the center of interest in scientific studies, especially due to its high photon-to-charge conversion efficiency or quantum yield under low light conditions, and the possibility of applying similar concepts to artificial light-harvesting and photo-conversion devices. Although over the past few decades such molecular mechanisms have been elucidated in detail,^{1–3} only recently two-dimensional electronic spectroscopy has reported long-lived exciton coherence in a variety of

pigment–protein photosynthetic complexes.^{4–6} These experimental results have generated a heated debate concerning the origin of the coherence and its impact on the excitonic energy transfer efficiency and pathways.^{7–9} Coherence effects have been observed even at room temperature in the phycoerythrin 545 (PE545) light-harvesting complex,⁶ as well as in the Fenna–Matthews–Olson (FMO) complex,¹⁰ which is an intriguing finding, since at first sight one would expect a quick dissipation of any coherent electronic state formed, due to the interactions with the apparent disordered environment, composed of the motions of the protein scaffold and solvent molecules. This simplistic view does not capture the full influence of the environment as it was originally pointed out by Lee, Cheng, and Fleming, who suggested that spatially correlated environmental fluctuations might support the coherent dynamics of electronic excitations.^{4,11,12} Studies combining classical molecular dynamics (MD) and quantum-chemical methods, however, indicate that spatial correlations do not translate into significant correlations among the site energies of the pigments in the Fenna–Matthews–Olson (FMO)^{12,13} and PE545¹¹ complexes. Recent theoretical and experimental studies indicate that the structured nature of the exciton–phonon interaction in light-harvesting antennae, and in particular, the coupling of electronic excitations to molecular vibrations, which do not relax quickly to thermal equilibrium, is fundamental for

^a Dipartimento di Chimica e Chimica Industriale, Università di Pisa,
via Risorgimento 35, 56126 Pisa, Italy. E-mail: lucas.viani@for.unipi.it,
benedetta.mennucci@unipi.it

^b Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona,
Av. Joan XXIII s/n, 08028 Barcelona, Spain. E-mail: carles.curutchet@ub.edu

^c Department of Physics and Astronomy, University College London, Gower Street,
London WC1E 6BT, UK

† Electronic supplementary information (ESI) available: Details of the methodology for the computation of exciton dynamics, tables reporting the individual rings of each bilin involved in the displacement of the QH modes, and figures comparing the distributions of modes obtained from the NMA analysis to those obtained at the DFT-B3LYP level. See DOI: 10.1039/c4cp01477d

effective energy distribution within these systems^{14–19} and may play a central role in the picosecond exciton coherence beating observed in the experiments.^{20–23} The coupling to vibrations in ranges that commensurate the energy difference between exciton states may even allow these systems to take on states that are truly quantum coherent with no classical analogue.²³

It is then clear that a full understanding of the origin and the role of coherence in the photosynthetic light-harvesting process can only be achieved *via* a combination of experimental and theoretical studies. From the theoretical point of view, the study of non-equilibrium energy transfer dynamics in a dissipative medium can be done using the density matrix formalism, which requires a detailed knowledge of the spectral density quantifying the exciton–phonon interaction.¹⁴ Spectral densities can, in principle, be obtained using experimental techniques, such as spectral hole-burning or fluorescence line-narrowing spectroscopy,^{24–27} or can be phenomenologically obtained *via* fitting spectroscopy results.²⁸ However, in multi-chromophoric systems often several low-energy chromophores contribute to the fluorescence of the complex, and in this case an averaged spectral density can only be estimated,^{14,15} thus missing the information related to the pigments' local environment. Alternatively, several groups have combined MD simulations with quantum chemical calculations^{8,9,13,29–34} or electrostatic calculations coupled to normal mode analysis³⁵ to theoretically derive spectral densities.

In this work we use hybrid quantum mechanics (QM)/polarizable molecular mechanics calculations combined to classical molecular dynamics (MD) to investigate exciton–phonon coupling in the PE545 antenna of the marine cryptophyte alga *Rhodomonas* sp. strain CS24.^{6,28,36–42} PE545 belongs to the family of the light-harvesting antenna proteins classified as phycoerythrins (PE), which are present in cryptophyte microalgae, ubiquitous members of freshwater and marine environments. These proteins are capable of absorbing the blue-green wavelengths of light that are available after broadband sunlight has passed through water. As in such environments the red wavelengths absorbed by chlorophyll are greatly attenuated, these proteins use tetrapyrrole pigments, called bilins, as the primary antennae.

Recently, PE545 has been studied both experimentally^{6,28,36,38,39,42} and computationally^{11,14,37,40,43} due to its capability to photosynthesize under low-light conditions, suggesting an efficient absorption of incident sunlight. This high efficiency is partially due to the flexible structural nature of the pigments which allows the modulation of their absorption energy through local pigment–protein interactions.⁴⁰ This feature maximizes the absorption of incident photons while creating an excitation funnel that drives the excitons formed inside of the complex to its exterior, facilitating its transfer to another complex or to the reaction center.⁴⁰ To understand the role of vibrations of the pigment–protein complex, eventually coupled to the solvent, we have computed the spectral density of the individual pigments and performed an analysis of the nuclear degrees of freedom of the pigments, the protein scaffold and the solvent to gain insights into the intra- and intermolecular origins of the features characterizing the frequency dependent coupling between the

electronic system and its environment. To get such a picture we have confronted quasi-harmonic (QH) and normal mode analysis (NMA). While NMA proves the vibrations around a single minimum, QH analysis aims at characterizing the global extent of the configurational space accessible to the system during an MD simulation. The modes obtained from the QH analysis can then be directly connected to the fluctuations of the energies experienced by the pigments along the MD trajectory.

Our results indicate that the protein and solvent significantly impact the distribution of motions sampled by the PE545 pigments, thus altering the structure of the individual spectral density experienced by each bilin, especially at low ($< 500 \text{ cm}^{-1}$) and mid-frequency ranges ($500\text{--}800 \text{ cm}^{-1}$). Because recent reports⁴⁰ indicate that several excitonic differences in PE545 will commensurate with the energies in this frequency range, accounting for such individual features of the spectral densities of pigments will be important to clarify the origin of quantum coherence. We also find that these low and mid-frequency features are largely smoothed out when the spectral density is averaged over the complex, something difficult to avoid in experiments that underscores the need to combine theory and experiment to understand the origin of quantum coherence in photosynthetic light-harvesting. We finally consider subunits within the antennae complex to illustrate some implications of pigment-dependent spectral densities and how averaging over the spectral densities may hide important features of the spatio-temporal distribution of energy within the complex. Our results open the door for future studies where the full implications of bilin-dependent spectral densities in ultrafast spectroscopy and full exciton dynamics can be thoroughly analysed.

Methodology

2.1 Spectral densities

The spectral density, $J_i(\omega)$, describes the frequency-dependent system–bath coupling strength of the pigment i . As it has been recently clarified, the spectral density can be expressed as^{33,44,45}

$$J_i(\omega) = \frac{\beta\omega}{\pi} \int_0^{\infty} C_i^{\text{cl}}(t) \cos(\omega t) dt \quad (1)$$

where $\beta = 1/(K_B T)$ and $C_i^{\text{cl}}(t)$ is the classical autocorrelation function of the fluctuation of the site energies:

$$C_i^{\text{cl}}(t_j) = \frac{1}{N-j} \sum_{k=1}^{N-j} \Delta E_i(t_j + t_k) \Delta E_i(t_k) \quad (2)$$

$C_i^{\text{cl}}(t)$ is here obtained using a methodology combining classical MD simulations and QM calculations of the site energies. In particular, a polarizable quantum mechanical/molecular mechanical (QM/MMPol) methodology we recently developed⁴⁶ has been used and applied to the configurations extracted from classical MD simulations. In this work we have used the same set of data generated in ref. 11, where we performed a 300 ps (after 10 ns equilibration) MD simulation of PE545 in water at 300 K (see ref. 40 for details), from which snapshots were extracted every 5 fs, thus totalizing 60 000 frames. In the next step,

QM/MMPol calculations of site energies were performed for each pigment along the classical trajectory, allowing the calculation of the autocorrelation function of their fluctuations. Due to the large number of calculations to be performed (480 000) and their computational cost, we adopted the ZINDO semiempirical method with Zerner's spectroscopic parameterization⁴⁷ to describe the excited states of the pigments, whereas the protein and solvent environments were described using a polarizable force field we developed from DFT calculations⁴⁰ (more details about the computational strategy are reported in ref. 11, 40 and 43). All PEBs and DBVs were modeled having unprotonated propionate chains as well as fully protonated central pyrrole rings, leading to a total -1 charge. Crystallographic data and the conservation of the Asp residue coordinating the central rings strongly suggest this protonation pattern for PEBs.³⁹ For DBVs, the crystal data are less conclusive, but they also indicate a neutral His residue coordinating the DBVs thus suggesting also this protonation pattern.

The autocorrelation functions were computed from the discrete time series following eqn (2), using fluctuations of the site energies, $\Delta E_k(t_k)$,²⁹ computed every 5 fs of the MD trajectory. In the PE545 complex, the fastest oscillations in the energies were found to have a period of around 20 fs similar to what has been found for the FMO complex, this time step thus provides a good compromise of reducing the amount of calculations without losing track of their oscillations at ambient temperature.¹³ Multiple autocorrelation functions of 2 ps length were computed using the 300 ps window by shifting its initial time by 1 ps. The final function was then obtained as the average of all functions computed for the same pigment. To minimize the effects of the noise on the spectral densities the final autocorrelations were multiplied by a Gaussian envelope of variance $\sigma^2 = 3.6 \times 10^{-5}$ fs² ensuring their convergence to zero.⁴⁵ The final autocorrelation functions are shown in Fig. 2. The high-frequency offset in the spectral densities was corrected with negligible effects in the low-frequency part by shifting the Fourier transform before its multiplication by the prefactor.⁴⁸ The spectral densities obtained from eqn (1) were finally fitted using an expression composed of two terms, one describing coupling to a continuous distribution of modes and a second one associated with discrete vibrational modes:

$$\begin{aligned} J(\omega) &= J_{\text{continuous}}(\omega) + J_{\text{peaks}}(\omega) \\ &= \frac{2\lambda}{\pi} \frac{\Omega\omega}{\omega^2 + \Omega^2} + \sum_q \frac{2\lambda_q}{\pi} \frac{\nu_q \omega_q^2 \omega}{(\omega_q^2 - \omega^2)^2 + \nu_q^2 \omega^2} \end{aligned} \quad (3)$$

$J_{\text{peaks}}(\omega)$ corresponds to the spectral densities of a set of damped quantum harmonic oscillators of frequencies ω_q with associated damping ν_q and reorganization energy $\lambda_q = s_q \omega_q$, whereas the continuous contribution has associated a reorganization energy λ and a bath cutoff frequency Ω .

2.2 Quasi-harmonic and normal mode analysis

The peaks in the spectral densities correspond to motions of the system coupled to the electronic transitions of the pigments. In order to identify the molecular origin of these motions,

we performed a quasi-harmonic (QH) analysis of the motions of each pigment in PE545 along the classical MD trajectory.⁴⁹ Such analysis requires diagonalization of the mass-weighted covariance matrix of atomic positions in Cartesian coordinates, and is typically used to determine configurational entropies of biomolecules. The corresponding eigenvectors describe motional modes (quasi-harmonic modes) around the average system configuration. To characterize the peaks on the spectral density of each individual pigment, a separate QH analysis was performed on the coordinates of the atoms of each pigment. Translational and rotational motions were removed from the analysis by superimposing each snapshot to the average pigment structure sampled along the 300 ps MD trajectory.

In order to assess the accuracy of the classical force field adopted in our simulation in describing the intrinsic vibrations of the pigments, we have also performed an NMA analysis of each pigment both at classical MM and QM levels. Classical MM-NMA analysis was performed using the same force field adopted in the MD simulation, whereas the QM-NMA analysis was performed at the B3LYP/6-31G(d) level of theory. In the latter case, we scaled the resulting vibrational frequencies by 0.952 in order to reproduce fundamental frequencies according to the Minnesota Database of Frequency Scale Factors for Electronic Model Chemistries.⁵⁰ In both MM and QM NMA analyses, the geometry of the pigments was minimized. In such a minimization, the cysteine links connecting the pigments with the protein were substituted by $-\text{SCH}_3$ groups. Finally, we note that in the QM-NMA analysis, the propionic acid groups in the bilin pigments were protonated to prevent folding and keep the biological conformation.

Results and discussion

The crystal structure of PE545, illustrated in Fig. 1, has been determined at 1.63 Å⁴¹ and later at 0.97 Å resolution.³⁹ It is organized as an $\alpha_1\alpha_2\beta\beta$ dimer, and contains eight bilins, characterized by a linear tetrapyrrole structure covalently linked to the protein scaffold. In particular, three types of bilins can be singled out: two 15,16-dihydrobiliverdin chromophores, labeled DBV_A and DBV_B, singly bound to the α subunits A and B, four phycoerythrobilin chromophores, labeled PEB_{158C}, PEB_{158D}, PEB_{82C}, and PEB_{82D}, singly bound to the β subunit C or D, and two other phycoerythrobilin chromophores, labeled PEB_{50/61C} and PEB_{50/61D}, doubly bound to the same subunits C and D, respectively. The overall PE545 structure (and the chromophores) displays a pseudosymmetry about the 2-fold axis relating the $\alpha_1\beta$ and $\alpha_2\beta$ monomers.

In a recent study,⁴⁰ we have shown that, in contrast to chlorophyll-based photosynthetic complexes, pigment composition and conformation play a major role in defining the energy ladder in the PE545 complex, rather than specific pigment–protein interactions. This is explained by the remarkable conformational flexibility of the eight bilin pigments in PE545, characterized by a quasi-linear arrangement of the four pyrrole units. Such conformational degrees of freedom are

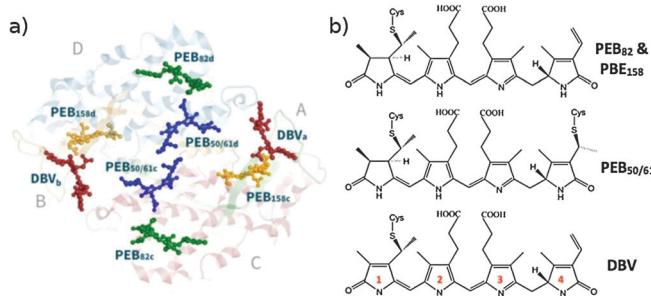


Fig. 1 (a) Structure of the $\alpha_1\alpha_2\beta\beta$ dimer and structure of the PE545 complex (α chains are in red and blue whereas the polypeptide β chains are in orange and cyan) and the relative position of the eight light-absorbing pigments; (b) structural view of the two types of bilins: dihydobiliverdins (DBV) and phycoerythrobilins (PEB) (the ring numbering used in the text is also shown). The 3D molecular representations were created using the MView software.⁵¹

therefore expected to significantly couple with motions in the protein–solvent environment and give rise to strong peaks in the spectral density of various pigments. In the following sections this investigation is reported and further supported by a detailed molecular explanation of the main features observed in the spectral densities.

3.1 Spectral densities

The autocorrelation functions computed from the site energies extracted from the MD trajectories are reported in Fig. 2. The convergence of the autocorrelations was tested against the number of frames, and no significant changes were observed for data sets larger than 25 000 frames.

As a general comment, we can see that the fluctuations of the autocorrelation functions for all the bilins are larger than those previously found for the bacteriochlorophylls in the FMO complex¹³ and the autocorrelation functions, particularly for the DBV bilins, decay much slower than those of the average FMO pigments. It is also worth noting that different features among pigments are evident. In particular, DBVs present much

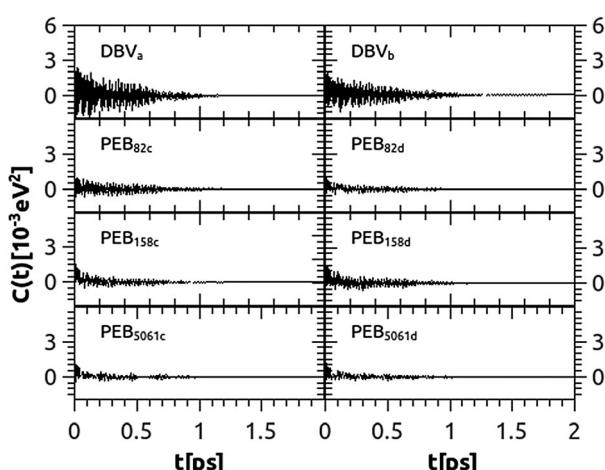


Fig. 2 Site energy autocorrelation functions for the eight bilins in the PE545 complex.

larger oscillations with respect to all the other pigments. This is a consequence of the additional double bond present in DBVs, which translates into a more delocalized $\pi-\pi^*$ electronic transition involving not only pyrrole rings 2 and 3 as in PEBs, but also ring 1 (see Fig. 1). This increased conjugation thus leads to an excited state more sensitive to structural deformations, as will be discussed later.

In Fig. 3 we report the resulting spectral densities averaged over all pigments (top) and over the pseudo-symmetric pairs (bottom). The spectral density averaged over all pigments is confronted against that reported in ref. 14, based on the linear spectra and spectroscopy studies,^{28,39} and we observe a reasonable agreement between the curves at low frequencies, representing the continuous component of the spectral density. Regarding the discrete peaks of the spectra, we also obtained reasonable peak intensities and amplitudes, although different frequency positions are predicted. High-frequency modes tend to be overpopulated in classical simulations, so the high-frequency part of the spectral density is expected to be less reliable compared to the low frequency region. The experimental spectral density, however, can also be affected by the way spectroscopic data were fitted, so the position of the peaks in this case must be taken with caution. Nevertheless, we computed the reorganization energies associated with each spectral density through the relation $\lambda = \int_0^\infty \frac{J(\omega)}{\omega} d\omega$ and we find an excellent agreement between our predicted estimate (557 cm^{-1}) and the value derived from the experimental spectral density (539 cm^{-1}).

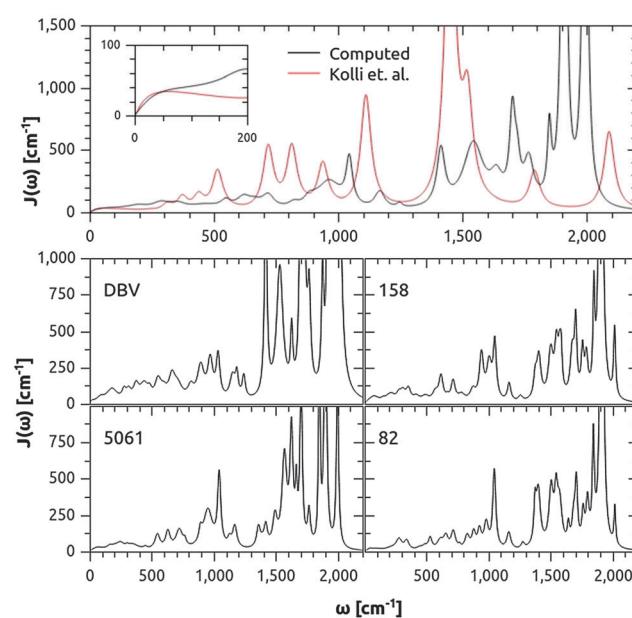


Fig. 3 Ambient temperature spectral densities averaged over all pigments confronted against the experimental one reported in Novoderezhkin²⁸ and used by Kolli et al.¹⁴ (top), and spectral densities averaged over pseudo-symmetric bilins (bottom) in the PE545 complex. The corresponding pigment number is shown in the top-left corner of the plot. The inset shows a close-up of the low frequency region.

As a general feature we see that the spectral densities are highly structured and the most recently reported electronic energies⁴⁰ indicate that several of the excitonic energy differences will commensurate with the energies at which the sharp peaks are observed with significant implications for the excitation dynamics as recently discussed.¹⁴ In addition to this general behavior, we also note that there are distinctive differences among the spectra of each of the bilins throughout the whole frequency regime. These differences are easily observed among non-symmetric pairs and indicate different local environments for the excitation of each bilin. Therefore the assumption of equal spectral densities for all pigments used in different reports to compute spectroscopic properties is an important approximation to the problem,^{28,40,52} and its impact on the excitonic dynamics is discussed in detail in Section 3.3. A major difference in the spectral densities of DBVs and PEBs comes from the effective coupling to the continuous distribution of harmonic modes in the low-frequency regime, which is reflected in the different values of the effective reorganization energies of this continuous component. The DBVs feature larger reorganization energies (λ is around 82 cm^{-1}) when compared to PEBs. Among the PEBs, the largest system-bath coupling appears to be for the PEB₁₅₈ pair followed by the PEB_{50/61} and PEB₈₂ pairs featuring λ values of ~ 48 , ~ 37 , and $\sim 20\text{ cm}^{-1}$ respectively. Such changes in λ are expected to manifest themselves in differences in the picosecond exciton dynamics as discussed below.

3.2 Origins of the peaks

To understand the differences in the peaks featured in the spectral densities of each pigment, it is important to determine the molecular nature of such vibrations. To this aim, we have determined the motions sampled by each pigment using both normal mode (NMA) and quasi-harmonic (QH) analyses based on the MD simulations. Before performing this analysis, the quality of the classical force field used to describe the bilin motions was checked by comparing the distributions of modes

obtained from the NMA analysis to those obtained at the DFT-B3LYP level.

Such a comparison is reported in Fig. S3 (ESI[†]). Overall, the distributions obtained with QM and classical methods are in reasonable agreement, although a slight shift at the high-frequency region ($1500\text{--}1700\text{ cm}^{-1}$) is observed which can be related to the fact that in the QM analysis, the propionic acid groups in the bilin pigments were protonated to prevent folding and keep the biological conformation. In addition, the MM distribution has a larger number of modes at very low frequencies ($< 50\text{ cm}^{-1}$). Although this analysis only provides a partial verification of the MM force field, as individual modes can be located at slightly shifted frequencies as compared to the QM description, overall, it indicates that the distribution of modes is reasonable. The validation of the classical description of the bilin motions against QM results allows us to apply the QH analysis along the trajectory and use the distribution and nature of the quasi-harmonic modes obtained to investigate the features of the spectral density at a molecular level. In Fig. 4, we show the distribution of QH modes determined for the individual bilins along with the corresponding spectral densities.

A high density of modes is observed in the $850\text{--}1200\text{ cm}^{-1}$ and $1400\text{--}1700\text{ cm}^{-1}$ regions, matching the largest peaks in the spectral densities. At frequencies larger than 800 cm^{-1} , all pigments show similar features, pointing to the intramolecular character of the motions, although the coupling strength of the peaks differs among pigments. Interestingly, however, pseudo-symmetric pairs of bilins show very similar coupling strengths, despite the different sequence of the α polypeptide chains in the $\alpha_1\alpha_2\beta\beta$ dimer structure of the PE545 complex, which slightly breaks the symmetry of the system (see Fig. 1). Because the vibrations of the pigments are coupled to different environments, the distribution of QH modes can change thus modulating the relative intensities of the peaks. Also in the $500\text{--}800\text{ cm}^{-1}$ range, all pigments show somewhat similar peaks. However, in this case the differences in the peak distribution and intensities are

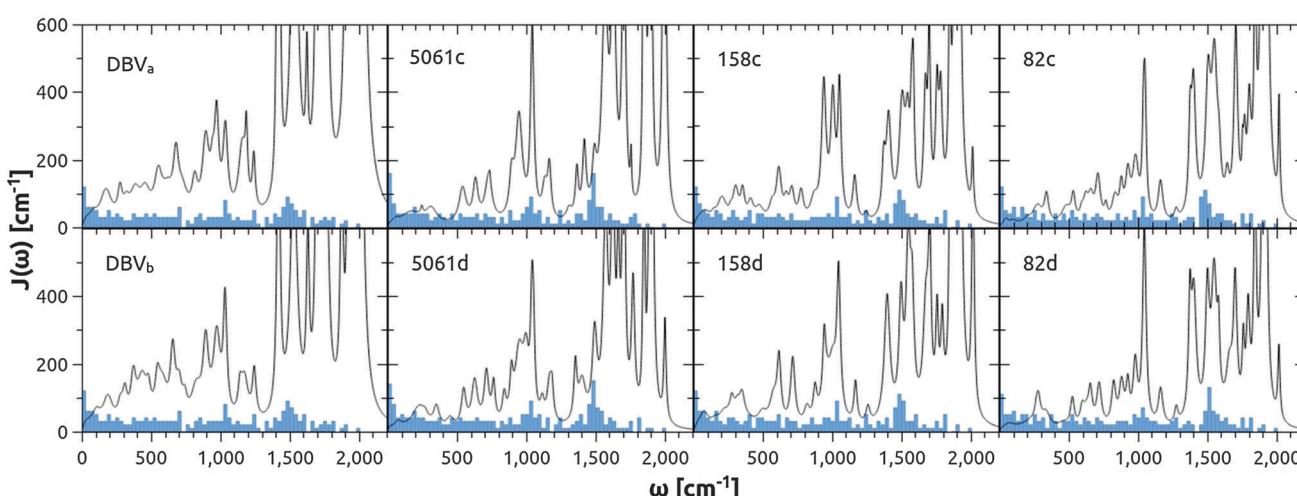


Fig. 4 Ambient temperature spectral densities of the eight bilins in the PE545 complex (black lines), and the respective distribution of MM normal modes (blue bars). The height of the bars was homogeneously scaled for visualization purposes.

enhanced compared to the higher-frequency region. In the low-frequency range ($< 500 \text{ cm}^{-1}$) instead all pigments exhibit significant differences both in position and coupling strength of the peaks, even among pseudo-symmetric pairs. This suggests that low-frequency motions of the pigments are strongly modulated by their local environment. This can be clearly appreciated in the spectral density averaged among all pigments shown in Fig. 3, as the different peaks apparent in the individual pigments are smoothed out. Among the different pigments, clearly the DBVs show the largest couplings to these low-frequency modes. Inspection of the atomic displacements involved in such modes indicates that many of them include bendings of the global linear tetrapyrrole arrangement of the bilins or torsional motions modifying the degree of planarity among the pyrrole rings; such displacements presumably modify the degree of π -conjugation among the rings and therefore are expected to couple significantly to the $\pi-\pi^*$ transition. Because the DBVs have a further double bond in the pyrrole ring 1 bonded to the Cys residue in the protein (see Fig. 1), the electronic excitation in this pigment extends over rings 1, 2 and 3, whereas in PEBs the excitation mainly involves rings 2 and 3. Such low-frequency motions modifying the degree of π -conjugation of ring 1 will couple much more strongly to the DBV excitations, thus explaining the previously observed high reorganization energy. It is also interesting to note that the PEBs of the central pair, PEB_{50/61C} and PEB_{50/61D}, show the smallest coupling to low-frequency motions. In contrast to the other PEBs, these bilins are linked to two Cys residues of the protein through rings 1 and 4, and such links could reduce the ability of the whole linear tetrapyrrole arrangement to bend or to stretch, compared to the other PEBs, thus reducing the coupling of low-frequency motions to the electronic excitation. These observations give an explanation for the previously observed low reorganization energies for these two sets of bilins.

It is important to note that the distribution of QH modes allows us to unveil the molecular basis of the peaks in the spectral density by associating them with modulated bilin intramolecular motions in the protein environment. The actual impact in the spectral density peaks can, however, either arise from fluctuating electrostatic pigment–protein interactions associated with a given motion or from oscillations in the bilin transition energies induced by changes in their internal geometries. It is remarkable that all spectral density peaks correspond to some distribution of normal modes, even the peaks at $\sim 2000 \text{ cm}^{-1}$. On the other hand, not all QH modes are expected to contribute to the spectral density, only those showing a strong exciton–phonon coupling.

The inspection of the atomic displacements involved in the modes thus allows explaining the molecular origin of the peaks. We have done such inspection for the modes in the $500\text{--}600 \text{ cm}^{-1}$, $950\text{--}1050 \text{ cm}^{-1}$ and $1400\text{--}1500 \text{ cm}^{-1}$ regions, which include the most remarkable peaks found in the overall averaged spectral density shown in Fig. 3. In addition, we have analyzed the modes in the $650\text{--}750 \text{ cm}^{-1}$ range, in order to shed light on the $\sim 710 \text{ cm}^{-1}$ peak which is better resolved in PEBs compared to DBVs. For the sake of simplicity, we only focus on the modes that are expected to couple more strongly to the $\pi-\pi^*$ transition

that extends over pyrrole rings 2 and 3 in the PEBs, and 1, 2 and 3 in the DBVs. Such modes involve torsional motions $\delta(\text{ring})$ distorting the planarity of the pyrrole rings, out-of-plane and in-plane bendings of the rings, $\gamma(\text{ring})$ and $\beta(\text{ring})$, respectively, and $\nu(\text{C-C})$ and $\nu(\text{C-N})$ stretchings leading to ring breathings. More in detail, in the $500\text{--}600 \text{ cm}^{-1}$ range, the interesting modes mostly involve $\delta(\text{ring})$, $\gamma(\text{ring})$ and $\beta(\text{ring})$ displacements whereas in the $950\text{--}1050 \text{ cm}^{-1}$ region, the vibrations are characterized by a mixing of $\beta(\text{ring})$, $\nu(\text{C-C})$ and $\nu(\text{C-N})$ distortions. Finally, modes in the $1400\text{--}1500 \text{ cm}^{-1}$ range mostly involve $\nu(\text{C-C})$ and $\nu(\text{C-N})$ stretchings. Because the DBVs have a further C=C double bond in ring 1, there is a larger number of modes strongly coupled to the DBVs in this region compared to PEBs, which explains why the $\sim 1450 \text{ cm}^{-1}$ peak is considerably more intense in the DBVs compared to the PEBs. In Table S1–S8 (ESI†), we report the individual rings of each bilin involved in the $\delta(\text{ring})$, $\gamma(\text{ring})$, $\beta(\text{ring})$, $\nu(\text{C-C})$ and $\nu(\text{C-N})$ displacements for each mode in these frequency ranges, thus allowing identification of those expected to be strongly coupled to the excitation.

In Fig. 5 we show a graphical representation of the displacements involved in modes found at about 550 cm^{-1} , 1050 cm^{-1} and 1450 cm^{-1} , which can be connected with the strongest peaks in the spectral density averaged over the pigments in PE545. In the same figure, we also provide a comparison of two modes observed at around 700 cm^{-1} for a PEB and a DBV, respectively.

Close inspection of the two modes at around 700 cm^{-1} indicates that, whereas for PEBs distortions of the central pyrrole rings 2 and 3 are evident, for DBVs the mode does not significantly distort rings 1, 2 or 3. This explains why for PEBs such modes show a stronger coupling compared to DBVs.

As previously discussed, the local protein environment translates into slightly different features of the spectral densities found for each pigment, especially at low frequencies. In order to shed light on the modulation of the vibrational modes induced by the protein environment, it is interesting to compare the distribution of vibrational modes obtained either from the

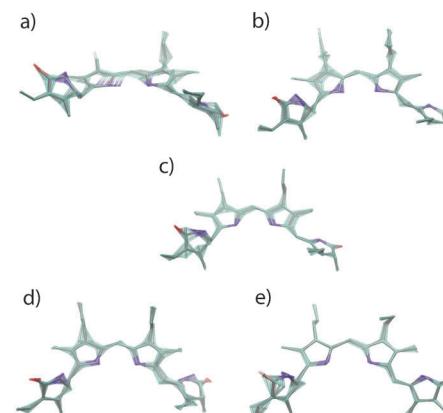


Fig. 5 Graphical representation of selected quasi-harmonic modes. (a) Mode at 555 cm^{-1} of PEB_{158D}, (b) mode at 1053 cm^{-1} of PEB_{158D}, (c) mode at 1428 cm^{-1} of PEB_{158D}, (d) mode at 711 cm^{-1} of PEB_{158D}, and (e) mode at 697 cm^{-1} of DBV₁₉₄.

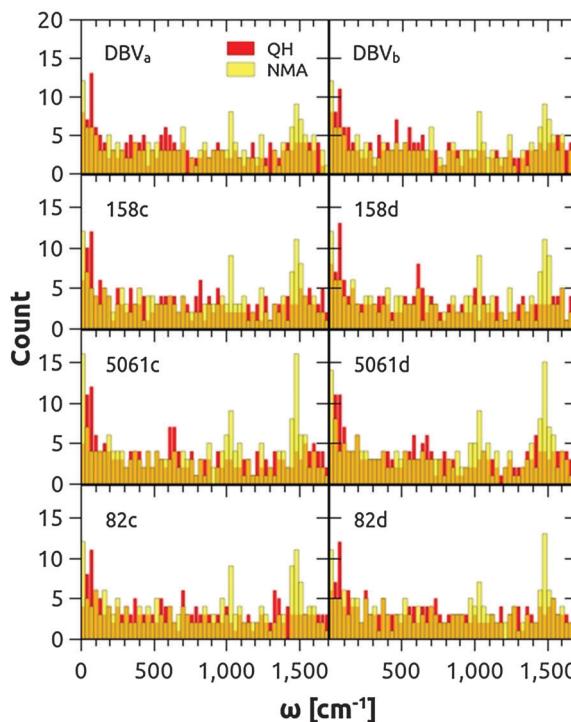


Fig. 6 Comparison between quasi-harmonic and normal mode analysis distributions of vibrational frequencies.

QH or NMA analysis based on the same potential energy function. This comparison is shown in Fig. 6.

The results indicate two interesting effects. First, the large accumulation of modes at around 1000 and 1500 cm^{-1} found in the NMA distribution is smoothed out when such modes couple to the protein motions, as shown in the QH distribution. On the other hand, there is a slight increase in the number of modes observed in the QH histograms at around 500 cm^{-1} compared to their NMA counterparts. These two effects are similar for all pigments, including the PEB_{50/61C} and PEB_{50/61D} bilins linked twice to the protein scaffold, thus suggesting that this effect is not induced by such covalent links but rather by the overall coupling between protein and pigment motions.

To investigate the impact of the solvent on the spectral density, the same quasi-harmonic analysis was repeated for a MD trajectory of the PE545 complex *in vacuo* (see ref. 11 for details). A direct comparison between the distributions of normal modes provides an insight into modes that are promoted or suppressed by the presence of the solvent. In Fig. 7 the difference between the distribution of normal mode frequencies in water and the gas phase is shown; positive/negative blue/red bars indicate modes that are promoted/suppressed by the solvent, while the bars around zero indicate modes that are not affected by the presence of the solvent. Interestingly, the solvent promotes a large accumulation of modes at particular high frequencies, where the spectral densities feature the larger couplings, whereas in a vacuum the modes spread over a larger range of values, indicating the important role of the solvent in modulating the exciton–phonon couplings.

The analysis of the plots suggests that the majority of the peaks in the mid-frequency region (900 – 1200 cm^{-1}) and

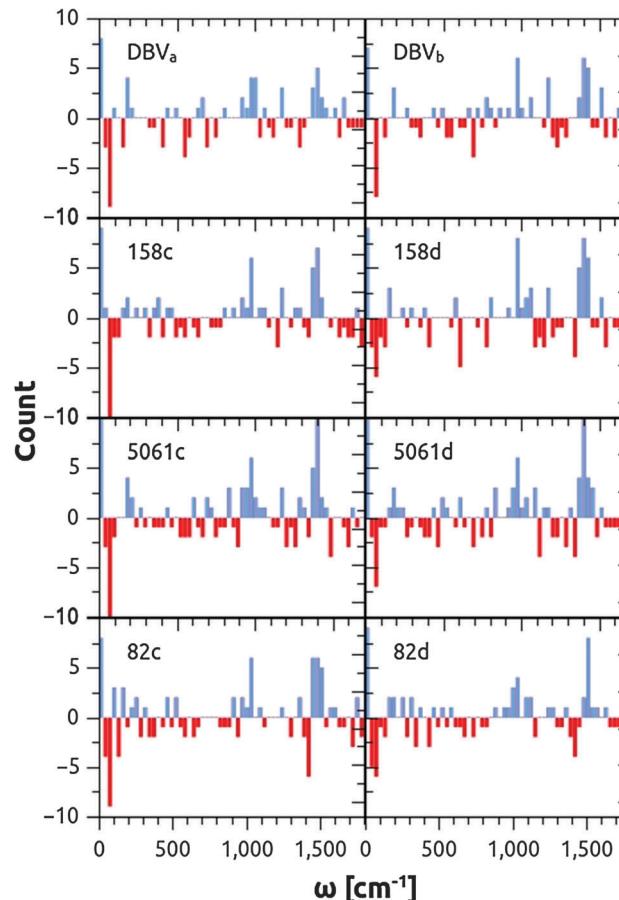


Fig. 7 Variation in molecular normal mode distribution in water against the *in vacuo* simulation. Blue/red bars indicate the modes that were promoted/suppressed by the solvent, and its height represents the intensity of the effect.

high-frequency region (1400 – 1600 cm^{-1}) present in the spectral densities are strongly promoted by the solvent, *i.e.* the solvent contributes to accumulate modes at these frequencies. Thus, the protein smoothening of the large accumulation of modes at around 1000 and 1500 cm^{-1} suggested by the comparison of NMA and QH mode distributions in Fig. 6 seems to be partially counterbalanced by the effect of the solvent. These results thus indicate that the solvent significantly impacts the structure of the spectral density. It is interesting to note that the PEB_{82C} and PEB_{82D} pigments are significantly less exposed to the solvent than the other chromophores. Also, the central PEB_{50/61D} bilin is more exposed to water compared to PEB_{50/61C} (see also Fig. 4 in ref. 40). Despite these differences, only PEB_{82D} shows a somewhat smaller impact of the solvent on the distribution of modes compared to the other pigments. This finding suggests that the solvent mostly impacts the bilin vibrations indirectly by modulating the overall protein motions.

3.3 Implications for excitation dynamics

The above analyses indicate that the main differences in the spectral densities appear between non-symmetric pairs: the PEBs have reduced reorganization energies in comparison to the DBVs while the latter have spectral densities that are closer

to the average fitted *via* ensemble measurements. In order to illustrate the implications of these bilin-dependent spectral densities we consider two subsets of chromophores in PE545 and investigate the exciton population dynamics within each subset comparing the case when all pigments have identical spectral densities with the case when the spectral densities are different as predicted by our QM/MMPol calculations. On the basis of the site energies reported in Table S4 (ESI†) of ref. 40, we defined one high-energy subset which includes the PEB_{50/61C} and PEB_{50/61D} and a low-energy subset which includes PEB_{82C} and PEB_{82D}; in both subsets DBV_a is included as a potential trapping site. Using the electronic couplings and site energies reported in ref. 40, we performed non-perturbative calculations of the exciton dynamics *via* numerical solution of a set of hierarchically coupled equation of motions^{53,54} considering spectral densities featuring a coupling to a continuous distribution and to a single relevant broaden vibrational mode as presented in eqn (3). The maximum tier considered in our calculations is 9 and more details of the methodology for computation of exciton dynamics are presented in Section S3 of the ESL.† The results shown in Fig. 8 and 9 indicate that averaging spectral densities among non-symmetric pairs results in qualitatively different features of the spatio-temporal distribution of excitation for the high- and low-energy subsets.

Our mode analysis and the corresponding fittings for the spectral densities predict that the three bilins in the high-energy subunit (PEB_{50/61C}, PEB_{50/61D} and DBV_a) exhibit a significant coupling to a mode around $\omega_q = 1050 \text{ cm}^{-1}$ with coupling strength about $\lambda_q = 18 \text{ cm}^{-1}$ and broadening $\nu_q = 31 \text{ cm}^{-1}$. In this high-energy subset, this vibrational mode is quasi-resonant with the energy gap between the intermediate-exciton state (quasi-localised in PEB_{50/61D}) and the lowest exciton (quasilocalised in the DBV_a) and then it dominates transfer to this exit site. As already discussed, the PEBs feature a reorganisation energy of $\lambda = 36 \text{ cm}^{-1}$ with cut off frequency $\Omega = 32 \text{ cm}^{-1}$, which is smaller than the parameters for the continuous spectral density for DBV_a which are $\lambda = 80 \text{ cm}^{-1}$ and $\Omega = 42 \text{ cm}^{-1}$. Considering that initially only the

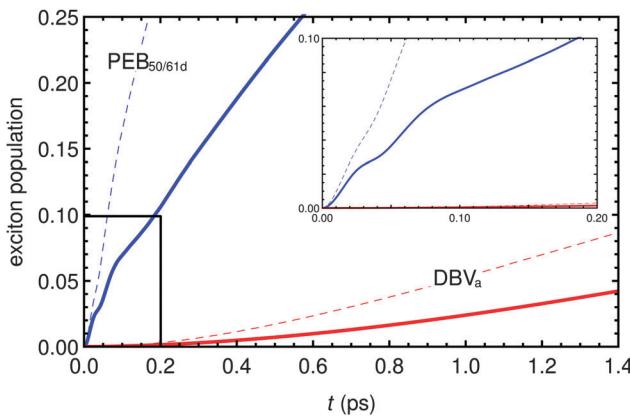


Fig. 8 Population dynamics for the intermediate (blue lines) and lowest (red lines) exciton states of the high-energy subunit that includes the PEB_{50/61C}, PEB_{50/61D} and DBV_a bilins. Solid lines denote the dynamics with different spectral densities for the three pigments while dashed lines denote the situation with identical spectral densities.

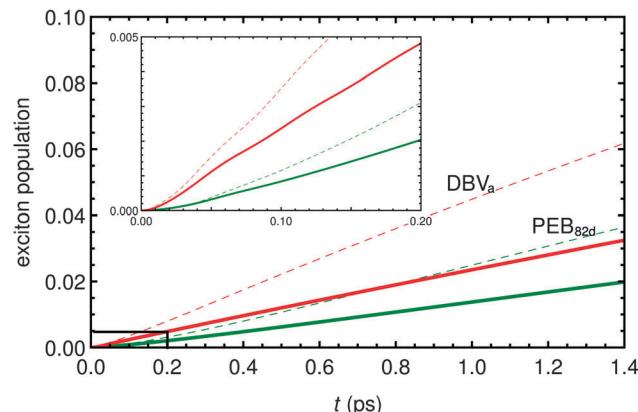


Fig. 9 Population dynamics for the intermediate (green lines) and lowest (red lines) exciton states of the low-energy subunit that includes the PEB_{82C}, PEB_{82D} and DBV_a bilins. Solid lines denote the dynamics with different spectral densities for the three pigments while dashed lines denote the situation with identical spectral densities.

highest exciton state is excited, we observe that non-equilibrium vibrational dynamics is manifested in population oscillations of the intermediate exciton state (quasilocalised in PEB_{50/61D}) that lasts up to a few hundred of femtoseconds, while the smaller reorganization energy leads to an overall slower transfer to the intermediate and lower exciton states as illustrated in Fig. 8 (top). The inset of Fig. 8 depicts the femtosecond coherent (oscillatory) population of the intermediate state, which arises from coupling to the underdamped mode that is comparable with the energy gap between PEB_{50/61D} and DBV_a (see parameters listed in eqn (S3) of the ESI†). Such coherent energy transfer is washed away when all the sites are considered to have the same continuous component of the spectral density as the DBV_a.

For the low-energy subset, our analysis indicates that PEB_{82C} and PEB_{82D} are coupled to a vibrational mode of frequency around $\omega_q = 530 \text{ cm}^{-1}$ with a strength of $\lambda_q = 8 \text{ cm}^{-1}$ and a broadening $\nu_q = 33 \text{ cm}^{-1}$ while the DBV_a is coupled to a mode of energy $\omega_q = 553 \text{ cm}^{-1}$ with $\lambda_q = 22 \text{ cm}^{-1}$ and $\nu_q = 65 \text{ cm}^{-1}$. A vibrational mode of this energy is quasi-resonant with the energy gap between the highest and intermediate energy exciton states (quasi-localised in PEB_{82C} and PEB_{82D}, respectively), and therefore is expected to influence exciton transport in this subunit. Moreover, the PEB₈₂ bilins exhibit weak coupling to a continuous component with $\lambda = 20 \text{ cm}^{-1}$ and $\Omega = 31 \text{ cm}^{-1}$, while the DBV_a is coupled to continuous spectral density with which $\lambda = 80 \text{ cm}^{-1}$ and $\Omega = 42 \text{ cm}^{-1}$ as before. Starting with an excitation in the highest exciton state, Fig. 9 depicts the populations of the intermediate and lowest exciton states of this subunit when spectral densities are different (solid lines) and identical to the DBV_a (dashed lines). The main effect of the reduced reorganisation energies for the PEBs is a decrease of the rate of transfer to both exciton states. Since the coupling to the localised vibrational mode is weak, its non-equilibrium dynamics only induces some non-exponential growth of population of the intermediate exciton state in the 200 fs time scale as clearly seen in the inset. Both the oscillatory and non-exponential growth of exciton population for intermediate states shown in Fig. 8(inset) and 9(inset) indicate

that energy dissipation into the thermal background is transiently prevented as has been recently discussed,²³ and therefore coherent energy transport is at play in this time scale.

From the experimental viewpoint, transient absorption could witness the influence of bilin-dependent spectral density in the dynamics. While the spectral density that has been experimentally fitted gives a good agreement with the fluorescence spectra of PE545, the agreement is not as good in the case of the transient absorption.²⁸ It is also important to highlight that the fitted spectral density in ref. 28 was made in the framework of modified Redfield that is known to overestimate rates of transfer. These two points justify once more the need for atomic-level and alternative determination of spectral densities. A complete study of the implications of bilin-dependent spectral densities for transient spectra and full exciton dynamics is out of the scope of this paper and will be presented in a separate publication.

Conclusions

In this work we have combined a recently developed QM/MMPol formulation with MD techniques to compute the spectral densities characterising exciton–phonon interactions in the PE545 light-harvesting complex. We have shown the existence of drastic differences in the spectral densities of individual pigments of PE545 when compared to those averaged over pseudo-symmetric pairs or over all pigments. We predicted some implications of such differences in the population dynamics for the intermediate and lowest exciton states of the pigments, concluding that using averaged spectral densities one cannot predict accurately the effects of non-equilibrium vibrational dynamics on the excitation energy transfer. We expect that such differences could be seen in transient absorption experiments and will in general lead to a richer dynamics of the coherences among exciton states though a complete study in this direction is out of the scope of this paper.

We have analysed the rich structure of the individual spectral densities as a function of energy and identified the prominent peaks that indicate preferential coupling to specific vibrational modes. The origin of such peaks was investigated and correlated with intramolecular normal modes using the quasi-harmonic and normal mode analyses. Equally, the impact of the solvent was analysed, leading to unexpected findings suggesting that it modulates the coupling of the lowest excited state to the environment by promoting or suppressing certain normal modes using the protein scaffold as a pathway.

The identification of the structural origins of the exciton–phonon interactions driving exciton transfer in photosynthetic systems as given in this study provides important guidelines for the design of new experiments that probe and exploit the largely unexplored correlated exciton-vibrational dynamics in light-harvesting complexes.

Acknowledgements

Support from the European Research Council (ERC) through the Starting Grant proposal no. 277755 (EnLight) is acknowledged by

B.M. and L.V. C.C. and M.C. acknowledge support from the Ministerio de Economía y Competitividad of Spain (grants CTQ2012-36195, RYC2011-08918 and BES-2013-064088) and computational resources provided by the Centre de Supercomputació de Catalunya. A.O.-C. and E.J.O. thank the Engineering and Physical Sciences Research Council of the UK (EPSRC) Grant EP/G005222/1 and the EU FP7 Project PAPETS for financial support (Grant Agreement no. 323901).

References

- 1 R. J. Cogdell, A. Gall and J. Köhler, *Q. Rev. Biophys.*, 2006, **39**, 227–324.
- 2 Y.-C. Cheng and G. R. Fleming, *Annu. Rev. Phys. Chem.*, 2009, **60**, 241–262.
- 3 V. Sundström, *Annu. Rev. Phys. Chem.*, 2008, **59**, 53–77.
- 4 H. Lee, Y.-C. Cheng and G. R. Fleming, *Science*, 2007, **316**, 1462–1465.
- 5 G. S. Engel, T. R. Calhoun, E. L. Read, T.-K. Ahn, T. Mancal, Y.-C. Cheng, R. E. Blankenship and G. R. Fleming, *Nature*, 2007, **446**, 782–786.
- 6 E. Collini, C. Y. Wong, K. E. Wilk, P. M. G. Curmi, P. Brumer and G. D. Scholes, *Nature*, 2010, **463**, 644–647.
- 7 I. Kassal, J. Yuen-Zhou and S. Rahimi-Keshari, *J. Phys. Chem. Lett.*, 2013, **4**, 362–367.
- 8 J. Huh, S. K. Saikin, J. C. Brookes, S. Valleau, T. Fujita and A. Aspuru-Guzik, *J. Am. Chem. Soc.*, 2014, **136**, 2048–2057.
- 9 T. Fujita, J. C. Brookes, S. K. Saikin and A. Aspuru-Guzik, *J. Phys. Chem. Lett.*, 2012, **3**, 2357–2361.
- 10 G. Panitchayangkoon, D. Hayes, K. A. Fransted, J. R. Caram, E. Harel, J. Wen, R. E. Blankenship and G. S. Engel, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 12766–12770.
- 11 L. Viani, C. Curutchet and B. Mennucci, *J. Phys. Chem. Lett.*, 2013, **4**, 372–377.
- 12 C. Olbrich, J. Strümpfer, K. Schulten and U. Kleinekathöfer, *J. Phys. Chem. B*, 2011, **115**, 758–764.
- 13 C. Olbrich, J. Strümpfer, K. Schulten and U. Kleinekathöfer, *J. Phys. Chem. Lett.*, 2011, 1771–1776.
- 14 A. Kolli, E. J. O'Reilly, G. D. Scholes and A. Olaya-Castro, *J. Chem. Phys.*, 2012, **137**, 174109.
- 15 A. Kolli, A. Nazir and A. Olaya-Castro, *J. Chem. Phys.*, 2011, **135**, 154112.
- 16 C. Kreisbeck and T. Kramer, *J. Phys. Chem. Lett.*, 2012, **3**, 2828–2833.
- 17 N. Christensson, H. F. Kauffmann, T. Pullerits and T. Mančal, *J. Phys. Chem. B*, 2012, **116**, 7449–7454.
- 18 V. Tiwari, W. K. Peters and D. M. Jonas, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 1203–1208.
- 19 A. W. Chin, J. Prior, R. Rosenbach, F. Caycedo-Soler, S. F. Huelga and M. B. Plenio, *Nat. Phys.*, 2013, **9**, 113–118.
- 20 J. M. Womick and A. M. Moran, *J. Phys. Chem. B*, 2011, **115**, 1347–1356.
- 21 G. H. Richards, K. E. Wilk, P. M. G. Curmi, H. M. Quiney and J. A. Davis, *J. Phys. Chem. Lett.*, 2012, **3**, 272–277.

- 22 D. B. Turner, R. Dinshaw, K.-K. Lee, M. S. Belsley, K. E. Wilk, P. M. G. Curmi and G. D. Scholes, *Phys. Chem. Chem. Phys.*, 2012, **14**, 4857–4874.
- 23 E. J. O'Reilly and A. Olaya-Castro, *Nat. Commun.*, 2014, **5**, 3012.
- 24 J. Pieper, J. Voigt and G. J. Small, *J. Phys. Chem. B*, 1999, **103**, 2319–2322.
- 25 E. J. Peterman, H. van Amerongen, R. van Grondelle and J. P. Dekker, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6128–6133.
- 26 E. J. G. Peterman, T. Pullerits, R. van Grondelle and H. van Amerongen, *J. Phys. Chem. B*, 1997, **101**, 4448–4457.
- 27 R. Jankowiak, M. Reppert, V. Zazubovich, J. Pieper and T. Reinot, *Chem. Rev.*, 2011, **111**, 4546–4598.
- 28 V. I. Novoderezhkin, A. B. Doust, C. Curutchet, G. D. Scholes and R. van Grondelle, *Biophys. J.*, 2010, **99**, 344–352.
- 29 A. Damjanović, I. Kosztin, U. Kleinekathöfer and K. Schulten, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2002, **65**, 031919.
- 30 L. Janosi, I. Kosztin and A. Damjanović, *J. Chem. Phys.*, 2006, **125**, 014903.
- 31 R. C. Walker, I. P. Mercer, I. R. Gould and D. R. Klug, *J. Comput. Chem.*, 2007, **28**, 478–490.
- 32 C. Olbrich and U. Kleinekathöfer, *J. Phys. Chem. B*, 2010, **114**, 12427–12437.
- 33 S. Shim, P. Rebentrost, S. Valleau and A. Aspuru-Guzik, *Biophys. J.*, 2012, **102**, 649–660.
- 34 E. Rivera, D. Montemayor, M. Masia and D. F. Coker, *J. Phys. Chem. B*, 2013, **117**, 5510–5521.
- 35 T. Renger, A. Klinger, F. Steinecker, M. Schmidt am Busch, J. Numata and F. Müh, *J. Phys. Chem. B*, 2012, **116**, 14565–14580.
- 36 A. B. Doust, I. H. M. van Stokkum, D. S. Larsen, K. E. Wilk, P. M. G. Curmi, R. van Grondelle and G. D. Scholes, *J. Phys. Chem. B*, 2005, **109**, 14219–14226.
- 37 H. Hossein-Nejad, C. Curutchet, A. Kubica and G. D. Scholes, *J. Phys. Chem. B*, 2011, **115**, 5243–5253.
- 38 C. Y. Wong, R. M. Alvey, D. B. Turner, K. E. Wilk, D. A. Bryant, P. M. G. Curmi, R. J. Silbey and G. D. Scholes, *Nat. Chem.*, 2012, **4**, 396–404.
- 39 A. B. Doust, C. N. J. Marai, S. J. Harrop, K. E. Wilk, P. M. G. Curmi and G. D. Scholes, *J. Mol. Biol.*, 2004, **344**, 135–153.
- 40 C. Curutchet, V. I. Novoderezhkin, J. Kongsted, A. Muñoz-Losa, R. van Grondelle, G. D. Scholes and B. Mennucci, *J. Phys. Chem. B*, 2013, **117**, 4263–4273.
- 41 K. E. Wilk, S. J. Harrop, L. Jankova, D. Edler, G. Keenan, F. Sharples, R. G. Hiller and P. M. G. Curmi, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 8901–8906.
- 42 A. B. Doust, K. E. Wilk, P. M. G. Curmi and G. D. Scholes, *J. Photochem. Photobiol. A*, 2006, **184**, 1–17.
- 43 C. Curutchet, J. Kongsted, A. Muñoz-Losa, H. Hosseini-Nejad, G. D. Scholes and B. Mennucci, *J. Am. Chem. Soc.*, 2011, **133**, 3078–3084.
- 44 V. May and O. Kühn, *Charge and Energy Transfer Dynamics in Molecular Systems*, Wiley-VCH Verlag GmbH, Weinheim, Germany, 2003.
- 45 S. Valleau, A. Eisfeld and A. Aspuru-Guzik, *J. Chem. Phys.*, 2012, **137**, 224103.
- 46 C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G. D. Scholes and B. Mennucci, *J. Chem. Theory Comput.*, 2009, **5**, 1838–1848.
- 47 M. C. Zerner, in *Rev. Comput. Chem.*, ed. K. B. Lipkowitz and D. B. Boyd, Wiley VCH, New York, II, 1994, p. 313.
- 48 M. Aghtar, J. Strümpfer, C. Olbrich, K. Schulten and U. Kleinekathöfer, *J. Phys. Chem. B*, 2013, **117**, 7157–7163.
- 49 R. Baron, W. F. van Gunsteren and P. H. Hünenberger, *Trends Phys. Chem.*, 2006, **11**, 87.
- 50 Minnesota Database of Frequency Scale Factors for Electronic Model Chemistries, 2013.
- 51 L. Viani, *MView: A tool for visualization and analysis of molecular properties*, www.mview-tools.com.
- 52 V. I. Novoderezhkin and R. van Grondelle, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7352–7365.
- 53 Y. Tanimura and R. Kubo, *J. Phys. Soc. Jpn.*, 1989, **58**, 101–114.
- 54 A. Ishizaki and G. R. Fleming, *J. Chem. Phys.*, 2009, **130**, 234111.

Determination of the protonation preferences of bilin pigments in cryptophyte antenna complexes

Physical Chemistry Chemical Physics, **2018** (Submitted)

Marina Corbella,¹ Zi S. D. Toa,² Gregory D. Scholes,² F. Javier Luque,³ and Carles Curutchet¹

¹ Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica i Institut de Química Teòrica i Computacional (IQTCUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona, Espanya.

² Department of Chemistry, Princeton University, Princeton, New Jersey, United States.

³ Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Institut de Biomedicina (IBUB) i Institut de Química Teòrica i Computacional (IQTCUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Santa Coloma de Gramenet, Espanya.

3.1.2. Determination of the protonation preferences of bilin pigments in cryptophyte antenna complexes.

Resum

Les proteïnes antena són les entitats encarregades de recol·lectar la llum i transferir-ne l'energia d'excitació resultant als centres de reacció dels diferents aparells fotosintètics. En les algues procedents del filum dominat per criptòfits, s'ha observat que aquestes proteïnes antena presenten una gran diversitat de colors al mateix temps que poden portar a terme aquesta transferència d'energia amb un rendiment màxim inclús en condicions de baixa lluminositat. Aquestes propietats han despertat l'interès en entendre'n els mecanismes de funcionament per tal de poder aplicar-los a sistemes artificials de captura de llum. Les proteïnes antena presents en aquestes algues són les anomenades ficobiliproteïnes, ja que estan formades per un únic tipus de pigment, les ficobilines. Les ficobilines presenten una estructura lineal tetrapirròlica i estan unides covalentment a l'estructura proteica a través de residus específics de cisteïna. El que confereix a aquest tipus d'antena la diversitat de colors és precisament la presència de diferents tipus de ficobilines, amb petites diferències estructurals, dins d'una mateixa proteïna. Tot i la gran varietat de treballs que s'han desenvolupat entorn a aquestes proteïnes antena, l'estat de protonació de les diferents ficobilines presents no ha estat mai estudiat en deteniment. La presència d'un residu d'àcid aspàrtic o glutàmic interaccionant directament amb els dos anells de pirrole centrals d'aquests pigments, ha portat a considerar-los més habitualment en la seva forma completament protonada, no obstant, en certes ocasions alguns pigments no presenten aquesta característica. Aquest tret diferencial, juntament amb els canvis de pH que pot experimentar el lumen (medi on es troben aquestes ficobiliproteïnes) en funció del temps d'exposició a la llum solar, ha portat a iniciar aquest estudi per tal d'elucidar el pK_a de cada una d'aquestes ficobilines dins de la seva estructura proteica, ja que el canvi de l'estat de protonació d'alguns d'aquests pigments podria tenir un impacte important en les corresponents energies d'excitació i en el procés de transferència d'energia. Mitjançant cicles termodinàmics i combinant càlculs quàntics en fase gas i en solució s'han obtingut els pK_{as} intrínsecos (en aigua) per a cada tipus de ficobilina (PEB, DBV, PCB, MBV)

present en cinc ficobiliproteïnes diferents (PE545, PC577, PC612, PC630, PC645). Posteriorment s'han emprat simulacions de dinàmica molecular juntament amb estimacions empíriques dels diferents pK_a s per tal de descriure l'impacte de l'entorn específic de cada pigment a l'interior de cada sistema proteic. Els resultats d'aquest estudi teòric s'han racionalitzat mitjançant la obtenció dels diferents espectres d'absorció de les ficobiliproteïnes a diferents valors de pH , entre 4.6 i 9.0, per tal d'observar l'impacte dels canvis del pH extern en les propietats espectrals d'aquestes proteïnes i correlacionar-los amb els canvis en l'estat de protonació dels pigments que les confereixen. Els nostres resultats reafirmen la hipòtesis de que tots els pigments es troben en la seva forma protonada en el rang de pH fisiològic (5-7), en les cinc proteïnes estudiades. No obstant, mentre que pels pigments que presenten el residu d'àcid aspàrtic o glutàmic interaccionant amb els dos anells centrals es suggereix un pK_a al voltant de 8, els pigments als quals els hi manca aquest residu, presenten un pK_a inferior, sobre 7; de manera que, tot i trobar-se majoritàriament en la seva forma protonada, certa part de la població estarà present també, en aquests casos, en la seva forma desprotonada.



Determination of the protonation preferences of bilin pigments in cryptophyte antenna complexes

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

The light-harvesting mechanisms of cryptophyte antenna complexes have attracted considerable attention due to their ability to exhibit maximal photosynthetic activity at very low-light conditions and to display several colors, as well as the observation of vibronic coherent features in their two-dimensional electronic spectra. However, detailed investigations on the interplay between the protein environment and their light-harvesting properties are hampered by the uncertainty related to the protonation state of the underlying bilin pigments. Here we study the protonation preferences of four types of bilin pigments including 15,16-dihydrobiliverdin (DBV), phycoerythrobilin (PEB), phycocyanobilin (PCB) and mesobiliverdin (MBV), which are found in phycoerythrin PE545 and phycocyanin PC577, PC612, PC630 and PC645 complexes. We apply quantum chemical calculations coupled to continuum solvation calculations to predict the intrinsic acidity of bilins in aqueous solution, and then combine molecular dynamics simulations with empirical pK_a estimates to investigate the impact of the local protein environment in the acidity of the pigments. We also report measurements of the absorption spectra of the five complexes measured in a wide range of pH in order to validate our simulations and investigate possible changes in the light harvesting properties of the complexes in the range of physiological pH found in the lumen ($pH \sim 5-7$). The results suggest a $pK_a > 7$ for DBV and MBV pigments in the α polypeptide chains of PE545 and PC630/PC645 complexes, which are not coordinated to a negatively charged amino acid. For the other PEB, DBV and PCB pigments, which interact with a Glu or Asp side chain, higher pK_a values ($pK_a > 8$) are estimated. Overall, the results support a preferential population of the fully protonated state for bilins in cryptophyte complexes at physiological conditions regardless of the specific type of pigment and local protein environment.

1 Introduction

In photosynthesis specialized light-harvesting pigment-protein complexes (PPCs) are used to capture incident sunlight and funnel its energy to the reaction center.^{1,2} Cryptophytes are eukaryotic algae that live in marine and freshwater environments and represent important primary producers, able to capture sunlight through tunable linear tetrapyrrole chromophores known as bilins (see Fig. 1). Despite the apparent simplicity of their light-harvesting antennae

compared to other classes of algae, cryptophytes display several colors and exhibit maximal photosynthetic activity at very low-light conditions.³⁻⁵ Depending on their absorption maximum they are divided in two classes, phycoerythrins (PE, red colored) and phycocyanins (PC, blue colored), and so far four types of PEs (PE545, PE555, two types of PE566) and five types of PCs (PC569, PC577, PC612, PC630 and PC645) have been reported.⁵ The variability in spectral properties of these phycobiliproteins mainly arises from differences in the underlying bilin composition and content. Indeed, several types of bilins are present in cryptophytes, which differ in the number of double bonds and the degree of conjugation of the π -electron system (Fig. 1): 15,16-dihydrobiliverdin (DBV), phycoerythrobilin (PEB), phycocyanobilin (PCB), mesobiliverdin (MBV), bilin 584 and bilin 618.⁵ Furthermore, the protein sequence also impacts the spectral properties of these complexes, as illustrated by differences in PC577 and PC612 spectra, which contain the same set of bilin pigments, six PCBs and two DBVs.

The remarkable photosynthetic activity of these algae has motivated a variety of studies aimed at unraveling their light

^a Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica and Institut de Química Teòrica i Computacional (IQTCUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona Spain.

^b Department of Chemistry, Princeton University, Washington Road, Princeton, New Jersey 08544, United State.

^c Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Institut de Biomedicina (IBUB) and Institut de Química Teòrica i Computacional (IQTCUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Santa Coloma de Gramenet, Spain.

† Electronic Supplementary Information (ESI) available. See DOI: 10.1039/x0xx00000x

harvesting mechanisms,^{2,6} in the quest for principles to be used in the development of artificial devices.⁷ Among them, the observation of coherence phenomena at room temperature for the PE545 and PC645 complexes⁸ and other photosynthetic systems has fueled an ongoing debate on the molecular origin of the observed oscillations and their role in the photosynthetic function.⁹ An emerging consensus indicates that vibronic coherences can be supported by resonant vibrations, although their electronic or vibrational nature is still under debate.^{10–12} Recent simulations on the PC645 complex, for example, suggest that energy transport occurs in an incoherent vibronic regime, a finding explained by the rather large reorganization energies of the pigments compared to the electronic couplings among them.¹³ Cryptophyte antenna complexes are also interesting because of the observation of persistent vibronic coherences in "closed" forms of these antenna proteins (PC630, PC645) compared to "open" ones (PC577, PC612), where the insertion of a single residue induces a rotation of ~73 degrees of the protein subunits leading to an almost two-fold symmetry in the quaternary structure.¹⁴

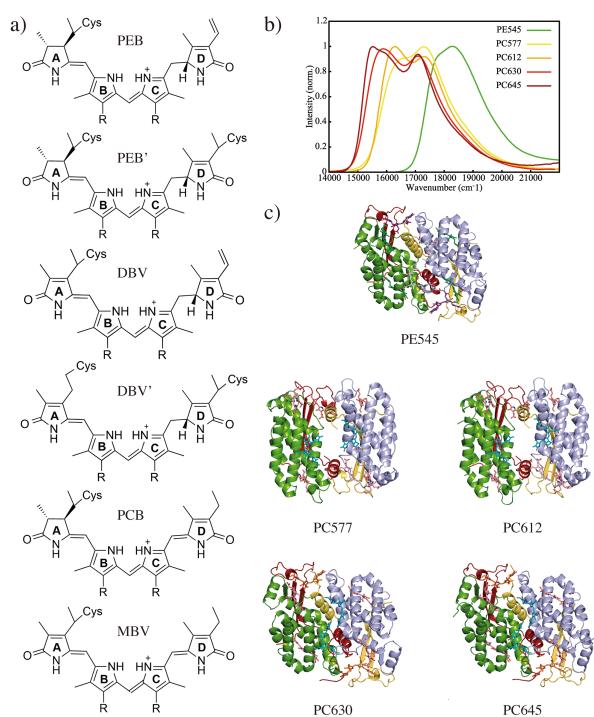


Figure 1. Bilin pigments present in the cryptophyte antenna complexes studied in this work. a) Bilin structures, b) absorption spectra, and c) protein structures of the PE545, PC577, PC612, PC630 and PC645 complexes. R stands for the propionic group.

Theoretical studies of energy transfer dynamics in antenna complexes and the role of coherence, however, still face considerable challenges related to the accurate determination of the ingredients needed to construct the exciton Hamiltonian of the system: the site energies, the electronic couplings and the spectral densities of the corresponding pigments.⁶ An additional complication in the study of phycobiliproteins is the need to understand the protonation preferences of bilin pigments in the complex, as they can considerably impact their

electronic transition properties, and also their degree of coupling to the environment.¹⁵ Moreover, protonation/deprotonation of the pigments is also important because, unlike other photosynthetic organisms where antenna complexes are bound to the thylakoid membrane, cryptophyte biliproteins are suspended in the lumen,¹⁶ where the pH varies on the range ~5–7, depending on the prolongation of the incident sunlight.¹⁷ The uncertainty regarding this issue leaves open questions: can the protonation state of the pigments change upon variations of available sunlight?, and can changes in the bilin protonation patterns alter the light-harvesting pathways and dynamics in the system?

Cryptophyte antennae are organized either as $\alpha_1\beta\alpha_2\beta$ heterodimers or $(\alpha\beta)_2$ homodimers. The crystal structures solved for several of these complexes (PE545, PE555, PC577, PC612, PC630 and PC645) indicate that all bilins in the highly-conserved β polypeptide chains present an Asp residue interacting with the central pyrrole rings B and C, thus suggesting a fully protonated form of the pyrrole rings.^{14,18,19} This would be consistent with research that supports a protonated form for the PCB in the C-Phycocyanin complex of cyanobacteria, where an Asp counterion also coordinates the B and C central pyrrole rings.^{20–22} Spectroscopic studies performed on the similar phytochromobilin pigment in plant phytochromes also support a fully protonated tetrapyrrole.²³ Compared to the β chains, however, the structures of the α subunits in cryptophyte antennae are much more divergent. In this case, the bilins in the α subunits of PE555, PC577 and PC612 interact with a Glu side chain, but the DBVs in PE545 and the MBVs in PC645 and PC630 lack the presence of a negatively charged amino acid. In PE545, indeed, difference electron density maps obtained from ultrahigh resolution data at 0.97 Å seem to suggest a protonated form for the PEBs in the complex, which interact with an Asp residue, but can not distinguish among protonated or neutral forms for the DBVs.¹⁹ The uncertainty regarding the protonation preferences of the bilins in cryptophyte antennae is reflected in a variety of theoretical studies performed on these systems, where the pigments are either assumed to be all protonated or unprotonated at physiological conditions.^{13,14,19,24–33} Recently, the Coker group has studied the protonation state of the MBVs in PC645 by comparing experimental and theoretical spectra derived from quantum chemical calculations.¹⁵ Their results indicate that the PC645 complex with unprotonated MBVs displays better maximum intensities in the two main absorption peaks compared to the protonated case. However, the low-energy edge of the predicted spectra is slightly too broad, so that the assignment of the protonation preferences still remains open.

In this contribution, we study the protonation preferences of the PEB, DBV, PCB and MBV pigments found in five cryptophyte phycobiliproteins (PE545, PC577, PC612, PC630 and PC645), shown in Fig. 1, by combining structure-based predictions of the pK_a 's of the pigments based on thermodynamic cycles with pH-dependent measures of their absorption spectra. Our results strongly suggest that all bilins

in these proteins are present in their protonated form at biologically relevant pH in the lumen ($pH \sim 5-7$). We find that the MBVs in PC630 and PC645 as well as the DBVs in PE545, located at the α polypeptide chains of the complexes and lacking the presence of a Glu counterion, are the most acid bilins in the complexes, but still have pK_a values > 7 . Nevertheless, the changes observed in the absorption spectra suggest that a small fraction of deprotonated species of these pigments could be populated at pH 's ~ 7 .

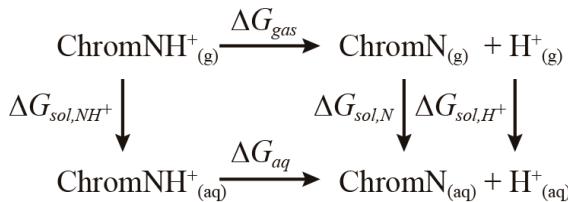
2 Methods

2.1 Thermodynamic cycles

Acid dissociation constants, known as pK_a values, are related to the reaction where an acid dissociates into its conjugate base and a proton in solution. The pK_a can thus be obtained from the Gibbs free energy change of this reaction, ΔG_{aq} :

$$pK_a = \frac{\Delta G_{aq}}{RT \ln(10)} \quad (1)$$

In order to estimate accurate values of ΔG_{aq} , we adopt the common thermodynamic cycle illustrated in Scheme 1, which combines quantum-mechanical (QM) calculations of the reaction free energy in vacuum, ΔG_{gas} , with solvation free energies in water estimated using quantum-chemical continuum solvation models, as described in the next section.^{34,35}



Scheme 1. Thermodynamic cycle used to calculate the aqueous deprotonation free energies of bilin pigments. ChromNH⁺ and ChromN stand for the protonated and deprotonated states of the chromophore.

Thus, we calculate ΔG_{aq} values using the following expression:

$$\Delta G_{aq} = \Delta G_{gas} + \Delta \Delta G_{sol} \quad (2)$$

where

$$\Delta G_{gas} = G_{gas}(ChromN) + G_{gas}(H^+) - G_{gas}(ChromNH^+) \quad (3)$$

and

$$\Delta \Delta G_{sol} = \Delta G_{sol}(ChromN) + \Delta G_{sol}(H^+) - \Delta G_{sol}(ChromNH^+) \quad (4)$$

We adopted the well-accepted value of $G_{gas}(H^+)$ equal to -6.28 kcal/mol defined for a standard state of 1 atm, as well as the value of $\Delta G_{sol}(H^+)$ equal to -265.9 kcal/mol, defined for the transfer from the gas-phase to aqueous solution at 1M standard state.³⁴ G_{gas} values computed for protonated/unprotonated forms of the bilins using QM methods also correspond to a standard state of 1 atm, whereas the corresponding ΔG_{sol} terms calculated using continuum solvation models are defined for a gas-phase 1M

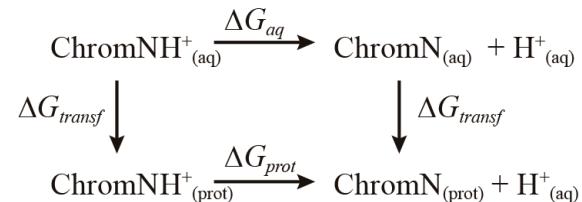
standard state. Thus, ΔG_{gas} values were converted to a 1M standard state by adding the usual 1.89 kcal/mol term.³⁴

Because the propionic acid groups can adopt multiple conformations with respect to the pyrrole rings, and therefore impact in different ways their deprotonation properties, the aqueous pK_a values corresponding to the deprotonation reaction on rings A, B, C or D of each type of bilin was computed for model pigments where the propionic groups were substituted by methyl units, as shown in Fig. 2. Later, the impact of the negatively charged propionic groups on the pK_a was introduced when computing the pK_a shift associated to the transfer of the bilin from water solution to the protein. Note that a similar strategy was recently adopted in order to compute the pK_a of biliverdin in the IFP2.0 fluorescent protein.³⁶

The prediction of the pK_a of titratable groups or ligands inside a protein environment is a challenging task, and a variety of approaches have been proposed.³⁷⁻³⁹ The usual strategy relies on estimating the free energies associated to the transfer of the titratable group from aqueous solution to the protein, as illustrated in Scheme 2. Then, these free energy contributions are combined to estimate the ΔpK_a shift compared to an aqueous reference value:

$$pK_a(prot) = pK_a(wat) + \Delta pK_a(wat \rightarrow prot) \quad (5)$$

where $pK_a(wat)$ is the reference value adopted for the group and the ΔpK_a term mainly involves electrostatic contributions due to interactions established with the protein environment as well as the desolvation cost.



Scheme 2. Thermodynamic cycle used to calculate the deprotonation free energies of bilin pigments in the protein environment.

Different strategies have been developed to compute the $\Delta pK_a(wat \rightarrow prot)$ term, mainly based on i) continuum solvation models, like Generalized Born (GB) or Poisson-Boltzmann (PB) electrostatic calculations, ii) molecular dynamics (MD)-based techniques coupled to free energy simulations, and iii) empirical models in which the shift is given as a sum of effective perturbation contributions, for example hydrogen bonds, charge-charge interactions or desolvation effects, whose weights are optimized by comparison to large sets of experimental data, like the popular PROPKA method.^{40,41} Here we adopt the fast PROPKA method to compute the ΔpK_a shifts, which has been shown to provide pK_a values with similar accuracy compared to more costly molecular dynamics/generalized Born/thermodynamic integration (MD/GB/TI) techniques, with root-mean-squared deviations of ~ 1.4 pK_a units.³⁸ In PROPKA, the ΔpK_a shift is

expressed as the sum of a Coulomb contribution due to charge-charge interactions with all other charged or ionizable groups, and a self-energy contribution, representing the remaining contributions obtained when all other charged or ionizable groups are kept in their neutral form. The latter involves desolvation and intrinsic electrostatic energy contributions, like hydrogen-bonding interactions and other unfavorable electrostatic interactions. The resulting coupled titrations are then solved iteratively using a Monte Carlo scheme. Moreover, we average pK_a estimates over conformational ensembles sampled from MD simulations of the complexes in order to account for thermal fluctuations.

2.2 Computational details

Structures of the cryptophyte complexes. Calculations were performed based on the X-ray crystal structures reported by Curmi and co-workers for phycoerythrin PE545 from *Rhodomonas* sp. CS24 (PDB code 1XG0, resolution 0.97 Å),^{18,19} phycocyanin PC645 from *Chroomonas* sp. CCMP270 (PDB code 4LMS, resolution 1.35 Å),¹⁴ phycocyanin PC612 from *Hemiselmis virescens* CCAC 1635 B (PDB code 4LM6, resolution 1.70 Å),¹⁴ as well as phycocyanin PC577 from *Hemiselmis pacifica* CCMP706 (resolution 1.00 Å) and PC630 from *Chroomonas* CCAC 1627 B (resolution 1.60 Å), which have not been deposited yet. PE545, PC630 and PC645 are organized as $\alpha_1\beta\alpha_2\beta$ heterodimers, whereas PC577 and PC612 are $(\alpha\beta)_2$ homodimers and display an almost perfect two-fold symmetry. In PE545, each α chain binds one DBV covalently linked to one Cys residue (DBV_{19A} and DBV_{19B}), whereas each β chain binds two PEBs linked to a single Cys (PEB_{158C}, PEB_{158D}, PEB_{82C} and PEB_{82D}) and one PEB linked to two Cys residues (PEB_{50/61C} and PEB_{50/61D}). In PC577 and PC612, on the other hand, each α chain binds one singly linked PCB (PCB_{20A} and PCB_{20C}) and each β chain binds two singly linked PCBs (PCB_{158B}, PCB_{158D}, PCB_{82B}

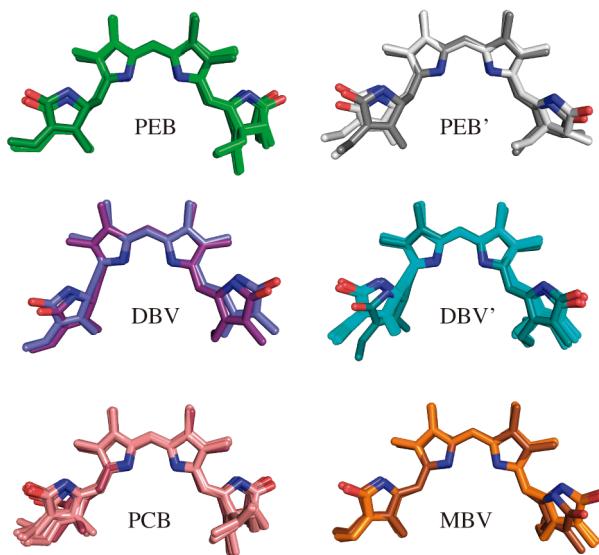


Figure 2. Structures of the bilin pigments found in PE545, PC577, PC612, PC630 and PC645 aligned to the quantum-chemical optimized geometries used for the prediction of the aqueous pK_a . The optimized geometries are displayed in slightly darker colours.

and PCB_{82D}) and a doubly linked DBV (DBV_{50/61B} and DBV_{50/61D}). Finally, in PC630 and PC645 the β chains bind the same pigments as in PC577 and PC612, whereas the α chains binds one singly linked MBV (MBV_{19A} and MBV_{19C}) instead of a PCB. In PE545, the missing residues in the crystal were added as reported previously,²⁶ whereas in the PC577 and PC612 ($\alpha\beta$)₂ homodimers missing residues were added by homology with the other symmetric polypeptide chain. In PC630 and PC645, we simply capped polypeptide chains using an acetyl group due to missing residues in the N-terminal end of the chains.

Quantum chemical calculations. Starting from the geometries of bilins as found in the crystal structures, we defined simplified models in order to compute their acidity/basicity in gas phase and in water solution. Such models included the full tetrapyrrolic structure, in which the propionic side chains of rings B and C were replaced by methyl groups to avoid the folding of the chromophore and thus maintain its biological conformation, whereas the thioether linkage was replaced by a hydrogen atom. In order to perform the aqueous pK_a calculations, we selected a single geometry for each type of pigment PEB, PEB', DBV, DBV', PCB and MBV, where the prime indicates bilins doubly-linked to the protein through Cys residues. The optimized geometries of each model bilin, aligned with the multiple conformations found for each type of pigment in the crystal structures, are shown in Fig. 2. As can be observed, these geometries keep the conformations of the pigments in their specific protein environments.

The geometries of fully protonated and deprotonated (at rings A, B, C or D) forms of each model chromophore (PEB, PEB', DBV, DBV', PCB and MBV) were optimized at the B3LYP/6-31G(d) level of theory. The resulting minima were verified by vibrational frequency analysis, and corresponding free-energy corrections (G_{corr}) at 298.15 K were obtained by scaling the vibrational frequencies by 0.977, as recommended for this level of theory.⁴² Subsequently, the electronic energies were refined by combining SCS-MP2 energies extrapolated at the complete basis set (CBS) limit with CCSD corrections obtained with a reduced basis set. The spin-component-scaled version of MP2 provides a significant improvement on ground state energies by scaling parallel and antiparallel-spin pair correlation energies:⁴³

$$E_{SCS-MP2} = E_{HF} + \frac{1}{3}(E_{corr(\alpha-\alpha)} + E_{corr(\beta-\beta)}) + \frac{6}{5}E_{corr(\alpha-\beta)} \quad (6)$$

The CBS energy was then extrapolated from SCS-MP2 correlation energies computed using Dunning's cc-pVDZ and cc-pVTZ basis sets according to the formula proposed by Halkier:⁴⁴

$$E_{SCS-MP2/CBS} = E_{HF/cc-pVTZ} + \frac{E_{SCS-MP2/cc-pVTZ}^{corr} - E_{SCS-MP2/cc-pVDZ}^{corr}}{3^3 - 2^3} \quad (7)$$

Finally, in order to recover the correlation energy between MP2 and CCSD, we added a correction given by the difference between CCSD/6-31G(d) and MP2/6-31G(d) calculations.

$$\Delta E_{CCSD} = E_{CCSD/6-31G(d)} - E_{SCS-MP2/6-31G(d)} \quad (8)$$

Thus, the final gas-phase free energies for the protonated and deprotonated species were given by the following expression:

$$G_{\text{gas}} = E_{\text{SCS-MP2/CBS}} + \Delta E_{\text{CCSD}} + G_{\text{corr}} \quad (9)$$

Gas-phase free energies were then combined with hydration free energies computed using the MST^{45,46} and SMD⁴⁷ continuum solvation models. MST and SMD calculations were performed at the B3LYP/6-31G(d) level of theory, in order to be consistent with the parametrization of these methods. In order to test the accuracy of the computational protocol we also computed the known pK_a 's of the heteroaromatic molecules pyridinium and imidazolium, which can also deprotonate an aromatic NH⁺ group like the bilins, following the same strategy. All quantum chemical calculations were performed using the Gaussian 09 software.⁴⁸

MD simulations and protein pK_a shifts. We performed MD simulations for PC577, PC612, PC630 and PC645 solvated in pre-equilibrated TIP3P water boxes (buffer zone 15 Å) using the Leap module of the Amber14 suite of programs.⁴⁹ The ff14SB⁵⁰ and GAFF⁵¹ force fields were used for the proteins and the bilins, respectively. Charges for PCB, MBV, and DBV' bilins in their fully protonated forms were derived from multiconformational RESP⁴⁹ fits performed at the HF/6-31G(d) level of theory on the multiple geometries of the pigments found in the complexes, where hydrogens were previously optimized and Cys links were treated as described previously for PE545.²⁶ Protein hydrogen atoms were added according to PROPKA predictions,⁴⁰ which indicated standard protonation states for all titratable residues except for His22A in PC630 and His21A in PC645, which were considered in their fully protonated form. The four systems were initially minimized and then thermalized running 50 ps NVT and 150ps NPT simulations applying constraints on the protein and bilin atoms with a harmonic potential (500.0 kcal/mol·Å²). Then, we thermalized the complete system from 50K to 300K by running 50 ps NVT and 50 ps NPT simulations. Finally, we extended the simulations for 10 ns for production purposes. All runs were performed with Amber 14⁴⁹ using an integration time step of 2 fs together with the SHAKE algorithm to restrain all bonds involving hydrogen, periodic boundary conditions, the particle-mesh Ewald approach to account for long-range electrostatics and a nonbonded cutoff equal to 10 Å. For PE545 we used the MD simulation described in ref.²⁶ We then extracted 50 snapshots equally spaced along the trajectories to estimate the bilin pK_a values in the complexes. Water-to-protein pK_a shifts were computed using the empirical rules implemented in the PROPKA3.0⁴⁰ server, which have been extended to pK_a shifts of active site residues and ionizable ligands/cofactors in PROPKA3.1.⁴¹ In these calculations, the propionic acid groups of the pigments were treated as additional titratable groups with a reference pK_a value of 4.5.

2.3 Absorption spectra

Concentrated samples of purified PE545, PC577, PC630 and PC645 isolated from *Rhodomonas* sp. CS24 (PE545), *Hemiselmis pacifica* CCMP706 (PC577), *Chroomonas* CCAC

1627 B (PC630) and *Chroomonas mesostigmatica* CCMP269 (PC645) were frozen at -20°C in 0.050 M sodium phosphate buffer (*pH* 6.5 for PE545; *pH* 7.5 for PC577, PC630 and PC645) until required for spectroscopic measurements.⁵² We then prepared buffer solutions at varying *pH* levels by adjusting relative amounts of Na₂HPO₄ and citric acid (*pH* 4.0, 4.6, 5.4), Na₂HPO₄ and NaH₂PO₄ (*pH* 6.5, 7.8) trizma HCl and trizma base (*pH* 8.2, 8.5, 9.0) and Na₂CO₃ and NaHCO₃ (*pH* 9.4). We measured the *pH* levels to be 4.0 ± 0.1, 4.6 ± 0.1, 5.4 ± 0.1, 6.5 ± 0.1, 7.8 ± 0.1, 8.2 ± 0.1, 8.5 ± 0.1, 9.0 ± 0.1 and 9.4 ± 0.1 using a calibrated microelectrode (Mettler Toledo). 100 µL aliquots of concentrated protein were diluted into the different 1.5 mL volume buffer solutions. The protein and buffer solutions were measured and dispensed using a micropipette to make the concentrations as equal as possible. Proteins were diluted to an optical density (O.D.) of ~0.05/mm. Solution-phase steady-state linear absorption spectra for each phycobiliprotein was then recorded using a Varian Cary 6000i UV-vis spectrometer with a resolution of 1 nm at room temperature.

3. Results and discussion

3.1 Aqueous pK_a

The first-principles determination of aqueous pK_a values is considerably challenging due to potential inaccuracies in the calculation of the gas-phase free energies of the acid and the conjugated base, as well as on their solvation free energies, as their counterparts for the proton, $G_{\text{gas}}(H^+)$ and $\Delta G_{\text{sol}}(H^+)$, are well tabulated. The main source of inaccuracies may be ascribed to the calculation of the solvation free energies of the ionic species, as the gas-phase free energies and the solvation free energies of neutrals can be computed with errors close to ~1 kcal/mol if high-level correlated methods are combined with modern parametrizations of solvation models.³⁴ In contrast to neutrals, solvation free energies of charged species are larger in magnitude due to strong electrostatic solute-solvent interactions, thus increasing associated errors. In addition, this can lead to persistent specific interactions, like hydrogen bonds, not properly accounted for in continuum solvation models. Thus, the hydration free energy of ions estimated from continuum models may lead to increased errors of ~4-5 kcal/mol, though prediction of relative free energies may benefit from cancellation of errors. Because of this, we have chosen to use two alternative solvation models, the SMD model developed by Marenich and co-workers⁴⁷ and the MST model developed in Barcelona.^{45,46}

In order to test the validity of our computational protocol, we first computed the pK_a of two aromatic heterocyclic compounds with known acidity, pyridinium and imidazolium, which, like the bilins, can be deprotonated on NH⁺ groups without breaking the aromaticity of the heterocycle. The results, reported in Table 1, support the reliability of the computational protocol, with absolute mean errors of ~1 pK_a unit for both solvation models (contributions to the proton dissociation free energies are reported in Table S1 of the Supporting Information). This indicates the absence of

ARTICLE

Journal Name

important systematic deviations in the prediction of absolute pK_a values, an error often found to depend on the nature of the functional group.^{34,53} Thus, pK_a predictions for the bilin chromophores can be roughly expected to have an accuracy of ~1–2 pK_a units.

Table 1. Gas-phase and aqueous proton dissociation free energies and corresponding pK_a values computed for pyridinium and imidazolium.

	MST			SMD		
	ΔG_{gas}° ^a	ΔG_{aq}^{*} ^a	pK_a	ΔG_{aq}^{*} ^a	pK_a	$pK_{a,exp}^b$
Pyridinium	215.6	5.4	3.9	5.4	3.9	5.2
Imidazolium	219.2	10.1	7.4	8.7	6.4	7.0

^aFree energies (kcal/mol) corresponding to standard states of 1 atm in gas phase (ΔG_{gas}^0) and 1 M in aqueous solution (ΔG_{aq}^*). ^bExperimental values from Ref.⁵⁴

For the bilins, we have computed the pK_a using the same protocol and adopting solvation free energies averaged over MST and SMD-based predictions. Table 2 shows the aqueous pK_a predicted for each pyrrole ring of the bilin chromophores, as well as the corresponding gas-phase and aqueous free energies (the individual contributions to these free energies are shown in Table S2). In all cases the outer pyrrole rings A and D present larger pK_a values (from 11.9 to 31.8) than those of the central B and C rings, which lie in the range 5.8–7.4. The amide character of the nitrogens in the outer rings explains this result, as deprotonation would prevent the delocalization of the nitrogen lone pair toward the carbonyl group and the corresponding resonance stabilization. However, the pK_a values of outer rings show three characteristic behaviours, depending on i) whether the ring is linked through a methine or a methylene bridge to the central pyrrole ring, and ii) the oxidation state of the ring. In the first group, the D rings of PEB and DBV (as well as doubly linked PEB' and DBV') are not conjugated with the other rings because of the methylene bridge, leading to a pK_a of 30–32. This can be explained by the lack of the positive charge delocalization onto this ring in PEB and DBV, whereas in PCB and MBV the positive charge can delocalize along the four pyrroles, including ring D (note here, however, that for the sake of simplicity only a single Lewis resonance structure is shown in Fig. 1). The A rings of PEB, PEB' and PCB, in contrast, are the least basic outer pyrroles, given their less oxidized state compared to the A rings in the other bilins, which leads to pK_a values of 12–13. Finally, the other outer rings (A ring of DBV, DBV' and MBV, D ring of PCB and MBV) display a pK_a of 16–17.

Overall, the results clearly indicate that the central B and C rings are the most acidic ones, with a pK_a of 6–7 in aqueous solution. The PCB pigment seems to be the least acidic, with a pK_a value of 7.1 (7.4) for the B (C) ring. The other pigments show similar acidities, the DBV one being slightly more acidic, especially the doubly linked DBV' one, in which the pK_a of ring C is lowered from 6.3 to 5.8 compared to the singly linked DBV. The results also indicate minor differences regarding the acidity of rings B and C, suggesting that at basic pH there could be a mixture of both deprotonated tautomers. On the other

hand, the protein environment might play a key role in determining the relative stability of these tautomers as well as of the fully protonated form, in a situation reminiscent of the influence exerted by the local environment on the pK_a of His side chains in proteins.

Table 2. Gas-phase and aqueous proton dissociation free energies and pK_a values computed for the deprotonation reaction on different pyrrole rings of bilin molecules.

	Ring	ΔG_{gas}° ^a	ΔG_{aq}^{*} ^a	pK_a
PEB	A	249.4	18.2	13.3
	B	235.4	9.3	6.8
	C	234.1	8.9	6.5
	D	290.4	41.0	30.0
PEB'	A	248.2	17.0	12.4
	B	234.9	9.5	6.9
	C	233.9	8.8	6.4
	D	292.1	42.3	31.0
DBV	A	255.3	23.7	17.4
	B	234.3	9.2	6.7
	C	233.1	8.6	6.3
	D	288.4	40.7	29.9
DBV'	A	255.9	24.1	17.6
	B	234.5	9.2	6.7
	C	232.6	7.9	5.8
	D	290.4	43.4	31.8
PCB	A	250.2	16.2	11.9
	B	238.4	9.7	7.1
	C	238.7	10.1	7.4
	D	258.1	23.1	17.0
MBV	A	256.6	22.9	16.7
	B	236.7	9.1	6.6
	C	236.6	9.0	6.6
	D	256.2	21.9	16.1

^aFree energies (kcal/mol) corresponding to standard states of 1 atm in gas phase (ΔG_{gas}^0) and 1 M in aqueous solution (ΔG_{aq}^*), the latter computed by using the solvation free energies averaged from MST and SMD calculations.

Recent predictions by Feliks and Field on the biliverdin pigment of the IFP2.0 fluorescent protein found slightly more acidic pK_a values by about ~4–5 pK_a units compared to our results for the similar MBV bilin.³⁶ Nevertheless, biliverdin differs from MBV by the presence of two vinyl groups instead of ethyl groups linked to rings A and D as well as by the orientation of the A ring. Moreover, their calculations included the propionic groups and involved an empirical correction calibrated on a training set of acids. Nevertheless, it is worth noting that a similar protocol was adopted in that study based

on DFT and single point MP2/CBS calculations. However, here we further refine the MP2/CBS values by adopting the SCS variant of MP2 and by adding CCSD corrections, two effects that contribute by ~50% each in increasing the pK_a estimates of the MBV central rings by ~3.5 pK_a units, thus illustrating the importance of such corrections and partially explaining the discrepancy with our results. For example, our predictions for pyridinium and imidazolium show absolute mean errors of ~1 pK_a unit, as discussed before, but degrade considerably to ~3 pK_a units if SCS and CCSD corrections are omitted.

3.2 Protonation preferences in the protein complex

In this section we investigate the impact of the specific environment of the bilins in five different cryptophyte complexes. Our predictions suggest that the model bilin pigments (lacking the propionic acid groups) have aqueous pK_a values around ~6–7, corresponding to deprotonation of either B or C central pyrrole rings. Thus, the protein environment and the relative orientation of the propionic acid groups may play a key role in determining the relative stability of the protonated form and of the two unprotonated tautomers. Table 3 reports the pK_a values computed for bilin pyrrole rings B and C in their protein environments, computed as the sum of the aqueous pK_a results discussed in the previous section and the MD-averaged ΔpK_a ($wat \rightarrow prot$) shifts computed by using PROPKA (Eq. 5). In addition, in Table S3 of the Supporting Information we report the bilin pyrrole pK_a values computed for the crystal structure, whereas in Tables S4 and S5 we show the pK_a values estimated for the bilin propionic groups and selected titratable residues directly interacting with the bilins.

The predicted pK_a values for the bilin pyrroles in the five cryptophyte complexes span a range between ~5 and 8 pK_a units. Thus, the protein environment and the propionic acid groups tune in slightly different ways the aqueous pK_a 's of bilins, which are in the range ~6–7 (Table 2). Interestingly, in most bilins the acidity of rings B and C is very similar, with differences less than 0.5 pK_a units. However, for doubly linked pigments DBV_{50/61B} and DBV_{50/61D} (PC577, PC612, PC630 and PC645 complexes) there is a clear trend indicating a pK_a smaller by ~1 unit for ring C compared to B, thus suggesting that deprotonation should only occur in pyrrole unit C, a preference that was already observed in the acidities of DBV' in water (Table 2). Note also that in the symmetric proteins PC577 and PC612, which are organized as $(\alpha\beta)_2$ homodimers, the symmetric pairs of chromophores located in the A/C and B/D polypeptide chains present very similar acidity constants, as one would expect. The impact of thermal fluctuations on the acidity constants is found to be small. Indeed, most pK_a values present standard deviations of ~0.1–0.2 units along the MD trajectories, and the estimates are close to those derived from the crystal structure of the complexes, with differences below 0.5 units in most cases (see Table S3). The only remarkable change is found for PEB_{158D} in PE545, in which the crystal structure leads to an unexpected pK_a of 5.2 compared to the MD-averaged value of 7.2, which is closer to the value of 7.5 found for the similar PEB_{158C} bilin. On the other hand, we find pK_a values in the range 2.5–5 for all bilin propionic acid

groups (Table S4), suggesting that they are unprotonated at physiological conditions.

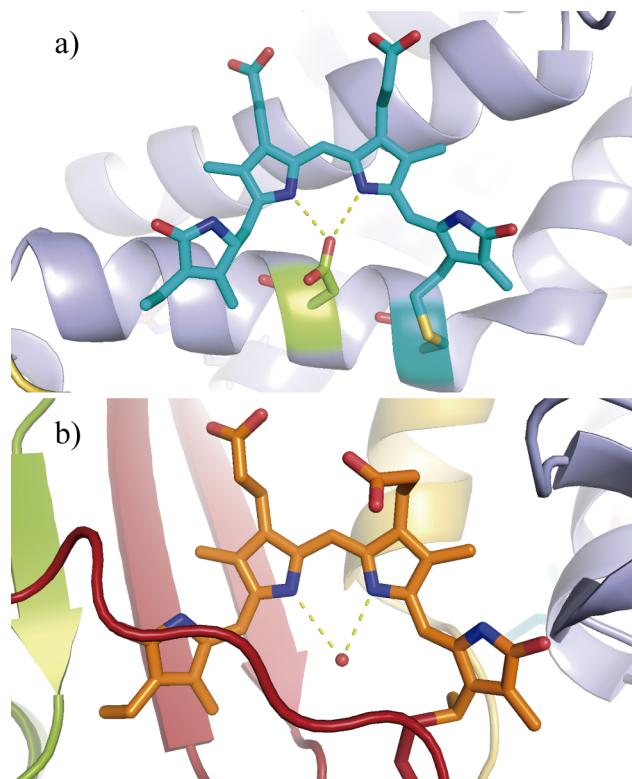


Figure 3. Representation of the environment surrounding bilin pigments. a) DBV_{50/61D} pigment in the PC645 complex with central pyrrole rings interacting with Asp54D. b) MBV_{19A} pigment in the PC645 complex with central pyrrole rings interacting with a crystallographic water.

By analyzing the details of the environment surrounding the bilin pigments on each complex, we can identify two different situations regardless of the nature of the chromophore that largely explain the trends in pK_a values. As it also happens in the C-PC complex of cyanobacteria,^{20–22} the central pyrrole rings of all PEB, PCB and DBV bilins located in the highly-conserved β polypeptide chains are coordinated with an Asp, which stabilizes the protonated form. A similar explanation applies to PCB₂₀ bilins in PC577 and PC612 located in the more divergent α subunits, which interact with a Glu. As reported in Table S5, our simulations indicate pK_a values in the range 3.3–6.9 for these Asp and Glu residues, in all cases lower than those estimated for the bilin pyrrole rings they interact with. In these situations, the bilin pK_a values are in the range ~6–8. In contrast, MBVs in PC630 and P645 and DBVs in PE545 lack the coordination with Asp or Glu, leading to pK_a values of ~5. In Fig. 3 we show the environment surrounding DBV_{50/61D} in PC645, with its central pyrrole rings interacting with Asp54D, and MBV_{19A}, in which the pyrroles are coordinated with a crystallographic water. It is also worth noting that in DBVs in PE545 and MBV19C in PC630 and PC645 there is a His residue interacting with the bilins, as shown in Fig. S1. In this case, a pK_a of ~4 is estimated for His16A/His16B in PE545, in agreement with the neutral state suggested by difference electron density maps of the complex.¹⁹ On the other hand,

ARTICLE

Journal Name

the pK_a of ~7.5 computed for His22A and His21A in PC630 and PC645 indicate a fully protonated form, consistent with its simultaneous interaction with a propionic group and a Glu observed in the crystal structure.

Table 3. MD-averaged acid dissociation constants computed using PROPKA for bilin pyrrole rings B and C in their specific protein environments (standard deviations in brackets).

	pK_a (B)	pK_a (C)
PE545 ($\alpha\beta\alpha\beta$)		
DBV _{19A}	5.6 (0.1)	5.7 (0.1)
DBV _{19B}	5.8 (0.1)	5.6 (0.1)
PEB _{50/61C}	7.3 (0.2)	7.1 (0.1)
PEB _{158C}	7.5 (0.1)	7.6 (0.1)
PEB _{82C}	6.7 (0.1)	6.9 (0.1)
PEB _{50/61D}	7.4 (0.1)	7.5 (0.1)
PEB _{158D}	7.2 (0.3)	7.2 (0.3)
PEB _{82D}	6.9 (0.1)	7.0 (0.2)
PC577 ($\alpha\beta\alpha\beta$)		
PCB _{20A}	8.0 (0.1)	7.8 (0.1)
DBV _{50/61B}	7.6 (0.1)	6.7 (0.2)
PCB _{158B}	7.9 (0.2)	8.2 (0.2)
PCB _{82B}	6.9 (0.1)	7.3 (0.2)
PCB _{20C}	7.9 (0.2)	7.8 (0.1)
DBV _{50/61D}	7.5 (0.2)	6.6 (0.3)
PCB _{158D}	7.9 (0.1)	8.2 (0.1)
PCB _{82D}	7.0 (0.1)	7.4 (0.1)
PC612 ($\alpha\beta\alpha\beta$)		
PCB _{20A}	8.1 (0.2)	7.9 (0.1)
DBV _{50/61B}	7.5 (0.3)	6.6 (0.4)
PCB _{158B}	7.9 (0.1)	8.1 (0.2)
PCB _{82B}	7.0 (0.1)	7.4 (0.1)
PCB _{20C}	8.1 (0.2)	8.0 (0.1)
DBV _{50/61D}	7.7 (0.1)	6.8 (0.2)
PCB _{158D}	7.9 (0.1)	8.1 (0.1)
PCB _{82D}	7.0 (0.1)	7.4 (0.1)
PC630 ($\alpha\beta\alpha\beta$)		
MBV _{19A}	5.6 (0.1)	5.0 (0.2)
DBV _{50/61B}	7.3 (0.1)	6.3 (0.4)
PCB _{158B}	7.9 (0.1)	8.3 (0.2)
PCB _{82B}	7.0 (0.1)	7.3 (0.1)
MBV _{19C}	5.4 (0.1)	4.9 (0.2)
DBV _{50/61D}	7.2 (0.2)	6.4 (0.1)
PCB _{158D}	7.6 (0.2)	7.9 (0.2)
PCB _{82D}	7.1 (0.1)	7.3 (0.1)
PC645 ($\alpha\beta\alpha\beta$)		
MBV _{19A}	5.9 (0.1)	5.3 (0.2)
DBV _{50/61B}	7.3 (0.2)	6.3 (0.5)
PCB _{158B}	7.8 (0.1)	8.0 (0.2)
PCB _{82B}	7.1 (0.1)	7.4 (0.1)
MBV _{19C}	5.5 (0.1)	5.0 (0.2)
DBV _{50/61D}	7.3 (0.2)	6.3 (0.1)
PCB _{158D}	7.6 (0.1)	8.0 (0.2)
PCB _{82D}	7.0 (0.1)	7.4 (0.1)

3.3 Absorption spectra

Because cryptophyte biliproteins are immersed in the lumen,¹⁶ where the pH varies in the range ~5-7 depending on the prolongation of the incident sunlight,¹⁷ light exposure could

potentially tune the light harvesting properties of these complexes given that protonation/deprotonation could be expected to occur in these conditions. However, as discussed in Sections 2.1 and 3.1, the accurate prediction of both absolute pK_a values in water and the corresponding shift upon transfer from water to a protein environment are still challenging, which may lead to uncertainties of ~1-2 pK_a units on the acidities reported in Table 3. The relative pK_a values for the pigments, however, are expected to be more accurate, with errors within 1 pK_a unit. In order to validate our simulations and identify potential systematic errors in the predicted acidity constants, we have recorded the absorption spectra of four of the complexes (PE545, PC577, PC630 and PC645) at pH values ranging from 4.0 to 9.4. Our aim is to identify variations in the absorption spectra upon varying conditions and correlate them with changes in the protonation pattern of the pigments inside the proteins. In addition, these measurements allow us to directly test whether pH changes in the range ~5-7 as those occurring in the lumen could impact the absorption spectra of the complexes and thus their photosynthetic function upon prolonged light exposure.

The pH -dependent absorption spectra measured for the different complexes are shown in Fig. 4, whereas Fig. 5 shows the difference spectra relative to the pH 6.5 curves in order to amplify the spectral changes. We additionally show the spectra obtained at pH 4.0 and 9.4, in which the trends observed are less regular, in Fig. S2 of the Supporting Information. Note that at these pH values protonation/deprotonation of protein residues and partial unfolding of the complexes may affect the absorption spectra, making it more difficult to correlate absorption changes to variations in the protonation pattern of the pigments.⁵⁵ Indeed, the absorption spectra of all the proteins varies at pH 4, matching the approximate pK_a of Asp and Glu side chains, but after the eventual deprotonation of these aminoacids all spectra remain mostly unchanged as pH is increased from 4.6 to 6.5. Thus, it seems unlikely that any change in protonation pattern occurs in this range of pH . In contrast, we find some appreciable changes in all absorption spectra in basic conditions as the pH increases until 9, especially for PC630 and PC645 complexes, but also to some extent in PE545 and PC577.

The changes observed in PC645 indicate that the relative intensity of the low energy band is reduced. This is consistent with the spectral changes predicted recently by Lee and co-workers arising from deprotonation of MBVs,¹⁵ which increases the energy of the $\pi \rightarrow \pi^*$ transition. As expected by the lack of a coordinating anionic residue, MBV pigments are predicted to be the least basic bilins in the complex. However, the predicted pK_a is ~5 for MBVs, whereas the spectral changes occur at $pH > 7$. Thus, the observed spectral changes suggest that our predictions are underestimating the pK_a of these pigments by at least 2 units, which we deem reasonable. Another possibility could be that MBVs are indeed already deprotonated in acidic conditions, as suggested by our predicted pK_a of ~5, so that the spectral changes might arise from an eventual deprotonation of the central DBV's in the complex, predicted to be slightly more basic than MBVs with a

$pK_a \sim 6.3$. This possibility, however, is inconsistent with the fact that the spectra remains almost unchanged in the 4.6 – 7.8 pH range, as this would suggest a larger difference in acidity among DBVs and MBVs. Moreover, the DBVs contribute to the high-energy band of the absorption,^{15,24,25} so we would expect the largest spectral changes to occur there instead of the low-energy band. Thus, all bilins, including MBVs lacking the anionic counterion that interacts with the central pyrrole rings, seem to be protonated at the pH range found in the lumen for PC645. These findings are in line with negligible changes in both spectra and energy transfer dynamics measured for PC645 using two-dimensional electronic spectroscopy at pH values 5.7, 6.5 and 7.4.⁵⁶

In the PC630 complex, we find a similar trend in spectral changes compared to PC645, with appreciable changes in the spectra occurring only at pH values > 7 . As for PC645, we expect that deprotonation of MBVs in the complex, rather than the central DBVs, is responsible for the changes observed, given that protonation of the DBVs and the consequent increase in transition energy should mostly impact the high-energy band of the spectra, whereas MBVs are expected to contribute to the low energy bands. In contrast to PC645, in PC630 the rise in pH decreases the relative intensities of both high and low-energy bands compared to the medium-energy region. In addition, there is a slight shift to shorter wavelengths both at the red and blue edges of the spectrum, leading to an increase in the overall broadening. Because the spectra in PC630 is more congested than for PC645, the shift in

MBV energies seems to have a larger impact on the total broadening of the absorption, thus causing an overall decrease of the intensity in the normalized spectra shown in Fig. 4, which affects both low and high-energy bands.

In contrast to PC630 and PC645, in PE545 and PC577 the absorption is found to be less sensitive to changes in acidity. Nevertheless, the trends are similar given that appreciable changes in the spectra are only found when the pH is increased over 7. This finding supports the idea that most bilins are fully protonated in acidic conditions, whereas some of them can deprotonate at pH values ~ 7 or larger. As reported in Table 3, the more acidic bilins in PE545 are the central DBV_{19A} and DBV_{19B} pigments, which lack an anionic amino acid counterion, with a pK_a 5.6. The variation found at basic conditions on the low energy band are in agreement with the expected impact of deprotonation of the DBVs, which mainly contribute to the red-edge.^{28,57} This interpretation again suggests that our pK_a predictions are underestimated by at least ~ 2 units. Otherwise, if most PEB pigments in PE545 would have a $pK_a \sim 7$ as reported in Table 3, we would expect more drastic changes in the spectra when the pH is increased from 4.6 to 9.0. As shown in Fig. S2, a drastic change is indeed observed at pH 9.4, which could be partially caused by deprotonation of most PEBs in the complex. Thus, our results support a fully protonated form of the DBV pigments lacking the anionic counterion at the pH ranges ~ 5 -7 found in the lumen, although some small fraction of these pigments could be present in their deprotonated form.

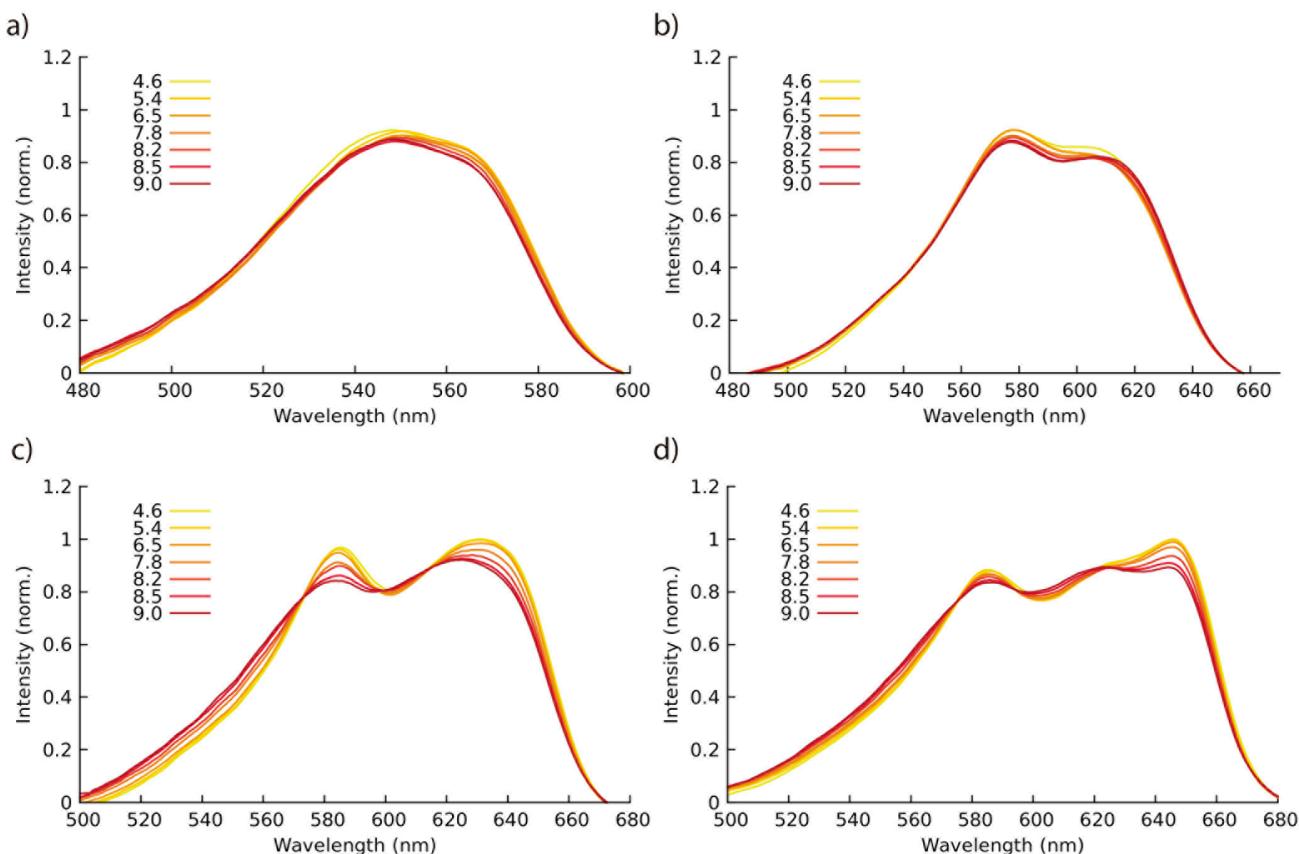


Figure 4. Normalized absorption spectra of cryptophyte antenna complexes at different pH values. a) PE545, b) PC577, c) PC630 and d) PC645.

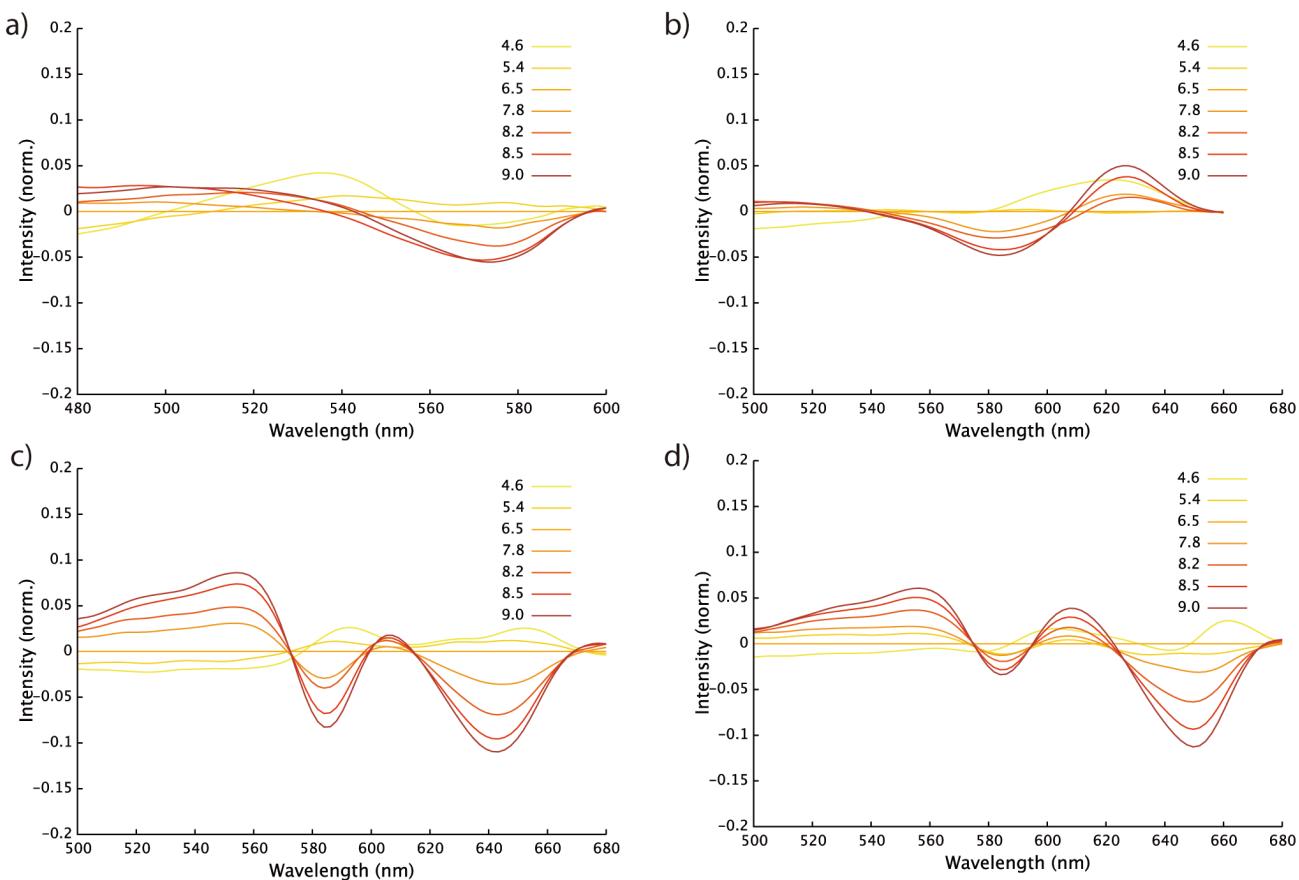


Figure 5. Normalized difference absorption spectra of cryptophyte antenna complexes at different *pH* values compared to the spectra at *pH* 6.5. a) PE545, b) PC577, c) PC630 and d) PC645.

On the other hand, in PC577, as in PC612, all pigments in the complex are coordinated either to Asp or Glu. In this case, the central DBV_{50/61B} and DBV_{50/61D} pigments are predicted to have the lowest *pK_a*, close to ~6.7, which may be somewhat underestimated as discussed for the other complexes. The changes observed in the PC577 spectra are again consistent with this interpretation, as we find almost no changes when the *pH* is increased from 4.6 to 6.5, and only a small decrease in the relative intensity of the maximum absorption band located at ~577 nm when the *pH* is further increased until 9.0. Because the central DBVs are expected to contribute to this high-energy band,⁵ these spectral changes are probably caused by a progressive deprotonation of the DBV pigments at basic conditions. Due to the high similarity between PC612 and PC577 in terms of structure, spectra and predicted *pK_a*, we expect the same situation to happen in PC612.

Conclusions

In this study we have addressed the protonation preferences of four types of bilin pigments (PEB, PCB, MBV and DBV) found in different antenna complexes from cryptophyte algae, including three "closed" (PE545, PC630 and PC645) and two "open" forms (PC577, PC612). We applied SCS-MP2 quantum chemical calculations extrapolated to the CBS limit, supplemented with CCSD corrections, as well as SMD and MST continuum solvation calculations to predict the intrinsic acidity

of model bilins (lacking the propionate side chains), leading to *pK_a* values of ~6-7 in aqueous solution. The results point out that deprotonation of bilins is expected to occur at any of the central B and C pyrrole rings, which are predicted to have similar acidities, given the amide character of the nitrogens in the outer rings (A and D). We then investigated the impact of the propionate side groups and of the specific protein environment, both as found in the crystal structure of the complexes, on the acidity constants using PROPKA combined with MD simulations, and critically assessed the predictions through comparison with the absorption spectra of the complexes measured in the *pH* range from 4.0 to 9.4. Our simulations predict a strong impact of the local environment in the protonation preferences of bilins, leading to both increases as well as decreases in their acidities. The results indicate approximate *pK_a* values ~7-8 for PCBs and ~6-7 for PEBs and DBVs in which the central pyrrole rings interact with an anionic Asp or Glu counterion, as also found for PCBs in the C-PC complex of cyanobacteria, whereas a *pK_a* ~5 is predicted for DBVs and MBVs lacking such counterion in PE545 and PC630/PC645 complexes, respectively. The analysis of the absorption spectral changes observed upon increasing *pH* values, however, suggest that the predictions are underestimated by ~2 *pK_a* units, given that appreciable variations in the spectra of the complexes are only observed at *pH* ~7 and found to be consistent with deprotonation of the DBV and MBV pigments lacking a coordinated anionic residue.

The present results thus suggest that most bilins in cryptophyte complexes are expected to be fully protonated when the *pH* varies on the range ~5–7 in the lumen, depending on the prolongation of light exposure, although some minor fraction of DBV and MBVs could also be deprotonated at *pH* ~7. Nevertheless, protonation/deprotonation is not expected to change the light harvesting pathways and dynamics in the complexes at physiological conditions. This information will be useful to derive improved models aimed at understanding the interplay between the protein environment and energy transfer dynamics in cryptophytes, including the observation of persistent vibronic coherences in "closed" forms of these antenna proteins.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

We are grateful to the Consorci de Serveis Universitaris de Catalunya for providing access to computational resources (Molecular Recognition project). Financial support from the Agència de Gestió d'Ajuts Universitaris i de Recerca from Generalitat de Catalunya (GENCAT; SGR2014-1189 and SGR2017-1746) and from the Spanish Ministerio de Economía y Competitividad (MINECO; grants CTQ2012-36195, CTQ2017-89924-P and SAF2017-88107-R) are acknowledged. C.C. is a Serra Húnter Fellow (Generalitat de Catalunya) and M.C. acknowledges a fellowship from the Spanish Ministerio de Economía y Competitividad (MINECO; grant BES-2013-064088). F.J.L. acknowledges the financial support from Icrea Academia. We are thankful to Paul M. G. Curmi for providing the crystal structures of the PC577 and PC630 complexes. Z.S.D.T. and G.D.S. would like to acknowledge the generous funding by Photosynthetic Antenna Research Center, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0001035.

References

- 1 Y.-C. Cheng and G. R. Fleming, *Annu. Rev. Phys. Chem.*, 2009, **60**, 241–262.
- 2 T. Mirkovic, E. E. Ostroumov, J. M. Anna, R. van Grondelle, Govindjee and G. D. Scholes, *Chem. Rev.*, 2017, **117**, 249–293.
- 3 E. Collini, C. Curutchet, T. Mirkovic and D. Gregory, in *Energy Transfer Dynamics in Biomaterial Systems*, eds. I. Burghardt, V. May, D. A. Micha and E. R. Bittner, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, vol. 93, pp. 3–34.
- 4 R. J. Ritchie, *Photosynth. Res.*, 2008, **96**, 201–215.
- 5 G. D. Scholes, T. Mirkovic, D. B. Turner, F. Fassioli and A. Buchleitner, *Energy Environ. Sci.*, 2012, **5**, 9374–9393.
- 6 C. Curutchet and B. Mennucci, *Chem. Rev.*, 2017, **117**, 294–343.
- 7 G. D. Scholes, G. R. Fleming, A. Olaya-Castro and R. van Grondelle, *Nat. Chem.*, 2011, **3**, 763–774.
- 8 E. Collini, C. Y. Wong, K. E. Wilk, P. M. G. Curmi, P. Brumer and G. D. Scholes, *Nature*, 2010, **463**, 644–647.
- 9 A. Chenu and G. D. Scholes, *Annu. Rev. Phys. Chem.*, 2015, **66**, 69–96.
- 10 J. C. Dean, T. Mirkovic, Z. S. D. Toa, D. G. Oblinsky and G. D. Scholes, *Chem.*, 2016, **1**, 858–872.
- 11 H.-G. Duan, V. I. Prokhorenko, R. J. Cogdell, K. Ashraf, A. L. Stevens, M. Thorwart and R. J. D. Miller, *Proc. Natl. Acad. Sci.*, 2017, **114**, 8493–8498.
- 12 B. S. Rolczynski, H. Zheng, V. P. Singh, P. Navotnaya, A. R. Ginzburg, J. R. Caram, K. Ashraf, A. T. Gardiner, S.-H. Yeh, S. Kais, R. J. Cogdell and G. S. Engel, *Chem.*, 2018, **4**, 138–149.
- 13 S. M. Blau, D. I. G. Bennett, C. Kreisbeck, G. D. Scholes and A. Aspuru-Guzik, *Proc. Natl. Acad. Sci.*, 2018, **115**, E3342–E3350.
- 14 S. J. Harrop, K. E. Wilk, R. Dinshaw, E. Collini, T. Mirkovic, C. Y. Teng, D. G. Oblinsky, B. R. Green, K. Hoef-Emden, R. G. Hiller, G. D. Scholes and P. M. G. Curmi, *Proc. Natl. Acad. Sci.*, 2014, **111**, E2666–E2675.
- 15 M. K. Lee, K. B. Bravaya and D. F. Coker, *J. Am. Chem. Soc.*, 2017, **139**, 7803–7814.
- 16 L. Spear-bernstein and K. R. Miller, *J. Phycol.*, 1989, **3**, 412–419.
- 17 B. Demming-Adams and W. W. Adams, *New Phytol.*, 2006, **172**, 11–21.
- 18 K. E. Wilk, S. J. Harrop, L. Jankova, D. Edler, G. Keenan, F. Sharples, R. G. Hiller and P. M. G. Curmi, *Proc. Natl. Acad. Sci.*, 1999, **96**, 8901–8906.
- 19 A. B. Doust, C. N. J. Marai, S. J. Harrop, K. E. Wilk, P. M. G. Curmi and G. D. Scholes, *J. Mol. Biol.*, 2004, **344**, 135–153.
- 20 H. Kikuchi, T. Sugimoto and M. Mimuro, *Chem. Phys. Lett.*, 1997, **274**, 460–465.
- 21 C. Kneip, P. Hildebrandt, F. Mark and K. Schaffner, *Chem. Phys. Lett.*, 1999, **311**, 479–484.
- 22 M. A. Mroginski, F. Mark, W. Thiel and P. Hildebrandt, *Biophys. J.*, 2007, **93**, 1885–1894.
- 23 F. Velazquez Escobar, C. Lang, A. Takiden, C. Schneider, J. Balke, J. Hughes, U. Alexiev, P. Hildebrandt and M. A. Mroginski, *J. Phys. Chem. B*, 2017, **121**, 47–57.
- 24 A. B. Doust, K. E. Wilk, P. M. G. Curmi and G. D. Scholes, *J. Photochem. Photobiol. A Chem.*, 2006, **184**, 1–17.
- 25 T. Mirkovic, A. B. Doust, J. Kim, K. E. Wilk, C. Curutchet, B. Mennucci, R. Cammi, P. M. G. Curmi and G. D. Scholes, *Photochem. Photobiol. Sci.*, 2007, **6**, 964–75.
- 26 C. Curutchet, J. Kongsted, A. Muñoz-Losa, H. Hosseini-Nejad, G. D. Scholes and B. Mennucci, *J. Am. Chem. Soc.*, 2011, **133**, 3078–3084.
- 27 H. Hosseini-Nejad, C. Curutchet, A. Kubica and G. D. Scholes, *J. Phys. Chem. B*, 2011, **115**, 5243–5253.
- 28 C. Curutchet, V. I. Novoderezhkin, J. Kongsted, A. Muñoz-Losa, R. van Grondelle, G. D. Scholes and B. Mennucci, *J. Phys. Chem. B*, 2013, **117**, 4263–4273.
- 29 L. Viani, C. Curutchet and B. Mennucci, *J. Phys. Chem. Lett.*, 2013, **4**, 372–377.

ARTICLE

Journal Name

- | | | | |
|----|---|----|--|
| 30 | L. Viani, M. Corbella, C. Curutchet, E. J. O'Reilly, A. Olaya-Castro and B. Mennucci, <i>Phys. Chem. Chem. Phys.</i> , 2014, 16 , 16302–16311. | 49 | D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, I. T.E. Cheatham, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P.A. Kollman, <i>AMBER 14</i> , University of California, San Francisco, 2014. |
| 31 | M. Aghtar, J. Strümpfer, C. Olbrich, K. Schulten and U. Kleinekathöfer, <i>J. Phys. Chem. Lett.</i> , 2014, 5 , 3131–3137. | | J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, <i>J. Chem. Theory Comput.</i> , 2015, 11 , 3696–3713. |
| 32 | M. Aghtar and U. Kleinekathöfer, <i>J. Lumin.</i> , 2016, 169 , 406–409. | | J. M. Wang, R. M. Wolf, J. W. Caldwell, P. a Kollman and D. a Case, <i>J. Comput. Chem.</i> , 2004, 25 , 1157–1174. |
| 33 | S. Chandrasekaran, K. R. Pothula and U. Kleinekathöfer, <i>J. Phys. Chem. B</i> , 2017, 121 , 3228–3236. | | C. C. Jumper, P. C. Arpin, D. B. Turner, S. D. McClure, S. Rafiq, J. C. Dean, J. A. Cina, P. A. Kovac, T. Mirkovic and G. D. Scholes, <i>J. Phys. Chem. Lett.</i> , 2016, 7 , 4722–4731. |
| 34 | K. S. Alongi and G. C. Shields, in <i>Annual Reports in Computational Chemistry</i> , Elsevier, 2010, pp. 113–138. | 50 | A. D. Bochevarov, M. A. Watson, J. R. Greenwood and D. M. Philipp, <i>J. Chem. Theory Comput.</i> , 2016, 12 , 6001–6019. |
| 35 | R. Casasnovas, J. Ortega-Castro, J. Frau, J. Donoso and F. Muñoz, <i>Int. J. Quantum Chem.</i> , 2014, 114 , 1350–1363. | 51 | P. M. Dewick, <i>Essentials of organic chemistry : for students of pharmacy, medicinal chemistry and biological chemistry</i> , J. Wiley, 2006. |
| 36 | M. Feliks and M. J. Field, <i>J. Chem. Inf. Model.</i> , 2015, 55 , 2288–2296. | 52 | A. J. Laos, J. C. Dean, Z. S. D. Toa, K. E. Wilk, G. D. Scholes, P. M. G. Curmi and P. Thordarson, <i>Angew. Chemie Int. Ed.</i> , 2017, 56 , 8384–8388. |
| 37 | M. N. Davies, C. P. Toseland, D. S. Moss and D. R. Flower, <i>BMC Biochem.</i> , 2006, 7 , 18–29. | | D. B. Turner, R. Dinshaw, K.-K. Lee, M. S. Belsley, K. E. Wilk, P. M. G. Curmi and G. D. Scholes, <i>Phys. Chem. Chem. Phys.</i> , 2012, 14 , 4857–4874. |
| 38 | C. L. Stanton and K. N. Houk, <i>J. Chem. Theory Comput.</i> , 2008, 4 , 951–966. | 53 | V. I. Novoderezhkin, A. B. Doust, C. Curutchet, G. D. Scholes and R. van Grondelle, <i>Biophys. J.</i> , 2010, 99 , 344–352. |
| 39 | E. Alexov, E. L. Mehler, N. Baker, A. M. Baptista, Y. Huang, F. Milletti, J. Erik Nielsen, D. Farrell, T. Carstensen, M. H. M. Olsson, J. K. Shen, J. Warwicker, S. Williams and J. M. Word, <i>Proteins Struct. Funct. Bioinforma.</i> , 2011, 79 , 3260–3275. | 54 | |
| 40 | M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, <i>J. Chem. Theory Comput.</i> , 2011, 7 , 525–537. | 55 | |
| 41 | C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski and J. H. Jensen, <i>J. Chem. Theory Comput.</i> , 2011, 7 , 2284–2295. | | |
| 42 | I. M. Alecu, J. Zheng, Y. Zhao and D. G. Truhlar, <i>J. Chem. Theory Comput.</i> , 2010, 6 , 2872–2887. | 57 | |
| 43 | S. Grimme, <i>J. Chem. Phys.</i> , 2003, 118 , 9095–9102. | | |
| 44 | A. Halkier, T. Helgaker, P. Jørgensen and W. Klopper, <i>Chem. Phys. Lett.</i> , 1998, 286 , 243–252. | | |
| 45 | I. Soteras, C. Curutchet, A. Bidon-Chanal, M. Orozco and F. J. Luque, <i>J. Mol. Struct. THEOCHEM</i> , 2005, 727 , 29–40. | | |
| 46 | C. Curutchet, A. Bidon-Chanal, I. Soteras, M. Orozco and F. J. Luque, <i>J. Phys. Chem. B</i> , 2005, 109 , 3565–3574. | | |
| 47 | A. V. Marenich, C. J. Cramer and D. G. Truhlar, <i>J. Phys. Chem. B</i> , 2009, 113 , 6378–6396. | | |
| 48 | M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V Ortiz, J. Cioslowski and D. J. Fox, <i>Gaussian 09, Revision D.1</i> , Gaussian Inc., Wallingford CT, 2009. | | |

**Spectral variability in phycocyanin cryptophyte antenna complexes is
controlled by changes in the α polypeptide chains**

(to be submitted)

Marina Corbella,¹ Lorenzo Cupellini,² Benedetta Mennucci,² Gregory D. Scholes³
and Carles Curutchet¹

¹ Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica i Institut de Química Teòrica i Computacional (IQT-CUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona, Espanya.

²Dipartimento di Chimica e Chimica Industriale, University of Pisa, Italy.

³Department of Chemistry, Princeton University, Princeton, New Jersey, United States.

3.1.3. Spectral variability in phycocyanin cryptophyte antenna complexes is controlled by changes in the α polypeptide chains

Resum

Les proteïnes antena dels diferents sistemes fotosintètics tenen un paper crucial, ja que són els encarregats de captar la llum i transportar-ne l'energia resultant als centres de reacció. Aquestes proteïnes antena presenten una diversitat molt amplia, tant en nombre com en l'estructura dels pigments que les constitueixen, per tal d'adaptar-se a la diferent disponibilitat de llum de l'habitat on es troben. L'interès en entendre l'eficàcia en el transport de l'energia en aquests complexes ha anat augmentant des de l'obtenció de les primeres estructures cristall, ja que les diferències en les longituds d'ona d'absorció de les diferents proteïnes antena sembla anar lligada, en part, a la naturalesa de l'estructura proteica. Tot i el gran nombre de treballs i estratègies desenvolupades per a la obtenció d'aquestes energies, avui en dia encara representen un repte considerable per les tècniques computacionals actuals degut a la gran dimensió dels sistemes. En aquest treball s'han estudiat quatre proteïnes antena presents en les algues del filum dels criptòfits, que es presenten en dues estructures quaternàries diferenciades, una estructura "oberta" (PC577 i PC612) i una "tancada" (PC630 i PC645), canvi que provoca la separació del dímer format per els pigments centrals i que es tradueix en una reducció en l'acoblament electrònic entre aquests i en una atenuació de les coherències observades en els sistemes "tancats". Per tal d'obtenir el Hamiltonià electrònic per descriure el procés de EET en aquests quatre sistemes, s'han calculat les energies d'excitació dels diferents pigments i els acoblaments electrònics mitjançant dues estratègies diferents. Per un cantó, una estratègia esta basada en simulacions de MD per tal de tenir en compte els efectes tèrmics de les diferents estructures, sobre les quals s'han realitzat a posteriori càlculs QM/MMPol sobre les estructures extretes de la dinàmica, mentre que la segona estratègia està basada en una pre-optimització de la geometria dels diferents pigments a nivell QM/MM, per tal de realitzar posteriorment un únic càlcul multiescala QM/MMPol/ddCOSMO de les energies d'excitació i dels diferents acoblaments electrònics. Tot i que ambdues estratègies presenten pros i contres, els resultats obtinguts semblen indicar que l'ús d'un *force field* clàssic en el mètode basat en les MD introduceix errors

significatius en les energies relatives entre els diferents tipus de pigments. D'altra banda, el mètode basat en les geometries optimitzades QM sembla predir energies relatives molt més raonables, tot i que en neglidir els efectes tèrmics el rang d'energies descrites resulta exagerat en comparació amb els espectres experimentals.

Spectral variability in phycocyanin cryptophyte antenna complexes is controlled by changes in the α polypeptide chains

Marina Corbella,¹ Lorenzo Cupellini,² Benedetta Mennucci,² Gregory D. Scholes³ and Carles Curutchet¹

¹Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica and Institut de Química Teòrica i Computacional (IQTC-UB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona Spain

²Dipartimento di Chimica e Chimica Industriale, University of Pisa, Italy.

³Department of Chemistry, Princeton University, Washington Road, Princeton, New Jersey 08544, United States

Abstract

Quantitative models of light harvesting in photosynthetic pigment-protein antenna complexes depend sensitively on the relative energetic ordering of the site energies of the pigments. The accurate determination of relative site energies is however still a considerable challenge for structure-based simulations. Here we analyze the light harvesting properties of four antennae from cryptophyte algae, phycocyanines PC577, PC612, PC630 and PC645. We compare two alternative theoretical strategies to derive the excitonic Hamiltonian. The first is based on molecular dynamics (MD) simulations and subsequent polarizable quantum/molecular mechanics (QM/MMPol) calculations of the relevant excitonic parameters (site energies and electronic couplings). The second more cost-effective approach is based on three-layer QM/MMPol/ddCOSMO calculations performed on geometries of the pigments optimized at the quantum-chemical level, where the water solvent is described using the ddCOSMO continuum model. Our results point to significant systematic errors between different types of bilins induced by the underlying classical force field adopted in the classical MD simulations, whereas the strategy based on QM-based geometries still underestimates the relative energies of MBVs due to the neglect of thermal effects. Our results strongly suggest that these four phycobiliproteins share a common energetic ordering $PCB_{82} < PCB_{158} < DBV_{51/61}$ for pigments located in the highly-conserved β chain. The spectral variability among PBPs thus seems to arise from the energy tuning of the bilins in the more divergent α chain, which in PC577 and PC612 are expected to contribute to the middle-energy region of the spectra, whereas in PC630 and PC645 contribute to the low-energy one. In addition, we predict a strong screening of the coupling among central DBVs in “open” form complexes PC577 and PC612 compared to “closed” form ones, which together with the increased interpigment separation explains the attenuation of coherence beatings observed for these complexes.

1. Introduction

Pigment-protein antenna complexes play an important role in the photosynthetic machinery by collecting sunlight and transporting it to reaction centers (RC), where the absorbed energy is used to drive charge separation events.¹ Whereas reaction center complexes are quite conserved, antenna complexes encompass a rich variety of structures and pigment compositions adapted to the needs of the photosynthetic organisms depending on their particular habitats. The remarkable quantum efficiency of the electronic energy transfer (EET) process that drives the energy to the RCs has inspired a continued effort aimed at understanding the subtle details that relate the EET process with the structure of the underlying biomolecule.²⁻⁴ However, since the first high-resolution crystal structure of an antenna complex, the Fenna-Matthews-Olson (FMO) complex, appeared 40 years ago,⁵ a particularly challenging aspect has been the precise determination of the site energies – the uncoupled excitation energies of the pigments.⁶ A rather successful strategy relies on the simultaneous fitting of a variety of optical spectra.⁷ However, those fittings are not free from ambiguities, even in antenna complexes with a relatively low number of sites.⁸ In addition, they do not allow drawing a relation between the light harvesting properties of the complex and its underlying structure, a desirable insight in order to establish structural blueprints for the design of artificial light harvesting systems. An attractive alternative to empirical models consists in the theoretical calculation of site energies from the structure of the complex, but this has proven to be a remarkable challenge, because the errors associated to present computational methods are similar to the relative site energy differences.

In the last decade a variety of groups have attempted the calculation of site energies in variety of photosynthetic systems, with variable degrees of success.²⁻⁴ The differences among these computational approaches are mainly related to the quality of description used to model the pigments and the surrounding protein and solvent environment, and thus their interactions. In addition, these strategies also differ on whether the solvatochromic shift arises purely from i) pigment-protein interactions, ii) deformation of the pigment internal geometries in the protein scaffold, or iii) both effects. Simulations of site energies on chlorophyll-containing antennae have been quite successful by limiting the mechanism of site-energy tuning to direct electrostatic pigment-protein interactions,⁴ even if contradicting reports indicate an important role played by the chlorophyll macrocycle deformation.⁹ In contrast, in phycobiliproteins (PBP), which contain quite flexible linear tetrapyrrole pigments called bilins, site energy tuning exerted by the protein has been previously adscribed mainly to the constrained conformation of the bilins in the protein pocket.¹⁰ Accounting for internal deformation effects is however hard to estimate in a reliable way, which partially explains why neglecting this effect has led to good results in chlorophyll-based systems. Indeed, estimating this effect from the pigment geometries in a crystal structure is in general not reliable due to the limited resolution of the crystal.¹¹

Accurate geometries with enough quality to carry out excited-state calculations can thus be derived, in principle, by using quantum-chemical (QM) methods, often coupled to a classical molecular mechanics (MM) description of the environment in so-called hybrid QM/MM approaches.¹² However, proteins are characterized by complex energy landscapes, so thermal effects can have an important impact on the results. In other words, performing calculations on a single structure can be a limited description of the properties of a thermally disordered biological system.

Here, we investigate the performance of alternative theoretical strategies to determine the light harvesting properties of photosynthetic complexes by studying four PBPs from cryptophyte algae, the phycocyanins PC577, PC612, PC630 and PC645.¹³ These systems are well-suited for this task because both the PC577/PC612 and PC630/PC645 pairs share the same pigment composition, therefore its changes in spectral properties arises from the changes in the underlying protein sequences. In particular, PC577 and PC612 contain six phycocyanobilins (PCB) and two 15,16-dihydrobiliverdins (DBV), whereas PC630 and PC645 contain 4 PCBs, 2 DBVs, and two mesobiliverdin (MBV). Moreover, single-residue insertion switches the “closed” quaternary structure from PC630/PC645 to an “open” structure for PC577/PC612,¹⁴ as shown in Fig. 1, leading to a reduction of the exciton coupling of the central DBV-DBV pair that explains the attenuation of coherence beatings observed for these PBPs.

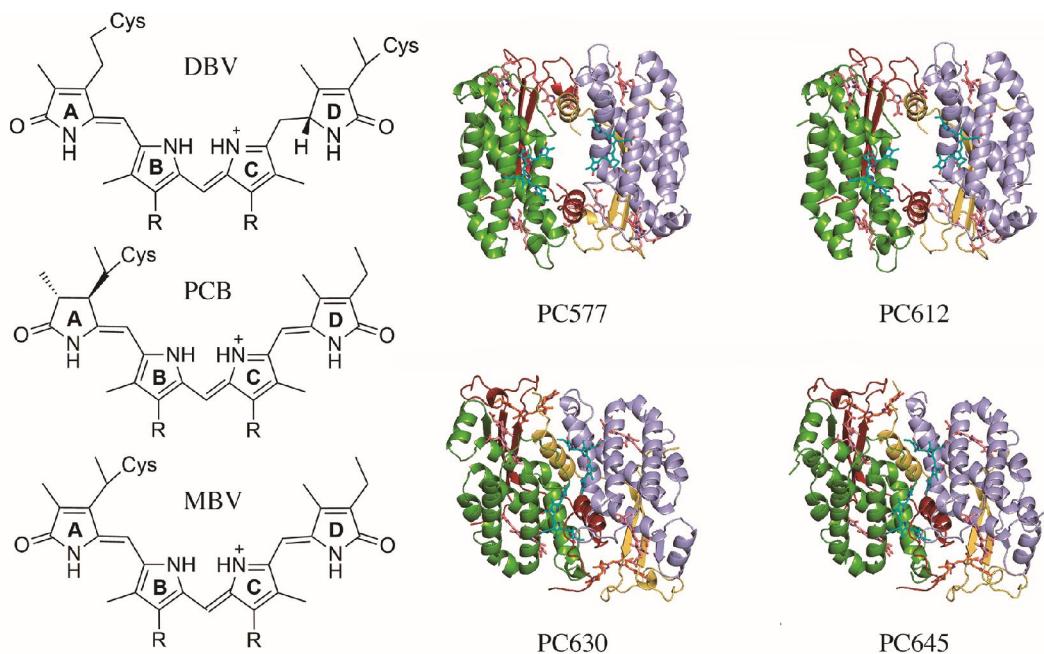


Figure 1. Structure of the cryptophyte antenna complexes PC577, PC612, PC630 and PC645 and the bilin pigments contained.

We compare two different theoretical strategies in order to derive the excitonic Hamiltonian of these systems, which mainly differ in the way the effect due to the internal geometries of the

pigments is handled. Our first strategy is based on classical molecular dynamics (MD) simulations of the complexes,¹⁵ and a post-processing of the sampled structures using polarizable QM/MMPol calculations¹⁶ of the relevant excited states and corresponding electronic coupling values. This strategy was shown to provide a good description of the properties of the similar phycoerythrin PE545 complex, although it required a different shift in order to correct for systematic errors in the site energy predictions for the phycoerythobilin (PEB) and 15,16-dihydrobiliverdin (DBV) pigments in the complex, probably owing to the limitations of the QM method used and the underlying force field adopted in the MD.^{10,16} A similar MD-QMMM strategy has been used by the Kleinekathöfer group to study the properties of the PE545 and PE555 antennae^{17,18} We then explore a cost-effective strategy based on QM/MM geometry optimization of the pigments in the protein scaffold, followed by multiscale three-layer QM/MMPol/ddCOSMO calculations of the excited states and electronic coupling values, in which the surrounding solvent, absent in the crystal, is modeled through the novel linear-scaling domain decomposition solution of the COSMO equations (ddCOSMO).¹⁹ Recently, the Coker group has shown that combining an intermediate strategy based on multiple geometry optimizations of the pigments along structures sampled from MD can provide a reliable description of the spectra in PE545 and PC645.²⁰ Our results show that the underlying force field used in the classical MDs introduce significant errors in the site energies, and significant systematic errors between DBVs and the other PCB and MBV pigments are found from these simulations. On the other hand, geometry optimization provides more reliable relative energies between PCB, MBV and DBV pigment types, but the neglect of thermal effects leads to an underestimation of the energies in the MBV pigments in PC630 and PC645. Finally, we compute the spectral densities from the Vertical Gradient (VG) approach,²⁰ based on a normal mode analysis of the pigment vibrations, which lead to excellent emission lineshapes, although for a few pigments fails to provide such accurate description. Finally, we compare electronic couplings and solvent screening factors computed from the MD-QM/MMPol and the QM/MMPol/ddCOSMO methods. Our results indicate a drastic screening of the coupling in the central pair of DBVs in “open” form PBPs, which together with the increased DBV-DBV separation explain the attenuation of coherence beatings observed for these PBPs. Moreover, our results suggest that these PBPs share a common ordering $\text{PCB}_{82} < \text{PCB}_{158} < \text{DBV}_{51/61}$ of the site energies for pigments located in the highly-conserved β chain. The spectral variability among PBPs thus seems to arise from the energetic location of the bilins in the more divergent α chain, which in PC577 and PC612 are expected to contribute to the middle-energy region of the spectra (PCB_{20A} and PCB_{20C}), whereas in PC630 and PC645 they contribute to the low-energy one (MBV_{19A} and MBV_{19C}).

2. Methods

2.1 MD simulations and geometry optimizations

The structural models and the details of the MD simulations are reported in Ref. 21. These are based on crystal structures solved for PC645 (PDB 4LMS, resolution 1.35 Å),¹⁴ PC612 (PDB 4LM6, resolution 1.70 Å),¹⁴ PC577 (resolution 1.00 Å) and PC630 (resolution 1.60 Å). PC577 and PC612 are organized as $(\alpha\beta)_2$ homodimers and contain 6 PCBs and 2 DBVs, whereas PC630 and PC645 are organized as $\alpha_1\beta\alpha_2\beta$ heterodimers and contain 2 MBVs, 4 PCBs and 2 DBVs. Note that DBVs, in contrast to the other singly-linked pigments, are linked through two Cys residues to the protein. Based on the analysis in Ref. 21, all amino acids were considered in their standard protonation state except His22A and Glu65C in PC630 and His21A and Glu65C in PC645, which were considered protonated, whereas bilin pigments were modeled with a fully protonated tetrapyrrole backbone and anionic propionic side chains. 10-ns MD simulations reported in Ref. 21 were extended to a total of 200 ns, and we extracted a total of 100 snapshots at regular time intervals from the last 50 ns to be used as input structures in QM/MMpol calculations. All MD runs were performed with the Amber 14 suite of programs.²²

Geometry optimizations of the bilins in their native protein environment were performed using the ONIOM method with electrostatic embedding,²³ using B3LYP/6-31G(d) and a classical Amber description for the QM and MM regions, respectively, as implemented in the Gaussian code.²⁴ The MM region was described using the same force field adopted in the MD. We first perform optimizations, starting from the crystal structures, including the full bilin pigments and selected amino acid interacting with them in the QM region, which was fully relaxed keeping the MM region frozen. In particular, for PCBs and DBVs we included the Asp or Glu side chains coordinating the central rings, whereas for MBVs we included selected amino acids (Lys22A and Lys23A for MBV_{19A} and His22C and Glu26C for MBV_{19C} in PC630; Asn22A for MBV_{19A} and His21C and Glu25C for MBV_{19C} in PC645). QM/MM boundaries were defined at residue-residue and Cys-bilin bonds using the link atom scheme.²⁵ We then further optimized the bilin structures limiting the QM region only to the the bilin tetrapyrrole backbone, and thus keeping all amino acids and the propionic side chains frozen in the MM region (capping the pyrroles with methyl groups) in order to limit the NMA vibrational analysis to the tetrapyrrole backbone and in order to prevent contamination of excited-state calculations from charge-transfer effects.

2.2 QM/MMPol and QM/MMPol/ddCOSMO calculations

Excited-state energies and electronic couplings were computed using the multiscale two-layer QM/MMPol²⁶ and the three-layer QM/MMPol/ddCOSMO¹⁹ models. In both models, a QM

description of the chromophores is complemented with a polarizable MM description of the environment based on point charges and induced dipoles. In the QM/MMPol/ddCOSMO, however, the solvent is described as a continuum medium characterized by its macroscopic dielectric constant, following the recently developed linear-scaling domain decomposition algorithm for COSMO.^{27–29} Both models account for full mutual polarization effects among the two or three layers. The QM/MMPol model is based on the following effective Hamiltonian:

$$\hat{H}_{eff} = \hat{H}_0 + \hat{H}_{QM/MM}^{el} + \hat{H}_{QM/MM}^{pol} + \hat{H}_{MM}^{el} + \hat{H}_{MM}^{pol} \quad (1)$$

which includes Hamiltonian of the isolated QM system (\hat{H}_0), electrostatic ($\hat{H}_{QM/MM}^{el}$) and polarization ($\hat{H}_{QM/MM}^{pol}$) QM/MM interaction energy terms, and the electrostatic self-energy of the MM charges (\hat{H}_{MM}^{el}) as well as the MM polarization energy (\hat{H}_{MM}^{pol}).

The QM/MMPol/ddCOSMO model extends the effective Hamiltonian in Eq. (1) adding further QM/ddCOSMO ($\hat{H}_{QM/ddCOSMO}$) and MM/ddCOSMO ($\hat{H}_{MM/ddCOSMO}$) interaction terms:

$$\hat{H}_{eff} = \hat{H}_0 + \hat{H}_{QM/MM}^{el} + \hat{H}_{QM/MM}^{pol} + \hat{H}_{MM}^{el} + \hat{H}_{MM}^{pol} + \hat{H}_{QM/ddCOSMO} + \hat{H}_{MM/ddCOSMO} \quad (2)$$

We use the extension of both approaches to linear response time-dependent density functional theory (TD-DFT) to perform excited-state calculations. The corresponding transition densities are then used to compute electronic couplings as a sum of Coulomb and environment-mediated terms.^{2,30}

$$V = V_{Coul} + V_{env} \quad (3)$$

The V_{env} , which in the present formulation includes contributions mediated by both the MMPol and the ddCOSMO environment, typically leads to an overall attenuation of the coupling, and an effective screening factor can be defined as:

$$s = \frac{V_{Coul} + V_{env}}{V_{Coul}} \quad (4)$$

Both QM/MMPol and QM/MMPol/ddCOSMO calculations were performed based on the optimized geometries of the pigments described in the previous section at the TD-DFT CAM-B3LYP/6-31G(d) level of theory.³¹ QM/MM boundaries in the bilin-S bonds were treated using the link atom scheme.²⁵ The MMPol region was described using the Amber pol12 AL parameters.^{32,33} Atomic charges for water were computed from RESP fits at the MP2/aug-cc-pVTZ level of theory on a TIP3P geometry, whereas polarization-consistent ESP charges for bilins were derived at the B3LYP/aug-cc-pVTZ level, based on the crystal geometries, using the Polchat tool.³⁴ Explicit polarization in the MM region was only included for atoms within a cutoff radius of 18 Å from the heavy atoms of the QM region. All calculations were performed using a locally modified development version of the Gaussian package.²⁴ In this work, the

magnitude of the transition dipole moments predicted for the different bilins was in the range ~12-15 D (see Tables S3-S6 in the Supporting Information). This is ~10-30% larger than an experimental estimate of 11.24 D,³⁵ so we did not rescale the electronic couplings to match the experimental value.

2.3 Spectral densities and modeling of steady-state spectra

Spectral densities of electronic-vibrational coupling were obtained with the Vertical Gradient (VG) approach.³⁶ This method assumes ground and excited-state potential energy surfaces (PES) described by the same harmonic oscillator but with displaced equilibrium position, and allows extracting the reorganization energy from the difference between the vertical and adiabatic excitation energies. Normal modes were computed based on the optimized geometries described in Section 2.1. Then, vertical gradients were calculated from the vertical excitation energies obtained at the same level of theory. Huang-Rhys factors S_k of each mode k were obtained by projecting the gradient of the excited state $\tilde{\mathbf{f}}$ on the ground-state normal mode k :

$$S_k = \frac{\tilde{f}_k^2}{2\hbar\omega_k^3} \quad (5)$$

thus obtaining the following expression for the spectral density:

$$\tilde{C}''(\omega) = \pi\omega \sum_{k=0}^M S_k \omega_k \quad (6)$$

where ω_k is the frequency of the mode k .

This model allows estimating the intramolecular part of the spectral density. The continuous part due to slow environmental motions was added *a posteriori* using an overdamped brownian oscillator, defined by λ and τ . While τ was fixed to a value of 50 fs, λ was fitted for each protein in order to match the experimental Stokes shift between absorption and emission.

Simulations of absorption (OD), circular dichroism (CD) and fluorescence (FL) spectra were performed using the site energies, electronic couplings and spectral densities, as reported in Ref.10. Static disorder was modeled by averaging realizations of the spectra over a random distribution of the site energies characterized by a given standard deviation σ , which was adjusted in order to reproduce the broadening of the experimental emission lineshapes for each complex. Lifetime broadening was accounted for based on transfer rates computed using the modified Redfield tensor.³⁷ We used the following excitonic Hamiltonian describing the multichromophoric system:

$$\hat{H} = \sum_{n=1}^N \varepsilon_n |n\rangle\langle n| + \sum_{n \neq m}^N V_{nm} |n\rangle\langle m| \quad (7)$$

where N is the number of interacting chromophores, ε_n is the excitation energy of the n^{th} chromophore and V_{nm} is the coupling between the n^{th} and m^{th} chromophores. Exciton states $|k\rangle$ were then obtained from diagonalization of the excitonic Hamiltonian:

$$\hat{H} = \sum_{k=1}^N \epsilon_k |k\rangle\langle k|; |k\rangle = \sum_{n=1}^N c_n^k |n\rangle$$

where ϵ_k is the energy of the k^{th} exciton state and c_n^k describes the participation of the n^{th} chromophore to the k^{th} exciton state.

3. Results and discussion

3.1 Spectral densities

In Figure 2 we show the spectral densities (SD) computed using the VG method for the four PBPs considered, whereas in Tables S1 and S2 of the Supporting Information we report the corresponding reorganization energies. On the other hand, in Figure 3, we show the total accumulated reorganization energies as a function of vibrational frequency. In the development of quantitative models of light harvesting, it is common to assume equal SDs for all pigments in a given complex. This can be a stronger approximation in systems with different kind of pigments, whose coupling to the environment can be different, as is the case of PBPs.³⁸ Our results clearly indicate a larger coupling of DBVs to the vibrations compared to PCBs, whereas for MBVs the results are less clear. This could be related to the fact that DBVs are covalently linked to the protein backbone through two Cys residues instead of one. Our simulations however provide some clear outliers on the general trends observed in the four complexes, characterized by total reorganization energies in the range $\lambda \sim 410\text{-}690 \text{ cm}^{-1}$: the two DBVs in PC612 and the MBV_{19A} in PC630, which display values $\lambda \sim 900\text{-}1300 \text{ cm}^{-1}$ (see Table 1 and 2). Also the value for MBV_{19C} in PC645 is somewhat low, with a $\lambda = 350 \text{ cm}^{-1}$. This suggests that in a few particular conformations of the pigments, as optimized in the protein scaffold, the vertical gradients computed suffer significant errors. Indeed, our calculation on the DBV_{50/61B} pigment in PC630 failed to provide reasonable results.

In Figure 3, one can observe that such outliers arise mostly from inaccuracies in the contribution to λ of low frequency vibrations, probably contaminated by the constrained optimization in the protein scaffold. For DBV_{50/61B} and MBV_{19A}, however unusual steps in λ are also observed at frequencies $\sim 1500\text{-}1600 \text{ cm}^{-1}$, the region characterized by C=C stretchings. Nevertheless, as we show in the next section, in general the computed SDs lead to simulated emission lineshapes in excellent agreement with experiment, once we use the SD of MBV_{19C} for the MBV_{19A} outlier pointed out above. Note that in the spectral simulations of PC612 we used the SD values for DBVs computed for PC577, but this does not impact the emission spectra of this complex as the DBVs have the higher energies in the complex.

Thus, although the calculated SDs cannot account for the low-frequency part due to the environmental motions, they provide an efficient and reliable method to estimate vibronic coupling compared to much more costly MD-QMMM approaches, which usually lead to a strong overestimation of high-frequency peaks when the MD is based on an approximate classical force field.^{15,17,38–47} A more reliable description consists in performing the MD at the QM/MM level instead of a purely MM potential,^{48,49} but then the computational cost associated to the calculation of SD is drastically increased.

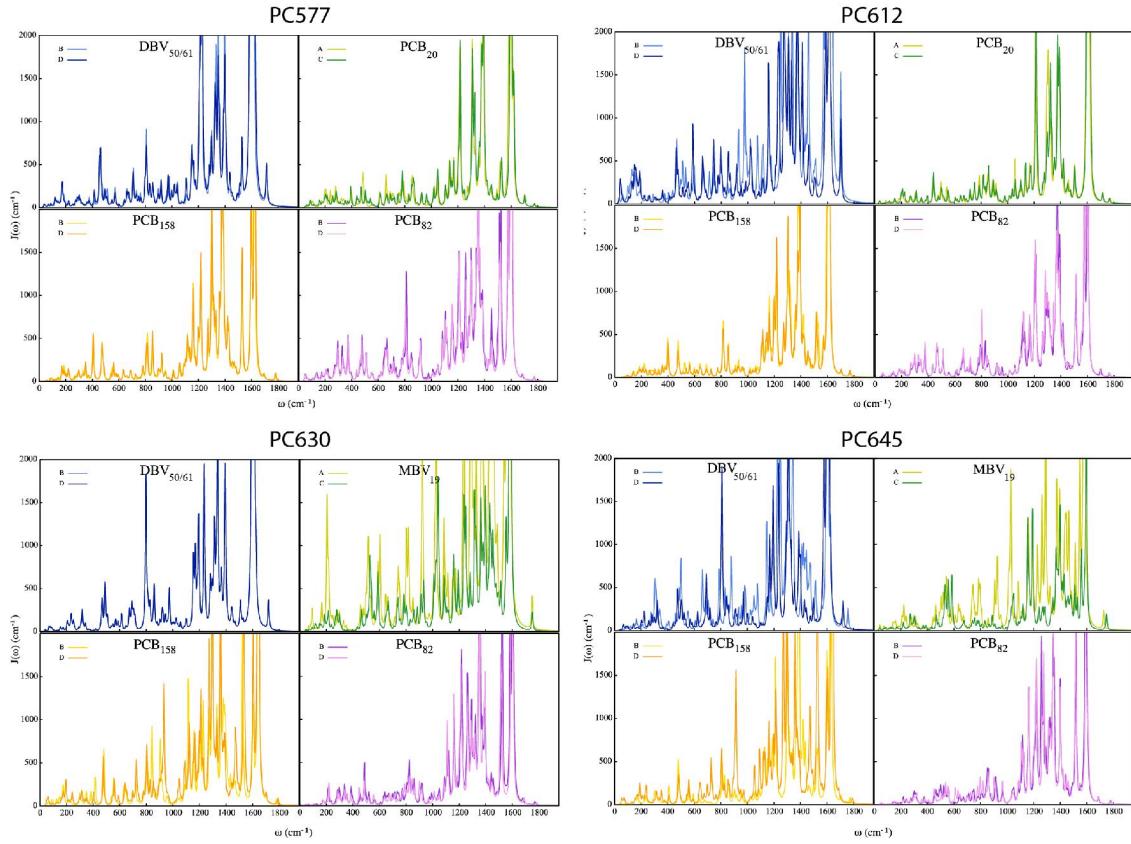


Figure 2. Spectral densities computed for bilin pigments in the PC577, PC612, PC630 and PC645 antenna complexes

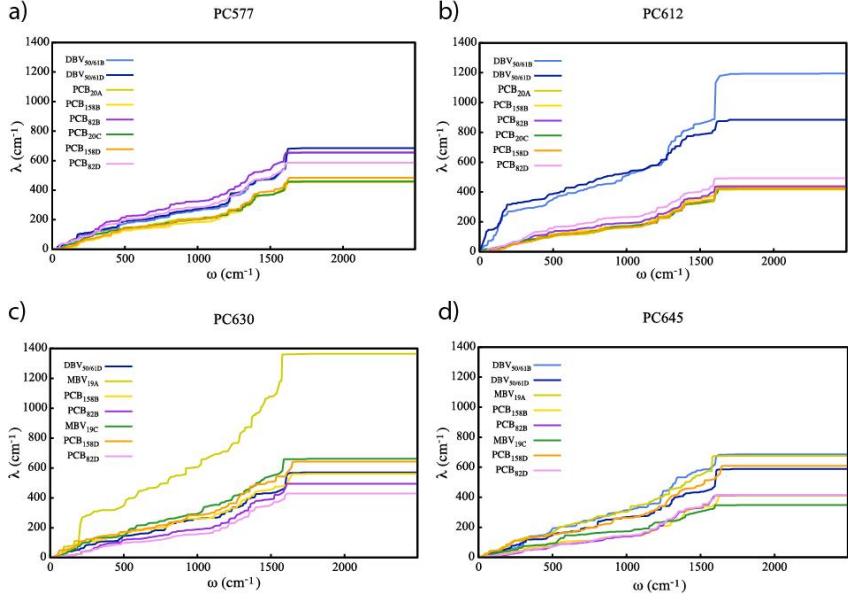


Figure 3. Total accumulated reorganization energies as a function of vibrational frequency estimated from the spectral densities computed for bilin pigments in the a) PC577, b) PC612, c) PC630 and d) PC645 antenna complexes.

3.2 Site energies

In Figure 4 we report the relative site energies for the bilins in the PBP complexes computed using MD-averaged QM/MMPol values compared with QM/MMPol/ddCOSMO calculations performed on the optimized geometries of the pigments in the protein scaffold. If we focus on PC577 and PC612, which only contain PCB and DBV chromophores, the results show a clear overestimation of the energy difference between these two kind of molecules when the calculations are based on structures extracted from the MD, whereas the use of QM optimized geometries leads to much more reasonable relative site energies. Similar systematic deviations between DBV and PEB pigments were observed previously when applying the same approach to the PE545 PBP complex.^{16,10} It seems clear that this problem arises from the use of an approximate classical force field in the MD, which is expected to lead to different systematic errors for different kind of pigments. For PC630 and PC645, this issue is further complicated by the presence of MBV bilins. Regarding DBVs, again the MD-based values are largely overestimated compared to the PCB energies as found for PC577 and PC612. On the other hand, MBVs have the largest degree of π conjugation along the tetrapyrrole backbone, and could thus be expected to be the lowest-lying pigments in these complexes. The use of QM optimized geometries indeed leads to such result, whereas MD-averaged values lead to MBV energies in ranges similar to those of PCBs, which could be seemingly ascribed again to systematic errors arising from the use of MD-derived geometries.

Overall, thus, the MD QM/MMPol protocol seems to be rather problematic in order to derive energy differences for pigments with diverse underlying chemical structure, whereas the use of QM-optimized geometries looks more reliable. This latter approach, however, suffers from the neglect of thermal effects. Indeed, whereas the relative energies between DBVs and PCBs computed in this way look reasonable, those of the MBVs are clearly underestimated in light of the experimental absorption spectra. Previous calculations based on the crystal structure of PC645, however, suggested that MBVs lie in between the energies of DBVs and PCBs,⁵⁰ and QM/MMPol calculations performed on crystal geometries lead to similar results (data not shown). Thus, it seems that thermal effects can be particularly important to properly describe the MBV properties because of its largest degree of π conjugation among all bilin types. In other words, thermal fluctuations are expected to distort the π conjugation that is otherwise exaggerated based on the optimized geometry. Indeed, recent calculations presented by the Coker group showed similar energies for MBVs and PCBs when QM-optimizations were done along different conformations of the protein complex sampled along an MD.²⁰ Thus, whereas it seems clear that DBVs populate the high-energy part the absorption spectra of these complexes, it is difficult from the present simulations to determine the precise relative positions of MBV and PCB bands.

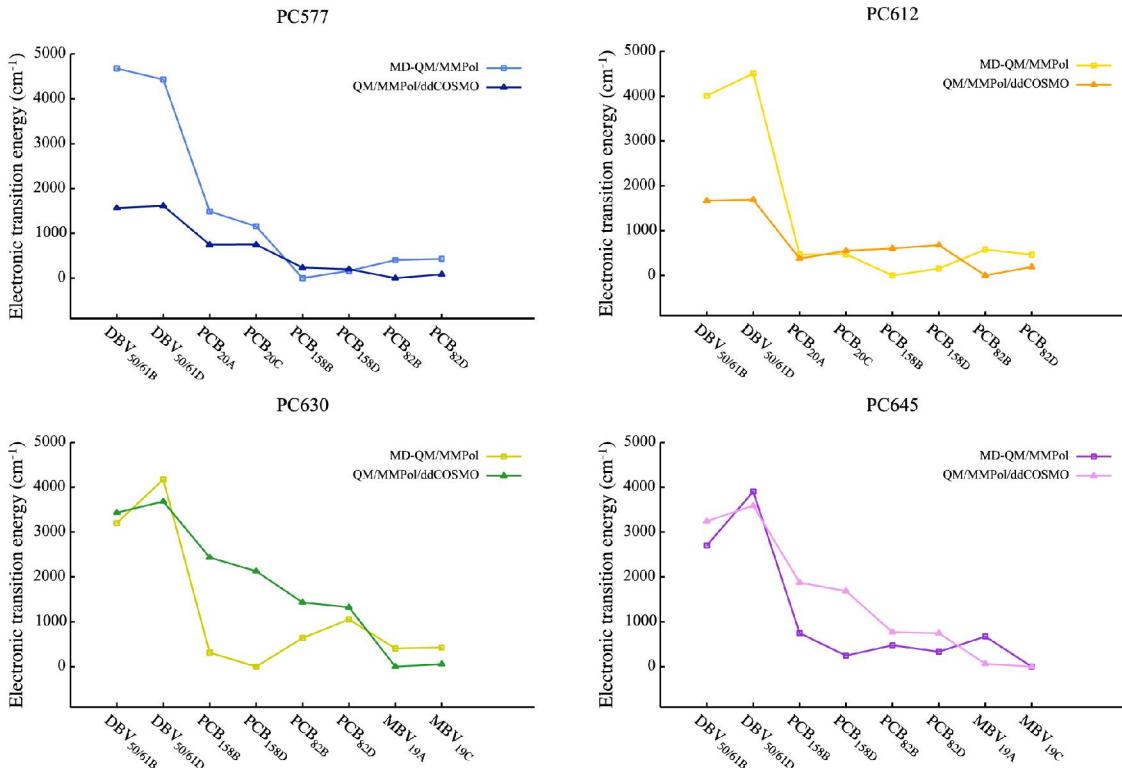


Figure 4. Site energies of the pigments in PC577, PC612, PC630 and PC645 computed using the MD-QM/MMPol and QM/MM/MMPol methods relative to the minimum energy site in the complex.

In Figures 5 and 6 we show the absorption (OD), fluorescence (FL) and circular dichroism (CD) spectra of the PBP complexes simulated using both theoretical approaches compared to experiments.^{14,51} As previously discussed, the OD spectra in Figure 5, based on MD-averaged site energies and couplings, clearly show the overestimated energies for DBVs. In addition, whereas the relative energies of the PCBs in PC577 look very reasonable, those for the other PBPs seem to span a range of energies that is too small, leading to a spectrum that is too narrow. On the other hand, the FL lineshapes are nicely reproduced by the SDs discussed in the previous section, as emission is not much affected by the relative site energies given that only populated lowest-lying states contribute to it. The main features of the CD spectra are also reproduced.

We remind here that PC577 and PC612, as well as PC630 and PC645, share the same pigment composition, thus spectral changes arise from changes in the amino acid sequence of the protein complex. In this light, the simulations in Figure 5 are in qualitative agreement with the displacement of intensity from the high-energy band to the low-energy one when passing from PC577 to PC612, which seems to arise from a lowering of the energies of PCB_{20A} and PCB_{20C}, located in the α polypeptide chains, whose structure is much more divergent compared to that of the highly-conserved β chains, identical in these complexes. A similar qualitative agreement is found if we compare the spectra of PC630 and PC645. In this case, the low-energy band is shifted to lower energies in PC645, a trend also qualitatively reproduced by our simulations, which in this case seems to arise from a stabilization of the MBV_{19C} as well as PCB_{82B} and PCB_{82D}. This suggests that the description of pigment-protein interactions based on the QM/MMPol method correctly describes this qualitative trends, so the main errors in the determination of the site energies seem to arise from the problem related to the internal pigment geometries use for excited-state calculations.

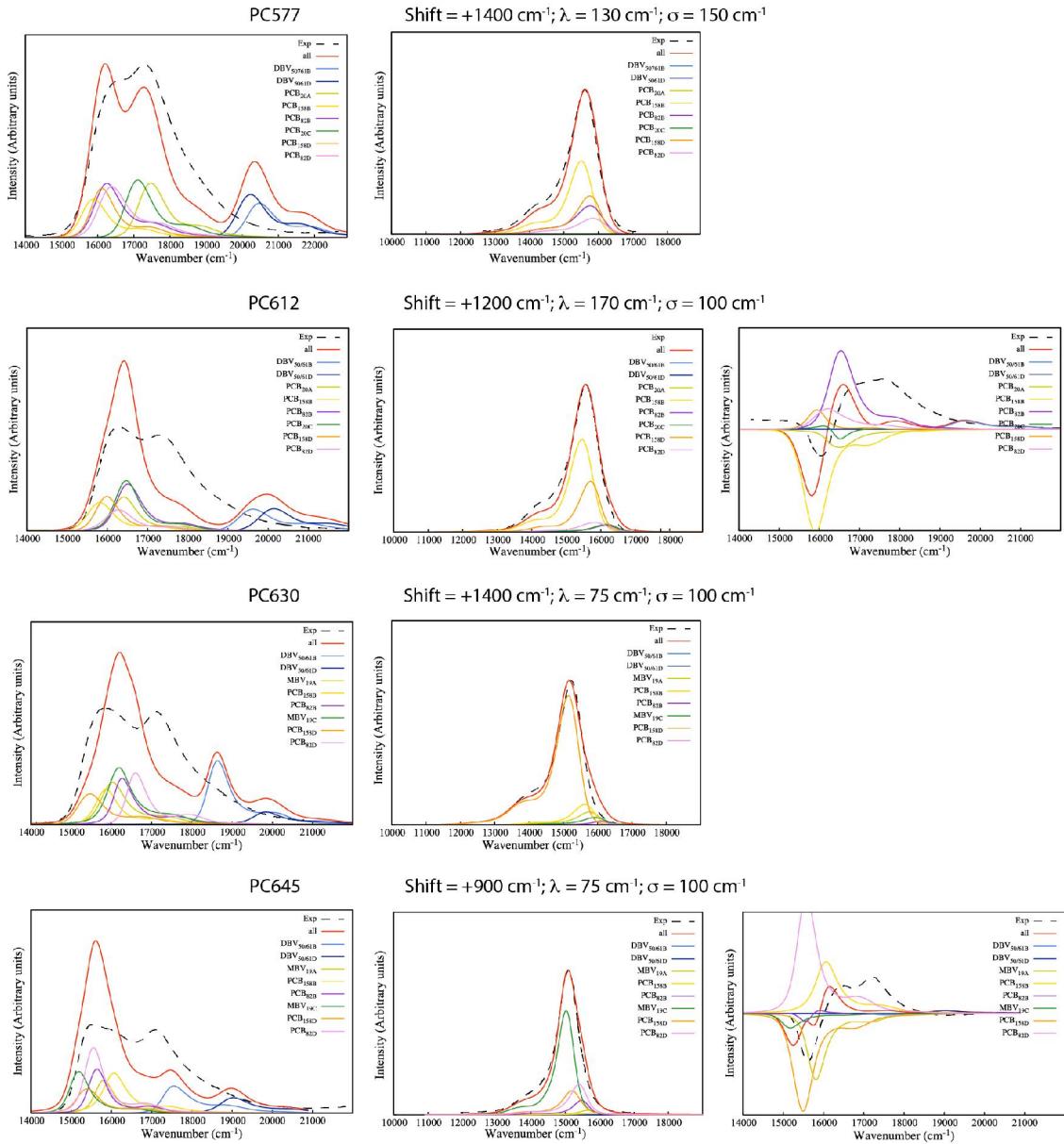


Figure 5. Experimental and simulated absorption (left), fluorescence (center) and circular dichroism (right) spectra of the PC577, PC612, PC630 and PC645 complexes (area normalized). Simulations were done using MD-averaged site energies and couplings computed from QM/MMPol calculations.

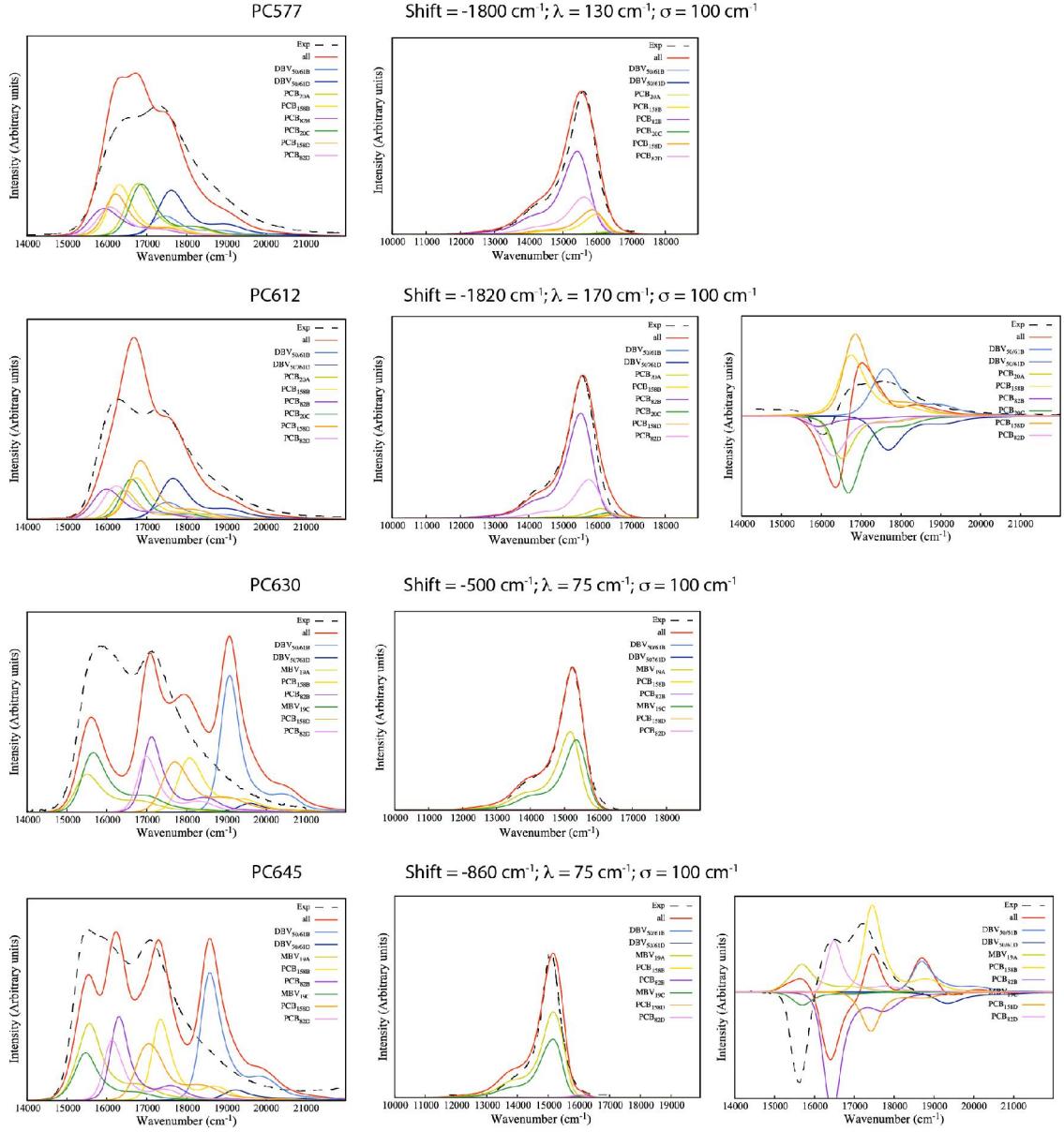


Figure 6. Experimental and simulated absorption (left), fluorescence (center) and circular dichroism (right) spectra of the PC577, PC612, PC630 and PC645 complexes (area normalized). Simulations were done using site energies and couplings computed from QM/MMPol/ddCOSMO calculations based on bilin optimized geometries.

If we focus instead on the spectral simulations based on QM-optimized geometries shown in Figure 6, the results for PC612, and especially for PC577, are in reasonable agreement with experiments, although in this case the intensity shift to lower energies when passing from PC577 to PC612 is not qualitatively reproduced, because the energies of PCB_{20A} and PCB_{20D} are not lowered in PC612. Instead, the energies of PCB_{158B} and PCB_{158D} are slightly increased. Overall, our MD results indicate an approximate energy ladder PCB₁₅₈ < PCB₈₂ < PCB₂₀ < DBV_{51/61} for PC577 and PC612, whereas results based on QM-optimizations exchange the lowest energy site for the PCB₈₂s, and moreover predict a different ordering for the intermediate

energy site, leading to an ordering $\text{PCB}_{82} < \text{PCB}_{158} < \text{PCB}_{20} < \text{DBV}_{51/61}$ for PC577 and $\text{PCB}_{82} < \text{PCB}_{20} < \text{PCB}_{158} < \text{DBV}_{51/61}$ for PC612. Thus, it is unclear if the lowest-energy sites on these complexes are the PCB_{158} or the PCB_{82} bilins.

On the other hand, in the spectra of PC630 and PC645 the lowest-lying pigments are the MBVs when QM-optimized geometries are used, whereas in MD-based results they show very similar energies compared to the low-energy PCBs, as found by Coker and co-workers for PC645.²⁰ Moreover, in this case the QM-based approach indicates a similar energy ordering $\text{MBV}_{19} < \text{PCB}_{82} < \text{PCB}_{158} < \text{DBV}_{51/61}$ for both complexes, whereas MD-based data estimate rather similar energies for all low-energy sites: $\text{MBV}_{19} \approx \text{PCB}_{158} \approx \text{PCB}_{82} < \text{DBV}_{51/61}$. Although the spectra based on QM-optimized structures is too broad, probably owing to the significant underestimation of MBV energies, the overall shape of the bands due to PCBs and DBVs is in much better agreement than that obtained from MD-based data. Thus, the energy ordering $\text{MBV}_{19} \approx \text{PCB}_{82} < \text{PCB}_{158} < \text{DBV}_{51/61}$ seems quite likely, once MBVs are shifted up in energy.

Taken globally, the results based on QM-optimized geometries in almost the four systems agree with the ordering $\text{PCB}_{82} < \text{PCB}_{158} < \text{DBV}_{51/61}$, which seems reasonable as these sites are located in the highly-conserved β chain. The spectral variability thus seems to arise from the energetic location of the bilins in the more divergent α chain, which in PC577 and PC612 seems to be in the middle-energy range (PCB_{20A} and PCB_{20C}), whereas in PC630 and PC645 seem to be in the low-energy region (MBV_{19A} and MBV_{19C}). Indeed, the β chains in PC577 and PC612 are identical. This overall picture further agrees with the recent simulations in the Coker group.²⁰

3.3 Electronic couplings

The spectra of “open” (PC577, PC612) and “closed” form (PC630, PC645) PBP_s discussed in the previous section have a fundamental difference related to the change in quaternary structure. In PC630 and PC645, the small center-to-center distance ($\sim 13\text{\AA}$) between central DBVs leads to strong excitonic splitting of the high-energy bands, whereas switch to the “open” form in PC577 and PC612 leads to a larger separation ($\sim 19\text{\AA}$) that drastically attenuates the coupling between DBVs, thus leading to mostly localized bands.¹⁴ According to previous CIS/cc-pVTZ coupling estimates based on the transition density cube method, this leads to a change in $\text{DBV}_{50/61B} - \text{DBV}_{50/61D}$ coupling from 647 cm^{-1} in PC645 to 29 cm^{-1} in PC612.¹⁴ The change in quaternary structure also induced variations in other interdimer couplings, whereas intradimer ones remained similar. Note here that these values did not include the screening effect exerted by the environment. Our Coulomb terms based on the crystal structures lead to similar values $515/559\text{ cm}^{-1}$ for PC630/PC645 and $19/21\text{ cm}^{-1}$ for PC577/PC612, as reported in Tables S11-S18 in the Supporting Information. The MD-based values shown in Table SXX are rather similar, except

for PC577, where thermal effects increase the V_{Coul} value to 67 cm^{-1} , still however much smaller than the corresponding value in “closed” PBPs.

The Coulomb term however is modulated by the additional environment-mediated term V_{env} in Eq. 3. Here we included this effect based on the MD-QM/MMPol and QM/MMPol/ddCOSMO methods. The latter differ from the MD-based values in two important aspects. First, the water solvent is described through the ddCOSMO continuum model instead of the atomistic MMPol description used for the protein. In addition, the coupling values are based on the crystal structure, so the protein is not relaxed like in the MD, a delicate issue that can lead to some exaggerated polarization interactions and thus induce some noise in the MMPol coupling term. For these two reasons, we thus consider the MD-based values to be more reliable and we only focus on those here. For PC630 and PC645, we predict screening factors s equal to 0.60 and 0.50, respectively, for the central DBV pair, slightly smaller than the 0.64 value predicted previously for PC645 based on a continuum model.⁵² In contrast, the environment leads to a striking attenuation of the coupling in PC577 and PC612, which screening factors 0.1 and -0.3, respectively, which correspond to $V_{Coul} = 67 \text{ cm}^{-1}$ and $V_{env} = -54 \text{ cm}^{-1}$ for PC577 and $V_{Coul} = 23 \text{ cm}^{-1}$ and $V_{env} = -30 \text{ cm}^{-1}$ for PC612. Thus, for PC612 the environment-mediated term is actually larger in magnitude than the Coulomb interaction between the DBVs. This strong screening effect thus further reduces the coupling in the central DBV pair beyond the increase in pigment separation, which explains the attenuation of coherence beatings observed for these PBPs.¹⁴ Beyond this unexpected strong screening effect, most other pairs in the PBPs experience reasonable screening values, as shown in Figure 7. However, some pairs in PC577 and PC612 deviate from the general $s \sim 0.5\text{-}0.7$ trend with s values close to 1 or even larger, showing a small screening effect or even an enhancement of the coupling mediated by the environment.

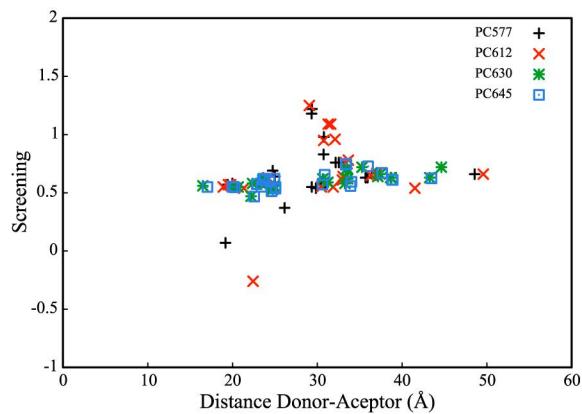


Figure 7. MD-averaged screening factors derived from QM/MMPol calculations for pigment pairs in the PC577, PC612, PC630 and PC645 complexes.

Conclusions

In this contribution we have provided an analysis of the light harvesting properties of four PBP_s from cryptophyte algae, PC577, PC612, PC630 and PC645. We compared two different theoretical strategies to derive the excitonic Hamiltonian of the system. The first is based QM/MMPol excited state calculations performed on geometries sampled from classical MD simulations. We then explored a cost-effective strategy based on multiscale QM/MMPol/ddCOSMO excited state calculations performed on geometries optimized for the bilins in the protein scaffolds. Our results indicate that the underlying force field used in the classical MDs introduce significant systematic errors in site energies predicted for the different type of bilins DBV, PCB and MBV. The second strategy based on QM-optimized geometries gives more reliable energy differences, although the neglect of thermal effects leads to a significant underestimation of the MBV energies. In addition, we compute the spectral densities of the pigments using the Vertical Gradient approach, which leads to emission lineshapes in excellent agreement with experiment. Regarding electronic interactions in the complexes, our results indicate a drastic screening of the coupling in the central pair of DBVs in “open” form PBPs, which together with the increased DBV-DBV separation explains the attenuation of coherence beatings observed for these PBPs. Overall, our results strongly suggests that these PBPs share a common ordering $\text{PCB}_{82} < \text{PCB}_{158} < \text{DBV}_{51/61}$ of the site energies for pigments located in the highly-conserved β chain. The spectral variability among PBPs thus seems to arise from the energy tuning of the bilins in the more divergent α chain, which in PC577 and PC612 are expected to contribute to the middle-energy region of the spectra (PCB_{20A} and PCB_{20C}), whereas in PC630 and PC645 they contribute to the low-energy one (MBV_{19A} and MBV_{19C}).

Supporting Information

Tables of reorganization energies, transition dipole moments, coordinates of pigment centers, site energies and electronic couplings for PC577, PC612, PC630 and PC645.

Acknowledgements

We are grateful to the Consorci de Serveis Universitaris de Catalunya for providing access to computational resources (Molecular Recognition project). Financial support from the Agència de Gestió d’Ajuts Universitaris i de Recerca from Generalitat de Catalunya (GENCAT; SGR2014-1189 and SGR2017-1746) and from the Spanish Ministerio de Economía y Competitividad (MINECO; grants CTQ2012-36195 and CTQ2017- 89924-P) are

acknowledged. C.C. is a Serra Húnter Fellow (Generalitat de Catalunya) and M.C. acknowledges a fellowship from the Spanish Ministerio de Economía y Competitividad (MINECO; grant BES-2013- 064088). We are thankful to Paul M. G. Curmi for providing the crystal structures of the PC577 and PC630 complexes.

References

1. Mirkovic, T. *et al.* Light Absorption and Energy Transfer in the Antenna Complexes of Photosynthetic Organisms. *Chem. Rev.* **117**, 249–293 (2017).
2. Curutchet, C. & Mennucci, B. Quantum Chemical Studies of Light Harvesting. *Chem. Rev.* **117**, 294–343 (2017).
3. König, C. & Neugebauer, J. Quantum Chemical Description of Absorption Properties and Excited-State Processes in Photosynthetic Systems. *ChemPhysChem* **13**, 386–425 (2012).
4. Renger, T., Madjet, M. E.-A., Schmidt am Busch, M., Adolphs, J. & Müh, F. Structure-based modeling of energy transfer in photosynthesis. *Photosynth. Res.* **116**, 367–388 (2013).
5. Fenna, R. E. & Matthews, B. W. Chlorophyll arrangement in a bacteriochlorophyll protein from *Chlorobium limicola*. *Nature* **258**, 573–577 (1975).
6. Muh, F. *et al.* Alpha-Helices direct excitation energy flow in the Fenna Matthews Olson protein. *Proc. Natl. Acad. Sci.* **104**, 16862–16867 (2007).
7. van Grondelle, R. & Novoderezhkin, V. I. Energy transfer in photosynthesis: experimental insights and quantitative models. *Phys. Chem. Chem. Phys.* **8**, 793–807 (2006).
8. Novoderezhkin, V. I., Doust, A. B., Curutchet, C., Scholes, G. D. & van Grondelle, R. Excitation Dynamics in Phycoerythrin 545: Modeling of Steady-State Spectra and Transient Absorption with Modified Redfield Theory. *Biophys. J.* **99**, 344–352 (2010).
9. Zucchelli, G., Santabarbara, S. & Jennings, R. C. The Q_y Absorption Spectrum of the Light-Harvesting Complex II As Determined by Structure-Based Analysis of Chlorophyll Macrocycle Deformations. *Biochemistry* **51**, 2717–2736 (2012).
10. Curutchet, C. *et al.* Energy Flow in the Cryptophyte PE545 Antenna Is Directed by Bilin Pigment Conformation. *J. Phys. Chem. B* **117**, 4263–4273 (2013).
11. Dreuw, A., Harbach, P. H. P., Mewes, J. M. & Wormit, M. Quantum chemical excited state calculations on pigment–protein complexes require thorough geometry re-optimization of experimental crystal structures. *Theor. Chem. Acc.* **125**, 419–426 (2010).
12. Nogueira, J. J. & González, L. Computational Photophysics in the Presence of an

- Environment. *Annu. Rev. Phys. Chem.* **69**, 473–497 (2018).
- 13. Scholes, G. D., Mirkovic, T., Turner, D. B., Fassioli, F. & Buchleitner, A. Solar light harvesting by energy transfer: from ecology to coherence. *Energy Environ. Sci.* **5**, 9374–9393 (2012).
 - 14. Harrop, S. J. *et al.* Single-residue insertion switches the quaternary structure and exciton states of cryptophyte light-harvesting proteins. *Proc. Natl. Acad. Sci.* **111**, E2666–E2675 (2014).
 - 15. Smith, D. E. & Dang, L. X. Computer simulations of NaCl association in polarizable water. *J. Chem. Phys.* **100**, 3757–3766 (1994).
 - 16. Curutchet, C. *et al.* Photosynthetic Light-Harvesting Is Tuned by the Heterogeneous Polarizable Environment of the Protein. *J. Am. Chem. Soc.* **133**, 3078–3084 (2011).
 - 17. Aghtar, M., Strümpfer, J., Olbrich, C., Schulten, K. & Kleinekathöfer, U. Different Types of Vibrations Interacting with Electronic Excitations in Phycoerythrin 545 and Fenna–Matthews–Olson Antenna Systems. *J. Phys. Chem. Lett.* **5**, 3131–3137 (2014).
 - 18. Chandrasekaran, S., Pothula, K. R. & Kleinekathöfer, U. Protein Arrangement Effects on the Exciton Dynamics in the PE555 Complex. *J. Phys. Chem. B* **121**, 3228–3236 (2017).
 - 19. Lipparini, F. *et al.* Quantum, classical, and hybrid QM/MM calculations in solution: General implementation of the ddCOSMO linear scaling strategy. *J. Chem. Phys.* **141**, 184108 (2014).
 - 20. Lee, M. K., Bravaya, K. B. & Coker, D. F. First Principles Models for Biological Light-harvesting: Phycobiliprotein Complexes from Cryptophyte Algae. *J. Am. Chem. Soc.* **139**, 7803–7814 (2017).
 - 21. Corbella, M., Toa, Z. S. D., Scholes, G. D., Luque, F. J. & Curutchet, C. Determination of the protonation preferences of bilin pigments in cryptophyte antenna complexes. (*Submitted*) (2018).
 - 22. Case, D. A. *et al.* AMBER 14. *AMBER 14* (2014).
 - 23. Vreven, T. *et al.* Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2**, 815–826 (2006).
 - 24. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Janesko, B. G.; Li, D. J. Gaussian Development Version, Revision H.36. *Gaussian, Inc. Wallingford CT* (2010).
 - 25. Senn, H. M. & Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chemie Int. Ed.* **48**, 1198–1229 (2009).
 - 26. Curutchet, C. *et al.* Electronic Energy Transfer in Condensed Phase Studied by a Polarizable QM/MM Model. *J. Chem. Theory Comput.* **5**, 1838–1848 (2009).
 - 27. Cancès, E., Maday, Y. & Stamm, B. Domain decomposition for implicit solvation

- models. *J. Chem. Phys.* **139**, (2013).
- 28. Lipparini, F., Stamm, B., Cancès, E., Maday, Y. & Mennucci, B. Fast domain decomposition algorithm for continuum solvation models: Energy and first derivatives. *J. Chem. Theory Comput.* **9**, 3637–3648 (2013).
 - 29. Lipparini, F. *et al.* Quantum Calculations in Solution for Large to Very Large Molecules: A New Linear Scaling QM/Continuum Approach. *J. Phys. Chem. Lett.* **5**, 953–958 (2014).
 - 30. Caprasecca, S., Curutchet, C. & Mennucci, B. Toward a Unified Modeling of Environment and Bridge-Mediated Contributions to Electronic Energy Transfer: A Fully Polarizable QM/MM/PCM Approach. *J. Chem. Theory Comput.* **8**, 4462–4473 (2012).
 - 31. Yanai, T., Tew, D. P. & Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **393**, 51–57 (2004).
 - 32. Wang, J. *et al.* Development of polarizable models for molecular mechanical calculations I: parameterization of atomic polarizability. *J. Phys. Chem. B* **115**, 3091–9 (2011).
 - 33. Wang, J. *et al.* Development of Polarizable Models for Molecular Mechanical Calculations II: Induced Dipole Models Significantly Improve Accuracy of Intermolecular Interaction Energies. *J. Phys. Chem. B* **115**, 3100–3111 (2011).
 - 34. Caprasecca, S., Curutchet, C., Jurinovich, S. & Mennucci, B. PolChat: A polarisation-consistent charge fitting tool. molecolab.dcci.unipi.it/tools
 - 35. Doust, A. B., Wilk, K. E., Curmi, P. M. G. & Scholes, G. D. The photophysics of cryptophyte light-harvesting. *J. Photochem. Photobiol. A Chem.* **184**, 1–17 (2006).
 - 36. Lee, M. K., Huo, P. & Coker, D. F. Semiclassical Path Integral Dynamics: Photosynthetic Energy Transfer with Realistic Environment Interactions. *Annu. Rev. Phys. Chem.* **67**, 639–668 (2016).
 - 37. Zhang, W. M., Meier, T., Chernyak, V. & Mukamel, S. Exciton-migration and three-pulse femtosecond optical spectroscopies of photosynthetic antenna complexes. *J. Chem. Phys.* **108**, 7763–7774 (1998).
 - 38. Viani, L. *et al.* Molecular basis of the exciton–phonon interactions in the PE545 light-harvesting complex. *Phys. Chem. Chem. Phys.* **16**, 16302–16311 (2014).
 - 39. Olbrich, C., Strümpfer, J., Schulten, K. & Kleinekathöfer, U. Theory and simulation of the environmental effects on FMO electronic transitions. *J. Phys. Chem. Lett.* **2**, 1771–1776 (2011).
 - 40. Rivera, E., Montemayor, D., Masia, M. & Coker, D. F. Influence of Site-Dependent Pigment–Protein Interactions on Excitation Energy Transfer in Photosynthetic Light Harvesting. *J. Phys. Chem. B* **117**, 5510–5521 (2013).

41. Janosi, L., Kosztin, I. & Damjanović, A. Theoretical prediction of spectral and optical properties of bacteriochlorophylls in thermally disordered LH2 antenna complexes. *J. Chem. Phys.* **125**, 14903 (2006).
42. Olbrich, C. & Kleinekathöfer, U. Time-Dependent Atomistic View on the Electronic Relaxation in Light-Harvesting System II. *J. Phys. Chem. B* **114**, 12427–12437 (2010).
43. Shim, S., Rebentrost, P., Valleau, S. & Aspuru-Guzik, A. Atomistic Study of the Long-Lived Quantum Coherences in the Fenna-Matthews-Olson Complex. *Biophys. J.* **102**, 649–660 (2012).
44. Valleau, S., Eisfeld, A. & Aspuru-Guzik, A. On the alternatives for bath correlators and spectral densities from mixed quantum-classical simulations. *J. Chem. Phys.* **137**, 224103 (2012).
45. Aghtar, M., Strümpfer, J., Olbrich, C., Schulten, K. & Kleinekathöfer, U. The FMO Complex in a Glycerol–Water Mixture. *J. Phys. Chem. B* **117**, 7157–7163 (2013).
46. Fujita, T., Huh, J., Saikin, S. K., Brookes, J. C. & Aspuru-Guzik, A. Theoretical characterization of excitation energy transfer in chlorosome light-harvesting antennae from green sulfur bacteria. *Photosynth. Res.* **120**, 273–289 (2014).
47. Huh, J. *et al.* Atomistic Study of Energy Funneling in the Light-Harvesting Complex of Green Sulfur Bacteria. *J. Am. Chem. Soc.* **136**, 2048–2057 (2014).
48. Rosnik, A. M. & Curutchet, C. Theoretical Characterization of the Spectral Density of the Water-Soluble Chlorophyll-Binding Protein from Combined Quantum Mechanics/Molecular Mechanics Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **11**, 5826–5837 (2015).
49. Blau, S. M., Bennett, D. I. G., Kreisbeck, C., Scholes, G. D. & Aspuru-Guzik, A. Local protein solvation drives direct down-conversion in phycobiliprotein PC645 via incoherent vibronic transport. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E3342–E3350 (2018).
50. Mirkovic, T. *et al.* Ultrafast light harvesting dynamics in the cryptophyte phycocyanin 645. *Photochem. Photobiol. Sci.* **6**, 964–75 (2007).
51. McClure, S. D., Turner, D. B., Arpin, P. C., Mirkovic, T. & Scholes, G. D. Coherent Oscillations in the PC577 Cryptophyte Antenna Occur in the Excited Electronic State. *J. Phys. Chem. B* **118**, 1296–1308 (2014).
52. Scholes, G. D., Curutchet, C., Mennucci, B., Cammi, R. & Tomasi, J. How Solvent Controls Electronic Energy Transfer and Light Harvesting. *J. Phys. Chem. B* **111**, 6978–6982 (2007).

3.2. Efectes de l'entorn en processos CT en l'ADN

Marina Corbella, Alexander A. Voityuk, and Carles Curutchet, *Single Amino Acid Mutation Controls Hole Transfer Dynamics in DNA-Methyltransferase HhaI Complexes*, Journal of Physical Chemistry Letters, **2015**, 6 (10), 3749-3753.

Marina Corbella, Alexander A. Voityuk, and Carles Curutchet, *How basic sites impact hole transfer Dynamics in GC-rich DNA sequences*, (submitted).

Single Amino Acid Mutation Controls Hole Transfer Dynamics in DNA-Methyltransferase *HhaI* Complexes

The Journal of Physical Chemistry Letters, 2015, 6 (18) 3749-3753

Marina Corbella,¹ Alexander A. Voityuk,^{2,3} and Carles Curutchet¹

¹ Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica i Institut de Química Teòrica i Computacional (IQT CUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona, Espanya.

² Institut de Química Computacional i Catalisi i Departament de Química, Universitat de Girona, Girona, Espanya.

³ Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Espanya..

3.2.1. Single Amino Acid Mutation Controls Hole Transfer Dynamics in DNA-Methyltransferase *HhaI* Complexes.

Resum

La oxidació de certes bases de l'ADN dóna lloc a la formació d'estats catió radical, aquests estats, que s'anomenen "forats" degut a l'absència d'un electró, poden ser precursors d'efectes mutagènics en el codi genètic. S'ha observat que aquests "forats" poden migrar a través de la doble hèlix fins a distàncies de 200 Å en funció de la seqüència de bases. S'han portat a terme múltiples estudis, tant experimentals com teòrics, destinats a entendre els factors que governen aquests processos de HT. No obstant, recentment, la idea de que aquests "forats" poden migrar a aminoàcids aromàtics de proteïnes que s'uneixen a l'ADN, protegint així del dany oxidatiu el genoma, ha obert la porta a un nou camp d'estudi encara inexplorat. El grup de J. K. Barton ha observat que en el lloc d'unió a l'ADN de l'enzim metiltrasferasa *HhaI*, que durant la seva acció enzimàtica extreu una base de la doble hèlix i inserta en el seu lloc la cadena lateral d'un aminoàcid, el procés de HT es veu alterat en funció de la naturalesa de l'aminoàcid inserit. L'estudi ha revelat que, mentre la unió a la proteïna nativa, que inserta la cadena lateral d'un residu de glutamina, redueix considerablement (42%) el dany oxidatiu en el costat oposat al punt de generació del catió radical en el lloc d'unió amb l'ADN, la unió a una proteïna mutant Q237W, en la que la glutamina inserida ha estat substituïda per un residu aromàtic de triptòfan, manté la presència de dany oxidatiu. La idea principal d'aquest treball és intentar entendre el paper que juga la deformació estructural, degut a la inserció d'un residu de la proteïna, en els diferents paràmetres que governen el procés de HT en el complex *HhaI*. En estudis anteriors ja s'ha advertit que el procés de HT a través de la doble hèlix és molt sensible a la deformació estructural, ja que depèn en gran part de l'acoblament electrònic, i aquest està estretament lligat a l'apilament entre les diferents bases. Per aquest motiu, s'han combinat simulacions de dinàmica molecular, per captar les fluctuacions estructurals, amb càlculs QM/MM de les energies i els acoblaments de les diferents bases en la seqüència d'unió a la proteïna. Els acoblaments electrònics entre les bases (i les energies dels estats de càrrega localitzada) s'han obtingut

mitjançant el mètode de FCD utilitzant un nivell de teoria semiempíric INDO/S. Posteriorment, s'ha emprat la teoria de Marcus per obtenir les velocitats de transferència entre cada parell de bases, i s'ha utilitzat l'algoritme de Kinetic Monte Carlo per simular el procés global de HT a través de totes les bases compreses dins de la seqüència d'unió. Els resultats obtinguts concorden amb les observacions experimentals prèvies i expliquen el mecanisme subjacent. La inserció de la cadena lateral de glutamina entre les bases de la seqüència desestabilitza els estats de les guanines contigües alentint el procés de HT a través de la seqüència d'unió i per tant inhibint la presència de dany oxidatiu a les guanines posteriors al punt d'inserció. Per contra, la inserció del residu aromàtic de triptòfan estabilitza els estats de les guanines contigües i n'augmenta l'acoblament electrònic, el que es tradueix en una acceleració del procés global de HT. D'altra banda, es confirma la formació del catió radical de triptòfan, no obstant, la seva participació en el procés de HT sembla alentir el procés global a través de la seqüència d'unió.

Single Amino Acid Mutation Controls Hole Transfer Dynamics in DNA-Methyltransferase *HhaI* Complexes

Marina Corbella,[†] Alexander A. Voityuk,^{*,‡,§} and Carles Curutchet^{*,†}

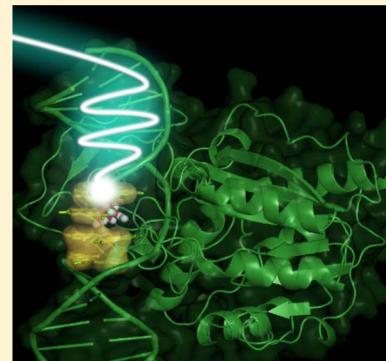
[†]Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain

[‡]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

[§]Institut de Química Computacional i Catalisi and Departament de Química, Universitat de Girona, Campus Montilivi, 17071 Girona, Spain

Supporting Information

ABSTRACT: Different mutagenic effects are generated by DNA oxidation that implies the formation of radical cation states (so-called holes) on purine nucleobases. The interaction of DNA with proteins may protect DNA from oxidative damage owing to hole transfer (HT) from the stack to aromatic amino acids. However, how protein binding affects HT dynamics in DNA is still poorly understood. Here, we report a computational study of HT in DNA complexes with methyltransferase *HhaI* with the aim of elucidating the molecular factors that explain why long-range DNA HT is inhibited when the glutamine residue inserted in the double helix is mutated into a tryptophan. We combine molecular dynamics, quantum chemistry, and kinetic Monte Carlo simulations and find that protein binding stabilizes the energies of the guanine radical cation states and significantly impacts the corresponding electronic couplings, thus determining the observed behavior, whereas the formation of a tryptophan radical leads to less efficient HT.



Charge transport through the DNA base stack has been a subject of considerable interest in recent years,^{1–5} with possible implications in areas as diverse as carcinogenesis and molecular electronics.⁶ It has been shown that radical cation states (electron holes) generated in DNA can migrate through nucleobases over distances up to 200 Å from the initial oxidized point. This ability of DNA can be employed for redox sensing and signaling in the genome, as long as it may provide new insight into the direction of mutations upon oxidative stress. Extensive studies, both theoretical and experimental, suggest that long-range HT occurs as a multistep hopping process between adjacent guanines.^{7–10} Although HT through DNA strands is well-known, how this process occurs in DNA-protein systems, where the hole can migrate from DNA to aromatic amino acid residues protecting the genome, is still poorly understood.^{7,11,12} The formation of tryptophan and tyrosine radicals arising from HT initiated in DNA has been observed using the flash-quench technique upon the addition of DNA-bound peptides to ruthenated DNA assemblies.^{13,14} Quantum chemical calculations suggest that both electronic coupling and the driving force of HT between nucleobases and aromatic amino acids depend strongly on the mutual arrangement of donor and acceptor sites.^{15,16}

The Barton group has probed charge migration and radical trapping in DNA assemblies in the presence of the specifically bound enzyme methyltransferase *HhaI* using a combination of gel electrophoretic assays and transient spectroscopy.^{11,12} The enzyme carries out an alkylation process after flipping out the central cytosine of its binding domain 5'-GCGC-3' and

inserting a glutamine side chain into the resultant DNA pocket (see Figure 1). In the present study we do not address this enzymatic reaction, but rather investigate the molecular factors that explain the interesting behavior unveiled by the experi-

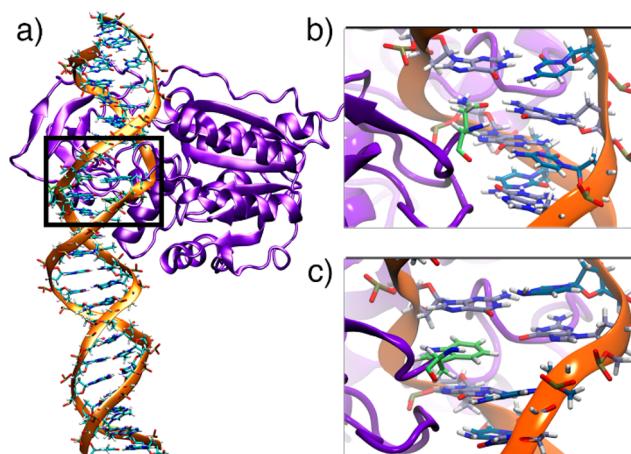


Figure 1. (a) Structure of the DNA-*HhaI* complex considered in this study. Detailed view of the binding mode of the (b) glutamine in the wild-type system and (c) of the tryptophan in the Q237W mutant.

Received: August 4, 2015

Accepted: September 2, 2015

ments performed in the Barton group. Such experiments indicated that binding of the wild type (WT) enzyme decreases long-range oxidative damage at distal 5'-GG-3' sites substantially, by 42%. In contrast, distal oxidative damage was retained in a Q237W mutant, which inserts a tryptophan (W) side chain instead of a glutamine (Q) within the DNA π -stack.¹¹ These studies also revealed the involvement of both guanine and tryptophan radicals in the process through detection of transient spectroscopic signals, and gel electrophoresis experiments indicated considerable oxidative damage at the G 5' to the W inserted amino acid, suggesting a stabilization mechanism of the G cation state due to stacking with the W reminiscent of the stabilization of 5'-GG-3' motifs.¹²

In order to establish the microscopic basis for the observed HT behavior, here we combine classical molecular dynamics simulations with INDO/S quantum mechanical calculations of HT energetics and electronic couplings in the binding domain (see Figure 2), to explore how the interaction of the DNA

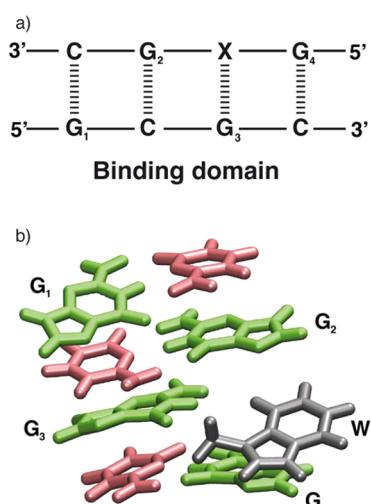


Figure 2. (a) Binding domain of the DNA-HhaI complex. (b) Structure of the model system, which includes the nucleobases in the binding domain (except the X abasic site) plus the Q (wild-type) or W (Q237W mutant) residues inserted into the DNA pocket.

duplex with the wild-type (WT) or mutant enzyme affects its HT properties. This approach allows a realistic description of the effect of structural fluctuations on the HT parameters,¹⁷ which are strongly dependent on the arrangement of donor and acceptor. We use Marcus theory and a Kinetic Monte Carlo (KMC) scheme to estimate HT rates and various pathways along the binding domain of both assemblies. In Figure 3 we show the HT reactions included in the simulations of the

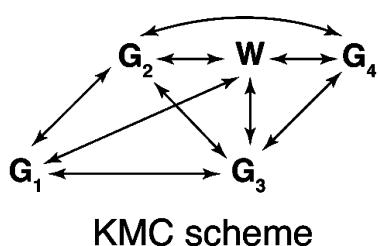


Figure 3. Scheme of relevant pathways and states included in the Kinetic Monte Carlo simulations of the HT dynamics.

overall HT kinetics. Our results indicate that protein binding strongly affects the HT characteristics of DNA. In particular, the modulation of G site energies and electronic couplings are responsible for the observed HT behavior, whereas the formation of a W⁺ radical leads to less efficient HT.

It is well-known that the HT properties of DNA are highly sensitive to disruptions of the π -stack. In HhaI-DNA complexes, the local structure of the DNA in the binding domain is significantly distorted.^{18,19} In the DNA–protein assemblies we address, the presence of an abasic site X, i.e., the absence of a nucleobase attached to the sugar, can further impact conformational changes within the π -stack. In order to understand the impact of these two effects in the HT dynamics, we have performed MD simulations on the DNA and DNA-X systems, as well as on the DNA-X system bound to the WT and Q237W mutant HhaI enzymes. Our results indicate that the structural parameters of the double helix change significantly upon the protein binding and the presence of the abasic site. Earlier, it was suggested that in contrast to the WT protein, the Q237W mutant, retains long-range oxidative damage at distal 5'-GG-3' sites because the inserted W fulfills the π -gap in the DNA stack by adopting a stacked configuration relative to the nucleobases.^{11,12} Our simulations confirm this hypothesis and indicate that W and G₃ compete for the abasic site space, as shown by the degree of insertion shown in Figure 4. We define the degree of insertion as the ratio between the distance among upper and lower stacked guanines and the sum of the distances among the inserted amino acid side chain and such guanines. Thus, this value ranges from 0 to 1, the latter indicating complete insertion of the amino acid in the π -stack (see

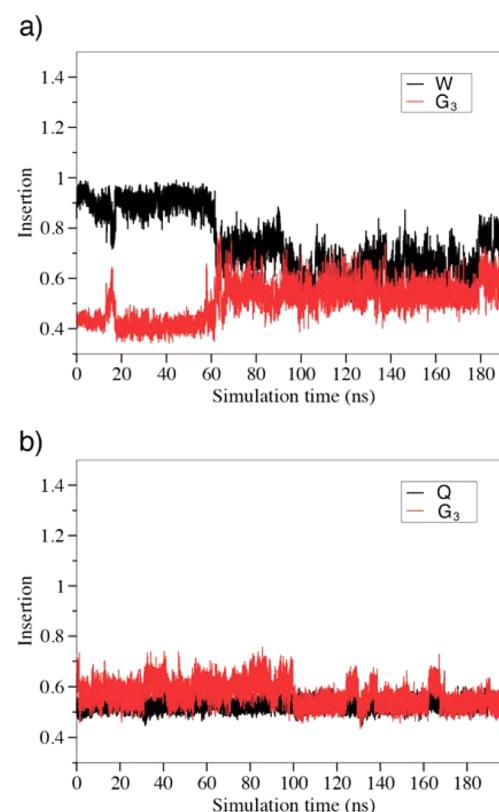


Figure 4. Degree of insertion of (a) W/G₃ and (b) Q/G₃ residues between the G₂ and G₄ sites in the Q237W and WT assemblies, respectively (see text for details).

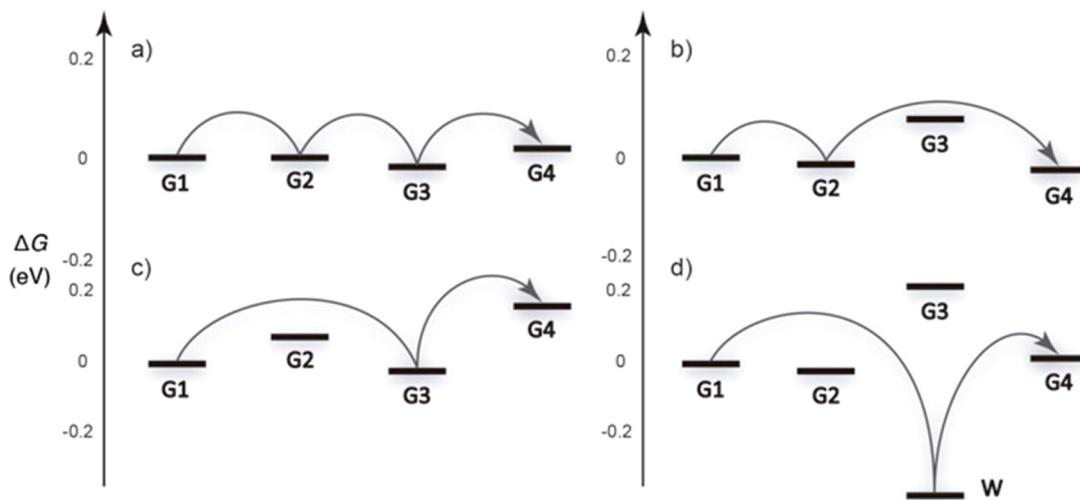


Figure 5. Site energies of the radical cation states localized on the guanines and the tryptophan for the (a) DNA, (b) DNA-X, (c) WT, and (d) Q237W systems. The most frequent HT pathways according to KMC simulations are indicated by arrows.

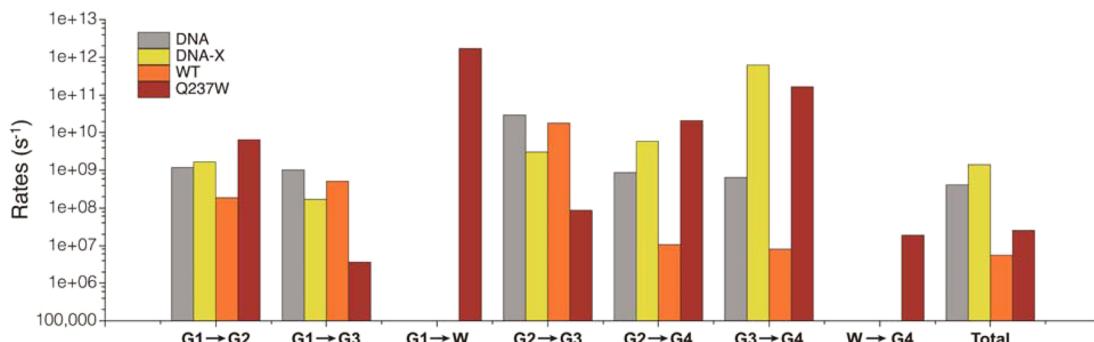


Figure 6. HT rates obtained from the Kinetic Monte Carlo simulations corresponding to the reference DNA stack, the DNA-X with an abasic site, and the DNA-X bound to the WT and Q237W *HhaI* protein.

Table 1. Time Constants for the $G_1 \rightarrow G_4$ HT Transfer and the Contribution of Preferred Pathways in the Studied Systems

	DNA	DNA-X	WT	Q237W
τ (ns)	2.41	0.71	178	0.22
$G_1G_2G_3G_4$ (%)	2%	54%	1%	90% ^a

^aTotal time and corresponding pathway in Q237W including all possible forward/backward HT reactions involving W.

Supporting Information for further details on the structural analysis).

Regarding the impact of the abasic site and protein binding on the parameters determining the HT behavior, we find that the formation of the abasic site in the DNA stack increases V_{24} and V_{34} couplings, which change from 7 to 15 meV and from ~8 to ~66 meV by passing from DNA to DNA-X. This dramatic change arises from the displacement of G_3 allowing a better stacking between guanines G_3 and G_4 (see Figure S9 in Supporting Information). The conformational changes also significantly destabilize the radical cation state of G_3 in contrast to the DNA, where all sites have similar energies. When, however, the DNA-X helix binds the WT *HhaI* protein the insertion of the Q residue in the π -gap precludes the rearrangement of G_3 site and thereby leads to a dramatic attenuation of V_{24} and V_{34} (2 and 4 meV). Also, the V_{23} coupling decreases from 31 to 12 meV. The energies of the G_2 and G_4 hole states are destabilized compared to the other sites (see Figure 5) causing a potential inhibition of long-range oxidative damage as observed experimentally. Interestingly, the

overall picture for HT in the Q237W mutant reminds that observed for the unbound DNA-X system. Compared to the WT complex, in Q237W both G_2 and G_4 sites are stabilized due to the insertion of the stacked tryptophan, thus recovering the fast HT toward the final G_4 site in the binding domain. Moreover, the mutation of the inserted Q residue to W enhances strongly the couplings between guanine sites, as shown in Table S1, due to the superexchange interaction of the nucleobases mediated by the W residue.

In the mutant, W237 may act also as an intermediate state in the HT process. Our additional simulations performed on single guanine and tryptophan in water predict that a correction of -0.61 eV to the INDO/S energy of the W state should be applied to reproduce the data derived from experimental oxidation potentials at neutral pH.²⁰ The correction has been applied in the simulation of HT within the Q237W complex.

In Figure 5 we show the preferred pathway for HT estimated from the KMC simulations for each system, which include the pathways illustrated in Figure 3 (more details are provided in Figures S11–S12). The rates for the individual HT steps and

the overall process in each system are presented in **Figure 6**. Due to a small energetic disorder among guanines in the DNA system, no particular pathway is preferred, and most transfer steps occur with a time constant, i.e., the inverse of the rate, of $\tau \approx 1$ ns, leading to the overall HT time of ~ 2.4 ns (**Table 1**). In the DNA-X system, the $G_1G_2G_4$ pathway is preferred in 54% of the KMC realizations, due to the G_3 site destabilization. However, a better stacking between guanines G_3 and G_4 leads to stronger electronic couplings in the binding domain resulting in faster overall HT with a characteristic time of ~ 710 ps.

The binding of DNA-X to the *HhaI* complex further modulates its HT dynamics. Binding to the WT enzyme, which destabilizes the G_4 site and attenuates V_{24} and V_{34} , leads to much slower hopping rates from G_2/G_3 to G_4 ($\tau = 94$ and 124 ns), resulting in a longer overall HT time of 178 ns. Because of the relatively close energies of G_1 , G_2 , and G_3 , a variety of paths contribute to the transfer in this case, the $G_1G_3G_4$ one being slightly preferred due to the lower energy of G_3 compared to G_2 (see **Supporting Information**). In Q237W, we can separate the impact of the binding into two effects: (1) modulation of site energies and electronic couplings change the rate of individual hopping steps, and (2) new pathways appear that imply the formation of W^+ . The insertion of the tryptophan clearly destabilizes the G_3 site, but it strongly mediates all couplings involving G_2 , G_3 , and G_4 . If we simulate the overall HT kinetics excluding the eventual formation of W^+ , the path $G_1G_2G_4$ is preferred in 90% of the realizations, and the overall HT process becomes much faster ($\tau = 220$ ps). Transient absorption experiments, however, confirm the participation of the W^+ state in the hopping process,¹² so the eventual formation of W^+ needs to be included in the simulations. In this case, our KMC simulations indicate that the G_1WG_4 pathway is preferred in 26% of the realizations, in agreement with experiments.

Here it is worth noting that gel electrophoresis experiments indicated that G_4 is a better trapping site than G_2 , in accord with results found for the 5'- G_4WG_2 -3' sequence.¹² Our simulations rather predict a slightly higher energy for G_4 . However, some damage was also experimentally observed in the G_2 site, suggesting that the site energies of G_2 and G_4 are comparable, in accord with our simulations. So, the insertion of the stacked tryptophan at the sequence 5'- G_4WG_2 -3' stabilizes the G_2 and G_4 sites compared to the WT system (**Figure 5**). This effect is similar to that observed in 5'-GG-3' or 5'-GGG-3' motifs, where the electrostatic interaction with neighboring GC pairs stabilizes a hole trap.^{21,22} In contrast, only minor damage at G_2 and no damage at G_4 is observed in the WT complex in agreement with the predicted destabilization G_2 and G_4 . We note here that these experiments only assessed the oxidative damage in the labeled strand carrying the G_2 and G_4 sites.

Because the energy of the W is predicted to be considerably lower than those of the guanine sites, its participation in the hopping process leads to a trapping effect, which slows down the $G_1 \rightarrow G_4$ transfer, leading to an overall HT time of 39 ns. As mentioned above, in this case the preferred path through the binding domain is G_1WG_4 . This estimate still describes a significantly faster HT process in the Q237W mutant compared to the WT enzyme, in agreement with experimental observations.

We conclude that structural deformation of the DNA stack upon protein binding strongly affects its HT characteristics. The Q residue inserted in the π -stack in the WT enzyme destabilizes neighboring guanine sites (G 3' and G 5' to Q) and thereby inhibits the long-range oxidative damage, in accord

with experimental findings. In contrast, the HT is estimated to be quite fast in a complex of the Q237W mutant where W237 stabilizes hole states on the adjacent G bases and enhances the electronic coupling of these sites. An alternative HT pathway that implies the formation of a W^+ radical also occurs in agreement with recent transient absorption experiments. However, it is found to slow overall HT dynamics along the DNA stack. Overall, our study provides a consistent molecular picture on how long-range HT in DNA is controlled by the nature of amino acid closely interacting with the nucleobases in DNA/*HhaI* complexes.

■ COMPUTATIONAL METHODS

We have performed extensive MD simulations at room temperature (300 K) of the DNA Ru-1 duplex described in ref 12, both including or not the abasic X site (DNA and DNA-X systems), as well as for the complex of DNA-X with the WT and Q237W mutant enzymes and a single guanine nucleobase and a tryptophan in water solution (see **Supporting Information** for a detailed description of the simulations). In the synthetic DNA-X system, the abasic site did not contain any nucleobase attached to the sugar. Because of the considerable size of the system studied, quantum-chemical calculations on the model system illustrated in **Figure 2** were performed using the semiempirical INDO/S method²³ on 1000 structures extracted along the MD trajectories. The INDO/S method has been shown to provide a reliable description of HT energies and electronic couplings of stacked nucleobases using Koopmans' approximation by comparison to high-level CASPT2 calculations, especially in the calculation of electronic couplings as compared to other semiempirical approaches.²⁴ However, our results indicated that it systematically overestimates the tryptophan radical cation state energy, which was corrected *a posteriori* based on the experimental one-electron oxidation potentials of guanine and tryptophan (see **Supporting Information** for details). Electrostatic embedding effects exerted by the environment, including protein, DNA, solvent, and counterions, were considered adopting the charges as defined in the force field used for the MD simulations.^{25–27} For structures of the WT and Q237W mutant systems, the hole states on the G sites and the W were represented by four and five highest-occupied molecular orbitals (HOMOs) of the neutral species ($HOMO \rightarrow HOMO-3/HOMO-4$), respectively. HT couplings between the states were estimated using the Fragment Charge Difference method (FCD).²⁸ From these calculations, HT rates were estimated for all forward and reverse HT reactions using Marcus theory. Finally, the effective rates and the paths for the overall HT processes G_1 to G_4 for each system were simulated using a Kinetic Monte Carlo algorithm.²⁹ In the KMC simulations, HT occurs in a stochastic manner, and the probability to find the hole in a given site is reproduced by averaging over multiple realizations. In some cases, like in the DNA and WT systems, the similar energies of several sites leads to multiple back and forth transfers of the hole among the sites, thus leading to many pathways that ultimately transfer the hole to the final G_4 site, as illustrated in Figures S11 and S12 in the **Supporting Information**. Further details on the methodology adopted and the analysis of structural and energetic parameters are provided in the **Supporting Information**.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jpclett.Sb01683](https://doi.org/10.1021/acs.jpclett.Sb01683).

Detailed methods description. RMSD deviations, stacking overlap, and base pair step parameters from the MD simulations. Structure of the binding domain in the DNA-X stack, distribution of HT pathways from KMC simulations, and table of electronic couplings, free energy differences, and time constants for the individual HT steps ([PDF](#))

AUTHOR INFORMATION

Corresponding Authors

*E-mail: carles.curutchet@ub.edu.

*E-mail: alexander.voityuk@icrea.cat.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

C.C. and M.C. acknowledge support from the Ministerio de Economía y Competitividad of Spain (grants CTQ2012-36195, RYC2011-08918, and BES-2013-064088), the Generalitat de Catalunya (SGR2014-1189), and computational resources provided by the Consorci de Serveis Universitaris de Catalunya. A.A.V. acknowledges financial support provided by grant CTQ 2011-26573.

REFERENCES

- (1) Giese, B. Long-Distance Electron Transfer through DNA. *Annu. Rev. Biochem.* **2002**, *71*, 51–70.
- (2) Wagenknecht, H. A., Ed. *Charge Transfer in DNA: From Mechanism to Application*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2005.
- (3) Genereux, J. C.; Barton, J. K. Mechanisms for DNA Charge Transport. *Chem. Rev.* **2010**, *110* (3), 1642–1662.
- (4) Venkatramani, R.; Keinan, S.; Balaeff, A.; Beratan, D. N. Nucleic Acid Charge Transfer: Black, White and Gray. *Coord. Chem. Rev.* **2011**, *255* (7–8), 635–648.
- (5) Siriwong, K.; Voityuk, A. A. Electron Transfer in DNA. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (5), 780–794.
- (6) Giese, B.; Amaudrut, J.; Köhler, A. K.; Spormann, M.; Wessely, S. Direct Observation of Hole Transfer through DNA by Hopping between Adenine Bases and by Tunnelling. *Nature* **2001**, *412*, 318–320.
- (7) Jortner, J.; Bixon, M.; Voityuk, A. A.; Rösch, N. Superexchange Mediated Charge Hopping in DNA. *J. Phys. Chem. A* **2002**, *106* (33), 7599–7606.
- (8) Berlin, Y. A.; Burin, A. L.; Ratner, M. A. Charge Hopping in DNA. *J. Am. Chem. Soc.* **2001**, *123*, 260–268.
- (9) Boon, E. M.; Barton, J. K. Charge Transport in DNA. *Curr. Opin. Struct. Biol.* **2002**, *12* (3), 320–329.
- (10) Meggers, E.; Michel-Beyerle, M. E.; Giese, B. Sequence Dependent Long Range Hole Transport in DNA. *J. Am. Chem. Soc.* **1998**, *120* (49), 12950–12955.
- (11) Rajski, S. R.; Kumar, S.; Roberts, R. J.; Barton, J. K. Protein-Modulated DNA Electron Transfer. *J. Am. Chem. Soc.* **1999**, *121* (23), 5615–5616.
- (12) Wagenknecht, H. A.; Rajski, S. R.; Pascaly, M.; Stemp, E. D.; Barton, J. K. Direct Observation of Radical Intermediates in Protein-Dependent DNA Charge Transport. *J. Am. Chem. Soc.* **2001**, *123* (19), 4400–4407.
- (13) Wagenknecht, H. A.; Stemp, E. D. A.; Barton, J. K. Evidence of Electron Transfer from Peptides to DNA: Oxidation of DNA- Bound Tryptophan Using the Flash-Quench Technique. *J. Am. Chem. Soc.* **2000**, *122*, 1–7.
- (14) Wagenknecht, H. A.; Stemp, E. D. A.; Barton, J. K. DNA-Bound Peptide Radicals Generated through DNA-Mediated Electron Transport. *Biochemistry* **2000**, *39*, 5483–5491.
- (15) Voityuk, A. A. Charge Transfer in DNA: Hole Charge Is Confined to a Single Base Pair due to Solvation Effects. *J. Chem. Phys.* **2005**, *122*, 204904.
- (16) Butchosa, C.; Simon, S.; Voityuk, A. A. Electron Transfer from Aromatic Amino Acids to Guanine and Adenine Radical Cations in Pi Stacked and T-Shaped Complexes. *Org. Biomol. Chem.* **2010**, *8* (8), 1870–1875.
- (17) Voityuk, A. A.; Siriwong, K.; Rösch, N. Environmental Fluctuations Facilitate Electron-Hole Transfer from Guanine to Adenine in DNA Pi Stacks. *Angew. Chem., Int. Ed.* **2004**, *43* (5), 624–627.
- (18) Cheng, X.; Kumar, S.; Posfai, J.; Pfugrath, J. W.; Roberts, R. J. Crystal Structure of the Hhal DNA Methyltransferase Complexed with S-Adenosyl-L-Methionine. *Cell* **1993**, *74*, 299–307.
- (19) O'Gara, M.; Klimasauskas, S.; Roberts, R. J.; Cheng, X. D. Enzymatic C5-Cytosine Methylation of DNA: Mechanistic Implications of New Crystal Structures for Hhal Methyltransferase-DNA-AdoHcy Complexes. *J. Mol. Biol.* **1996**, *261*, 634–645.
- (20) Steenken, S.; Jovanovic, S. V.; Curie, M.; Kn, C. How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution. *J. Am. Chem. Soc.* **1997**, *119* (3), 617–618.
- (21) Voityuk, A. A.; Jortner, J.; Bixon, M.; Rösch, N. Energetics of Hole Transfer in DNA. *Chem. Phys. Lett.* **2000**, *324* (5–6), 430–434.
- (22) O'Neill, M. A.; Barton, J. K. Effects of Strand and Directional Asymmetry on Base-Base Coupling and Charge Transfer in Double-Helical DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (26), 16543–16550.
- (23) Ridley, J.; Zerner, M. An Intermediate Neglect of Differential Overlap Technique for Spectroscopy: Pyrrole and the Azines. *Theor. Chim. Acta* **1973**, *32* (2), 111–134.
- (24) Voityuk, A. A. Intermediate Neglect of Differential Overlap for Spectroscopy. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3* (5), 515–527.
- (25) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48* (7), 1198–1229.
- (26) Akimov, A. V.; Prezhdo, O. V.; Fields, F. Large-Scale Computations in Chemistry: A Bird's Eye View of a Vibrant Field. *Chem. Rev.* **2015**, *115*, 5797–5890.
- (27) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115* (12), 6217–6263.
- (28) Voityuk, A. A.; Rösch, N. Fragment Charge Difference Method for Estimating Donor–acceptor Electronic Coupling: Application to DNA II-Stacks. *J. Chem. Phys.* **2002**, *117* (12), 5607.
- (29) Singh, J.; Bittner, E. R.; Beljonne, D.; Scholes, G. D. Fluorescence Depolarization in poly[2-Methoxy-5-((2-Ethylhexyl)-oxy)-1,4-Phenylenevinylene]: Sites versus Eigenstates Hopping. *J. Chem. Phys.* **2009**, *131* (19), 194905.

How Abasic Sites Impact Hole Transfer Dynamics in GC-rich DNA Sequences

Marina Corbella,¹ Alexander A. Voityuk,^{2,3} and Carles Curutchet¹

(Submitted)

¹ Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica i Institut de Química Teòrica i Computacional (IQT-CUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona, Espanya.

² Institut de Química Computacional i Catalisi i Departament de Química, Universitat de Girona, Girona, Espanya.

³ Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Espanya..

3.2.1. How Abasic Sites Impact Hole Transfer Dynamics in GC-rich DNA Sequences

Resum

L'excisió d'una base de la seqüència d'ADN es un procés que té lloc centenars de vegades al dia en una cèl·lula humana. Aquest procés pot tenir lloc de manera espontània o degut a agents físics com ara la radiació UV o Y. Generalment, les seqüències que contenen posicions abàsiques, degudes a la excisió d'una base, tendeixen a adoptar conformacions fràgils i inestables. Les cèl·lules estan equipades amb sofisticats mecanismes de reparació, entre ells, l'anomenat BER (*base excision repair*), per tal de prevenir la proliferació de defectes citotòxics. Aquest mecanisme de reparació sol iniciar-se mitjançant la formació d'una posició abàsica a través de l'acció de l'enzim ADN glicosilassa sobre la base danyada. El monitoratge electroquímic a través de canvis en les propietats CT degut a la creació de posicions abàsiques és un dels processos més emprats per a l'estudi dels diferents passos involucrats en el procés de reparació de l'ADN, ja que aquest tipus de lesions soLEN alentir el procés CT global a través de la seqüència. D'altra banda, la modulació de les propietats CT degut a la formació de posicions abàsiques també té aplicacions importants en la síntesi de biosensors. Ara bé, tant la seqüència de bases com la naturalesa de la base desaparellada tenen un impacte important en les diferents conformacions que poden adoptar. S'han portat a terme múltiples estudis, tant experimentals com teòrics, amb l'objectiu de relacionar la naturalesa de la seqüència i de la base desaparellada amb els diferents canvis conformacionals observats. La idea principal d'aquest treball és, en primer lloc, realitzar un estudi conformacional de diferents sistemes model de 15-mer poliG i poliGC, en els que s'ha escindit la base central, ja sigui una guanina o una citosina, per tal de contrastar els canvis estructurals observats amb estudis previs basats en RMN. En segon lloc, es vol investigar l'impacte d'aquestes diferències conformacionals en el procés global de transferència de càrrega. Per fer-ho, s'han combinat simulacions de dinàmica molecular, per tal de captar els diferents canvis conformacionals, amb càlculs QM/MM de les energies i els acoblaments electrònics entre les bases en una seqüència central de cinc parells de bases. Els acoblaments electrònics entre les

bases (i les energies dels estats de càrrega localitzada) s'han obtingut mitjançant el mètode de FCD utilitzant un nivell de teoria semiempíric INDO/S. Posteriorment, s'ha emprat la teoria de Marcus per obtenir les velocitats de transferència entre cada parell de bases, i s'ha utilitzat l'algoritme de Kinetic Monte Carlo per simular el procés global de HT. Els resultats obtinguts a partir de les simulacions de dinàmica molecular concorden amb els diferents estudis de RMN previs, en què l'estructura global de la doble hèlix es manté, de manera que la majoria de processos de transferència es veuen alentits degut a l'absència d'una base en la seqüència. Ara bé, en les seqüències en què la base desaparellada és una citosina flanquejada per citosines, aquesta sol adoptar conformacions fora de la doble hèlix, desestabilitzant l'energia dels diferents cations radicals i reduint-ne l'acoblament. Tots aquests factors, fan que el procés global de HT a través d'aquestes seqüències tingui un comportament asimètric, ja que el procés de HT de 5' a 3' a través de la cadena de guanines es veu alentit con en la majoria de sistemes, mentre que el procés invers de 3' a 5' es manté similar a la seqüència sense lesions. Per tant, s'ha de tenir especial cura amb les seqüències que puguin adoptar conformacions fora de la doble hèlix per al seu ús com a biosensors, o en el monitoratge electroquímic de processos de reparació de l'ADN.



Journal Name

ARTICLE

How abasic sites impact hole transfer dynamics in GC-rich DNA sequences

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Marina Corbella^a, Alexander A. Voityuk^{b,c,*}, and Carles Curutchet^{a,*}

Changes in DNA charge transfer properties upon the creation of apurinic and apyrimidinic sites has been used to monitor DNA repair processes, given that such lesions generally reduce charge transfer yields. However, because these lesions translate into distinct intra and extrahelical conformations depending on the nature of the unpaired base and its DNA context, it is unclear the actual impact of such diverse conformations on charge transfer. Here we combine classical molecular dynamics, quantum/molecular mechanics (QM/MM) calculations and Kinetic Monte Carlo simulations to investigate the impact of abasic sites on the structure and hole transfer (HT) properties of DNA. We consider both apurinic and apyrimidinic sites in polyG and polyGC sequences, and find that most situations lead to intrahelical conformations where HT rates are significantly slowed down due to the energetic disorder induced by the abasic void. In contrast, the presence of an unpaired C flanked by C bases leads to an extrahelical conformation where stacking among G sites is reduced, leading to an attenuation of electronic couplings and a destabilization of hole states. Interestingly, this leads to an asymmetric HT behavior, given that the 5' to 3' transfer along the G strand is slowed down by one order of magnitude while the opposite 3' to 5' transfer remains similar to that estimated for the reference polyG sequence. Our simulations thus suggest that electrochemical monitoring of DNA repair process following changes in charge transfer properties can miss repair events linked to abasic sites adopting extrahelical conformations.

1. Introduction

The loss of a purine or pyrimidine base in DNA, resulting in the creation of an abasic site, is one of the most frequent DNA lesions, with a ratio of formation reaching thousands of times per day in a human cell.¹ Abasic sites are formed by spontaneous or enzymatic hydrolysis of the N-glycosidic bond,^{2–4} by chemical modifications of nucleic bases, or by physical agents like UV or γ radiation.³ Events leading to abasic sites often involve guanines, and the structure and stability of G-rich quadruplex-forming DNA sequences, for example, has been shown to be strongly affected by such defects.^{5–8} In general, sequences containing abasic sites can adopt unusual conformations, making them fragile and susceptible to breaks.^{9,10} Luckily, cells are equipped with an elaborate repair machinery to neutralize the proliferation of cytotoxic and promutagenic defects. Among such repair mechanisms, base excision repair (BER) starts with the formation of an abasic site due to the excision of a damaged or mispaired nucleobase by the action of a DNA glycosilase.^{11,12} The process then continues

with the recognition of the abasic site by an apurinic/apyrimidinic (AP) endonuclease^{13,14} and finishes with the replacement of the missing nucleotide. The study of the different steps involved in the repair process is crucial and may have vast consequences in cancer research.

On the other hand, there is a growing interest in the synthetic incorporation of abasic sites due to its application to the development of biosensors. Abasic site biosensors have been used to detect analytes using fluorescent and electrochemical probes by selectively binding to the unpaired nucleobase,^{15,16} while electrochemical monitoring of DNA repair processes is also possible following the change on the charge transfer (CT) properties associated with the creation of abasic sites.¹⁷ Indeed, the CT properties of DNA have been found to be strongly sensitive to the presence of abasic sites, which in general lead to a decrease in the CT yield.^{18,19} How abasic sites modulate CT kinetics on DNA duplexes, however, has been only briefly explored.²⁰ Many electrochemical studies, for example, have used abasic sites as a brake for the CT process. The Slinker group proved the DNA repair function of two DNA glycosylases by measuring the current though a DNA monolayer after and before the damaged base excision, showing a substantial decrease in redox signal after the formation of the abasic site.¹⁷ This group also demonstrated an attenuation of the CT yield in the presence of an abasic site in a 17-mer AT-rich duplex.²⁰ Given the well-known dependence of CT properties on DNA sequence, the absence of a nucleobase is expected to directly modulate CT rates and

^a Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica and Institut de Química Teòrica i Computacional (IQT-C-UB), Universitat de Barcelona, 08028 Barcelona, Spain.

^b Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

^c Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Campus Montilivi, 17071 Girona, Spain.

† Electronic Supplementary Information (ESI) available: See DOI: 10.1039/x0xx00000x

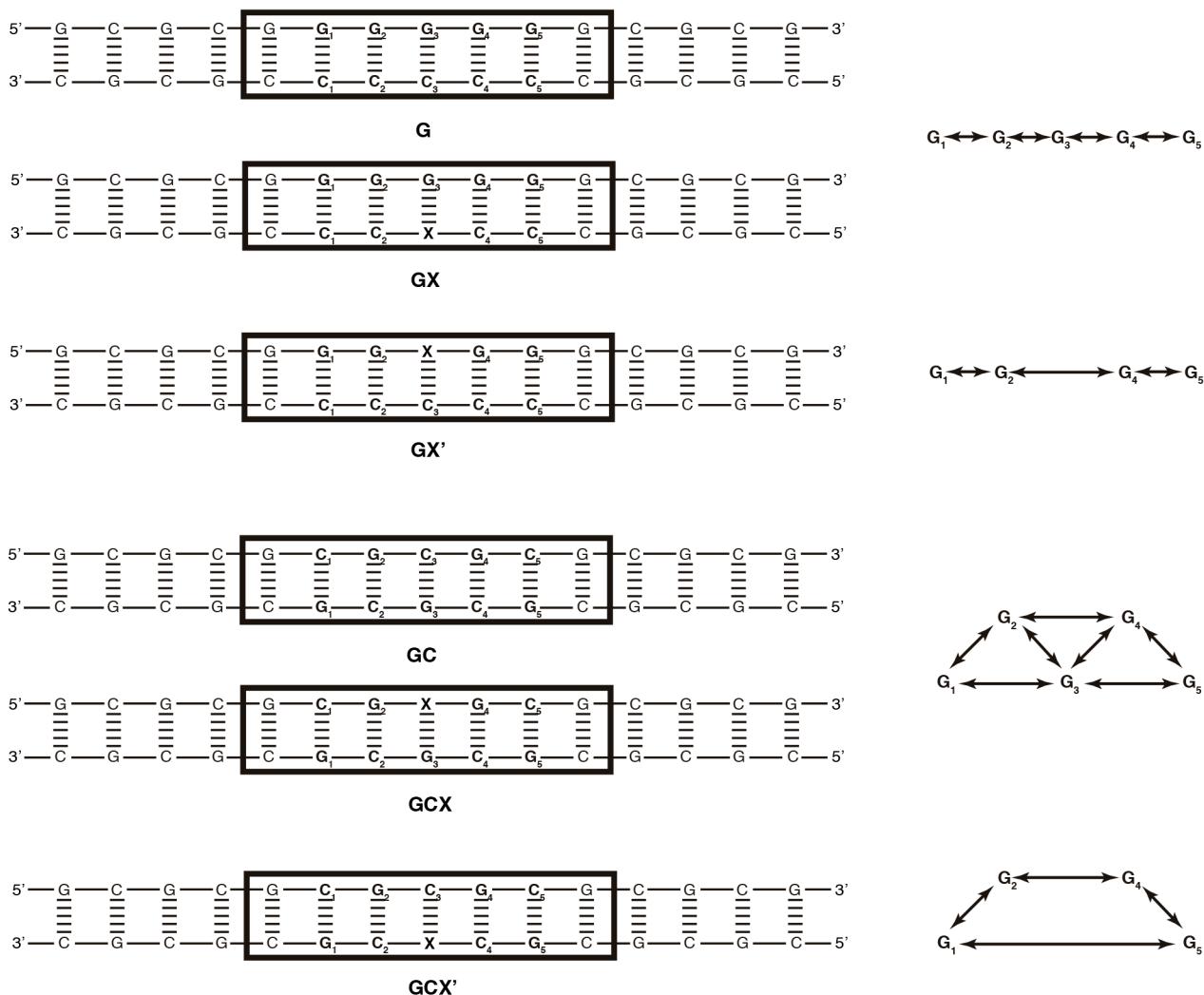


Fig. 1 Sequences of the duplex DNA systems considered (left) and corresponding Kinetic Monte Carlo schemes (right) used to simulate hole transfer dynamics. Nucleobases enclosed in the boxes were included in the quantum-mechanical region in QM/MM calculations.

pathways. However, the unusual conformations induced by the defect can lead to concomitant changes in the CT parameters (site energies and couplings) of proximate nucleobases, thus indirectly modulating CT kinetics. Indeed, a variety of studies indicate that the unpaired nucleobase can adopt intra and extrahelical conformations, depending on the nature of the unpaired base and its context in the DNA sequence.²¹⁻³³

In this study, we combine classical molecular dynamics (MD) simulations with semiempirical INDO/S quantum/molecular mechanics (QM/MM) calculations to investigate how an abasic site influences the structure and hole transfer (HT) kinetics of six model GC-rich DNA duplexes. In particular, we investigate the impact of a guanine (G) or cytosine (C) deletion in 5'-GXG-3' and 5'-CXC-3' contexts and compare our results with the analogous unmodified DNA 15-mers. We aim at understanding how these defects distort the double helix structure, the site energies and electronic couplings of the proximate nucleobases, and the overall HT process, depending on the DNA sequence. Our results indicate that in situations where

the unpaired nucleobase adopts intrahelical conformations, the overall HT process is slowed down by ~1-2 orders of magnitude. However, when an unpaired C adopts an extrahelical conformation the transfer becomes similar to the reference DNA in the 5' to 3' transfer direction along the G strand, but it is slowed down by 1 order of magnitude in the opposite direction, thus introducing a preferential route for oxidative damage.

2. Methods

2.1 MD simulations

MD simulations at room temperature (300 K) were performed for the six 15-mer DNA duplexes shown in Fig. 1. These systems represent polyG and polyGC central sequences where either a G or a C base has been deleted leading to an abasic site. The sequences were capped with GCGC moieties to keep base pairing in the ending regions and thus prevent strand slippage.³⁴ We adopted the prototypical abasic site,

tetrahydrofuran (X). MD simulations were performed starting from standard B-DNA fiber geometries created using the NAB module of the Amber12 software,³⁵ eventually modified in order to delete a C or G nucleobase and generate the abasic site. The systems were neutralized adding Na⁺ ions³⁶ and solvated in a truncated octahedron box of TIP3P water (buffer region 14 Å). Two sets of simulations were run either adopting the parmbsc0³⁷ and the recently developed parmbsc1³⁸ force fields for DNA, the latter introducing some improvements, for example, related to excessive terminal fraying. The tetrahydrofuran residue was modelled using parmbsc0 and parmbsc1 parameters for the phosphate and sugar moieties, the latter capped with a hydrogen, which charge was adjusted to give a global charge of -1. The solvent and counterions were first energy minimized during 1000 steps, followed by 2500 steps of further minimization of the full system. Then, the systems were gradually thermalized up to 300 K during 200 ps with weak constrains in the nucleic acids (10 kcal/mol·Å²), followed by 200 ns production runs at 1 bar. For the GX' system, production runs were extended up to 350 ns. All simulations were done using the Amber 12 suite of programs³⁵ using an integration time step of 2 fs, the SHAKE algorithm to restrain bonds involving hydrogen, periodic boundary conditions, the Particle Mesh Ewald approach, and a nonbonded cutoff of 10 Å. For each system, we extracted a total of 1000 structures at equal intervals from the last 100 ns of MD trajectories to be used as input in QM/MM calculations. In addition, we performed a cluster analysis based on the root-mean square deviations (RMSD) along the last 100 ns of trajectory. From this analysis, a representative structure for each system was extracted from the most populated cluster.

2.2 QM/MM calculations

QM/MM calculations were performed using the semiempirical INDO/S method on the structures extracted from the MD trajectories. Application of Koopman's approximation using this method has been shown to provide a reliable description of CT energies and electronic couplings of stacked nucleobases compared to high-level CASPT2.³⁹ In all cases the QM region included the 7 base pairs in the central region of the duplex (excluding the sugar and backbone atoms), as indicated in Fig. 1, where the C1' atom of the sugar unit was replaced by a hydrogen atom with a 1.08 Å bond length. Electrostatic embedding effects exerted by the environment (DNA, solvent, and counterions) were considered adopting the charges as defined in the force field used for the MD simulations. Hole states associated to the G sites in each duplex were represented by six or seven highest-occupied molecular orbitals (HOMOs) of the neutral species (HOMO → HOMO-5/HOMO-6), while CT couplings were estimated using the Fragment Charge Difference method (FCD):⁴⁰

$$V_{DA} = \frac{(E_2 - E_1)|\Delta q_{12}|}{\sqrt{(\Delta q_1 - \Delta q_2)^2 + 4\Delta q_{12}^2}} \quad (1)$$

For hole transfer (HT), adiabatic splittings $E_2 - E_1$ were estimated using the one-electron energies of the HOMOs of neutral systems. In Eq. (1), Δq_1 and Δq_2 are the difference of hole charges on the donor and acceptor sites in the adiabatic states of interest and Δq_{12} is the corresponding off-diagonal term. The calculation of these quantities and the underlying approximations are described in previous studies.^{40,41}

2.2 Hole transfer kinetics

HT rates for the forward and reverse reactions shown in Fig. 1 were estimated using Marcus theory:

$$k_{HT} = \frac{2\pi}{\hbar} |V|^2 \frac{1}{\sqrt{4\pi k_B T \lambda}} e^{-(\lambda + \Delta G^0)^2 / 4\lambda k_B T} \quad (2)$$

The driving force of the HT reaction (ΔG^0) was estimated from the energy difference in donor/acceptor ionization potentials, and the reorganization energy (λ), which includes internal and solvent terms, was assumed to be 0.7 eV in all cases.⁴² The latter value is a compromise between an experimental estimate 0.6±0.1 eV derived for hole transfer in oligonucleotides⁴³ and values ~ 0.8-1.5 eV obtained in several computational studies.⁴⁴⁻⁴⁶ Whereas a different choice of λ would change the absolute HT timescales we predict, the relative HT rates between standard DNA and the analogous sequences including an abasic site are more robust and essentially unaffected by the value of λ . All rates were computed using the energies and squared couplings averaged over the MD trajectories.

The effective rates corresponding to the overall HT process $G_1 \rightarrow G_5$ and $G_5 \rightarrow G_1$ were simulated using a Kinetic Monte Carlo (KMC) algorithm.⁴⁷ HT rates were used to generate a cumulative probability distribution function (CPD) for each possible step. This function is associated with a transition $i \rightarrow n$ according to Eq. (3), where $\sum_m k_{im}$ is the sum of transfer rates from the same origin i .

$$CPD_{i(n)} = \frac{\sum_{p=1}^n k_{ip}}{\sum_m k_{im}} \quad (3)$$

The Monte Carlo hopping algorithm consists of the following steps: 1) Compute the CPD function for each allowed transfer. 2) Start at a given site i (G_1 or G_5), choose a random number $0 < r_1 \leq 1$ and hop to site j , where $CPD_{i(j)} \geq r_1 > 0$. 3) Compute the waiting time t_w at site i , generating a new random number $0 < r_2 \leq 1$ as $t_w = -(ln(r_2)/\sum_m k_{im})$. 4) Choose a new random number and repeat steps 2 and 3 recording the total accumulated waiting time. 5) Stop if the final site has been reached or if the total time exceeds a fixed value t_{max} . For each system, we averaged HT times over 10^6 realizations of the KMC algorithm.

3. Results and discussion

3.1 DNA structure

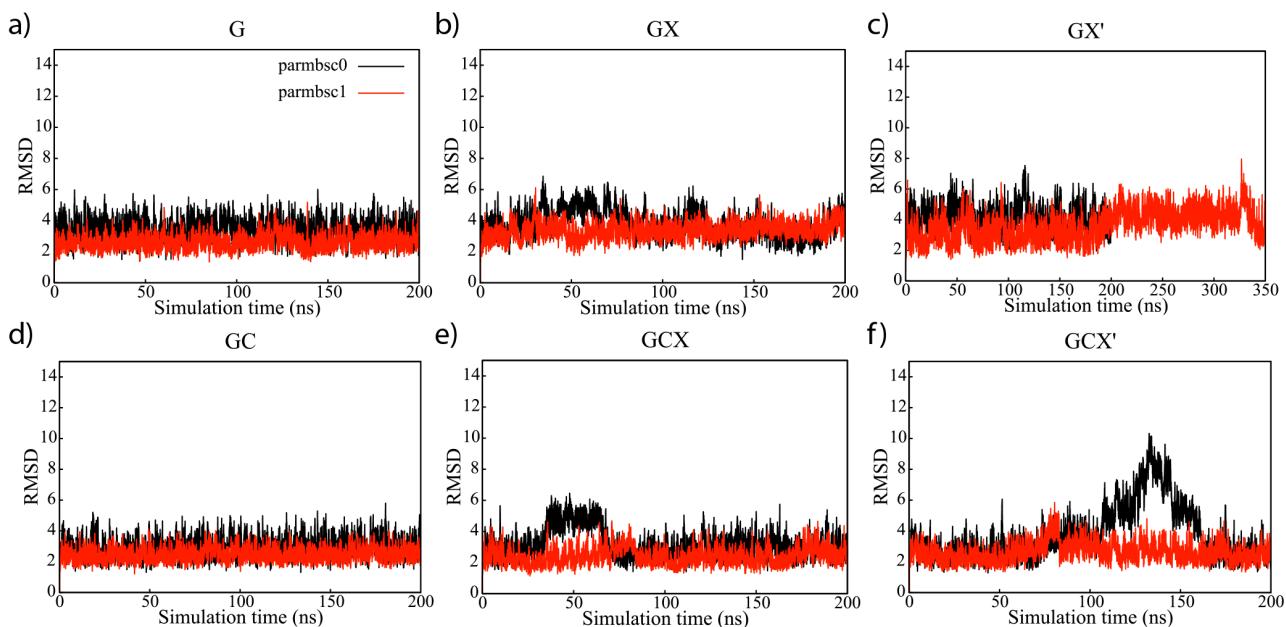


Fig. 2 Positional root-mean square deviations of DNA duplexes along MD trajectories: a) G, b) GX', c) GX, d) GC, e) GCX' and f) GCX.

The impact of abasic sites on the structure and flexibility of DNA has been investigated both using theoretical models and NMR spectroscopy.²¹⁻³³ The bulk of these studies indicate that the overall B-DNA structure of the duplex is maintained in the presence of an abasic site. However, the lesion increases structural flexibility and the local conformation of the unpaired base depends on its nature and its sequence context. Thus, whereas purine unpaired nucleobases adopt intrahelical conformations that preserve π stacking, pyrimidines stack poorly and tend to be extrahelical, although intrahelical conformations have also been identified when the pyrimidine is flanked by two guanines. Thus, it is important to assess the reliability of the structural models obtained here from MD simulations before investigating the impact of abasic sites on hole transfer dynamics.

In Fig. 2 we show the positional RMSD of the DNAs computed with respect to the initial B-DNA structure along the MD simulations performed using either the parmbsc1 or the parmbsc0 force fields. In all cases, neglecting terminal base pairs only lead to minor variations in RMSD values, indicating that terminal fraying was small. If we focus on the standard DNA sequences without the lesion (G and GC systems), simulations performed using the refined parmbsc1 potential leads to RMSD values $\sim 1 \text{ \AA}$ lower than those obtained using the parmbsc0, thus keeping the structures closer to the initial B-DNA. The presence of an abasic site (GX, GX', GCX and GCX' systems) also leads to differences in the structures described by both force fields, again with parmbsc1 displaying more stable structures characterized by lower RMSD values.

For the GCX system, both force fields lead to similar intrahelical position for the unpaired guanine site, as shown in Fig. 3, with the sugar of the abasic site flipping in and out of

the duplex, in accord with NMR data and theoretical simulations performed for an unpaired G in the same CGC^{21,23} or similar CGT contexts.³² When the unpaired guanine is flanked by other G bases (GX system), both force fields again predict intrahelical conformations, in accord with previous studies of unpaired guanines in GGG^{23,33} or GGA sequences.²⁴ In this case, however, the simulation performed using parmbsc1 shows a strong disruption of the Watson-Crick hydrogen bonds between the G₂ and C₂ base pairs adjacent to X, with the unpaired G₃ and C₂ forming non-Watson-Crick hydrogen bonds. This pattern is also briefly exchanged with similar interactions between G₃ and C₄ (the base 5' to X). The observation of this interaction between the unpaired base and a nucleobase adjacent to X agrees with previous predictions for this GGG sequence²³ and was also observed for an unpaired cytosine in a GCA context.²⁴ In a recent study, NMR and constrained MD simulations also pointed to such non-Watson-Crick pattern for the same GGG sequence when a 2'-deoxyribose abasic site was considered, whereas, for its chemically stable analogue tetrahydrofuran, as used here, the standard Watson-Crick pattern was kept.³³ We note however that the sequence of the DNA in that study differs from the one considered here and contained also an additional 8-oxoguanine lesion.

On the other hand, when the unpaired site is a cytosine in a GCG context (GCX' system) the parmbsc1 potential describes a stable intrahelical conformation, in agreement with previous predictions,²³ whereas parmbsc0 leads to both intrahelical and extrahelical conformations characterized by large distortions of the overall stacking pattern. Intrahelical conformations have previously been observed for an unpaired C in a similar GCA context²⁴ or for an unpaired T flanked by guanines.²⁷

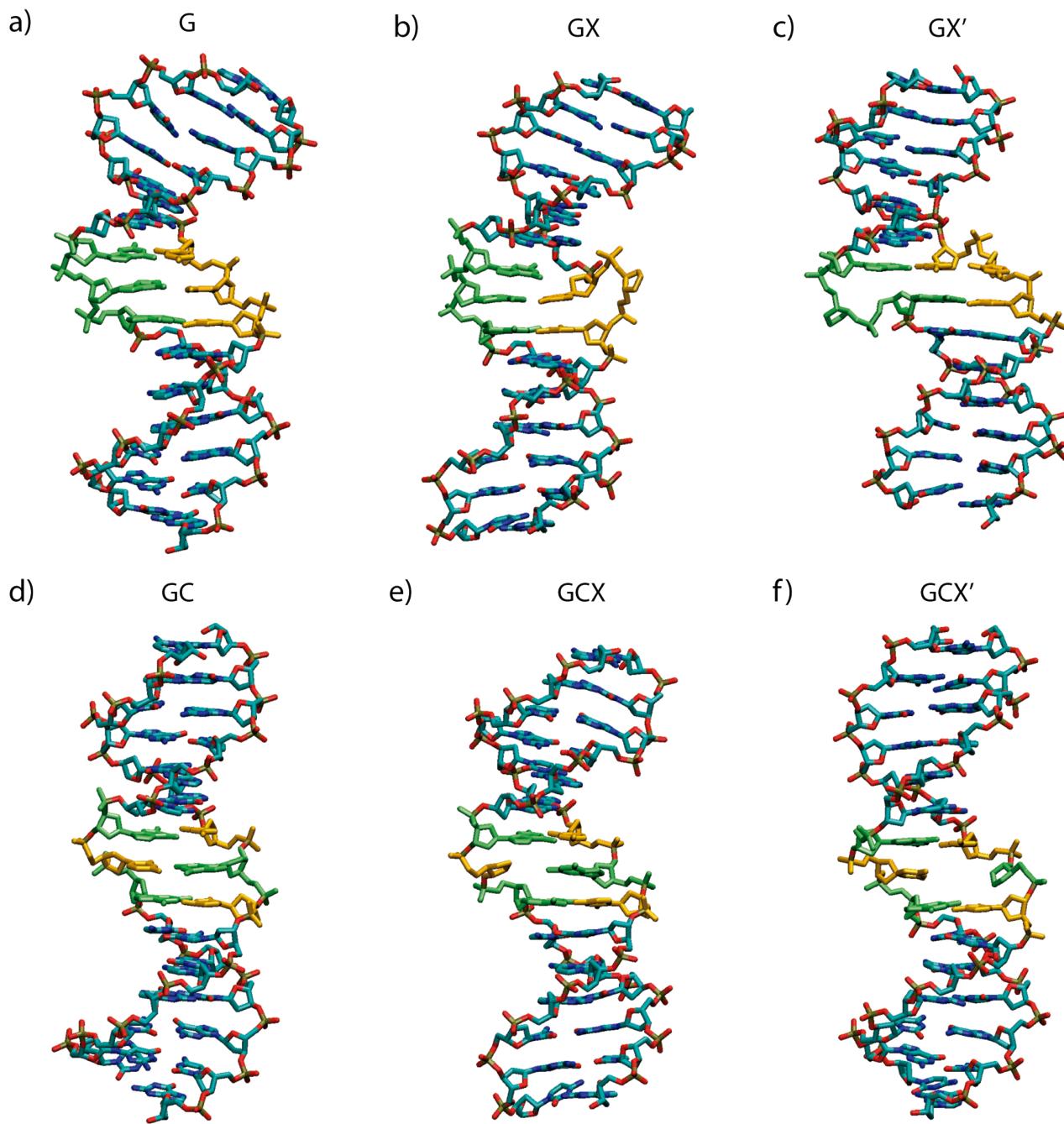


Fig. 3 Representative structures of DNA duplexes extracted from MD simulations based on the parmbsc1 force field: a) G, b) GX, c) GX', d) GC, e) GCX and f) GCX'.

Finally, when the unpaired cytosine is located in a CCC sequence (GX' system), both force fields initially predict a stable intrahelical conformation. However, with the parmbsc1 potential at ~ 180 ns, the orphan base C₃ flips out of the helix, which collapses allowing the flanking bases to stack. More in detail, the carbonyl group of the extrahelical C₃ establishes a hydrogen bond with the amino group of C₂, which keeps its Watson-Crick hydrogen bonds with G₂ thus leading to a triplex-like G₂:C₂:C₃ structure, in which the G₄:C₄ base pair is mainly stacked with the C₂:C₃ pair. Extension of the simulations for a total of 350 ns confirmed the stability of this conformation, in agreement with previous NMR data²¹ and theoretical predictions²³ pointing out to such extrahelical configuration.

Overall, our analysis indicates that the parmbsc1 force field leads to more stable structures that nicely agree with previous theoretical and experimental data on the intra and extrahelical position of the unpaired nucleobase.²¹⁻³³ Moreover, in Fig. 4 we show a superimposition of the backbone of the DNA duplexes studied, indicating that the abasic site in the GX, GX', GCX and GCX' duplexes does not alter the global B-like conformation, in accord again with previous studies. Thus the description of the structural impact of abasic sites seems to be quite robust when the refined parmbsc1 force field is adopted, and in the following, we use the trajectories sampled with this potential to investigate the impact of these lesions on hole transfer dynamics.

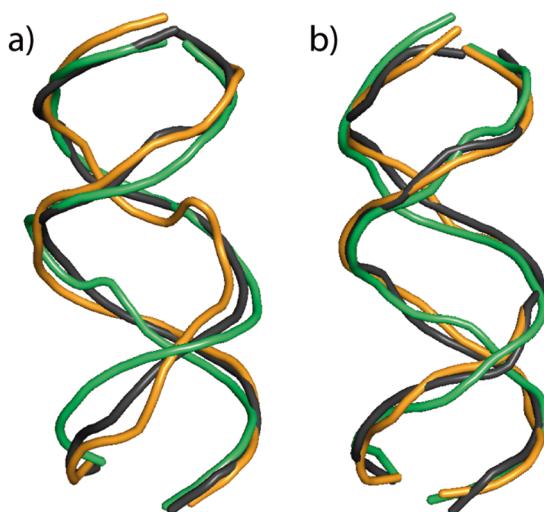


Fig. 4 Backbone overlay of representative structures of DNA duplexes extracted from MD simulations based on the parmbsc1 force field. a) G (black), GX (orange) and GX' (green), b) GC (black), GCX (orange) and GCX' (green).

3.2 Charge transfer kinetics

HT in DNA is known to be rather sensitive to structural deformations of the double helix, especially those modifying the degree of stacking between nucleobases. We have thus computed the energies of the hole states as well as the corresponding HT couplings for the guanines located in the central regions of the duplexes, as shown in Fig. 1, in order to investigate how this lesion affects HT dynamics. In Fig. 4 we show schematically the relative energetic ordering of the different guanine sites in each system, whereas in Table 1 we report the energy differences among sites and the

corresponding electronic coupling parameters and HT times. Based on the HT times for individual transfer steps, we then used KMC simulations to simulate the overall HT times reported in Table 2. We simulated both forward and backward processes, starting from G_1 and recording the time needed to reach G_5 or the inverse transfer process.

In the absence of any lesion, the HT transfer process in the GC system follows several pathways involving both intra and interstrand hops, given the relatively similar energies and electronic couplings involved, although interstrand hops are slightly favored. In this case, the overall forward ($G_1 \rightarrow G_5$) and backward ($G_5 \rightarrow G_1$) times are ~ 2 ns. In the G system the transfer only involves intrastrand hops, characterized by much larger electronic coupling values ~ 100 meV, thus leading to a faster transfer with times 0.2 and 0.04 ns for forward and backward processes.

In the polyGC sequence, the presence of the abasic site (GCX and GCX' systems) does not lead to major structural changes in the double helix, as shown in Fig. 3, with the unpaired base adopting an intrahelical conformation. In the GCX' sequence, indeed, the electronic couplings relevant for the HT process (V_{12} , V_{24} and V_{45}) remain very close to the GC reference system, and the main impact of the lesion is a significant stabilization ~ 150 meV of G_4 with respect of G_2 . This stabilization slows down hops from G_2 to the neighboring sites, and together with the lack of the intermediate G_3 leads to a slower overall time of ~ 300 ns compared to the GC duplex.

In the GCX system, in contrast, the unpaired G_3 is partially displaced toward the abasic site void (see Fig. 3), leading to a better stacking with G_2 and G_4 which increases V_{23} and V_{34} interstrand couplings from 9 to 21 meV and from 24 to 32 meV, respectively, as found previously for a similar sequence.⁴⁸ On the other hand, V_{24} decreases from 8 to 2 meV, because in this case there is no C_3 mediating the coupling between G_2 and G_4 . In this case, however, the abasic site

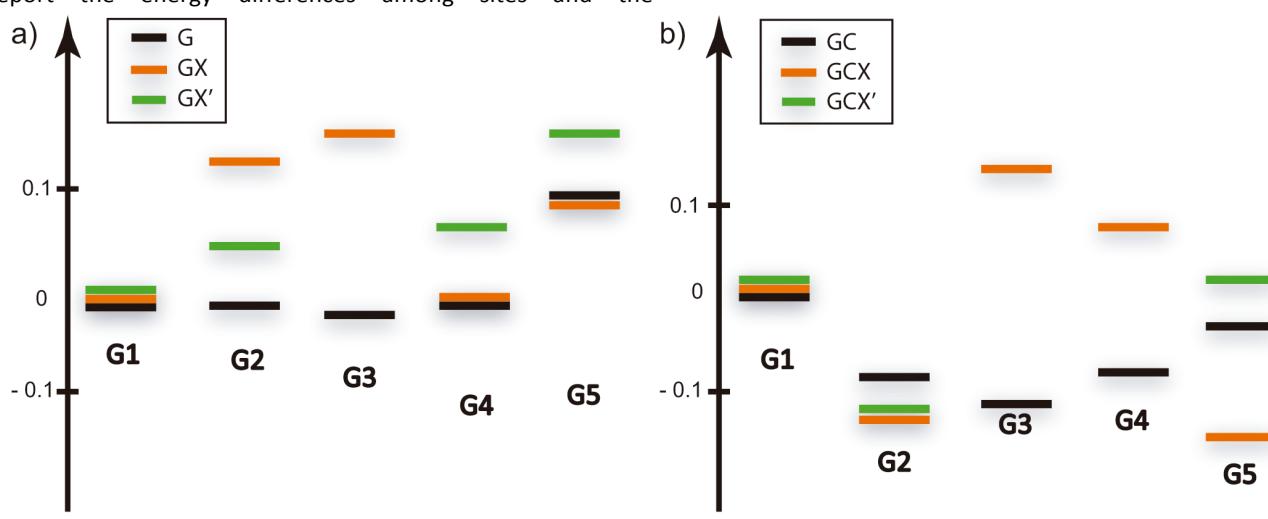


Fig. 5 Representation of the energies computed using QM/MM calculations for the radical cation states localized on guanine sites. a) G, GX and GX' systems, b) GC, GCX and GCX' systems.

induces an important destabilization of the hole states on G_3 and G_4 . Thus, although interstrand couplings are enhanced compared to the reference GC system, the HT process is again considerably slower with total times $\sim 100\text{-}150$ ns due to the larger time needed to access those sites.

Table 1. Electronic couplings, free energy differences and forward and backward HT times computed for the DNA sequences.

	V (meV)	ΔG (meV)	τ_F (ns)	τ_B (ns)
GC				
$G_1 - G_2$	24	-80	1.9×10^{-2}	0.42
$G_1 - G_3$	10	-100	6.9×10^{-2}	3.39
$G_2 - G_3$	9	-18	0.40	0.82
$G_2 - G_4$	8	7	0.83	0.64
$G_3 - G_4$	24	21	0.12	5.4×10^{-2}
$G_3 - G_5$	11	69	1.60	0.11
$G_4 - G_5$	10	54	1.33	0.17
GCX				
$G_1 - G_2$	22	-119	1.1×10^{-2}	1.13
$G_1 - G_3$	6	123	15.59	0.13
$G_2 - G_3$	21	248	29.59	1.9×10^{-3}
$G_2 - G_4$	2	188	1085.78	0.71
$G_3 - G_4$	32	-67	1.23×10^{-2}	0.17
$G_3 - G_5$	9	-261	8.47×10^{-3}	222.07
$G_4 - G_5$	8	-185	3.31×10^{-2}	44.39
GCX'				
$G_1 - G_2$	23	-106	1.2×10^{-2}	0.75
$G_2 - G_4$	9	-153	4.4×10^{-2}	16.54
$G_4 - G_5$	10	275	281	6.3×10^{-3}
$G_1 - G_5$	0.3	13	848	505.82
G				
$G_1 - G_2$	91	-7	4.8×10^{-3}	6.3×10^{-3}
$G_2 - G_3$	91	-10	4.6×10^{-3}	6.6×10^{-3}
$G_3 - G_4$	102	10	5.3×10^{-3}	3.6×10^{-3}
$G_4 - G_5$	86	94	4.3×10^{-2}	1.1×10^{-3}
GX				
$G_1 - G_2$	92	135	9.5×10^{-2}	5.1×10^{-4}
$G_2 - G_3$	73	23	1.3×10^{-2}	5.5×10^{-3}
$G_3 - G_4$	117	-157	2.2×10^{-4}	9.8×10^{-2}
$G_4 - G_5$	106	87	2.4×10^{-2}	8.2×10^{-4}
GX'				
$G_1 - G_2$	94	43	1.2×10^{-2}	2.3×10^{-3}
$G_2 - G_4$	40	27	4.9×10^{-2}	1.7×10^{-2}
$G_4 - G_5$	61	89	7.7×10^{-2}	2.5×10^{-3}

In the G system, the abasic site leads to qualitatively different structural deformations in GX and GX'. As discussed previously, in GX the Watson-Crick hydrogen bonds between the G_2 and C_2 base pairs are disrupted, with the unpaired G_3 forming non-Watson-Crick hydrogen bonds with C_2 3' to the abasic site, as shown in Fig. 3. This conformational change does not alter much the couplings among guanines compared to the G system, which remain close to ~ 100 meV, but leads to an important energetic destabilization of G_2 and G_3 hole states and a corresponding increase in the times needed to reach these sites. Thus, the overall HT times, both for the forward and backward pathways, are increased to ~ 2 ns compared to the 0.04–0.2 ns obtained for the same sequence without the lesion.

Table 2. Overall hole transfer times computed using Kinetic Monte Carlo simulations for $G_1 \rightarrow G_5$ and $G_5 \rightarrow G_1$ transfers.

τ (ns)	G	GX'	GX	GC	GCX'	GCX
$G_1 \rightarrow G_5$	0.21	1.84	2.73	2.20	282.93	104.70
$G_5 \rightarrow G_1$	0.038	0.023	2.71	1.74	302.95	149.26

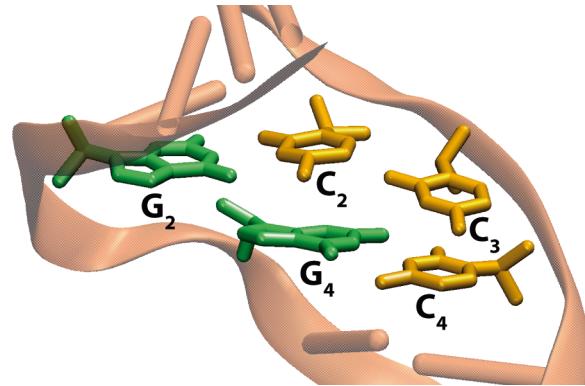


Fig. 6 Structure of the abasic site region for the GX' system.

In the GX' system, on the other hand, the lesion induces an important conformational change, with unpaired C_3 adopting an extrahelical conformation and the sugar of the abasic site being extruded from the helix. As discussed in the previous section, this leads to the triplex like $G_2:C_2:C_3$ structure shown in Fig. 6, in which the $G_4:C_4$ base pair is mainly stacked with the $C_2:C_3$ pair, but also to some degree with the $G_2:C_2$ pair. Because the stacking between G_1 and G_2 is not significantly modified, the V_{12} intrastrand coupling remains similar to the value estimated for G and GX, around 90 meV. The V_{24} and V_{45} couplings, however, are approximately halved compared to the standard ~ 100 meV coupling value for intrastrand G-G contacts because the stacking between G_2/G_4 and G_4/G_5 is reduced for these pairs. This reduced stacking explains the partial destabilization of G_2 , G_4 and G_5 shown in Fig. 4, given the well-known ability of GG motifs to stabilize hole states.⁴⁹ KMC simulations indicate that destabilization of these sites, specially of G_5 , induces a significant asymmetry in overall HT times for the forward $G_1 \rightarrow G_5$ and the backward $G_5 \rightarrow G_1$ process, for which we estimate time constants of 1.8 ns and 23 ps, respectively, compared to 210 ps and 38 ps estimated for the G system without lesion. Overall, thus, the partial disruption of the staking interactions along the G strand in GX' leads to fast forward HT times similar to the reference G system, whereas the backward process is significantly slowed down by an order of magnitude, inducing a directional asymmetry in HT dynamics.

Conclusions

Apurinic and apyrimidinic sites constitute one of the most frequent DNA lesions. Indeed, the repair of other lesions by the BER machinery leads also to the formation of an abasic site after enzymatic hydrolysis of the N-glycosidic bond. Electrochemical monitoring of DNA repair processes is thus

ARTICLE

Journal Name

often performed following the modulation of charge transfer properties associated to the creation of abasic sites. Despite the fact that several studies indicate a general attenuation of charge transfer yield in the presence of abasic sites, the fact that such lesions lead to diverse alterations in the DNA structure depending on the nature of the unpaired nucleobase and the DNA context suggests that the consequences on charge transfer dynamics are also diverse. In this study we have investigated the impact of unpaired cytosine and guanine sites on the structure of polyG and polyGC sequences in atomic detail using MD simulations. Individual HT rates are then derived from Marcus theory based on MD-averaged energies and couplings estimated using QM/MM calculations, and the impact of abasic sites on overall HT dynamics are modeled using a Kinetic Monte Carlo scheme.

We find that MD simulations based on the parmbsc0 force field lead to large structural fluctuations and unusual conformations for the DNA sequences considered. In contrast, those performed using the recently refined parmbsc1 potential point to small alterations on the global B-DNA structure, with abasic sites adopting both intrahelical (GX, GCX and GCX') and extrahelical (GX') conformations, in agreement with previous studies based on NMR spectroscopy and theoretical models. When the DNA adopts intrahelical conformations, the electronic interactions among guanines remain similar to the reference sequence without lesion. However, the abasic void induces a significant disorder in the energies of the hole states, which translate into HT longer times by ~1-2 orders of magnitude. In the GX' sequence adopting an extrahelical conformation, in contrast, the unpaired cytosine and the abasic sugar are extruded from the DNA and the double helix collapses. This leads to a reduced stacking among sites in the G strand, leading to a destabilization of the hole states over G₂, G₄ and G₅ and a significant attenuation of electronic coupling values between these sites. Interestingly, this translates into an asymmetric behavior in HT kinetics, given that the 5' to 3' transfer along the guanine strand (G₁ → G₅) is slowed down by one order of magnitude, whereas HT in the opposite 3' to 5' direction remains similar to that estimated for the reference sequence.

Overall, our results thus show that abasic sites giving rise to intrahelical conformations generally slow down HT dynamics due to energetic disorder induced by the presence of the abasic void, whereas extrahelical conformations lead to less trivial implications on HT dynamics, including asymmetries in the directionality of HT. Thus, our results suggest that electrochemical monitoring of DNA repair process following changes in charge transfer properties can miss repair events linked to abasic sites adopting extrahelical conformations.

Conflicts of interest

There are no conflicts to declare.

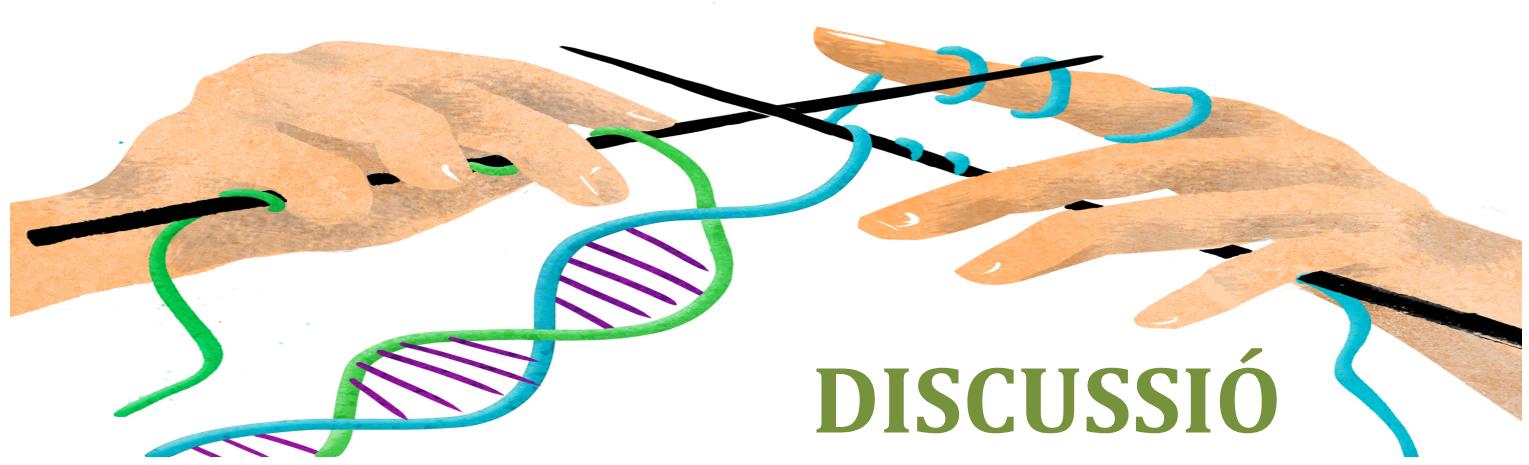
Acknowledgements

We are grateful to the Consorci de Serveis Universitaris de Catalunya for providing access to computational resources (Molecular Recognition project). Financial support from the Agència de Gestió d'Ajuts Universitaris i de Recerca from Generalitat de Catalunya (GENCAT; SGR2014-1189, 2014SGR931, and SGR2017-1746) and from the Spanish Ministerio de Economía y Competitividad (MINECO; grants CTQ2012-36195, CTQ2015-69363-P, and CTQ2017-89924-P) are acknowledged. C.C. is a Serra Húnter Fellow (Generalitat de Catalunya) and M.C. acknowledges a fellowship from the Spanish Ministerio de Economía y Competitividad (MINECO; grant BES-2013-064088).

References

- 1 S. Boiteux and M. Guillet, *DNA Repair (Amst.)*, 2004, **3**, 1–12.
- 2 T. Lindahl and B. Nyberg, *Biochemistry*, 1972, **11**, 3610–3618.
- 3 T. Lindahl, *Nature*, 1993, **26**, 709–715.
- 4 B. Demple and L. Harrison, *Annu. Rev. Biochem.*, 1994, **63**, 915–948.
- 5 V. Esposito, L. Martino, G. Citarella, A. Virgilio, L. Mayol, C. Giancola and A. Galeone, *Nucleic Acids Res.*, 2010, **38**, 2069–2080.
- 6 P. Školáková, K. Bednárová, M. Vorlíčková and J. Sagi, *Biochem. Biophys. Res. Commun.*, 2010, **399**, 203–208.
- 7 A. Virgilio, L. Petraccone, V. Esposito, G. Citarella, C. Giancola and A. Galeone, *Biochim. Biophys. Acta - Gen. Subj.*, 2012, **1820**, 2037–2043.
- 8 J. Beckett, J. Burns, C. Broxson and S. Tornaletti, *Biochemistry*, 2012, **51**, 5257–5268.
- 9 J. Malina and V. Brabec, *Chem. - A Eur. J.*, 2014, **20**, 7566–7570.
- 10 M. M. Greenberg, *Acc. Chem. Res.*, 2014, **47**, 646–655.
- 11 H. E. KROKAN, R. STANDAL and G. SLUPPCHAUG, *Biochem. J.*, 1997, **325**, 1–16.
- 12 O. D. Schärer and J. Jiricny, *BioEssays*, 2001, **23**, 270–281.
- 13 G. Barzilay and I. D. Hickson, *BioEssays*, 1995, **17**, 713–719.
- 14 H. Gattuso, E. Durand, E. Bignon, C. Morell, A. G. Georgakilas, E. Dumont, C. Chipot, F. Dehez and A. Monari, *J. Phys. Chem. Lett.*, 2016, **7**, 3760–3765.
- 15 J. K. Ahn, K. S. Park, B. Y. Won and H. G. Park, *Biosens. Bioelectron.*, 2015, **67**, 590–594.
- 16 Y. Pang, Z. Xu, Y. Sato, S. Nishizawa and N. Teramae, *ChemBioChem*, 2012, **13**, 436–442.
- 17 M. a. McWilliams, F. H. Anka, K. J. Balkus and J. D. Slinker, *Biosens. Bioelectron.*, 2014, **54**, 541–546.
- 18 A. K. Boal, E. Yavin, O. A. Lukianova, V. L. O'Shea, S. S. David and J. K. Barton, *Biochemistry*, 2005, **44**, 8397–8407.
- 19 M. C. Buzzeo and J. K. Barton, *Bioconjug. Chem.*, 2008, **19**, 2110–2112.
- 20 M. a. McWilliams, R. Bhui, D. W. Taylor and J. D. Slinker, *J. Am. Chem. Soc.*, 2015, **137**, 11150–11155.
- 21 P. Cuniasse, G. V. Fazakerley, W. Guschlbauer, B. E. Kaplan and L. C. Sowers, *J. Mol. Biol.*, 1990, **213**, 303–314.
- 22 C. A. Gelfand, G. E. Plum, A. P. Grollman, F. Johnson and K. J. Breslauer, *Biochemistry*, 1998, **37**, 7321–7327.
- 23 L. Ayadi, C. Coulombeau and R. Lavery, *Biophys. J.*, 1999, **77**, 3218–3226.
- 24 D. Barsky, N. Foloppe, S. Ahmadia, D. M. Wilson III and A. D. MacKerell, *Nucleic Acids Res.*, 2000, **28**, 2613–2626.
- 25 M. P. Singh, G. C. Hill, D. Peoc'h, B. Rayner, J.-L. Imbach and J. W. Lown, *Biochemistry*, 1994, **33**, 10271–10285.
- 26 I. Goljer, S. Kumar and P. H. Bolton, *J. Biol. Chem.*, 1995, **270**, 22980–22987.

- 27 Y. Coppel, N. Berthet, C. Coulombeau, C. Coulombeau, J. Garcia and J. Lhomme, *Biochemistry*, 1997, **36**, 4817–4830.
- 28 K. Y. Wang, S. A. Parker, I. Goljer and P. H. Bolton, *Biochemistry*, 1997, **36**, 11629–11639.
- 29 R. D. Beger and P. H. Bolton, *J. Biol. Chem.*, 1998, **273**, 15565–15573.
- 30 L. Ayadi, D. Forget, A. Martelli, J.-F. Constant, M. Demeunynck and C. Coulombeau, *Theor. Chem. Accounts*, 2000, **104**, 284–289.
- 31 S. T. Hoehn, *Nucleic Acids Res.*, 2001, **29**, 3413–3423.
- 32 J. Chen, F.-Y. Dupradeau, D. A. Case, C. J. Turner and J. Stubbe, *Nucleic Acids Res.*, 2008, **36**, 253–262.
- 33 J. Zálešák, M. Lourdin, L. Krejčí, J.-F. Constant and M. Jourdan, *J. Mol. Biol.*, 2014, **426**, 1524–1538.
- 34 R. Y. Au, K. S. Ng, L. M. Chi and S. L. Lam, *J. Phys. Chem. B*, 2012, **116**, 14781–14787.
- 35 D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, I. T.E. Cheatham, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P.A. Kollman, *AMBER 12*, University of California, San Francisco, 2014.
- 36 D. E. Smith and L. X. Dang, *J. Chem. Phys.*, 1994, **100**, 3757–3766.
- 37 A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophys. J.*, 2007, **92**, 3817–3829.
- 38 I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case and M. Orozco, *Nat. Methods*, 2016, **13**, 55–8.
- 39 A. A. Voityuk, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2013, **3**, 515–527.
- 40 A. A. Voityuk and N. Rösch, *J. Chem. Phys.*, 2002, **117**, 5607–5616.
- 41 N. Rösch and A. A. Voityuk, in *Top. Curr. Chem.*, 2004, vol. 237, pp. 37–72.
- 42 K. Siriwong and A. A. Voityuk, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 780–794.
- 43 W. B. Davis, S. Hess, I. Naydenova, R. Haselsberger, A. Ogrodnik, M. D. Newton and M.-E. Michel-Beyerle, *J. Am. Chem. Soc.*, 2002, **124**, 2422–2423.
- 44 D. N. LeBard, M. Lilichenko, D. V. Matyushov, Y. A. Berlin and M. A. Ratner, *J. Phys. Chem. B*, 2003, **107**, 14509–14520.
- 45 K. Siriwong, A. A. Voityuk, M. D. Newton and N. Rösch, *J. Phys. Chem. B*, 2003, **107**, 2595–2601.
- 46 T. Kubář and M. Elstner, *J. Phys. Chem. B*, 2009, **113**, 5653–5656.
- 47 J. Singh, E. R. Bittner, D. Beljonne and G. D. Scholes, *J. Chem. Phys.*, 2009, **131**, 194905.
- 48 M. Corbella, A. A. Voityuk and C. Curutchet, *J. Phys. Chem. Lett.*, 2015, **6**, 3749–3753.
- 49 H.-A. Wagenknecht, S. R. Rajski, M. Pascaly, E. D. A. Stemp and J. K. Barton, *J. Am. Chem. Soc.*, 2001, **123**, 4400–4407.



4. Discussió

Al llarg d'aquesta tesi doctoral s'ha explorat el rol de l'entorn en la modelització de processos CT i EET en sistemes biològics a través de dues línies d'investigació complementàries. D'una banda, s'ha investigat com l'entorn heterogeni complex és capaç de modificar les propietats espectrals i el procés EET en sistemes fotosintètics de recaptació de llum. D'altra banda, s'ha estudiat l'impacte de les fluctuacions estructurals, degut a agents externs o defectes en la seqüència, en el procés de transferència de càrrega a través de l'ADN.

En les últimes dècades s'han produït grans avenços pel que fa al desenvolupament de mètodes i a la capacitat de càlcul. Avui en dia és possible calcular excitacions de cromòfors a través de mètodes QM amb una gran precisió. Ara bé, quan es volen estudiar excitacions en sistemes biològics complexes s'ha de recórrer encara a la utilització de mètodes híbrids QM/MM per tal d'incloure l'efecte de l'entorn. La majoria de les implementacions d'aquests mètodes però, estan basades només en interaccions electrostàtiques entre la part QM i la part MM a través de l'ús de models basats en càrregues puntuals. Aquests mètodes han donat excel·lents resultats en una àmplia varietat de sistemes biològics, no obstant, són incapços de descriure, per exemple, els efectes de la reorganització electrònica que té lloc després d'una transició electrònica degut a la absorció d'un fotó.

D'altra banda, en la modelització de certs processos de transferència, un model estàtic, per acurat que sigui, no és suficient ja que cal incloure també els efectes dinàmics presents en qualsevol entorn biològic. Aquests efectes dinàmics són importants tant des d'un punt de vista estructural, ja que l'acoblament electrònic entre donador i acceptor depèn estretament de la seva orientació i separació relatives, com des d'un punt de vista vibracional. La recent hipòtesis de que els efectes de coherència persistents observats en certs sistemes fotosintètics,²²⁻²⁴ estan relacionats amb l'acoblament vibrònic a certes freqüències que coincideixen amb les diferències d'energies entre estats excitònics,³¹ ha posat de manifest la

importància d'incloure d'una manera acurada, tant els efectes estàtics com dinàmics de l'entorn.

L'objectiu principal que ha perseguit aquesta tesi doctoral ha sigut investigar l'impacte de l'entorn, així com avaluar i millorar els mètodes de simulació actuals per tal de descriure'n l'efecte, en relació a processos CT i EET en sistemes biològics. Els elements claus en tals simulacions són les energies dels estats electrònics involucrats, els acoblaments electrònics i els acoblaments vibrònics. Per fer-ho s'han emprat diverses estratègies computacionals en funció del sistema i dels objectius de cada treball en particular.

4.1. Efectes de l'entorn en processos EET en ficobiliproteïnes

En la primera línia d'investigació d'aquesta tesi s'ha estudiat l'efecte de l'entorn en els paràmetres necessaris per a la simulació de processos EET en sistemes fotosintètics. L'estudi i la comprensió d'aquests processos a nivell molecular és fonamental per a la construcció d'anàlegs artificials de gran eficiència. Davant la complexitat d'aquests sistemes i la impossibilitat d'usar mètodes purament quàntics, la tria d'estratègies basades en càlculs QM/MM, o més concretament QM/MM polaritzables, ha guanyat força durant els últims anys. Per això, per a l'estudi d'aquestes ficobiliproteïnes, en el transcurs d'aquesta tesi s'han emprat dues estratègies diferents, ambdues basades en la metodologia QM/MMPol recentment desenvolupada.⁵⁷

Concretament, s'han estudiat cinc ficobiliproteïnes diferents, la PE545, àmpliament analitzada anteriorment en el grup,^{27,61,115} i quatre nous sistemes, PC577, PC612, PC630 i PC645, l'estructura dels quals ha estat resolta en els darrers anys.³² Aquests quatre sistemes, estructuralment molt similars i amb pràcticament el mateix tipus de pigments, presenten dues estructures quaternàries diferenciades, degut a la inserció d'un sol aminoàcid que provoca una rotació de ~73° d'una part de l'estructura, generant una configuració "oberta" amb la presència d'un canal interior, en front de la configuració "tancada" observada fins aleshores en aquests

tipus de sistemes.³² Una de les característiques més rellevants és que aquest canvi conformacional és capaç de modular el efectes de coherència persistents en el temps observats en aquests i en d'altres sistemes fotosintètics de recaptació de llum, que es creu que poden tenir un impacte crucial en el procés de transferència d'energia.²²⁻²⁴

Dins d'aquesta línia d'investigació s'han portat a terme tres treballs diferenciats que tenen com a element comú aprofundir en la comprensió del funcionament d'aquestes cinc proteïnes recol·lectores de llum. En un primer treball s'ha estudiat la interacció entre les vibracions de l'entorn i les energies de transició dels diferents pigments presents en el complex PE545, mitjançant la metodologia mencionada anteriorment MD-QM/MMPol desenvolupada conjuntament entre el nostre grup i el de la Dra. Mennucci de Pisa.⁵⁷ A posteriori, s'ha analitzat l'origen molecular de les vibracions amb un acoblament major, ja que són les que més impacte tindran en el procés EET. En un segon treball, s'han estudiat els estats de protonació dels diferents pigments que es troben en aquests cinc ficobiliproteïnes, essencials per a una simulació realista de les propietats espectrals. Finalment, s'han calculat els ingredients necessaris per a la descripció dels processos de transferència d'energia en els complexes (PC577, PC612, PC630 i PC645) a través de dues metodologies diferents. La primera, basada en simulacions de MD i posteriors càlculs QM/MM d'estats excitats sobre les trajectòries (MD-QM/MMPol). La segona, en optimitzacions QM/MM dels pigments i posteriors càlculs multiescala d'estats excitats en presència d'un solvent continu (QM/MMPol/ddCOSMO). A partir dels resultats obtinguts s'ha realitzat la simulació dels diferents espectres d'estat estacionari i s'han comparat amb els espectres experimentals a fi de validar els diferents models i interpretar-ne els pics principals.

4.1.1. Anàlisi de les funcions de densitat espectral en el complex PE545

Les funcions de densitat espectral descriuen la força de l'acoblament entre una excitació electrònica i les vibracions de l'entorn. Així, són un paràmetre fonamental a l'hora de descriure el paper de l'entorn en processos EET. Recentment, la

implementació de la tècnica d'espectroscòpia electrònica bidimensional ha permès observar efectes de coherència persistents en el temps en la ficobiliproteïna PE545,²² entre d'altres complexos fotosintètics.¹¹⁶ Aquests descobriments han posat de manifest la importància de descriure de manera acurada aquestes interaccions amb l'entorn, ja que es creu que l'acoblament a vibracions amb una freqüència que coincideix amb la diferència d'energia entre estats excitònics del sistema pot ser la responsable dels efectes de coherència mencionats. No obstant, la obtenció experimental d'aquestes funcions de densitat espectral és encara complicada, especialment en complexos multicomofòrics, on es perd la informació individual de cada pigment.³⁰

Com s'ha comentat anteriorment, aquestes funcions de densitat espectral s'han obtingut mitjançant simulacions de dinàmica molecular, a partir de les quals s'han realitzat càlculs QM/MMPol de les energies d'excitació dels diferents pigments cada 5 fs. A partir de les fluctuacions en el temps de les diferents energies d'excitació s'han obtingut les funcions d'autocorrelació per a cada un dels pigments, i a partir d'aquestes s'han simulat les funcions de densitat espectral individuals. En comparar la funció de densitat espectral obtinguda per a tot el sistema amb una funció de densitat espectral obtinguda experimentalment,^{30,117,118} s'ha observat com la zona de baixa freqüència d'ambdues funcions, corresponents a la part continua de la SD, concorden de manera raonable. Pel que fa als pics corresponents a la part discreta de l'espectre, s'hi observen intensitats i amplituds similars malgrat discrepàncies en les posicions dels pics. No obstant, la forma com ha estat obtinguda la funció experimental no dóna una imatge realista de la posició dels pics degut a que prové d'un ajust empíric a l'espectre d'emissió. Altrament, la regió d'alta freqüència de l'espectre simulat no és capaç de proporcionar una representació fidel, degut a la ja coneguda sobre població dels modes d'alta freqüència en simulacions clàssiques. Tanmateix, l'energia de reorganització calculada a partir de l'espectre simulat (557 cm^{-1}) casa de manera excel·lent amb el valor provinent de l'espectre experimental (539 cm^{-1}).

En analitzar les diferents SDs i els diferents pigments presents en el complex PE545 s'ha pogut comprovar l'estreta relació que existeix entre l'estructura

química de cada pigment i la seva SD. El complex PE545 està constituït per vuit pigments, dos DBVs i sis PEBs, dos dels quals es troben doblement units a l'estructura proteica. Com es pot observar a la Fig. 9, en els DBVs la conjugació del sistema π s'estén en tres dels quatre anells de pirrole, mentre que en els PEBs només en dos. Aquest augment en la conjugació del pigment fa que l'estat excitat sigui més sensible a les deformacions estructurals i per tant, que tant les fluctuacions de la funció d'autocorrelació com la distribució de pics de la SD siguin majors. D'altra banda, aquest efecte també s'observa en els resultats dels PEBs doblement units a la proteïna, que per contra es troben més restringits i presenten menors oscil·lacions. Aquestes diferències en la distribució de pics entre pigments es fan evidents especialment a la zona de baixa freqüència, fet que suggereix que les fluctuacions en aquest regió estan més relacionades amb l'entorn local de cada pigment.

Per tal d'aprofundir en l'origen molecular dels pics més rellevants de les diferents SD s'ha realitzat un anàlisi dels diferents modes vibracionals. Concretament, s'han portat a terme dos tipus d'anàlisis, per un cantó dels modes normals de vibració, que representen les fluctuacions al voltant d'un mínim energètic i donen una imatge de les vibracions intrínseques de cada pigment. Per l'altre, dels modes quasi-harmònics, que representen les fluctuacions al voltant de la conformació mitjana obtinguda al llarg de la simulació de MD. Per tant, mentre que l'anàlisi de modes quasi-harmònics dóna una idea de les vibracions responsables dels diferents pics presents en les SD, la comparació d'ambdues distribucions dóna una imatge de quins modes vibracionals estan promoguts o atenuats degut a la interacció amb l'entorn. A través d'aquesta comparació s'ha observat que entre 1000-1500 cm^{-1} , hi ha una sèrie de modes vibracionals que es veuen atenuats degut a la interacció amb la proteïna, mentre que al voltant de 500 cm^{-1} s'observa un increment en la distribució de modes vibracionals associats a aquesta interacció. Els modes associats a la zona de baixa freqüència $\sim 500 \text{ cm}^{-1}$ estan regits bàsicament per torsions globals de l'estructura tetrapirròlica o torsions que afecten el grau de planaritat d'alguns dels anells de pirrole. Aquestes modes vibracionals modifiquen el grau de conjugació π , de manera que s'acoblen més intensament amb les transicions $\pi-\pi^*$. Aquest fet que es veu reflectit en una major

intensitat dels pics a la funció de densitat espectral, especialment en el cas dels DBVs, als quals la deslocalització π s'estén sobre un anell extra de pirrole. Aquesta diferència en l'estructura dels DBVs, amb la presència d'un doble enllaç C=C extra en un dels anells, també s'ha associat a la major distribució de modes vibracionals en aquests pigments a la regió de $\sim 1450 \text{ cm}^{-1}$, corresponents a estiraments $\nu(\text{C-C})$ i $\nu(\text{C-N})$.

D'altra banda, el dissolvent forma part de l'anomenat “bany” i sol jugar un paper important en la modulació de l'acoblament vibrònic. Per aprofundir en aquesta qüestió s'ha realitzat el mateix anàlisi de modes quasi-harmònics sobre una simulació de MD del complex en buit, per tal de comparar ambdues distribucions. S'ha observat que la presència del dissolvent augmenta la distribució de modes bàsicament a les regions de mitja i alta freqüència, mentre que l'atenuació a la regió entre $1000\text{-}1500 \text{ cm}^{-1}$ mencionada anteriorment, sembla ser fruit de la interacció amb el dissolvent. No obstant, tot i observar-se diferències en l'exposició al dissolvent per part dels diferents pigments, no s'observa un impacte en les distribucions individuals dels diferents modes. Per tant, es podria conoure que el dissolvent impacte en els modes vibracionals dels diferents pigments d'una manera indirecta, a través de les vibracions de la proteïna.

Pel que fa a la dinàmica de transferència d'energia, s'han investigat les implicacions de la utilització d'una funció de densitat espectral global o l'ús de les diferents funcions individuals per a cada pigment. Per fer-ho s'han considerat dos subconjunts de pigments dins de la proteïna i se n'ha estudiat la dinàmica de transferència en ambdues situacions mitjançant càlculs no pertorbatius a través d'equacions jeràrquiques del moviment.⁴⁹ S'han dividit els càlculs en un subconjunt d'alta energia, format pel PEB_{50/61C}, el PEB_{50/61D} i el DBV_A, i un de baixa energia format pel PEB_{82C}, el PEB_{82D} i el DBV_A, mentre que per les energies i els acoblaments electrònics s'han emprat els valors ajustats anteriorment en la ref.¹¹⁵. S'ha observat que els diferents pigments del subconjunt d'alta energia presenten un acoblament important a un mode vibracional al voltant de 1050 cm^{-1} , coincident amb la diferència d'energia entre els estats excitònics quasi-localitzats en el PEB_{50/61D} i el DBV_A, de manera que aquest mode conduirà la transferència

d'energia entre aquests dos estats. D'altra banda, en el subconjunt de baixa energia s'observa l'acoblament a un mode al voltat de ~ 500 cm $^{-1}$ coincident amb la diferència d'energia entre els estats excitònics quasi-localitzats en el PEB_{82C} i el PEB_{82D}. En representar la dinàmica de la població dels diferents estats excitònics en els dos subconjunts s'observa la presència d'oscil·lacions que desapareixen en considerar la mateixa SD en els diferents pigments.

Per tant, els nostres resultats recolzen la implicació d'efectes de coherència en la transferència d'energia en la ficobiliproteïna PE545 observats experimentalment, al mateix temps que posen de manifest la necessitat de determinar les funcions de densitat espectral a nivell atòmic, per tal d'incloure-hi els efectes de l'entorn heterogeni complex que es veuen difuminats a nivell experimental.

4.1.2. Determinació de l'estat de protonació de les ficobilines presents en els complexes PE545, PC577, PC612, PC630 i PC645

Malgrat el gran nombre d'estudis publicats sobre la relació entre l'entorn proteic i les propietats de recaptació de llum en aquests tipus de sistemes, l'estat de protonació de les diferents bilines que els constitueixen no havia estat encara estudiat en detall. Fins al moment, diversos treballs teòrics han considerat indistintament un estat desprotonat o protonat al sistema tetrapirrole, més enllà de l'estat aniónic de les dues cadenes laterals àcides que també presenten a l'estructura. La hipòtesi de l'estat protonat es veia recolzada per certs indicis suggerits per mapes de densitat electrònica d'alta resolució, així com per la presència d'un residu d'àcid aspàrtic o glutàmic de la proteïna interaccionant amb el grup tetrapirrole en la majoria d'aquests pigments. No obstant, la determinació precisa dels diferents pK_a s d'aquests pigments no només és important per l'efecte que pot tenir l'estat de protonació en els estats excitats sinó que, a diferència d'altres organismes fotosintètics que tenen les seves proteïnes antena unides a la membrana tilacoide, aquestes ficobiliproteïnes es troben suspeses en el lumen, on el pH pot variar entre $\sim 5\text{-}7$ en funció de l'exposició a la llum solar.¹¹⁹ Així, en funció del pK_a de les bilines l'estat de protonació i les propietats EET d'aquests complexes es podria veure afectat per aquests canvis en el grau d'exposició a la llum.

La determinació dels diferents estats de protonació en aigua d'aquests pigments és difícil a nivell experimental, degut a la pèrdua de la conformació bioactiva en solució. Així, primer s'ha obtingut l'estat de protonació en aigua per a cada pigment a través de cicles termodinàmics. S'ha calculat la variació de l'energia lliure de desprotonació de cada un dels quatre anells de pirrole de cada pigment en fase gas, mitjançant càlculs SCS-MP2/CBS amb correccions CCSD, i posteriorment s'han calculat les corresponents energies de solvatació a través de models de solvatació continus com són l'MST^{53,54} i l'SMD.⁵⁵ Mentre que els càlculs en fase gas s'espera que tinguin una precisió suficient, el càlcul d'energies de solvatació, especialment en el cas d' ions, conté errors associats de l'ordre de varies kcal/mol. Per tant, paral·lelament s'ha validat el protocol computacional emprat en dos compostos heterocíclics aromàtics dels quals es coneix el pK_a experimental, obtenint uns errors absoluts de ~1 unitat de pK_a . Els resultats obtinguts per aquests estats de protonació indiquen clarament que els anells de pirrole més acídics en tots els pigments són els dos anells centrals, amb valors d'entre 6 i 7 unitats de pK_a , mentre que els anells externs que presenten cert caràcter d'amida, presenten pK_a s que oscil·len entre 11 i 31.

La inclusió de l'entorn proteic en el càlcul dels diferents pK_a s és un factor fonamental en la determinació dels estats de protonació dels diferents pigments, ja que és el que marca la diferència entre pigments amb la mateixa estructura que es troben en diverses ficobiliproteïnes. Els valors de pK_a per als diferents entorns proteics s'han obtingut a través del càlcul de l'energia lliure associada al pas d'aigua a l'interior de la proteïna, mitjançant el mètode PROPKA,^{120,121} a partir dels valors en aigua obtinguts prèviament. S'han calculat els diferents pK_a s corresponents únicament als dos anells de pirrole centrals, tant per l'estructura cristall com per diverses estructures al llarg d'una simulació de dinàmica molecular, obtenint resultats més relaxats a partir dels valors mitjanats al llarg de la MD. Els resultats obtinguts s'han dividit en dos grups, els pigments en que els dos anells centrals interaccionen amb un residu d'àcid aspàrtic o glutàmic de la proteïna i els que no. Pel que fa al primer grup s'han obtingut valors de pK_a que oscil·len entre 6 i 8 unitats, mentre que els aminoàcids amb què interaccionen

presenten valors de pK_a entre 3.3 i 6.9, en tots els casos inferiors als valors predicts per les bilines. D'altra banda, el segon grup de pigments presenta valors de pK_a al voltant de ~5.

Tot i els progressos en el desenvolupament de mètodes per al càlcul d'energies lliures, aquests encara porten un error associat que es pot traduir en una desviació d'entre 1-2 unitats de pK_a . Si es considera el rang de pH a l'interior del lumen, entorn fisiològic per aquestes ficobiliproteïnes, els diferents valors de pK_a obtinguts a través dels diferents cicles termodinàmics i l'error associat als mètodes emprats, l'estat de protonació dels diferents pigments al pH de treball segueix essent incert. Per aquest motiu, per tal d'aprofundir en els valor absoluts de pK_a , s'han obtingut els espectres d'absorció per a cada proteïna a diferents valor de pH entre 4.0 i 9.4. La idea principal és identificar variacions en els espectres d'absorció degudes a un canvi en el pH , que es puguin correlacionar amb canvis en l'estat de protonació dels diferents pigments.

En comparar els espectres d'absorció a diferents valor de pH per a cada proteïna, s'ha observat un element en comú, que és l'absència de canvis en el rang de pH que va de 4.6 a 6.5. D'altra banda, en la PC645 s'observa una disminució de la intensitat relativa de la banda de baixa freqüència, consistent amb la desprotonació dels MBVs, als que els manca el residu àcid interaccionant amb els anells de pirrole centrals, i per als quals el pK_a calculat és menor. El mateix patró s'ha pogut observar en la PC630, tot i que en presentar un espectre més congestionat, els canvis observats a la banda de baixa freqüència semblen tenir un impacte més general en l'eixamplament global de l'espectre. Ara bé, aquests canvis en la banda de baixa freqüència no s'observen fins a un $pH \geq 7$, de manera que si s'havien obtingut una valors de $pK_a \sim 5$ per als MBVs, sembla raonable considerar que els càlculs realitzats subestimen els valor d'aquests pigments en al menys 2 unitats de pK_a . Pel que fa a l'espectre de la PE545 i la PC577, ambdues semblen menys sensibles als canvis en el pH , i no és fins a valors superiors a 7 que s'observen petits canvis en els espectres a la zona de més alta freqüència. Aquests canvis, per a la PE545, es poden associar a la desprotonació dels pigments DBVs, que tampoc interaccionen amb cap residu àcid de la proteïna i per als que s'ha estimat un pK_a

de ~5.6, fet que un cop més recolza una subestimació del pK_a d'unes 2 unitats en les nostres simulacions. D'altra banda, pel que fa a la PC577, on tots els pigments interaccionen directament amb un residu àcid de la proteïna, les variacions en l'espectre són menors i es poden relacionar també amb l'estat de protonació dels DBVs, pels que s'ha estimat un valor menor que per la resta de pigments.

En conclusió, els resultats obtinguts experimentalment, suggereixen que les nostres prediccions subestimen el valor de pK_a dels diferents pigments en ~2 unitats, ja que no s'observen canvis apreciables en els espectres d'absorció fins al voltant de pH 7, consistents amb la desprotonació dels DBVs i MBVs als que els manca el residu àcid de la proteïna interaccionant amb els dos anells de pirrole centrals. Per tant, els nostres resultats recolzen estudis prèvis en que les bilines presents en aquestes proteïnes antena s'han considerat en la seva forma completament protonada en el pH de treball del lumen (5-7), tot i que certa fracció dels DBVs i dels MBVs es pot trobar en la seva forma desprotonada a un $pH \sim 7$. Ara bé, l'estat de protonació dels pigments no sembla que pugui tenir conseqüències en les propietats de captura de llum d'aquests complexos en condicions fisiològiques.

4.1.3. Propietats de captura de llum dels complexes PC577, PC612, PC630 i PC645

Recentment s'ha descobert una nova estructura quaternària en FBPs, en què la inserció d'un sol residu esdevé en una estructura "oberta", on la major separació entre els dos pigments centrals provoca una reducció de l'acoblament excitònic entre aquests, fet que es tradueix en una atenuació de les coherències observades.³² Això ha fet augmentar l'interès en el desenvolupament de models teòrics capaç de descriure de manera precisa els processos EET en aquests sistemes, a fi d'establir les bases moleculars que sustenen tals efectes de coherència. La correcta descripció d'aquests processos rau essencialment en una correcta descripció de l'escala energètica dels diferents pigments que les constitueixen, així com dels acoblaments entre ells i amb l'entorn. La determinació d'aquestes energies ha estat un dels grans reptes durant els últims anys.^{27,34,61,122} En aquesta tesi s'ha estudiat les energies relatives en quatre PBPs, de les quals

dues parelles comparteixen la mateixa composició en quant a pigments, PC577/PC612 i PC630/PC645, amb dues estructures quaternàries diferents, oberta i tancada, respectivament. Aquests dos conjunts de proteïnes permeten comparar tant les diferències espectrals degudes a canvis en la seqüència proteica, com les diferències ocasionades degut al canvi en l'estructura quaternària.

S'han emprat dues estratègies, que difereixen bàsicament en el tractament de les geometries internes dels diferents cromòfors. Per un cantó s'ha emprat la metodologia MD-QM/MMPol (secció 1.2.1.4.), mitjançant la qual es van obtenir resultats satisfactoris anteriorment en el grup per al sistema PE545.^{61,115} Per l'altra banda, s'ha emprat una estratègia basada en una prèvia optimització dels diferents pigments a nivell QM/MM, seguida de càlculs multiescala dels estats excitats i els acoblaments electrònics a nivell QM/MMPol/ddCOSMO. En aquests últims, el solvent ha estat descrit mitjançant un nou algoritme per a la resolució de les equacions del model COSMO, anomenat ddCOSMO,⁶⁸ que permet el tractament de biosistemes de mida considerable ja que escala linealment amb el tamany. Per tal d'avaluar la precisió dels resultats obtingut, s'han simulat els diferents espectres d'estat estacionari (absorció, emissió i dicroisme circular), per contrastar-los amb els diferents espectres experimentals.

Com ja s'ha comentat anteriorment, l'acoblament vibrònic és també un element clau a l'hora de simular els diferents espectres, de manera que s'han calculat les funcions de densitat espectral per a cada pigment mitjançant el mètode del gradient vertical (secció 1.2.2.3.2.). Els resultats obtinguts mostren una tendència similar per als diferents tipus de pigments que constitueixen les quatre proteïnes, tot i què els DBVs presenten acoblaments majors que la resta, obtenint un rang global per a l'energia de reorganització (λ) al voltant de 410-690 cm⁻¹. No obstant, el model falla en la obtenció de les SD d'alguns dels pigments, degut probablement a errors en la determinació del mínim correcte d'energia en les optimitzacions QM/MM realitzades a l'interior de proteïna. Cal remarcar que els MBVs són els pigments que presenten més inconsistències en el valor de λ , essent també els pigments més conjugats i per als que l'obtenció d'un mínim energètic podria exagerar aquest efecte, en neglidir els efectes tèrmics.

Pel que fa a les energies d'excitació dels pigments, l'ús del mètode basat en MD clarament sobreestima les diferències d'energia entre els diferents tipus de bilines (DBVs i PCBs) en tots els casos, mentre que els càlculs basats en les geometries optimitzades QM semblen aportar resultats en un rang d'energies relatives més raonable. Aquesta sobreestimació sembla ser fruit de l'ús d'un *force field* clàssic en la MD, fet que induceix errors sistemàtics diferents en pigments amb diferent estructura. Pel que fa a l'energia dels MBVs, que són els pigments més conjugats, hom esperaria que fóssin els pigments de més baixa energia. Ara bé, mentre que mitjançant els càlculs basats en MD semblen estar al mateix rang d'energia que els PCBs de més baixa energia, els càlculs basats en les optimitzacions QM, tot i situar-los en la posició de més baixa energia, clarament n'infraestimen el valor, probablement degut novament a negligir els efectes tèrmics i exagerar-ne el grau de conjugació en la geometria optimitzada.

En vista de tot el conjunt de resultats, sembla evident que els DBVs constitueixen els pigments de més alta energia, però la posició espectral tant dels MBVs com les energies relatives entre diferents PCBs és més complicada de determinar. Pel que fa a les dues proteïnes que no contenen MBVs, la PC577 i la PC612, les dues cadenes β són idèntiques, de manera que les diferències espectrals observades provenen únicament degut a les diferències en les cadenes α que contenen els pigments PCB_{20A/C}. Per tant, el canvi, a més baixa energia, del màxim d'absorció observat entre la PC577 i PC612, ha de ser degut a una estabilització dels pigments PCB_{20A/C}. Si es comparen els dos mètodes emprats per a la PC577 i la PC612 no sembla clar quins haurien de ser els pigments de més baixa energia (PCB_{82B/D}, PCB_{158B/D}), ja que en els càlculs basats en les MD corresponen als PCB_{158B/D}, mentre que pels càlculs basats en les optimitzacions QM corresponen als PCB_{82B/D}. D'altra banda, per als complexes PC630 i PC645 que contenen també MBVs, els càlculs basats en les optimitzacions QM prediuen energies en l'ordre MBV_{19A/C} < PCB_{82B/D} < PCB_{158B/D} < DBV_{50/61/B/D}, mentre que els càlculs basats en les MD prediuen un rang d'energies massa similar i difícil de discernir MBV_{19A/C} \approx PCB_{158B/D} \approx PCB_{82B/D} < DBV_{50/61/B/D}. Els càlculs basats en geometries QM, però, donen un eixamplament de banda molt més coherent amb els valors experimentals, especialment en quan a

la diferència d'energia entre DBVs i PCBs, tot i que l'energia dels MBVs sembla lleugerament infraestimada. L'ordre energètic predit per aquestes simulacions, a més, concorda amb els resultats recentment publicats pel grup del Prof. Coker sobre la PC645: $\text{MBV}_{19\text{A/C}} \approx \text{PCB}_{82\text{B/D}} < \text{PCB}_{158\text{B/D}} < \text{DBV}_{50/61\text{B/D}}$.³⁴

Per tant, els càlculs basats en les optimitzacions QM semblen aportar resultats més raonables pel que fa a les energies relatives, donant un mateix ordre a totes 4 FBPs per als pigments PCBs i DBVs situats a la cadena β , fortament conservada en aquestes algues: $\text{PCB}_{82\text{B/D}} < \text{PCB}_{158\text{B/D}} < \text{DBV}_{50/61\text{B/D}}$. Així, la diversitat espectral sembla originar-se bàsicament per la posició dels pigments presents en la cadena α , molt més divergent a nivell evolutiu. Aquests pigments semblen donar lloc a bandes en un rang d'energia mig en la PC577 i la PC612 ($\text{PCB}_{20\text{A/C}}$) i en la zona de baixa energia en la PC630 i PC645 ($\text{MBV}_{19\text{A/C}}$).

4.2. Efectes de l'entorn en processos CT en l'ADN.

En aquesta segona línia d'investigació s'ha estudiat l'impacte de l'entorn en processos CT a través de l'ADN. En particular, degut a la interacció amb proteïnes o a la introducció de defectes estructurals tant comuns com l'excisió d'una base de la seqüència. Per fer-ho, s'han emprat simulacions de dinàmica molecular per tal d'incloure les fluctuacions dinàmiques de l'entorn. Sobre diverses estructures extretes al llarg de la MD s'han realitzat càlculs puntuals QM/MM a fi d'obtenir uns valors mitjanats per a les energies dels cations radicals formats sobre les guanines i per als acoblaments electrònics entre els diferents parells de guanines. Per als acoblaments electrònics s'ha emprat el mètode FCD⁹⁵ a un nivell de teoria semiempíric INDO/S. Tot i el seu caràcter semiempíric, s'ha demostrat, en base a càlculs d'alt nivell CASPT2,⁹³ que el mètode INDO/S juntament amb l'aproximació de Koopmans és capaç de descriure de manera realista les energies i acoblaments electrònics que defineixen el procés HT entre les bases nitrogenades d'una seqüència d'ADN. Posteriorment, s'ha aplicat la teoria de Marcus als valors mitjanats sobre la MD per a l'estimació de les velocitats HT entre els diferents parells de bases de la seqüència. A partir de les velocitats individuals, s'ha simulat

el procés global de transferència de càrrega a través de la seqüència d'interès aplicant l'algoritme de KMC. Aquest procediment ha permès obtenir les velocitats HT des del punt inicial al punt final preestablerts per a cada sistema, al mateix temps que ens ha proporcionat la distribució de “camins” més favorables a través de la seqüència de bases.

Concretament, en el transcurs d'aquesta tesi s'ha estudiat el procés HT en el complex DNA-metiltransferasa *HhaI* i en diverses seqüències model d'ADN riques en parells de bases G-C. L'enzim metiltransferasa *HhaI* s'uneix de manera específica a una seqüència de la doble hèlix per portar a terme una reacció de metilació a una citosina. Per fer-ho extreu momentàniament la base de la cadena i inserta en el seu lloc la cadena lateral d'un residu de glutamina. Aquesta interrupció en l'apilament entre les bases de la seqüència frena el procés global de HT, reduint l'acoblament entre les guanines anterior i posterior a la inserció, al mateix temps que en desestabilitza els cations radicals formats. No obstant, si s'uneix a un enzim mutant Q237W, que en lloc d'un residu de glutamina inserta un residu aromàtic de triptòfan, el procés HT a través de la seqüència d'unió no només no es frena, sinó que es veu accelerat. Paral·lelament, s'ha estudiat també l'efecte de la excisió d'una base de la seqüència d'ADN, en primer lloc en la mateixa seqüència d'ADN del complex ADN-*HhaI* i posteriorment en una sèrie de sistemes model.

4.2.1. Sistemes d'ADN en els que s'ha escindit una base de la seqüència.

La excisió d'una base de la seqüència d'ADN no és un procés aïllat, pot tenir lloc bé de manera espontània, degut a agents físics com la radiació UV o γ , o bé mitjançant una acció enzimàtica, generalment com a punt de partida en el procés de reparació d'una base danyada.¹²³ Comprendre els mecanismes de reparació de l'ADN és un pas fonamental en la recerca contra el càncer, i un dels mètodes utilitzat en aquest context consisteix en el seguiment electroquímic del procés de reparació fent us de les propietats de transport d'electrons a través de la seqüència de bases.¹²⁴⁻¹²⁶ Ara bé, mentre que la transferència de càrrega a través de l'ADN ha estat àmpliament estudiada, com es veu afectat aquest procés en absència d'una base és encara un

camp en desenvolupament. En primer lloc, des d'un punt de vista estructural, l'excisió d'una base dóna lloc a conformacions inusuales i molt variables en funció de la seqüència de bases, mentre que aquestes deformacions estructurals afecten dràsticament les interaccions π - π entre les diverses bases apilades, modificant-ne així el procés HT.

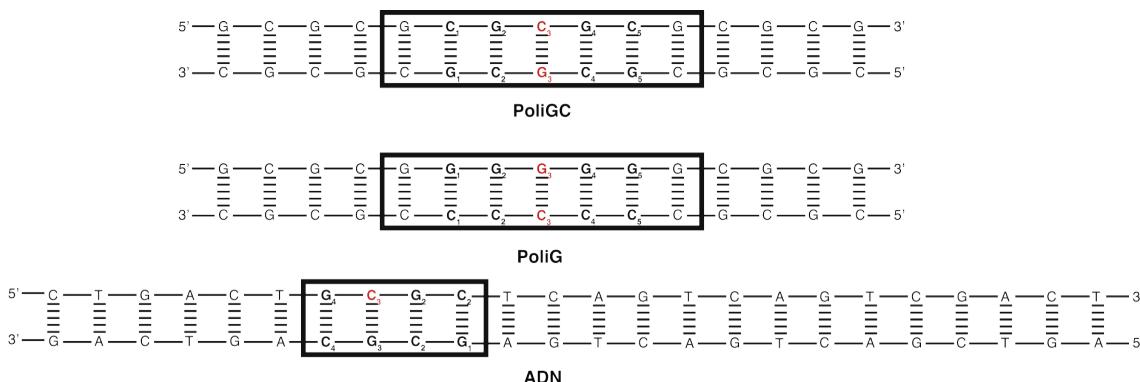


Figura 13. Seqüències estudiades on es ressalta la part que s'ha descrit mitjançant un mètode QM i en vermell les bases que es poden escindir en funció del sistema.

4.2.1.1. Paràmetres estructurals

Per tal d'estudiar la deformació estructural que experimenta l'ADN davant l'absència d'una base s'han portat a terme simulacions de dinàmica molecular dels diferents sistemes d'interès. En un primer treball es va estudiar la seqüència d'ADN que s'uneix a l'enzim *HhaI*,¹²⁷ però en absència d'aquest enzim, tant la seqüència completa d'ADN com el sistema que es va anomenar ADN-X. En aquest últim s'havia eliminat la citosina que s'extreu en el procés enzimàtic, generant una posició abàsica. Ja en aquest primer sistema es va poder observar, a través d'un ànalisi de RMSD de les posicions atòmiques, com les fluctuacions que experimentava el sistema ADN-X eren molt superiors que les del seu anàleg sense modificar. En examinar-ne la dinàmica en detall, es va observar com l'absència de la citosina permetia que la guanina amb la que s'aparellava (G_3) s'inserís en la posició abàsica facilitant així un apilament millor amb les guanines anterior (G_2) i posterior (G_4) al "forat", i provocant l'expulsió de la citosina amb la que una d'elles s'aparellava inicialment (C_4). Aquest comportament no és inesperat, ja que en les bases púriques (A i G) la deslocalització del sistema π s'estén als nou àtoms dels

anells, mentre que en les bases pirimidíniques (C i T) l'aromaticitat queda limitada als sis àtoms de l'anell de pirimidina, fent que la interacció sigui més pobra.

En el segon treball d'aquesta tesi dins de la present línia d'investigació, es van estudiar diversos sistemes d'ADN model rics en parelles de bases G-C. Concretament es van confeccionar sistemes model de 15-mers poli-G i poli-GC, i els seus anàlegs als quals se'ls va eliminar la guanina o la citosina oposada en la posició central, obtenint un total de sis sistemes. Un cop més, els dos sistemes que contenen totes les bases de la seqüència presenten RMSDs molt més robustos que els seus homòlegs amb una posició abàsica. Pel que fa a un anàlisi purament estructural, varis grups han treballat ja amb la idea d'intentar establir relacions entre la deformació estructural i la naturalesa de les bases desaparellades.¹²⁸⁻¹³¹ La principal hipòtesi és que quan la base desaparellada és del tipus púrica, la conformació natural de la hèlix es manté gràcies a una millor interacció entre les bases anterior i posterior, mentre que quan aquesta és del tipus pirimidínica té tendència a adoptar una conformació fora de la doble hèlix degut a un apilament més pobre, excepte si es troba situada entre dues guanines.

Aquestes observacions, tot i dependre molt de la resta de bases de la seqüència, semblen confirmar-se en els diferents sistemes estudiats. Quan la base desaparellada és una citosina, en el cas del sistema GCX', que es troba flanquejada per dues guanines, la conformació inicial es manté (intrahèlix) i l'anàlisi d'RMSD és força estable. D'altra banda, en el sistema GX' l'anàlisi d'RMSD presenta fluctuacions més importants, ja que durant pràcticament 200 ns la citosina desaparellada, tot i mantenir-se dins de la doble hèlix, es troba en ocasions fora de la cadena, per finalment adoptar una forma tipus "triplex", fora de la doble hèlix, amb la guanina G₂ i la citosina C₂, que es manté durant al menys 150 ns més. Aquesta conformació inusual permet que les guanines que envolten la posició abàsica es tanquin recuperant l'apilament de la cadena, de manera que el grup carbonil de la citosina desaparellada estableix un pont d'hidrogen amb el grup amino de la citosina que la precedeix, quedant fora de la doble hèlix com s'havia observat en estudis anteriors.^{129,130} D'altra banda, pel que fa als sistemes on la base desaparellada és una guanina, la conformació inicial d'aquesta es manté al llarg de

les simulacions, no obstant, en el cas del sistema GX, on la posició abàsica esta flauejada per dues citosines, la interacció entre aquestes es veu afeblida, provocant una major distorsió i el trencament dels ponts d'hidrogen tipus Watson-Crick entre la G₂ i la C₂ adjacents a la posició abàsica, establint-se nous ponts d'hidrogen de tipus no-Watson-Crick entre la guanina desaparellada G₃ i la C₂, patró que ja s'havia observat anteriorment en seqüències similars.^{129,131,132} Respecte al sistema GCX, la zona corresponent a la posició abàsica i les bases posteriors i anteriors, es correspon al sistema estudiat prèviament ADN-X. No obstant, no s'observa el mateix patró, ja que en aquest cas, les citosines que flanquegen la base desaparellada no es veuen desplaçades fora de la cadena quan aquesta s'introdueix lleugerament en el forat de la posició abàsica, en consonància amb estudis previs.^{129,130} Aquest comportament diferenciat, pot ser degut a les diferències en la seqüència de bases, o bé a una descripció errònia del primer sistema ADN-X mitjançant el *force field* parmbsc0,¹³³ ja que en el segon treball, on es compara l'anàlisi estructural d'ambdós camps de forces, s'observa que el nou parmbsc1¹³⁴ descriu estructures més estables i pròximes a l'estruatura B-ADN inicial.

4.2.1.2. Paràmetres HT

La deformació estructural de l'ADN mencionada anteriorment té conseqüències importants en relació a les energies dels cations radicals formats durant el procés HT i en els seus acoblaments electrònics. Així, per tal de capturar aquests paràmetres dinàmics s'han realitzat càlculs QM/MM en diverses estructures al llarg de la simulació MD, els quals han permès obtenir uns valors mitjanats de les energies, dels acoblaments electrònics i de les velocitats de transferència entre guanines. A partir de les velocitats individuals s'han simulat les velocitats de transferència globals mitjançat simulacions KMC. En absència de lesions, en els sistemes GC, tant en el poliGC com en l'ADN del primer treball, s'observen velocitats de transferència al voltant de 2 ns en ambdós sentits, amb una marcada preferència pels "salts" entre guanines de diferents cadenes, en què l'acoblament és superior. D'altra banda, en el sistema poliG, els "salts" només tenen lloc entre les guanines contigües en una sola cadena on l'acoblament és marcadament superior,

de l'ordre de ~100 meV, de manera que el procés de transferència és més ràpid, amb temps de 200 i 40 ps pels processos de $G_1 \rightarrow G_5$ i de $G_5 \rightarrow G_1$, respectivament.

La presència de posicions abàsiques impacta les energies, els acoblament i les velocitats de transferència globals. Pel que fa als sistemes poliGC (GCX i GCX'), la presència de la lesió, tot i no donar lloc a grans canvis estructurals, afecta en les propietats HT d'aquests sistemes. En la seqüència en la que s'ha escindit una guanina, els acoblaments es mantenen similars als de la seqüència de referència, mentre que l'energia del catió radical en la G_4 s'estabilitza ~150 meV respecte el de la G_2 . Aquesta estabilització, juntament amb l'absència de la C_3 , fan que el procés global de transferència es vegi alentit respecte el sistema de referència (300 ns). D'altra banda, en els sistemes en què s'ha escindit una citosina, tant en el sistema GCX com en l'ADN-X, la lleugera inserció de la guanina desaparellada a l'interior de la posició abàsica fa augmentar considerablement els acoblaments V_{23} i V_{34} . Ara bé, les diferents conformacions induïdes degut a la posició abàsica fan que els temps de transferència globals es vegin alterats de diferent manera. En el sistema GCX, la posició abàsica desestabilitza els cations radicals de les guanines G_3 i G_4 , fent que el temps de transferència es vegi augmentat respecte el sistema de referència, i obtenint temps de l'ordre de 100-150 ns. Per contra, en el sistema ADN-X només es veu desestabilitzat el catió radical format en la G_3 respecte el sistema de referència ADN, creant un camí preferent $G_1G_2G_4$ que fa que el procés global de transferència passi de 2.4 a 0.7 ns. Ara bé, cal remarcar, com s'ha comentat anteriorment, que la simulació de MD per el sistema ADN i ADN-X es van portar a terme mitjançant el *force field* parmbsc0, per el que s'han demostrat certes diferències conformacionals amb el nou *force field* parmbsc1 en aquests tipus de sistemes.

Pel que fa als sistemes poliG, les diferents conformacions adoptades impacten de diferents maneres el procés HT. Pel sistema GX en el que s'ha escindit la citosina central, els acoblaments entre guanines no es veuen pràcticament alterats, mentre que els cations radicals corresponents a les G_2 i G_3 es desestabilitzen, fent augmentar el temps de transferència globals al voltant de 2 ns en ambdós sentits. D'altra banda, pel sistema GX' en el que s'ha escindit la guanina central, l'estructura adoptada en forma de "triplex" impacta de forma asimètrica el procés global de HT.

L'apilament parcial de la parella de bases G₄:C₄ tant amb la parella G₂:C₂ com amb la parella C₂:C₃, fan que els acoblaments V₂₄ i V₄₅ es vegin reduïts pràcticament a la meitat en comparació amb l'acoblament observat per el sistema de referència. Aquest apilament parcial provoca també una desestabilització dels diferents cations radicals del sistema induint un procés de HT asimètric, en què el procés de G₅ → G₁ manté una velocitat de transferència igual que la del sistema de referència, mentre que el procés G₁ → G₅ es veu alentit, obtenint temps de transferència al voltant de 2 ns.

En conclusió, davant la creixent aplicació de mètodes de monitoratge electroquímic per a processos de reparació de l'ADN, basats en el seguiment dels canvis en les propietats de transferència de càrrega associats a la creació de posicions abàsiques, s'ha observat que tot i que múltiples estudis apunten cap a una atenuació del procés de transferència en presència de posicions abàsiques, la dependència de la seqüència i la naturalesa de les bases desparellades juga un paper determinant. En els nostres estudis es demostra una atenuació del procés global de HT en tots aquells sistemes en què la l'estruatura global de la seqüència es manté, mantenint-se la base desparellada dins de la doble hèlix. Ara bé, en les situacions en què alguna base de la seqüència es col·loca en una posició fora de la doble hèlix, el procés de transferència es pot veure inclús accelerat. Així, el monitoratge electroquímic dels mecanismes de reparació de l'ADN basat en el seguiment d'aquests processos pot obviar processos de reparació de llocs abàsics en que la seqüència adopti conformacions amb bases fora de la doble hèlix.

4.2.2. Sistemes ADN-proteïna

L'ADN s'uneix a determinades proteïnes per tal de portar a terme funcions tan essencials com són la replicació, la transcripció o la reparació de danys. En algunes ocasions aquesta unió es dóna de manera específica, quan la proteïna s'uneix a seqüències concretes de l'ADN, o no específica, com és el cas dels mecanismes de reparació. En els últims anys s'ha observat que la interacció de l'ADN amb determinades proteïnes és capaç d'alterar-ne el procés de transferència de càrrega, degut a la migració d'aquesta cap a residus aromàtics de la proteïna. S'ha observat,

mitjançant tècniques experimentals, la formació de cations radicals en residus de tirosina i triptòfan provinents de processos iniciats en l'ADN.^{111,112} Com s'ha comentat anteriorment, en aquesta tesi s'ha estudiat el procés HT a través de la seqüència d'unió del complex ADN-proteïna *HhaI*. En estudis experimentals anteriors, s'havia observat que la unió de l'enzim natiu feia desaparèixer la presència de dany oxidatiu a les guanines posteriors al punt d'inserció del residu provenint de la proteïna, mentre que a través de la unió d'un enzim mutant Q237W el dany oxidatiu es restablia.^{14,15} D'altra banda, aquests mateixos experiments revelaven la participació de cations radicals localitzats tant a les guanines com al triptòfan. A través dels mètodes mencionats anteriorment, s'ha estudiat l'impacte de la deformació estructural sobre el procés HT, al mateix temps que s'han determinat les velocitats de transferència, els "camins" preferents i la implicació del triptòfan en el procés HT.

4.2.2.1. Paràmetres estructurals

Com en el cas discutit en el punt 4.2.1.1., mitjançant simulacions de MD s'ha estudiat la deformació estructural ocasionada per la unió de la proteïna a la seqüència d'unió 5'-GCGC-3'. A través d'un anàlisi de RMSD de les posicions atòmiques es va poder observar com, a diferència del sistema ADN-X, les fluctuacions en la regió d'unió eren força suaus, fet que posa de manifest una unió estable en el temps de simulació. També es va realitzar un anàlisi de *Base Pair Step Parameters* (BPSP, Paràmetres de pas de parell de bases) mitjançant el programa X3DNA,¹³⁵ que determinen la orientació mútua entre parelles de bases veïnes, caracteritzades per set paràmetres (tres translacions, tres rotacions i el solapament degut a l'apilament). Aquest anàlisi va permetre observar que el sistema més afectat estructuralment, en comparació al sistema model d'ADN i als paràmetres de referència per la configuració B-ADN,¹³⁶ era l'ADN-X, especialment en el parell de bases CA/TG₄ a causa de la expulsió de la C₄ mencionada anteriorment. D'altra banda, la unió a la proteïna, tot i tenir cert impacte estructural, conserva més la conformació original de la doble hèlix.

No obstant, aquest anàlisi de BPSP només és vàlid en les parelles de bases completes, de manera que paral·lelament es va analitzar en profunditat la

orientació i el grau d'inserció, tant de la cadena lateral de glutamina com de triptòfan. Aquests paràmetres tenen un efecte clau en el procés HT ja que afecten directament les interaccions π - π entre les bases de la seqüència. La hipòtesis formulada pel grup de la Dra. Barton arran dels treballs experimentals realitzats amb el complex *HhaI*^{14,15} postulava que la presència de dany oxidatiu en les guanines posteriors a la inserció de l'aminoàcid en el complex mutant Q237W era deguda a que la cadena lateral de triptòfan omplia l'espai de la citosina extreta, mantenint-ne l'apilament i les interaccions π . Els nostres resultats recolzen aquesta hipòtesis, i indiquen que l'anell de triptòfan i la G₃ competeixen per omplir la posició de la citosina extreta. Durant els primers 60 ns de la simulació l'anell de triptòfan es troba omplint completament el buit de la citosina. No obstant, passat aquest temps tant el triptòfan com la G₃ semblen arribar a una posició d'equilibri, en què ambdós residus es troben ocupant l'espai mencionat, situació que es manté fins al final de la simulació. D'altra banda, pel que fa al complex amb la proteïna nativa, la inserció de la cadena lateral de glutamina es manté estable al llarg de tota la MD gràcies a la formació de tres ponts d'hidrogen entre la glutamina, la G₃ i la Ser87, determinants a l'hora de mantenir l'estructura del complex durant l'acció catalítica.¹³⁷

4.2.2.2. Paràmetres de HT

Tal com s'ha mencionat en l'apartat 4.2.1.2. s'han mitjanat els valors de les energies dels diferents cations radicals i dels acoblaments electrònics al llarg de la simulació de MD per tal de tenir en compte les fluctuacions estructurals dels sistemes. En els nostres resultats s'observa com la unió a la proteïna nativa i la introducció de la cadena lateral de glutamina entre la G₂ i la G₄ provoca una clara disminució de l'acoblament electrònic entre ambdues bases al interrompre's l'apilament, igual que entre aquestes i la G₃, que passen de ~30 a ~12 meV (V₂₃), de ~7 a ~2 meV (V₂₄) i de ~8 a ~4 meV (V₃₄). També les energies dels cations radicals contigus a la inserció dels residu de glutamina es veuen fortament desestabilitzats, especialment la G₄. Ambdós factors apunten cap a una clara reducció de la velocitat de HT a través de la seqüència d'unió, que explica la desaparició del dany oxidatiu observada experimentalment a les guanines posteriors a la inserció de la cadena lateral de glutamina. A través de les

simulacions de KMC a partir de les velocitats obtingudes en aplicar la teoria de Marcus, s'ha observat com les velocitats de HT de la G₂/G₃ a la G₄ es veuen especialment reduïdes, fet que culmina amb una disminució de la velocitat de HT global a través de tota la seqüència d'unió, que passa de ~2.41 (ADN) a ~178 ns. Si s'analitzen els resultats de les simulacions de KMC s'observa importants diferències respecte el sistema ADN-X no complexat a la proteïna. Al sistema d'ADN, per exemple, totes les bases tenen una energia similar i per tant hi ha una gran varietat de camins. Al sistema ADN-X, per contra, la desestabilització de la G₃ dóna clara preferència al camí G₁G₂G₄ (54%). En el cas de l'ADN-X unit a la proteïna nativa, la desestabilització de la G₄ juntament amb la reducció de l'acoblament electrònic entre G₂/G₃ i G₄, i la imposició del pas per G₄ com a punt final de l'algoritme, proporciona també una gran varietat de camins que contribueixen al procés de HT, en els que la càrrega migra de G₂ a G₃ i viceversa, ja que la transferència a G₄ queda clarament desfavorida, essent el camí preferent G₁G₃G₄ (1%).

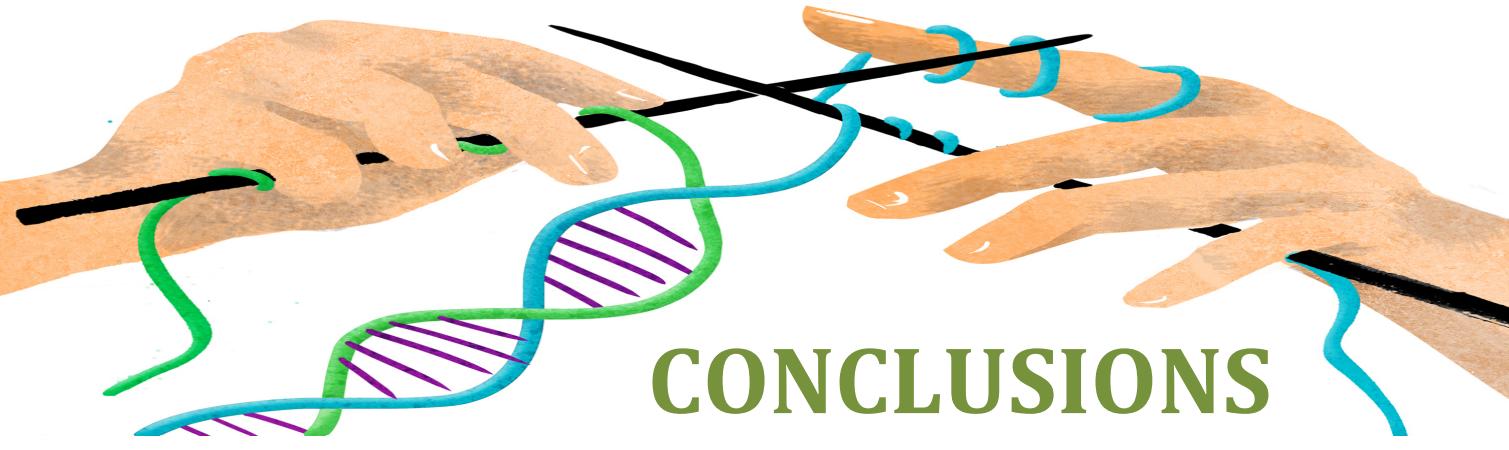
D'altra banda, la unió a la proteïna mutant provoca un increment en els valors resultants dels acoblaments electrònics, especialment V₂₄ i V₃₄ si es comparen amb els valors obtinguts per al sistema d'ADN, que passen de ~7 a ~44 meV i de ~8 a ~16 meV, respectivament. La variació en l'acoblament entre la G₂ i la G₄ és degut probablement a la millora de les interaccions π-π que proporciona la substitució de la citosina, en què la càrrega està deslocalitzada al voltant de l'anell de piridina de sis àtoms, per un anell de triptòfan en el que la deslocalització es pot estendre al voltant dels nou àtoms dels anells. Tanmateix, l'augment, en menor mesura, de l'acoblament V₃₄ es pot deure a la competència mencionada anteriorment entre la G₃ i la cadena lateral de triptòfan inserida, veient-se millorat l'apilament entre la G₃ i la G₄. Pel que fa a les energies, igual que en el sistema ADN-X, l'energia de la G₃ es veu clarament desestabilitzada, de manera que en simular el procés de HT s'obtenen resultats similars amb un clara acceleració en el procés si es compara amb el sistema model d'ADN o amb el complex amb la proteïna nativa, arribant a l'ordre de ~220 ps. Igual que el sistema amb l'ADN-X, la desestabilització de la G₃ i l'augment de l'acoblament electrònic V₂₄ dóna una preferència evident al camí format per G₁G₂G₄ en un 90% de les realitzacions. En conjunt, tots aquests

resultats donen suport als resultats experimentals en que s'observava la reaparició de dany oxidatiu a les guanines posteriors al punt d'inserció de l'aminoàcid en el cas de la proteïna mutant, i a la hipòtesis de que l'origen d'aquest comportament és degut a la regeneració i a l'enfortiment de les interaccions π - π , mitjançant l'anell de triptòfan, interrompudes amb la inserció del residu de glutamina en unir-se a la proteïna nativa.

Per tal de estudiar i corroborar la implicació del catió radical de triptòfan en el procés de HT en el cas de la unió a la proteïna mutant, s'ha repetit la metodologia emprada en un únic residu de guanina i de triptòfan en aigua. Com s'ha comentat, el model semiempíric INDO/S utilitzat juntament amb l'aproximació de Koopman's ha estat contrastat amb càlculs CASPT2 per a l'obtenció de les energies i els acoblaments electrònics en el cas de bases apilades. A partir dels valors obtinguts per aquests residus aïllats, s'ha comprovat que els nostres resultats sobreestimaven en ~ 0.61 eV l'energia del catió radical de triptòfan en comparar-los amb els resultats experimentals dels potencials d'oxidació de la guanina i el triptòfan a pH neutre. La simulació del procés global de HT en el complex amb la proteïna mutant amb la participació del triptòfan prediu una velocitat de HT inferior a la observada quan aquest no hi juga cap paper, però tot i així, superior que en el cas de la unió a la proteïna nativa (~ 39.2 ns). Aquest alentiment és degut a que l'energia del triptòfan es troba molt per sota de la de la resta de guanines, fet que, tot i el fort acoblament observat entre el triptòfan i les guanines contigües, dificulta la transferència a la G₄. Tanmateix, aquests resultats concorden amb els diferents resultats experimentals, en què s'observa la formació d'un cató radical de triptòfan en el procés de HT, al mateix temps que es postula que el procés de HT és veu accelerat en unir-se a la proteïna mutant Q237W.

En resum, l'estrategia emprada, que combina simulacions de MD amb càlculs QM/MM utilitzant un mètode semiempíric INDO/S juntament amb l'aproximació de Koopman's i el mètode de FCD, és capaç de proporcionar un conjunt d'energies i acoblaments electrònics per a les diferents bases d'una seqüència d'ADN. Aquests paràmetres poden ser utilitzats a posteriori per obtenir unes velocitats de HT a través de varíes bases d'una seqüència, que proporcionen una imatge realista i

acurada del procés de HT en diversos sistemes d'ADN no convencionals, i ADN-proteïna.



A stylized illustration of two hands, one brown-skinned and one light-skinned, reaching towards each other from opposite sides. They are holding a large, colorful DNA double helix. The DNA is composed of four strands: a green strand on the left, a blue strand with purple diagonal stripes in the center, a black strand at the top, and a teal strand on the right. The hands are positioned as if they are about to meet or are already joined, symbolizing genetic connection or research.

CONCLUSIONS

5. Conclusions

Anàlisi de les funcions de densitat espectral en el complex PE545

- Les funcions de densitat espectral simulades per a cada pigment en el complexe PE545 presenten diferències importants en comparació amb la funció de densitat espectral mitjanada sobre tots els pigments, impactant de manera rellevant la dinàmica de EET.
- L'acoblamet a modes vibracionals de baixa energia és el que es veu més afectat per les diferències en els entorns locals de cada pigment.
- Les simulacions MD-QM/MMPol proporcionen resultats raonables per a l'energia de reorganització (557 cm^{-1}) en comparació amb els resultats experimentals (539 cm^{-1}).

Determinació de l'estat de protonació de les bilines presents en els complexes PE545, PC577, PC612, PC630 i PC645.

- Els $pKas$ estimats per a les bilines en aigua indiquen que la desprotonació es produeix en els anells de pirrole centrals, amb valors de pKa 5.8 – 7.4.
- L'entorn proteic té un impacte important en el pKa de les bilines, donant lloc a diferències en les propietats de protonació d'un mateix pigment en diferents llocs a les ficobiliproteïnes estudiades.
- Els $pKas$ calculats semblen estar infraestimats ~2 unitats de pKa , ja que no s'observen variacions apreciables en els espectres experimentals d'absorció fins al voltant de pH 7, essent aquests canvis atribuibles a la desprotonació de MBVs i DBVs no coordinats a un Glu o un Asp.
- Els resultats suggereixen que totes les bilines als complexes PE545, PC577, PC612, PC630 i PC645 es troben protonats al rang de pH fisiològic ~5-7 del lumen.

Propietats de captura de llum en els complexes PC577, PC612, PC630 i PC645.

- L'ús d'un camp de forces clàssic a l'estrategia MD-QM/MMPol induceix errors sistemàtics significatius entre diferents tipus de pigments, resultant en un rang d'energies relatives massa ampli.
- L'estrategia basada en optimitzacions QM de les geometries dels pigments aporta resultats més raonables, tot i que la no inclusió d'efectes tèrmics exagera el grau de conjugació dels pigments de més baixa energia infraestimant-ne les energies relatives.
- En conjunt, els càlculs basats en optimitzacions QM suggereixen una mateixa escala energètica en les 4 PBPs estudiades per als pigments situats en la cadena β , fortament conservada en aquestes algues: $\text{PCB}_{82\text{B/D}} < \text{PCB}_{158\text{B/D}} < \text{DBV}_{50/61\text{B/D}}$. Així, els pigments i la seqüència de les cadenes α semblen ser l'origen de la variabilitat espectral en aquests complexos.

Efectes de l'entorn en processos de CT en sistemes ADN-proteïna.

- La unió de l'enzim metiltransferassa *HhaI* a l'ADN té un impacte important en el procés HT, ja que la inserció de la Gln desestabilitza l'energia dels cations radicals de les guanines contigües inhibint el dany oxidatiu a llargues distàncies.
- La unió a un enzim mutant que inserta un residu aromàtic de Trp enllloc de la Gln, per contra, estabilitza l'energia dels cations radicals de les guanines contigües a la inserció i n'augmenta l'acoblament electrònic, accelerant el procés global d'HT.
- La participació directa del catió radical Trp resulta en una velocitat d'HT inferior a la observada quan aquest no hi participa explícitament, però tot i així superior a la observada en el complexe amb la proteïna nativa.

Efectes de l'entorn en processos de CT en sistemes d'ADN amb posicions abàsiques.

- El camp de forces parmbsc1 utilitzat a les simulacions de MD és capaç de descriure les diferents conformacions adoptades pels ADN amb llocs abàsics, en acord amb varis estudis experimentals i teòrics previs.
- En seqüències d'ADN que contenen posicions abàsiques que adopten conformacions intrahèlix, el procés d'HT es veu alentit degut al desordre energètic induït per la presència de la posició abàsica.
- Per contra, l'adopció de conformacions extrahèlix comporta que l'apilament entre les bases es vegi afeblit, desestabilitzant els cations radicals de les guanines contigües i desencadenant asimetries en el procés d'HT, el qual es relentitza en la direcció 5' a 3' i s'accelera en la direcció oposada.

5. Conclusions

Analysis of spectral densities in the PE545 light-harvesting complex.

- Individual and averaged spectral densities simulated for pigments in PE545 display large differences with an important impact on EET dynamics.
- Coupling to low-energy vibrational modes is most affected due to differences in the local environment of each pigment.
- MD-QM/MMPol simulations provide a reasonable reorganization energy (557 cm^{-1}) in comparison with the experimental value (539 cm^{-1}).

Determination of the protonation preferences of bilins in the PE545, PC577, PC612, PC630 and PC645 complexes.

- The $pKas$ estimated for bilins in water indicate that deprotonation occurs at the central pyrrole rings, with pKa values between 5.8 and 7.4.
- The protein environment has an important impact on the bilin $pKas$, giving rise to differences in the protonation properties of a given pigment in different locations of the PBPs studied.
- The calculated $pKas$ seem to be underestimated by ~ 2 pKa units, given that no appreciable variations in the spectra of the complexes are observed until $pH \sim 7$, being such changes attributable to the deprotonation of DBVs and MBVs lacking a coordinated Glu or Asp.
- Our results suggest that all bilins in PE545, PC577, PC612, PC630 and PC645 are protonated at physiological pH ranges $\sim 5-7$ of the lumen.

Light-harvesting properties of the complexes PC577, PC612, PC630 and PC645.

- The use of a classical force field in the MD-QM/MMPol strategy induces significant systematic errors between different types of pigments, leading to an exaggerated range of relative energies.
- The strategy based on QM geometry optimizations of the pigments leads to more reliable results, although the neglect of thermal effects exaggerates the π -conjugation of the low-energy pigments, thus underestimating their relative energies.
- Overall, the results based on QM geometry optimizations suggest a common energy ladder in the 4 PBP studied for the pigments located in the β chains, highly conserved in these algae: $\text{PCB}_{82\text{B/D}} < \text{PCB}_{158\text{B/D}} < \text{DBV}_{50/61\text{B/D}}$. Thus, the pigments and the sequence of the α chains seem to be the origin of the spectral variability in these complexes.

Environmental effects on CT processes in DNA-protein complexes.

- The binding of DNA to the methyltransferase *HhaI* enzyme has an important impact on the HT process, as the insertion of the Gln destabilizes radical cation energies of the surrounding guanines thus inhibiting long-range oxidative damage.
- The binding to a mutant enzyme that inserts an aromatic Trp instead of Gln in contrast, stabilizes the energy of the radical cations on the surrounding guanines and enhances their electronic coupling, leading to an accelerated HT process.
- The direct participation of the Trp radical cation leads to an HT process slower compared to the situation where it does not participate explicitly, but still faster than that observed for the complex with the wild-type protein.

Environmental effects on CT processes in DNA containing abasic sites.

- The parmbsc1 force field used in MD simulations is capable of describing the different conformations adopted by DNAs with abasic sites, in agreement with previous theoretical and experimental studies.
- In DNA sequences containing abasic sites that adopt intrahelical conformations, the HT process is slowed down due to the energetic disorder induced by the presence of the abasic void.
- In contrast, the adoption of extrahelical conformations leads to a weakening of the stacking between bases, destabilizing guanine radical cations and leading to asymmetries in the HT process, which is slowed down in the 5' to 3' direction and accelerated in the opposite direction.

Bibliografia

1. Sisto, A. *et al.* Atomistic non-adiabatic dynamics of the LH₂ complex with a GPU-accelerated ab initio exciton model. *Phys. Chem. Chem. Phys.* **19**, 14924–14936 (2017).
2. Curutchet, C. & Mennucci, B. Quantum Chemical Studies of Light Harvesting. *Chem. Rev.* **117**, 294–343 (2017).
3. Blumberger, J. Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chem. Rev.* **115**, 11191–11238 (2015).
4. Giese, B. Long-Distance Electron Transfer Through DNA. *Annu. Rev. Biochem.* **71**, 51–70 (2002).
5. *Highlights in Bioorganic Chemistry*. (Wiley-VCH Verlag GmbH & Co. KGaA, 2004).
6. Genereux, J. C. & Barton, J. K. Mechanisms for DNA Charge Transport. *Chem. Rev.* **110**, 1642–1662 (2010).
7. Siriwong, K. & Voityuk, A. A. Electron transfer in DNA. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 780–794 (2012).
8. Berlin, Y. A., Burin, A. L. & Ratner, M. A. Charge Hopping in DNA. *J. Am. Chem. Soc.* **123**, 260–268 (2001).
9. Jortner, J., Bixon, M., Voityuk, A. A. & Rösch, N. Superexchange Mediated Charge Hopping in DNA. *J. Phys. Chem. A* **106**, 7599–7606 (2002).
10. Boon, E. M. & Barton, J. K. Charge transport in DNA. *Curr. Opin. Struct. Biol.* **12**, 320–329 (2002).
11. Voityuk, A. A., Rösch, N., Bixon, M. & Jortner, J. Electronic Coupling for Charge Transfer and Transport in DNA. *J. Phys. Chem. B* **104**, 9740–9745 (2000).
12. Hutter, M. & Clark, T. On the enhanced stability of the guanine-cytosine base-pair radical cation. *J. Am. Chem. Soc.* **118**, 7574–7577 (1996).
13. Voityuk, A. A., Jortner, J., Bixon, M. & Rösch, N. Energetics of hole transfer in DNA. *Chem. Phys. Lett.* **324**, 430–434 (2000).
14. Rajski, S. R., Kumar, S., Roberts, R. J. & Barton, J. K. Protein-Modulated DNA Electron Transfer. *J. Am. Chem. Soc.* **121**, 5615–5616 (1999).

15. Wagenknecht, H. a, Rajski, S. R., Pascaly, M., Stemp, E. D. & Barton, J. K. Direct observation of radical intermediates in protein-dependent DNA charge transport. *J. Am. Chem. Soc.* **123**, 4400–7 (2001).
16. Butchosa, C., Simon, S. & Voityuk, A. A. Electron transfer from aromatic amino acids to guanine and adenine radical cations in π stacked and T-shaped complexes. *Org. Biomol. Chem.* **8**, 1870 (2010).
17. Krokan, H. E., Standal, R. & Slupphaug, G. DNA glycosylases in the base excision repair of DNA. *Biochem. J.* **325**, 1–16 (1997).
18. Schärer, O. D. & Jiricny, J. Recent progress in the biology, chemistry and structural biology of DNA glycosylases. *BioEssays* **23**, 270–281 (2001).
19. Mirkovic, T. *et al.* Light Absorption and Energy Transfer in the Antenna Complexes of Photosynthetic Organisms. *Chem. Rev.* **117**, 249–293 (2017).
20. Fenna, R. E. & Matthews, B. W. Chlorophyll arrangement in a bacteriochlorophyll protein from *Chlorobium limicola*. *Nature* **258**, 573–577 (1975).
21. Scholes, G. D., Mirkovic, T., Turner, D. B., Fassioli, F. & Buchleitner, A. Solar light harvesting by energy transfer: from ecology to coherence. *Energy Environ. Sci.* **5**, 9374–9393 (2012).
22. Collini, E. *et al.* Coherently wired light-harvesting in photosynthetic marine algae at ambient temperature. *Nature* **463**, 644–647 (2010).
23. Engel, G. S. *et al.* Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature* **446**, 782–786 (2007).
24. Lee, H., Cheng, Y.-C. & Fleming, G. R. Coherence Dynamics in Photosynthesis: Protein Protection of Excitonic Coherence. *Science* **316**, 1462–1465 (2007).
25. Wong, C. Y. *et al.* Electronic coherence lineshapes reveal hidden excitonic correlations in photosynthetic light harvesting. *Nat. Chem.* **4**, 396–404 (2012).
26. Beljonne, D., Curutchet, C., Scholes, G. D. & Silbey, R. J. Beyond Förster Resonance Energy Transfer in Biological and Nanoscale Systems. *J. Phys. Chem. B* **113**, 6583–6599 (2009).
27. Viani, L., Curutchet, C. & Mennucci, B. Spatial and Electronic Correlations in the PE545 Light-Harvesting Complex. *J. Phys. Chem. Lett.* **4**, 372–377 (2013).
28. Olbrich, C., Strümpfer, J., Schulten, K. & Kleinekathöfer, U. Quest for Spatially

- Correlated Fluctuations in the FMO Light-Harvesting Complex. *J. Phys. Chem. B* **115**, 758–764 (2011).
29. Olbrich, C., Strümpfer, J., Schulten, K. & Kleinekathöfer, U. Theory and simulation of the environmental effects on FMO electronic transitions. *J. Phys. Chem. Lett.* **2**, 1771–1776 (2011).
30. Kolli, A., O'Reilly, E. J., Scholes, G. D. & Olaya-Castro, A. The fundamental role of quantized vibrations in coherent light harvesting by cryptophyte algae. *J. Chem. Phys.* **137**, 174109 (2012).
31. O'Reilly, E. J. & Olaya-Castro, A. Non-classicality of the molecular vibrations assisting exciton energy transfer at room temperature. *Nat. Commun.* **5**, 3012 (2014).
32. Harrop, S. J. *et al.* Single-residue insertion switches the quaternary structure and exciton states of cryptophyte light-harvesting proteins. *Proc. Natl. Acad. Sci.* **111**, E2666–E2675 (2014).
33. McClure, S. D., Turner, D. B., Arpin, P. C., Mirkovic, T. & Scholes, G. D. Coherent Oscillations in the PC577 Cryptophyte Antenna Occur in the Excited Electronic State. *J. Phys. Chem. B* **118**, 1296–1308 (2014).
34. Lee, M. K., Bravaya, K. B. & Coker, D. F. First Principles Models for Biological Light-harvesting: Phycobiliprotein Complexes from Cryptophyte Algae. *J. Am. Chem. Soc.* **139**, 7803–7814 (2017).
35. Sisto, A., Glowacki, D. R. & Martinez, T. J. Ab Initio Nonadiabatic Dynamics of Multichromophore Complexes: A Scalable Graphical-Processing-Unit-Accelerated Exciton Framework. *Acc. Chem. Res.* **47**, 2857–2866 (2014).
36. Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **437**, 55–75 (1948).
37. Marcus, R. A. On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. I. *J. Chem. Phys.* **24**, 966–978 (1956).
38. Krueger, B. P., Scholes, G. D. & Fleming, G. R. Calculation of Couplings and Energy-Transfer Pathways between the Pigments of LH2 by the ab Initio Transition Density Cube Method. *J. Phys. Chem. B* **102**, 5378–5386 (1998).
39. Van Der Meer, B. W. Kappa-squared: from nuisance to new sense. *Reviews Mol. Biotechnol.* **82**, 181–196 (2002).
40. Kocherzhenko, A. A., Grozema, F. C. & Siebbeles, L. D. A. Single molecule

- charge transport: from a quantum mechanical to a classical description. *Phys. Chem. Chem. Phys.* **13**, 2096–2110 (2011).
41. Redfield, G. On the Theory of Relaxation Processes. *IBM J Res Dev* **1**, 19–31 (1957).
 42. Renger, T., May, V. & Kühn, O. Ultrafast excitation energy transfer dynamics in photosynthetic pigment–protein complexes. *Phys. Rep.* **343**, 137–254 (2001).
 43. Leegwater, J. A., Durrant, J. R. & Klug, D. R. Exciton Equilibration Induced by Phonons: Theory and Application to PS II Reaction Centers. *J. Phys. Chem. B* **101**, 7205–7210 (1997).
 44. Kühn, O. & Sundström, V. Energy Transfer and Relaxation Dynamics in Light-Harvesting Antenna Complexes of Photosynthetic Bacteria. *J. Phys. Chem. B* **101**, 3432–3440 (1997).
 45. Cheng, Y.-C. & Fleming, G. R. Dynamics of Light Harvesting in Photosynthesis. *Annu. Rev. Phys. Chem.* **60**, 241–262 (2009).
 46. Zhang, W. M., Meier, T., Chernyak, V. & Mukamel, S. Exciton-migration and three-pulse femtosecond optical spectroscopies of photosynthetic antenna complexes. *J. Chem. Phys.* **108**, 7763–7774 (1998).
 47. Yang, M. & Fleming, G. R. In Influence of phonons on exciton transfer dynamics: comparison of the Redfield, Förster, and modified Redfield equations. *Chem. Phys.* **275**, 355–372 (2002).
 48. Tanimura, Y. Stochastic Liouville, Langevin, Fokker–Planck, and Master Equation Approaches to Quantum Dissipative Systems. *J. Phys. Soc. Japan* **75**, 82001 (2006).
 49. Ishizaki, A. & Fleming, G. R. Unified treatment of quantum coherent and incoherent hopping dynamics in electronic energy transfer: Reduced hierarchy equation approach. *J. Chem. Phys.* **130**, 234111–234110 (2009).
 50. Orozco, M. & Luque, F. J. Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems. *Chem. Rev.* **100**, 4187–4226 (2000).
 51. Cramer, C. J. & Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **99**, 2161–2200 (1999).
 52. Tomasi, J., Mennucci, B. & Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **105**, 2999–3094 (2005).

53. Soteras, I., Curutchet, C., Bidon-Chanal, A., Orozco, M. & Luque, F. J. Extension of the MST model to the IEF formalism: HF and B3LYP parametrizations. *J. Mol. Struct. THEOCHEM* **727**, 29–40 (2005).
54. Curutchet, C., Bidon-Chanal, A., Soteras, I., Orozco, M. & Luque, F. J. MST continuum study of the hydration free energies of monovalent ionic species. *J. Phys. Chem. B* **109**, 3565–3574 (2005).
55. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).
56. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976).
57. Curutchet, C. *et al.* Electronic Energy Transfer in Condensed Phase Studied by a Polarizable QM/MM Model. *J. Chem. Theory Comput.* **5**, 1838–1848 (2009).
58. Lipparini, F. & Barone, V. Polarizable Force Fields and Polarizable Continuum Model: A Fluctuating Charges/PCM Approach. 1. Theory and Implementation. *J. Chem. Theory Comput.* **7**, 3711–3724 (2011).
59. Boulanger, E. & Thiel, W. Solvent Boundary Potentials for Hybrid QM/MM Computations Using Classical Drude Oscillators: A Fully Polarizable Model. *J. Chem. Theory Comput.* **8**, 4527–4538 (2012).
60. Wang, J. *et al.* Development of polarizable models for molecular mechanical calculations I: parameterization of atomic polarizability. *J. Phys. Chem. B* **115**, 3091–9 (2011).
61. Curutchet, C. *et al.* Photosynthetic Light-Harvesting Is Tuned by the Heterogeneous Polarizable Environment of the Protein. *J. Am. Chem. Soc.* **133**, 3078–3084 (2011).
62. Daday, C., Curutchet, C., Sinicropi, A., Mennucci, B. & Filippi, C. Chromophore-Protein Coupling beyond Nonpolarizable Models: Understanding Absorption in Green Fluorescent Protein. *J. Chem. Theory Comput.* **11**, 4825–4839 (2015).
63. Mennucci, B. Polarizable continuum model. *Wiley Interdiscip. Rev. Comput.*

- Mol. Sci.* **2**, 386–404 (2012).
- 64. Chipman, D. M. Charge penetration in dielectric models of solvation. *J. Chem. Phys.* **106**, 10194–10206 (1997).
 - 65. Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 699–709 (2011).
 - 66. Lipparini, F. *et al.* Quantum Calculations in Solution for Large to Very Large Molecules: A New Linear Scaling QM/Continuum Approach. *J. Phys. Chem. Lett.* **5**, 953–958 (2014).
 - 67. Caprasecca, S., Jurinovich, S., Lagardère, L., Stamm, B. & Lipparini, F. Achieving Linear Scaling in Computational Cost for a Fully Polarizable MM/Continuum Embedding. *J. Chem. Theory Comput.* **11**, 694–704 (2015).
 - 68. Lipparini, F. *et al.* Quantum, classical, and hybrid QM/MM calculations in solution: General implementation of the ddCOSMO linear scaling strategy. *J. Chem. Phys.* **141**, 184108 (2014).
 - 69. Voityuk, A. A., Siriwhong, K. & Rösch, N. Charge transfer in DNA. Sensitivity of electronic couplings to conformational changesDedicated to Professor F. Dörr on the occasion of his 80th birthday. *Phys. Chem. Chem. Phys.* **3**, 5421–5425 (2001).
 - 70. Troisi, A. & Orlandi, G. Hole Migration in DNA: a Theoretical Analysis of the Role of Structural Fluctuations. *J. Phys. Chem. B* **106**, 2093–2101 (2002).
 - 71. Pérez, A., Luque, F. J. & Orozco, M. Frontiers in Molecular Dynamics Simulations of DNA. *Acc. Chem. Res.* **45**, 196–205 (2012).
 - 72. Lindahl, E. *et al.* *Molecular Modeling of Proteins*. (2015).
 - 73. Ivani, I. *et al.* Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **13**, 55–58 (2016).
 - 74. Pinheiro, S. & Curutchet, C. Can Förster Theory Describe Stereoselective Energy Transfer Dynamics in a Protein–Ligand Complex? *J. Phys. Chem. B* **121**, 2265–2278 (2017).
 - 75. You, Z.-Q. & Hsu, C.-P. Theory and calculation for the electronic coupling in excitation energy transfer. *Int. J. Quantum Chem.* **114**, 102–115 (2014).
 - 76. Dexter, D. L. A Theory of Sensitized Luminescence in Solids. *J. Chem. Phys.* **21**, 836–850 (1953).
 - 77. Scholes, G. D. & Ghiggino, K. P. Rate expressions for excitation transfer I.

- Radiationless transition theory perspective. *J. Chem. Phys.* **101**, 1251–1261 (1994).
78. Muñoz-Losa, A., Curutchet, C., Krueger, B. P., Hartsell, L. R. & Mennucci, B. Fretting about FRET: failure of the ideal dipole approximation. *Biophys. J.* **96**, 4779–88 (2009).
79. Madjet, M. E., Abdurahman, A. & Renger, T. Intermolecular Coulomb Couplings from Ab Initio Electrostatic Potentials: Application to Optical Transitions of Strongly Coupled Pigments in Photosynthetic Antennae and Reaction Centers. *J. Phys. Chem. B* **110**, 17268–17281 (2006).
80. Hsu, C.-P., Fleming, G. R., Head-Gordon, M. & Head-Gordon, T. Excitation energy transfer in condensed media. *J. Chem. Phys.* **114**, 44506–10521 (2001).
81. Curutchet, C., Scholes, G. D., Mennucci, B. & Cammi, R. How Solvent Controls Electronic Energy Transfer and Light Harvesting: Toward a Quantum-Mechanical Description of Reaction Field and Screening Effects. *J. Phys. Chem. B* **111**, 13253–13265 (2007).
82. Scholes, G. D., Curutchet, C., Mennucci, B., Cammi, R. & Tomasi, J. How Solvent Controls Electronic Energy Transfer and Light Harvesting. *J. Phys. Chem. B* **111**, 6978–6982 (2007).
83. Jurinovich, S., Curutchet, C. & Mennucci, B. The Fenna-Matthews-Olson Protein Revisited: A Fully Polarizable (TD)DFT/MM Description. *ChemPhysChem* **15**, 3194–3204 (2014).
84. Jurinovich, S., Viani, L., Prandi, I. G., Renger, T. & Mennucci, B. Towards an ab initio description of the optical spectra of light-harvesting antennae: application to the CP29 complex of photosystem II. *Phys. Chem. Chem. Phys.* **17**, 14405–14416 (2015).
85. Cupellini, L. *et al.* An Ab Initio Description of the Excitonic Properties of LH2 and Their Temperature Dependence. *J. Phys. Chem. B* **120**, 11348–11359 (2016).
86. Padula, D., Jurinovich, S., Di Bari, L. & Mennucci, B. Simulation of Electronic Circular Dichroism of Nucleic Acids: From the Structure to the Spectrum. *Chem. - A Eur. J.* **22**, 17011–17019 (2016).
87. Loco, D., Jurinovich, S., Bari, L. Di & Mennucci, B. A fast but accurate excitonic

- simulation of the electronic circular dichroism of nucleic acids: how can it be achieved? *Phys. Chem. Chem. Phys.* **18**, 866–877 (2016).
88. Rosnik, A. M. & Curutchet, C. Theoretical Characterization of the Spectral Density of the Water-Soluble Chlorophyll-Binding Protein from Combined Quantum Mechanics/Molecular Mechanics Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **11**, 5826–5837 (2015).
89. Jing, Y., Zheng, R., Li, H.-X. & Shi, Q. Theoretical Study of the Electronic-Vibrational Coupling in the Q_y States of the Photosynthetic Reaction Center in Purple Bacteria. *J. Phys. Chem. B* **116**, 1164–1171 (2012).
90. Renger, T. *et al.* Normal Mode Analysis of the Spectral Density of the Fenna-Matthews-Olson Light-Harvesting Protein: How the Protein Dissipates the Excess Energy of Excitons. *J. Phys. Chem. B* **116**, 14565–14580 (2012).
91. Davis, W. B. *et al.* Distance-Dependent Activation Energies for Hole Injection from Protonated 9-Amino-6-chloro-2-methoxyacridine into Duplex DNA. *J. Am. Chem. Soc.* **124**, 2422–2423 (2002).
92. Félix, M. & Voityuk, A. A. DFT performance for the hole transfer parameters in DNA π stacks. *Int. J. Quantum Chem.* **111**, 191–201 (2011).
93. Voityuk, A. A. Assessment of semiempirical methods for the computation of charge transfer in DNA π-stacks. *Chem. Phys. Lett.* **427**, 177–180 (2006).
94. Cave, R. J. & Newton, M. D. Generalization of the Mulliken-Hush treatment for the calculation of electron transfer matrix elements. *Chem. Phys. Lett.* **249**, 15–19 (1996).
95. Voityuk, A. A. & Rösch, N. Fragment charge difference method for estimating donor–acceptor electronic coupling: Application to DNA π-stacks. *J. Chem. Phys.* **117**, 5607–5616 (2002).
96. Subotnik, J. E., Yeganeh, S., Cave, R. J. & Ratner, M. A. Constructing diabatic states from adiabatic states: Extending generalized Mulliken–Hush to multiple charge centers with Boys localization. *J. Chem. Phys.* **129**, 244101 (2008).
97. Rabinowitch, E. & Govindjee. *Photosynthesis*. (American Association for the Advancement of Science, 1969).
98. Blankenship, R. E. *Molecular mechanisms of Photosynthesis*. (Blackwell Science Ltd., 2002).

99. Nitschke, W. & Rutherford, A. W. Photosynthetic reaction centres: variations on a common structural theme? *Trends Biochem. Sci.* **16**, 241–5 (1991).
100. Lee, R. E. *Phycology*. Cambridge University Press (1989).
101. Apt, K. E., Collier, J. L. & Grossman, A. R. Evolution of the Phycobiliproteins. *J. Mol. Biol.* **248**, 79–96 (1995).
102. Ishizaki, A. & Fleming, G. R. Quantum Coherence in Photosynthetic Light Harvesting. *Annu. Rev. Condens. Matter Phys.* **3**, 333–361 (2012).
103. Spear-bernstein, L. & Miller, K. R. Unique location of the phycobiliprotein light-harvesting pigment in the cryptophytae. *J. Phycol.* **3**, 412–419 (1989).
104. WATSON, J. D. & CRICK, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
105. Eley, D. D. & Spivey, D. I. Semiconductivity of organic substances. Part 9.— Nucleic acid in the dry state. *Trans. Faraday Soc.* **58**, 411–415 (1962).
106. Giese, B., Amaudrut, J., Köhler, A.-K., Spormann, M. & Wessely, S. Direct observation of hole transfer through DNA by hopping between adenine bases and by tunnelling. *Nature* **412**, 318–320 (2001).
107. Halliwell, B. & Gutteridge, J. M. C. *Free radicals in biology and medicine*. (Oxford University Press, 1999).
108. Giese, B. & Spichty, M. Long Distance Charge Transport through DNA: Quantification and Extension of the Hopping Model. *ChemPhysChem* **1**, 195–198 (2000).
109. Ring, S. J. Electron transfer in DNA Bernd Giese. *Pept. Nucleic Acid* **6**, 612–618 (2002).
110. Cullis, P. M., Jones, G. D. D., Symons, M. C. R. & Lea, J. S. Electron transfer from protein to DNA in irradiated chromatin. *Nature* **330**, 773–774 (1987).
111. Wagenknecht, H.-A., Stemp, E. D. A. & Barton, J. K. Evidence of Electron Transfer from Peptides to DNA: Oxidation of DNA-Bound Tryptophan Using the Flash-Quench Technique. *J. Am. Chem. Soc.* **122**, 1–7 (2000).
112. Wagenknecht, H.-A., Stemp, E. D. A. & Barton, J. K. DNA-Bound Peptide Radicals Generated through DNA-Mediated Electron Transport †. *Biochemistry* **39**, 5483–5491 (2000).
113. Shih, C. *et al.* Tryptophan-Accelerated Electron Flow Through Proteins. *Science (80-.).* **320**, 1760–1762 (2008).

114. Wittekindt, C., Schwarz, M., Friedrich, T. & Koslowski, T. Aromatic Amino Acids as Stepping Stones in Charge Transfer in Respiratory Complex I: An Unusual Mechanism Deduced from Atomistic Theory and Bioinformatics. *J. Am. Chem. Soc.* **131**, 8134–8140 (2009).
115. Curutchet, C. *et al.* Energy Flow in the Cryptophyte PE545 Antenna Is Directed by Bilin Pigment Conformation. *J. Phys. Chem. B* **117**, 4263–4273 (2013).
116. Panitchayangkoon, G. *et al.* Long-lived quantum coherence in photosynthetic complexes at physiological temperature. *Proc. Natl. Acad. Sci.* **107**, 12766–12770 (2010).
117. Doust, A. B. *et al.* Developing a Structure–Function Model for the Cryptophyte Phycoerythrin 545 Using Ultrahigh Resolution Crystallography and Ultrafast Laser Spectroscopy. *J. Mol. Biol.* **344**, 135–153 (2004).
118. Novoderezhkin, V. I., Doust, A. B., Curutchet, C., Scholes, G. D. & van Grondelle, R. Excitation Dynamics in Phycoerythrin 545: Modeling of Steady-State Spectra and Transient Absorption with Modified Redfield Theory. *Biophys. J.* **99**, 344–352 (2010).
119. Demming-Adams, B. & Adams, W. W. Photoprotection in an ecological context: the remarkable complexity of thermal energy dissipation. *New Phytol.* **172**, 11–21 (2006).
120. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
121. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K a Values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).
122. Aghtar, M., Strümpfer, J., Olbrich, C., Schulten, K. & Kleinekathöfer, U. Different Types of Vibrations Interacting with Electronic Excitations in Phycoerythrin 545 and Fenna–Matthews–Olson Antenna Systems. *J. Phys. Chem. Lett.* **5**, 3131–3137 (2014).
123. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **26**, 709–715 (1993).

124. McWilliams, M. a., Anka, F. H., Balkus, K. J. & Slinker, J. D. Sensitive and selective real-time electrochemical monitoring of DNA repair. *Biosens. Bioelectron.* **54**, 541–546 (2014).
125. Schibel, A. E. P. *et al.* Nanopore Detection of 8-Oxo-7,8-dihydro-2'-deoxyguanosine in Immobilized Single-Stranded DNA via Adduct Formation to the DNA Damage Site. *J. Am. Chem. Soc.* **132**, 17992–17995 (2010).
126. Zhang, J. *et al.* Role of the translationally controlled tumor protein in DNA damage sensing and repair. *Proc. Natl. Acad. Sci.* **109**, E926–E933 (2012).
127. Corbella, M., Voityuk, A. a. & Curutchet, C. Single Amino Acid Mutation Controls Hole Transfer Dynamics in DNA-Methyltransferase Hha I Complexes. *J. Phys. Chem. Lett.* **6**, 3749–3753 (2015).
128. Gelfand, C. A., Plum, G. E., Grollman, A. P., Johnson, F. & Breslauer, K. J. Thermodynamic Consequences of an Abasic Lesion in Duplex DNA Are Strongly Dependent on Base Sequence. *Biochemistry* **37**, 7321–7327 (1998).
129. Ayadi, L., Coulombeau, C. & Lavery, R. Abasic Sites in Duplex DNA: Molecular Modeling of Sequence-Dependent Effects on Conformation. *Biophys. J.* **77**, 3218–3226 (1999).
130. Cuniasse, P., Fazakerley, G. V., Guschlauer, W., Kaplan, B. E. & Sowers, L. C. The abasic site as a challenge to DNA polymerase. *J. Mol. Biol.* **213**, 303–314 (1990).
131. Barsky, D., Foloppe, N., Ahmadi, S., Wilson III, D. M. & MacKerell, A. D. New insights into the structure of abasic DNA from molecular dynamics simulations. *Nucleic Acids Res.* **28**, 2613–2626 (2000).
132. Zálešák, J., Lourdin, M., Krejčí, L., Constant, J.-F. & Jourdan, M. Structure and Dynamics of DNA Duplexes Containing a Cluster of Mutagenic 8-Oxoguanine and Abasic Site Lesions. *J. Mol. Biol.* **426**, 1524–1538 (2014).
133. Pérez, A. *et al.* Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **92**, 3817–3829 (2007).
134. Ivani, I. *et al.* Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **13**, 55–58 (2016).
135. Lu, X.-J. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*

- 31**, 5108–5121 (2003).
136. Curutchet, C. & Voityuk, A. A. Triplet-Triplet Energy Transfer in DNA: A Process that Occurs on the Nanosecond Timescale. *Angew. Chemie Int. Ed.* **50**, 1820–1822 (2011).
137. Mi, S., Alonso, D. & Roberts, R. J. Functional analysis of Gln-237 mutants of Hhal methyltransferase. *Nucleic Acids Res.* **23**, 620–7 (1995).

APÈNDIXS

Supporting Information

Molecular basis of the exciton-phonon interactions in the PE545 light-harvesting complex

Lucas Viani,¹ Marina Corbella,² Carles Curutchet,² Edward J. O'Reilly³, Alexandra Olaya-Castro,³ Benedetta Mennucci¹

¹Dipartimento di Chimica e Chimica Industriale, Università di Pisa, via Risorgimento 35, 56126 Pisa, Italy.

²Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain.

³Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom

1. Molecular origin of quasi-harmonic modes

In the following Tables S1-S8, we report the individual rings of each bilin involved in the $\delta(\text{ring})$, $\gamma(\text{ring})$, $\beta(\text{ring})$, $\nu(\text{C-C})$ and $\nu(\text{C-N})$ displacements for each mode analyzed, thus allowing identification of those expected to be strongly coupled to the excitations of the pigments. Such motions involve torsional motions $\delta(\text{ring})$ distorting the planarity of the pyrrole rings, out of plane and in plane bendings of the rings, $\gamma(\text{ring})$ and $\beta(\text{ring})$, respectively, and $\nu(\text{C-C})$ and $\nu(\text{C-N})$ stretchings leading to ring breathing modes. In the 500-600 and 650-750 cm^{-1} range, such motions mostly involve $\delta(\text{ring})$, $\gamma(\text{ring})$ and $\beta(\text{ring})$ displacements. In the 950-1050 cm^{-1} region, the vibrations are characterized by a mix of $\beta(\text{ring})$, $\nu(\text{C-C})$ and $\nu(\text{C-N})$ distortions leading to ring breathing modes. Finally, modes in the 1400-1500 cm^{-1} range mostly involve $\nu(\text{C-C})$ and $\nu(\text{C-N})$ stretchings, again leading to ring breathing modes. The correspondence between the ring numbering and the structure of PEBs and DBVs is shown in Figure S1 and S2.

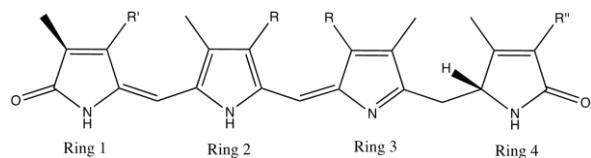


Figure S1: Correspondence between ring numbering and the structure of DBVs.

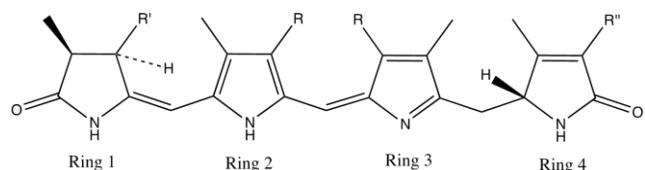


Figure S2: Correspondence between ring numbering and the structure of PEBs.

Table S1. Quasi-harmonic modes of pigment DBV_{19A}.

Frequency (cm ⁻¹)	Mode	$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$
493,66	79	3,4
521,93	80	
534,69	81	3,4
544,75	82	3,4
552,56	83	2
557,15	84	3,4
563,05	85	3,4
568,99	86	
573,88	87	3
576,87	88	
582,79	89	
585,11	90	
591,68	91	
594,56	92	
602,85	93	3,4
$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$		
667,26	102	2,3
677,86	103	1,2
683,33	104	3
694,92	105	1
697,02	106	
699,05	107	
711,91	108	
720,87	109	
733,33	110	1,2,3,4
738,06	111	4
756,35	112	1,2
$\beta(\text{ring}), \nu(\text{C-C})$ and $\nu(\text{C-N})$		
947,34	127	2
957,78	128	2
958,69	129	1,2
963,29	130	1
981,23	131	1
989,92	132	1,2
1005,68	133	2,3
1009,43	134	1,2,3
1028,61	135	1,2,3
1037,16	136	4
1055,17	137	1,2,3
$\nu(\text{C-C})$ and $\nu(\text{C-N})$		
1399,43	165	2,4
1410,20	166	1,2,3,4
1417,81	167	1,2,3,4
1426,94	168	1,2,3,4
1437,39	169	1,2,3,4
1442,27	170	1,2,3,4
1455,31	171	2,3,4
1468,32	172	1,2,3,4
1474,35	173	2,3,4

1478,78	174	1,2,3
1483,22	175	1,2,3
1493,94	176	1,3,4

Table S2. Quasi-harmonic modes of pigment DBV_{19B}

Frequency (cm ⁻¹)	Mode	$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$
501,12	80	3,4
519,93	81	3
529,12	82	3
529,34	83	3,4
534,35	84	
546,37	85	
552,23	86	3,4
559,41	87	
564,00	88	1,2,3,4
567,43	89	1,2,3,4
569,37	90	
573,61	91	
578,99	92	2,3
591,10	93	
595,15	94	
604,46	95	2,3,4
$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$		
659,30	103	2,3,4
663,77	104	1,2,3,4
677,90	105	1,2,3,4
697,68	106	1,2,3,4
711,50	107	4
719,64	108	4
729,43	109	
736,25	110	1
744,67	111	1,2,3
776,86	112	1,2
$\beta(\text{ring}), \nu(\text{C-C})$ and $\nu(\text{C-N})$		
954,40	127	2,3
977,15	128	4
989,26	129	1,2,3,4
997,22	130	1,3,4
1018,10	131	2,3,4
1022,46	132	1,2,3
1030,55	133	3,4
1038,32	134	1,2,3,4
1051,51	135	2,3,4
1059,05	136	4
$\nu(\text{C-C})$ and $\nu(\text{C-N})$		
1401,73	159	1,4
1404,90	160	1,2,4
1421,35	161	1,2,3
1426,99	162	1,2,3,4
1438,36	163	1,2,3,4
1441,63	164	1,2,4
1462,17	165	1,3,4
1462,65	166	1,2,3,4
1469,12	167	1,2,3,4
1482,64	168	1,2,3,4
1488,28	169	1,2,3
1492,94	170	1,2,3,4
1501,85	171	1,2,3

Table S3. Quasi-harmonic modes of pigment PEB_{50/61C}

Frequency (cm ⁻¹)	Mode	$\delta(\text{ring}), \gamma(\text{ring})$ and β (ring)
511,57	79	1,2,3
517,81	80	1,2
529,18	81	1,2
539,31	82	2,3,4
548,81	83	1,2,3
556,98	84	1,3,4
563,02	85	1,2
583,20	86	1,3
595,52	87	
600,39	88	4
$\delta(\text{ring}), \gamma(\text{ring})$ and β (ring)		
660,91	102	1,2,3
666,91	103	
678,22	104	1,2
681,58	105	1,2,4
694,64	106	2,3,4
702,16	107	1,2,3,4
722,70	108	
732,69	109	1,2,3,4
752,77	110	1,2,3
757,98	111	1
765,25	112	1,2,3
β (ring), $\nu(\text{C-C})$ and $\nu(\text{C-N})$		
948,90	126	1,2,3,4
953,52	127	1
963,78	128	2,4
986,60	129	1,3
999,66	130	1,2
1011,51	131	
1014,02	132	1,2,4
1020,63	133	1,4
1031,22	134	1,4
1042,89	135	1,2
1055,20	136	1,2,3,4
$\nu(\text{C-C})$ and $\nu(\text{C-N})$		
1401,87	167	1,2,4
1419,33	168	1,2,4
1423,79	169	1,3,4
1441,21	170	1,2,3
1451,93	171	1,2,3,4
1472,75	172	1,2,3,4
1475,23	173	2,3
1480,75	174	1,3,4
1486,74	175	1,2,3,4
1504,22	176	1,2,3,4

Table S4. Quasi-harmonic modes of pigment PEB_{50/61D}

Frequency (cm ⁻¹)	Mode	$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$
510,23	79	1,4
517,45	80	3,4
525,94	81	2,3
530,17	82	2
543,02	83	2
551,56	84	3
556,48	85	3
568,08	86	2,3
570,12	87	2,3,4
579,56	88	1,2,3,4
582,51	89	1,2
588,46	90	
591,19	91	3,4
598,52	92	1,2,3
$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$		
663,54	101	1,4
667,51	102	
675,41	103	1,2
682,73	104	2
686,77	105	1,2,4
688,62	106	4
694,60	107	
704,72	108	2,3,4
711,11	109	
716,27	110	2,3,4
733,04	111	1,4
759,68	112	1,4
$\beta(\text{ring}), \nu(\text{C-C})$ and $\nu(\text{C-N})$		
942,10	128	2,3,4
971,53	129	2,4
979,21	130	2,4
994,50	131	2,3,4
1017,83	132	1
1024,10	133	1,3
1031,13	134	2,3
1039,29	135	1,2,4
1052,23	136	1,2,4
$\nu(\text{C-C})$ and $\nu(\text{C-N})$		
1402,27	166	1,2,3
1413,23	167	2,3,4
1416,37	168	4
1423,93	169	1,2,3,4
1427,60	170	1,4
1433,68	171	1,2,3,4
1437,55	172	1,2,4
1443,73	173	2,3,4
1452,18	174	2,3,4
1458,68	175	2,4
1460,06	176	2
1475,81	177	1,2,3,4
1482,54	178	1,2,3,4
1494,60	179	2,3,4
1496,96	180	1,3,4
1509,85	181	1,2,3,4

Table S5. Quasi-harmonic modes of pigment PEB_{158C}

Frequency (cm ⁻¹)	Mode	$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$
504,25	76	3
514,14	77	1,2,3
518,27	78	1,2
531,49	79	2,4
542,99	80	
546,38	81	2,3
556,06	82	
567,90	83	
572,06	84	1
576,78	85	3,4
585,64	86	1,2,3
595,54	87	1,2,3,4
602,38	88	2,4
$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$		
658,43	95	3
676,24	96	2,3,4
684,33	97	4
692,23	98	1,2,3
700,95	99	
709,95	100	
711,13	101	2,3,4
721,89	102	1,2,3,4
741,77	103	1,2,3,4
744,57	104	1,2,3,4
756,49	105	1,2,3,4
$\beta(\text{ring}), v(\text{C-C})$ and $v(\text{C-N})$		
954,37	126	1
959,14	127	1
966,61	128	1
977,20	129	1,2
984,21	130	1,3
998,23	131	1,3
1000,96	132	1,2,3
1029,93	133	1,2,3
1032,77	134	3
1062,44	135	2
$v(\text{C-C})$ and $v(\text{C-N})$		
1403,12	162	1,2,3
1406,13	163	1,3,4
1413,13	164	1,2,3
1423,01	165	1,2,3,4
1427,02	166	1,3,4
1434,66	167	1,2,3,4
1448,91	168	1,2,3,4
1454,70	169	
1469,46	170	2,3,4
1470,47	171	1,2,3,4
1487,23	172	2,3,4
1494,04	173	1,2,4
1502,91	174	1,2,3,4

Table S6. Quasi-harmonic modes of pigment PEB_{82C}

Frequency (cm ⁻¹)	Mode	$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$
507,25	74	1,2,3
516,61	75	1,2
534,31	76	3,4
541,30	77	
546,66	78	1
549,30	79	2,4
556,52	80	1,2,3
559,65	81	
577,63	82	3,4
585,73	83	1,3,4
596,32	84	1,2,3
597,12	85	1,3,4
$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$		
659,17	91	1,2,3,4
670,96	92	1,2,3,4
691,17	93	2
696,70	94	1,2,3
701,10	95	
704,24	96	
708,06	97	1
711,41	98	
725,44	99	2,3,4
732,35	100	2,3
749,93	101	
753,36	102	2,3,4
764,48	103	1,2,4
$\beta(\text{ring}), v(\text{C-C})$ and $v(\text{C-N})$		
947,14	122	1,3,4
962,03	123	1,4
964,13	124	1,2,3
996,12	125	1,4
1004,78	126	2,3,4
1008,34	127	1,2,3
1016,25	128	4
1033,26	129	1
1049,23	130	2,3,4
$v(\text{C-C})$ and $v(\text{C-N})$		
1414,09	163	1,3,4
1423,19	164	1,2,3,4
1426,92	165	1,2,3,4
1435,91	166	1
1440,30	167	1,2,3,4
1457,18	168	1,2,3,4
1468,17	169	1
1478,04	170	1,2,3,4
1493,60	171	1,2,3,4
1499,41	172	4
1507,28	173	1,2,3,4

Table S7. Quasi-harmonic modes of pigment PEB_{158D}

Frequency (cm ⁻¹)	Mode	$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$
507,90	75	3,4
512,78	76	1,2,3
521,75	77	2,3
531,92	78	1,2,3,4
534,19	79	1,3,4
555,91	80	1,2,3,4
561,69	81	1,3,4
565,78	82	1
570,99	83	3
586,22	84	1,2,3,4
590,19	85	1
595,93	86	
601,43	87	2
$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$		
662,46	100	3
681,66	101	2,3,4
690,97	102	1,3,4
693,02	103	
699,84	104	1,2,3,4
711,18	105	1,2,3,4
726,59	106	1,2,3,4
733,19	107	1
743,60	108	1,2,3,4
771,49	109	1,2,3
$\beta(\text{ring}), \nu(\text{C-C})$ and $\nu(\text{C-N})$		
946,49	124	1,2
975,70	125	1,2,3*
979,39	126	
997,60	127	2
1003,23	128	2,3,4
1004,82	129	1,2,3,4
1009,66	130	1,2
1019,73	131	1,2,4
1025,28	132	1,3,4
1053,96	133	2,3,4
$\nu(\text{C-C})$ and $\nu(\text{C-N})$		
1400,27	163	1,2,3,4
1413,32	164	1,2,3,4
1418,70	165	2,3,4
1428,58	166	1,2,3,4
1444,91	167	3,4
1446,91	168	1,3,4
1452,29	169	1,3,4
1471,47	170	1,3
1478,99	171	1,2,3
1486,84	172	1,2,3
1502,77	173	1,3,4

Table S8. Quasi-harmonic modes of pigment PEB_{82D}

Frequency (cm ⁻¹)	Mode	$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$
502,38	76	1,3,4
515,83	77	1,2,3,4
532,29	78	1,3,4
535,59	79	1,2
548,00	80	2
554,62	81	
559,38	82	3,4
562,60	83	
574,04	84	2,3,4
589,12	85	3
590,59	86	1,2
595,42	87	1,2
$\delta(\text{ring}), \gamma(\text{ring})$ and $\beta(\text{ring})$		
662,36	94	1
664,08	95	1,2,3,4
672,91	96	1,4
684,89	97	
692,66	98	2,3
697,95	99	
706,27	100	1,2,3,4
712,72	101	1,2,3,4
720,65	102	
726,33	103	1,2
731,84	104	1,2,3,4
737,66	105	1,2,3
746,54	106	1,2,3,4
761,89	107	2,3
$\beta(\text{ring}), \nu(\text{C-C})$ and $\nu(\text{C-N})$		
954,36	126	1,3
975,12	127	1,3
982,89	128	2,3
994,27	129	1,3
1008,27	130	3
1014,71	131	1,2,3
1032,34	132	1,2,3,4
1042,89	133	1,2,3,4
1054,58	134	2,3,4
$\nu(\text{C-C})$ and $\nu(\text{C-N})$		
1405,30	165	1,2,3
1411,49	166	2,3,4
1432,37	167	1,4
1445,49	168	2,3,4
1452,28	169	1,2,3
1456,20	170	1,4
1466,07	171	1,2,3,4
1472,79	172	1,3,4
1493,30	173	1,2,3,4
1497,23	174	1,2,3,4
1500,58	175	1,3

2. Comparison of the distribution of frequencies found with the NMA analysis at the MM and DFT levels

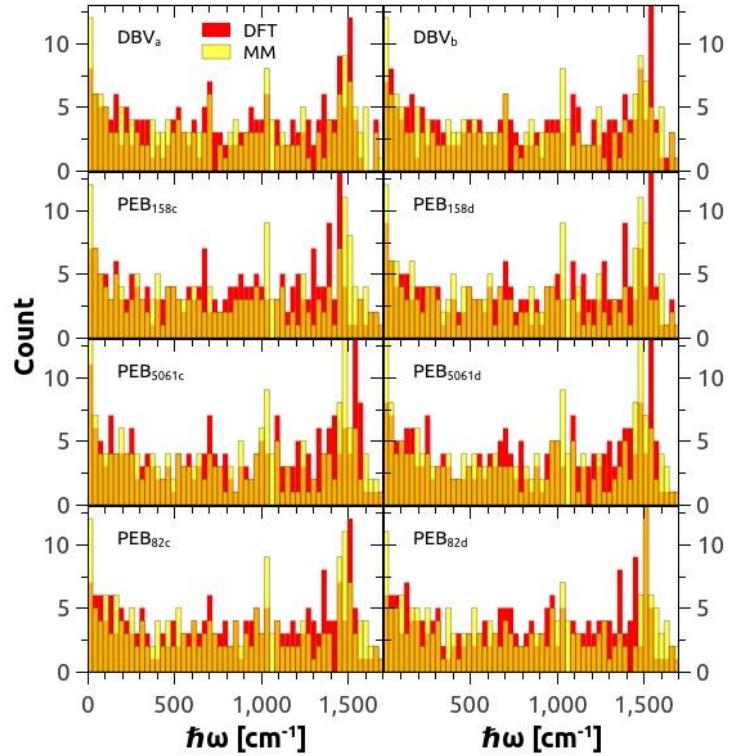


Figure S3: Comparison between QM (red) and MM (yellow) vibrational frequencies population.

3. Hierarchical expansion of exciton dynamics

The hierarchical expansion of exciton dynamics^{1,2} for an environment described by both over and underdamped Brownian oscillators^{3,4}, results in the coupled differential equations

$$\begin{aligned} \partial_t \tilde{\rho}_{\mathbf{n}}(t) = & -[H, \tilde{\rho}_{\mathbf{n}}(t)] - \sum_{i,k} n_{ik} \gamma_{ik} \tilde{\rho}_{\mathbf{n}}(t) \\ & - \sum_i \left(\sum_{k=1}^{\infty} \frac{c_{ik}}{\gamma_{ik}} - \sum_{k=1}^K \frac{c_{ik}}{\gamma_{ik}} \right) [Q_i, [Q_i, \tilde{\rho}_{\mathbf{n}}(t)]] \\ & - \sum_{i,k} \sqrt{(n_{ik} + 1) \sqrt{|c_{ik} c_{ik}^*|}} [Q_i, \tilde{\rho}_{\mathbf{n}_{ik}^+}(t)] \\ & - \sum_{i,k} \sqrt{n_{ik} / \sqrt{|c_{ik} c_{ik}^*|}} (c_{ik} Q_i \tilde{\rho}_{\mathbf{n}_{ik}^-} - c_{ik}^* \tilde{\rho}_{\mathbf{n}_{ik}^-} Q_i). \end{aligned} \quad (\text{S1})$$

The rescaled⁵ auxiliary matrices $\tilde{\rho}_{\mathbf{n}}(t)$, are indexed by the multi-index \mathbf{n} which has entries n_{ik} and $n = \sum_{ik} n_{ik}$. \mathbf{n}_{ik}^\pm is \mathbf{n} with entry $n_{ik} \rightarrow n_{ik} + 1$. The reduced density matrix of the electronic system ρ (with corresponding Hamiltonian H) is given by the operator $\tilde{\rho}_0(t)$ with all $n_{ik} = 0$. The operator Q_i describes the coupling of the excited state of pigment i to its local vibrational environment. The parameters c_{ik} , $c_{i\bar{k}}$ and γ_{ik} are those appearing in an expansion of the correlation function characterizing the local environment of pigment i , $C_i(t) = \sum_i c_{ik} e^{-\gamma_{ik} t}$. For an underdamped ($\nu_q < 2\omega_q$) Brownian oscillator this reads⁶

$$C^{BO}(t) = c_+^{BO} e^{-\gamma^+ t} + c_-^{BO} e^{-\gamma^- t} + \sum_{k=1}^K c_k^{BO} e^{-\gamma_k t}, \quad (\text{S2})$$

with $c_\pm^{BO} = \pm i \lambda_q \frac{\omega_q^2}{2\zeta_q} \left[\cot\left(\frac{\beta\gamma^\pm}{2}\right) - i \right]$, $c_{ik}^{BO} = -\frac{4\lambda_q\omega_q^2}{\beta} \frac{\gamma_k}{(\omega_q^2 + \gamma_k^2)^2 - \nu_q^2 \gamma_k^2}$, $\gamma^\pm = \frac{\nu_q}{2} \pm i \zeta_q$,

Matsubara frequencies $\gamma_k = 2\pi k/\beta$ and shifted frequency $\zeta_q \sqrt{\omega_q^2 - \nu_q^2}/4$. In equation (S1) $c_{i\bar{k}}$ refers to the coefficient of the corresponding term with complex-conjugate exponent. Here if $c_k = c_\pm$, $c_{\bar{k}} = c_\mp$ and $c_{\bar{k}} = c_k$ if the exponent is real.

In the over-damped limit ($\nu_q \gg 2\omega_q$), this reduces to the Drude form $C^D(t) = c_0^D e^{\Omega_D t} + \sum_{k=1}^K c_k^D e^{-\gamma_k t}$ with $c_0^D = \lambda_D \Omega_D [\cot(\beta\Omega_D/2) - i]$, $c_k^D = -\frac{4\lambda_D\Omega_D}{\beta} \frac{\gamma_k}{\Omega_D^2 - \gamma_k^2}$ and cut off frequency $\Omega_D = \omega_q^2/\nu_q$. When including both Brownian and Drude type components the Matsubara part of the expansions can be combined, The Matsubara

expansion is truncated at level K , while the hierarchy of equations itself is truncated at level $N = \sum_{ik} n_{ik}$.

4. Electronic and vibrational parameters

The electronic parameters for the high and low energy subunits are taken from the results of QM/MMPol calculations with site energies adjusted to better reproduce linear spectra⁷

$$H_{high} = \begin{pmatrix} \text{PEB}_{50/60c} & \text{PEB}_{50/60d} & \text{DBV}_a \\ 19450 & 71.7 & 2.2 \\ 71.7 & 19050 & -39.3 \\ 2.2 & -39.3 & 18000 \end{pmatrix} \quad (\text{S3})$$

$$H_{low} = \begin{pmatrix} \text{PEB}_{82c} & \text{PEB}_{82d} & \text{DBV}_a \\ 18550 & 4.0 & -11.4 \\ 4.0 & 18050 & 34.3 \\ -11.4 & 34.3 & 18000 \end{pmatrix} \quad (\text{S4})$$

For the continuous component part of the spectral density we use $\{\lambda_D, \Omega_D\} = \{36, 32\} \text{ cm}^{-1}$ in the PEB_{50/61} dimer, $\{20, 31\} \text{ cm}^{-1}$ in the PEB₈₂ dimer and $\{80, 42\} \text{ cm}^{-1}$ for the DBV_a. The bilins in the high energy unit are all coupled to a vibrational mode with parameters $\{\lambda_q, \omega_q, \gamma_q\} = \{17.68, 1044, 31\} \text{ cm}^{-1}$.

In the low energy subset, the PEB₈₂ pair is coupled to an underdamped mode with parameters $\{\lambda_q, \omega_q, \gamma_q\} = \{8, 529, 33\} \text{ cm}^{-1}$ while the DBV_a is coupled to a mode with parameters $\{22, 552.5, 65.5\} \text{ cm}^{-1}$.

References

- (1) Tanimura, Y.; Kubo, R. *J. Phys. Soc. Japan* **1989**, *58*, 101.
- (2) Ishizaki, A.; Fleming, G. R. *J. Chem. Phys.* **2009**, *130*, 234111.
- (3) Tanaka, M.; Tanimura, Y. *J. Phys. Soc. Japan* **2009**, *78*, 073802.
- (4) Ding, J.-J.; Xu, R.-X.; Yan, Y. *J. Chem. Phys.* **2012**, *136*, 224103.
- (5) Shi, Q.; Chen, L.; Nan, G.; Xu, R.-X.; Yan, Y. *J. Chem. Phys.* **2009**, *130*, 084105.
- (6) Mukamel, S.; *Principles of nonlinear optical spectroscopy*. Oxford series in optical and imaging sciences **1995** (Oxford University Press)

- (7) Curutchet, C.; Novoderezhkin, V. I.; Kongsted, J.; Muñoz-Losa, A.; van Grondelle, R.; Scholes, G. D.; Mennucci, B. *J. Phys. Chem. B* **2013**, *117*, 4263.

Supporting Information for
Determination of the protonation preferences of bilin pigments in
cryptophyte antenna complexes

Marina Corbella,¹ Zi S. D. Toa,² Gregory D. Scholes,² F. Javier Luque³ and Carles Curutchet¹

¹Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica and Institut de Química Teòrica i Computacional (IQT CUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona Spain

²Department of Chemistry, Princeton University, Washington Road, Princeton, New Jersey 08544, United States

³Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Institut de Biomedicina (IBUB) and Institut de Química Teòrica i Computacional (IQT CUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Santa Coloma de Gramenet, Spain.

Table S1. Electronic energies, free energy corrections and solvation free energies computed for pyridine, imidazole and their conjugated acids.

X	$E_{SCS-MP2/CBS}$ ^a	ΔE_{CCSD} ^a	G_{corr} ^a	$\Delta G_{sol,MST}$ ^b	$\Delta G_{sol,SMD}$ ^b
Pyridine					
X	-247.8279	-0.0568	0.0596	-4.3	-4.6
XH ⁺	-248.1942	-0.0579	0.0733	-58.0	-58.4
Imidazole					
X	-225.8370	-0.0461	0.0433	-10.0	-8.5
XH ⁺	-226.2080	-0.0482	0.0571	-64.9	-62.1

^aEnergies in hartree. ^bSolvation energies in kcal/mol.

Table S2. Electronic energies, free energy corrections and solvation free energies computed for deprotonated bilins on different pyrrole rings (A-D) as well as the fully protonated form (H^+).

	Ring	$E_{SCS-MP2/CBS}$ ^a	ΔE_{CCSD} ^a	G_{corr} ^a	ΔG_{solMST} ^b	ΔG_{solSMD} ^b
PEB	A	-1532.7499	-0.3571	0.5389	-17.1	-20.9
	B	-1532.7769	-0.3533	0.5397	-12.0	-15.6
	C	-1532.7802	-0.3535	0.5413	-11.7	-14.2
	D	-1532.6879	-0.3540	0.5391	-34.1	-40.3
	H^+	-1533.1724	-0.3557	0.5525	-49.8	-53.8
PEB'	A	-1572.7225	-0.3737	0.6034	-16.8	-20.5
	B	-1572.7477	-0.3700	0.6037	-10.7	-14.9
	C	-1572.7507	-0.3702	0.6046	-10.5	-13.5
	D	-1572.6558	-0.3710	0.6040	-34.4	-40.0
	H^+	-1573.1437	-0.3726	0.6179	-49.1	-53.7
DBV	A	-1531.5222	-0.3505	0.5150	-18.0	-21.9
	B	-1531.5608	-0.3456	0.5152	-11.8	-15.2
	C	-1531.5641	-0.3462	0.5171	-11.6	-14.2
	D	-1531.4713	-0.3469	0.5132	-33.0	-39.1
	H^+	-1531.9536	-0.3484	0.5274	-50.4	-54.4
DBV'	A	-1571.9913	-0.3672	0.5791	-17.0	-22.0
	B	-1572.0302	-0.3624	0.5791	-10.8	-15.2
	C	-1572.0335	-0.3630	0.5799	-10.5	-14.0
	D	-1571.9406	-0.3636	0.5798	-30.9	-38.4
	H^+	-1572.4240	-0.3653	0.5920	-49.2	-54.0
PCB	A	-1532.7702	-0.3533	0.5395	-16.4	-19.1
	B	-1532.7942	-0.3497	0.5412	-10.9	-14.0
	C	-1532.7940	-0.3494	0.5412	-10.8	-13.9
	D	-1532.7569	-0.3541	0.5396	-17.4	-19.9
	H^+	-1533.1937	-0.3525	0.5535	-45.7	-49.8
MBV	A	-1531.5408	-0.3463	0.5155	-16.5	-19.5
	B	-1531.5777	-0.3417	0.5163	-10.4	-13.5
	C	-1531.5776	-0.3417	0.5158	-10.3	-13.5
	D	-1531.5394	-0.3463	0.5137	-17.3	-20.0
	H^+	-1531.9740	-0.3449	0.5285	-46.6	-50.1

^aEnergies in hartree. ^bSolvation energies in kcal/mol.

Table S3. Acid dissociation constants computed using PROPKA for bilin pyrrole rings B and C in their specific protein environments based on the crystal structure.

		$pK_a(B)$	$pK_a(C)$
PE545 ($\alpha\beta\alpha\beta$)	DBV _{19A}	5.6	5.1
	DBV _{19B}	5.5	5.0
	PEB _{50/61C}	8.1	6.8
	PEB _{158C}	7.4	6.9
	PEB _{82C}	6.8	6.5
	PEB _{50/61D}	7.6	7.2
	PEB _{158D}	5.7	5.2
	PEB _{82D}	6.9	6.6
PC577 ($\alpha\beta\alpha\beta$)	PCB _{20A}	7.9	7.8
	DBV _{50/61B}	7.6	6.7
	PCB _{158B}	7.9	8.1
	PCB _{82B}	7.1	7.4
	PCB _{20C}	7.9	7.8
	DBV _{50/61D}	7.7	6.7
	PCB _{158D}	7.8	8.1
	PCB _{82D}	7.1	7.3
PC612 ($\alpha\beta\alpha\beta$)	PCB _{20A}	8.0	8.0
	DBV _{50/61B}	7.5	6.5
	PCB _{158B}	7.8	8.1
	PCB _{82B}	7.0	7.4
	PCB _{20C}	8.0	8.1
	DBV _{50/61D}	7.7	6.6
	PCB _{158D}	7.9	8.2
	PCB _{82D}	7.2	7.4
PC630 ($\alpha\beta\alpha\beta$)	MBV _{19A}	5.7	4.9
	DBV _{50/61B}	7.3	6.7
	PCB _{158B}	7.8	7.9
	PCB _{82B}	7.1	7.3
	MBV _{19C}	5.1	4.7
	DBV _{50/61D}	7.1	6.4
	PCB _{158D}	7.7	8.1
	PCB _{82D}	7.0	7.3
PC645 ($\alpha\beta\alpha\beta$)	MBV _{19A}	5.9	5.2
	DBV _{50/61B}	7.3	6.7
	PCB _{158B}	7.8	8.1
	PCB _{82B}	7.2	7.4
	MBV _{19C}	5.2	4.7
	DBV _{50/61D}	6.9	6.1
	PCB _{158D}	7.7	8.2
	PCB _{82D}	6.9	7.4

Table S4. MD-averaged acid dissociation constants computed using PROPKA for the propionic groups linked to pyrrole rings B and C of the bilins in their specific protein environments.

		$pK_a(B)^a$	$pK_a(C)^a$
PE545 ($\alpha,\beta\alpha,\beta$)	DBV _{19A}	3.7	5.0
	DBV _{19B}	3.6	4.3
	PEB _{50/61C}	4.9	4.6
	PEB _{158C}	4.3	2.7
	PEB _{82C}	2.7	3.6
	PEB _{50/61D}	4.8	4.2
	PEB _{158D}	4.5	4.1
	PEB _{82D}	3.5	3.4
PC577 ($\alpha\beta\alpha\beta$)	PCB _{20A}	3.0	2.7
	DBV _{50/61B}	3.7	4.3
	PCB _{158B}	4.3	3.7
	PCB _{82B}	3.8	3.2
	PCB _{20C}	3.5	2.7
	DBV _{50/61D}	3.6	4.0
	PCB _{158D}	4.3	3.8
	PCB _{82D}	3.8	3.4
PC612 ($\alpha\beta\alpha\beta$)	PCB _{20A}	4.8	2.5
	DBV _{50/61B}	4.2	4.5
	PCB _{158B}	4.4	4.1
	PCB _{82B}	3.4	3.5
	PCB _{20C}	4.9	2.8
	DBV _{50/61D}	4.3	4.6
	PCB _{158D}	4.4	3.8
	PCB _{82D}	3.6	3.5
PC630 ($\alpha,\beta\alpha,\beta$)	MBV _{19A}	3.6	2.7
	DBV _{50/61B}	4.5	4.6
	PCB _{158B}	4.4	3.6
	PCB _{82B}	3.7	3.3
	MBV _{19C}	3.8	3.0
	DBV _{50/61D}	4.0	4.4
	PCB _{158D}	4.3	3.6
	PCB _{82D}	2.9	3.4
PC645 ($\alpha,\beta\alpha,\beta$)	MBV _{19A}	4.7	3.0
	DBV _{50/61B}	4.7	4.7
	PCB _{158B}	4.3	3.9
	PCB _{82B}	3.9	3.4
	MBV _{19C}	3.9	2.9
	DBV _{50/61D}	4.7	4.3
	PCB _{158D}	4.4	3.7
	PCB _{82D}	3.7	3.2

^aValues are averaged over PROPKA calculations where the central pyrrole ring is assumed to deprotonate either on ring B or C, which present minor differences below 0.2 pK_a units in all cases.

Table S5. MD-averaged acid dissociation constants computed using PROPKA for residues interacting with bilin central pyrrole rings in their specific protein environments.

			$pK_a(B)^a$	$pK_a(C)^b$
PE545 ($\alpha\beta\alpha\beta$)				
	DBV _{19A}	His16A ^c	4.6	4.5
	DBV _{19B}	His16B ^c	3.8	3.7
	PEB _{50/61C}	Asp54C	4.6	4.8
	PEB _{158C}	Asp39C	5.1	5.0
	PEB _{82C}	Asp85C	5.0	5.1
	PEB _{50/61D}	Asp54D	3.8	3.9
	PEB _{158D}	Asp39D	6.9	6.8
	PEB _{82D}	Asp85D	5.1	5.2
PC577 ($\alpha\beta\alpha\beta$)				
	PCB _{20A}	Glu16A	4.3	4.3
	DBV _{50/61B}	Asp54B	4.6	4.6
	PCB _{158B}	Asp39B	5.3	5.2
	PCB _{82B}	Asp85B	4.9	5.0
	PCB _{20C}	Glu16C	4.6	4.5
	DBV _{50/61D}	Asp54D	4.7	4.7
	PCB _{158D}	Asp39D	5.3	5.3
	PCB _{82D}	Asp85D	4.8	5.0
PC612 ($\alpha\beta\alpha\beta$)				
	PCB _{20A}	Glu16A	4.0	4.0
	DBV _{50/61B}	Asp54B	4.8	4.8
	PCB _{158B}	Asp39B	5.3	5.2
	PCB _{82B}	Asp85B	4.9	5.0
	PCB _{20C}	Glu16C	4.6	4.6
	DBV _{50/61D}	Asp54D	4.8	4.9
	PCB _{158D}	Asp39D	5.6	5.5
	PCB _{82D}	Asp85D	4.9	5.0
PC630 ($\alpha\beta\alpha\beta$)				
	MBV _{19A}	-	-	-
	DBV _{50/61B}	Asp54B	4.6	5.0
	PCB _{158B}	Asp39B	6.8	6.7
	PCB _{82B}	Asp85B	4.8	4.9
	MBV _{19C}	His22A ^d	7.4	7.4
	DBV _{50/61D}	Asp54D	3.3	3.7
	PCB _{158D}	Asp39D	4.6	4.5
	PCB _{82D}	Asp85D	4.9	5.0
PC645 ($\alpha\beta\alpha\beta$)				
	MBV _{19A}	-	-	-
	DBV _{50/61B}	Asp54B	4.8	5.3
	PCB _{158B}	Asp39B	5.6	5.5
	PCB _{82B}	Asp85B	4.8	5.0
	MBV _{19C}	His21A ^d	7.6	7.6
	DBV _{50/61D}	Asp54D	4.0	4.3
	PCB _{158D}	Asp39D	4.4	4.3
	PCB _{82D}	Asp85D	4.9	5.0

^aValues computed assuming deprotonation on bilin pyrrole ring B. ^bValues computed assuming deprotonation on bilin pyrrole ring C. ^cHistidines in PE545 interacting with DBV central pyrrole rings through a water molecule as shown in Fig. S1a. ^d Histidines in PC630 and PC645 interacting with a propionic group of MBV in chain C and Glu26A/Glu25A, respectively, as represented in Fig. S1b.

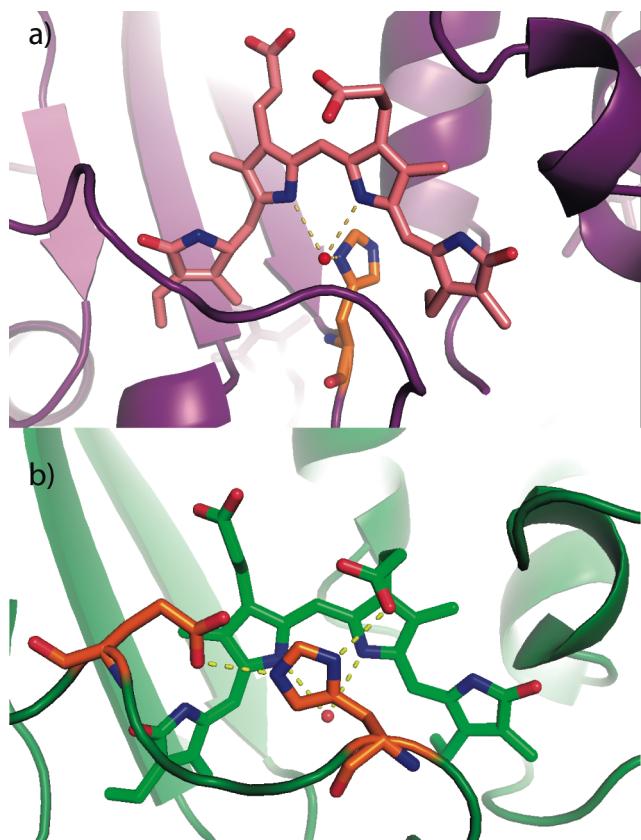


Figure S1. Representation of the environment surrounding bilin pigments. a) DBV_{19A} pigment in the PE545 complex with central pyrrole rings interacting with His16A through a water molecule. b) MBV_{19C} pigment in the PC645 complex with a propionic group and Glu25A interacting with His21A and central pyrrole rings interacting with a crystallographic water.

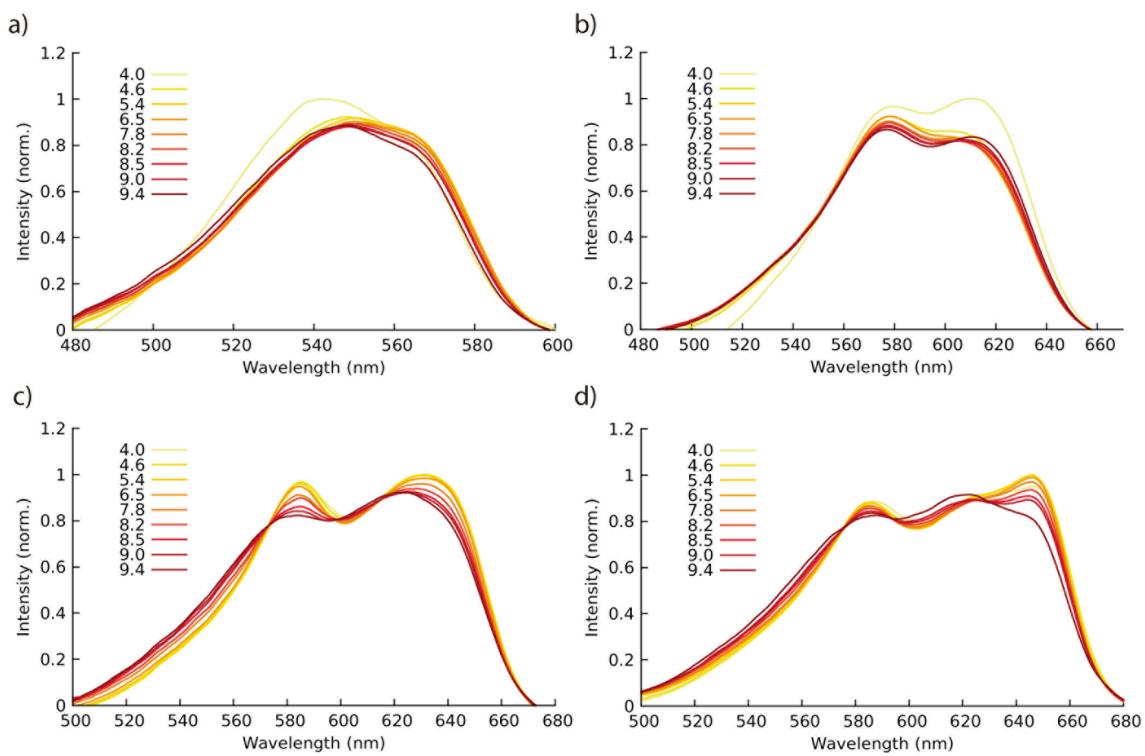


Figure S2. Normalized absorption spectra of cryptophyte antenna complexes at different *pH* values. a) PE545, b) PC577, c) PC630 and d) PC645.

Supporting Information for
Spectral variability in phycocyanin cryptophyte antenna complexes is
controlled by changes in the α polypeptide chains

Marina Corbella,¹ Lorenzo Cupellini,² Benedetta Mennucci,² Gregory D. Scholes³ and
Carles Curutchet¹

¹Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica and Institut de
Química Teòrica i Computacional (IQTC-UB), Facultat de Farmàcia i Ciències de
l'Alimentació, Universitat de Barcelona, Barcelona Spain

²Dipartimento di Chimica e Chimica Industriale, University of Pisa, Italy.

³Department of Chemistry, Princeton University, Washington Road, Princeton, New
Jersey 08544, United States

Table S1. Reorganization energies of the pigments in the open conformation PBP. All values in cm^{-1} .

PC577		PC612	
Chromophore	λ	Chromophore	λ
DBV _{50/61B}	658	DBV _{50/61B}	1199 ^a
DBV _{50/61D}	688	DBV _{50/61D}	891 ^a
PCB _{20A}	464	PCB _{20A}	418
PCB _{158B}	458	PCB _{158B}	444
PCB _{82B}	656	PCB _{82B}	440
PCB _{20C}	460	PCB _{20C}	427
PCB _{158D}	487	PCB _{158D}	426
PCB _{82D}	588	PCB _{82D}	494

^aSpectral simulations performed using values for PC577.

Table S2. Reorganization energies of the pigments in the close conformation PBP. All values in cm^{-1} .

PC630		PC645	
Chromophore	λ	Chromophore	λ
DBV _{50/61B}	-- ^b	DBV _{50/61B}	689
DBV _{50/61D}	572	DBV _{50/61D}	590
PCB _{158B}	563	PCB _{158B}	414
PCB _{82B}	496	PCB _{82B}	416
PCB _{158D}	645	PCB _{158D}	612
PCB _{82D}	431	PCB _{82D}	417
MBV _{19A}	1366 ^a	MBV _{19A}	677
MBV _{19C}	664	MBV _{19C}	350

^aSpectral simulations performed using values for MBV_{19C}.

^bVertical gradient calculation failed for this pigment.

Table S3. Site energies (cm^{-1}), transition electric dipole moments (debies) and centers (\AA) for the PC577 using the MD-QM/MMPol approach.

Chromophore	Site energy	μ_x	μ_y	μ_z	$\bar{\mu}$	x	y	z
DBV _{50/61B}	19605.9	-9.955	5.165	-5.105	12.322	60.979	43.200	38.444
DBV _{50/61D}	19357.7	5.018	-2.139	11.275	12.525	45.748	34.894	30.300
PCB _{20A}	16416.5	-2.036	-13.138	-5.128	14.249	37.047	59.154	44.300
PCB _{158B}	16083.4	10.880	7.205	-5.977	14.353	41.177	38.345	59.578
PCB _{82B}	14926.6	9.493	-5.488	-10.412	15.121	51.021	65.274	32.272
PCB _{20C}	15088.3	7.505	-6.981	-10.911	14.971	44.348	20.793	50.566
PCB _{158D}	15330.1	4.150	-14.060	-4.188	15.246	59.942	44.841	61.771
PCB _{82D}	15358.0	8.583	5.259	-11.450	15.245	26.084	47.294	32.173

Table S4. Site energies (cm^{-1}), transition electric dipole moments (debies) and centers (\AA) for the PC612 using the MD-QM/MMPol approach.

Chromophore	Site energy	μ_x	μ_y	μ_z	$\bar{\mu}$	x	y	z
DBV _{50/61B}	18983.5	6.456	9.817	-4.801	12.693	37.881	41.900	61.910
DBV _{50/61D}	19478.6	-8.494	-8.624	-2.733	12.409	59.145	42.269	54.808
PCB _{20A}	15437.2	12.397	-7.359	-2.730	14.673	41.160	60.076	35.926
PCB _{158B}	15445.4	1.987	-2.091	14.649	14.900	41.295	31.198	32.715
PCB _{82B}	14971.1	6.837	-10.504	8.244	14.930	35.283	64.975	55.834
PCB _{20C}	15125.6	-8.965	5.466	10.571	15.055	55.538	21.710	42.612
PCB _{158D}	15549.9	6.516	-1.938	13.433	15.001	26.272	31.032	44.177
PCB _{82D}	15436.1	-0.279	7.579	13.119	15.154	60.509	58.009	37.266

Table S5. Site energies (cm^{-1}), transition electric dipole moments (debies) and centers (\AA) for the PC630 using the MD-QM/MMPol approach.

Chromophore	Site energy	μ_x	μ_y	μ_z	$\bar{\mu}$	x	y	z
DBV _{50/61B}	17751.6	-5.355	-3.779	11.295	13.059	49.835	41.492	38.991
DBV _{50/61D}	18724.3	7.465	2.587	-9.891	12.659	48.607	37.401	54.929
PCB _{158B}	14862.1	8.626	-10.752	-6.161	15.099	27.976	33.150	47.428
PCB _{82B}	14550.5	15.264	-1.627	-1.562	15.322	70.938	37.144	44.280
PCB _{158D}	15188.5	-6.575	-13.824	0.647	15.430	44.565	63.788	34.552
PCB _{82D}	15603.0	-14.565	-4.431	-0.470	15.231	50.242	56.591	68.608
MBV _{19A}	14955.4	11.533	-2.405	9.617	15.208	28.533	48.913	59.734
MBV _{19C}	14977.4	-11.772	0.778	-9.634	15.232	67.723	57.114	40.020

Table S6. Site energies (cm^{-1}), transition electric dipole moments (debies) and centers (\AA) for the PC645 using the MD-QM/MMPol approach.

Chromophore	Site energy	μ_x	μ_y	μ_z	$\bar{\mu}$	x	y	z
DBV _{50/61B}	17280.2	8.815	0.119	5.943	10.632	37.553	42.879	49.564
DBV _{50/61D}	18484.4	-7.980	-1.487	-10.096	12.955	52.487	39.046	56.829
PCB _{158B}	15329.0	-7.690	-12.728	-1.096	14.911	40.615	49.241	73.492
PCB _{82B}	14826.0	4.530	-10.304	-10.147	15.193	45.668	25.852	37.200
PCB _{158D}	15055.2	-2.716	-7.513	12.923	15.155	37.419	63.739	38.908
PCB _{82D}	14912.5	-4.794	5.979	13.544	15.562	70.155	52.153	48.364
MBV _{19A}	15252.7	10.969	-8.614	-5.721	15.075	55.625	61.063	66.758
MBV _{19C}	14580.6	-10.679	8.938	6.526	15.379	45.830	43.194	27.272

Table S7. Site energies (cm^{-1}), transition electric dipole moments (debies) and centers (\AA) for the PC577 using the QM/MMPol/ddCOSMO approach.

Chromophore	Site energy	μ_x	μ_y	μ_z	$\bar{\mu}$	x	y	z
DBV _{50/61B}	19780.4	4.791	12.217	1.489	13.207	-0.285	-5.817	-10.358
DBV _{50/61D}	19831.2	4.224	-12.469	-0.793	13.189	-0.671	5.899	7.771
PCB _{20A}	18962.5	-8.25	2.58	11.302	14.229	24.573	11.079	-8.369
PCB _{158B}	18963.3	-7.299	-1.972	-11.968	14.156	23.251	-12.261	7.567
PCB _{82B}	18452.0	-13.831	-3.202	-1.451	14.271	11.271	12.117	-23.012
PCB _{20C}	18414.9	-13.643	3.988	0.656	14.229	9.044	-12.485	21.345
PCB _{158D}	18215.6	-13.109	-2.318	6.6	14.859	16.826	-20.867	-7.923
PCB _{82D}	18299.5	-12.253	2.816	-7.813	14.802	17.232	20.102	6.798

Table S8. Site energies (cm^{-1}), transition electric dipole moments (debies) and centers (\AA) for the PC612 using the QM/MMPol/ddCOSMO approach.

Chromophore	Site energy	μ_x	μ_y	μ_z	$\bar{\mu}$	x	y	z
DBV _{50/61B}	19875.6	11.72	5.89	2.282	13.314	-59.764	26.742	-11.038
DBV _{50/61D}	19900.6	-5.735	-11.65	-2.934	13.312	-51.264	33.199	9.192
PCB _{20A}	18585.9	-3.827	4.847	13.071	14.456	-29.398	22.173	-9.357
PCB _{158B}	18758.5	-7.164	4.351	-11.551	14.272	-46.010	3.649	3.448
PCB _{82B}	18812.5	-11.893	7.29	-1.142	13.996	-39.138	34.317	-22.022
PCB _{20C}	18886.7	-7.525	11.444	3.185	14.062	-57.241	11.275	18.458
PCB _{158D}	18209.2	-11.359	5.402	7.797	14.799	-56.860	3.955	-11.998
PCB _{82D}	18398.7	-6.348	11.87	-5.937	14.712	-28.370	31.353	7.460

Table S9. Site energies (cm^{-1}), transition electric dipole moments (debies) and centers (\AA) for the PC630 using the QM/MMPol/ddCOSMO approach.

Chromophore	Site energy	μ_x	μ_y	μ_z				
DBV _{50/61B}	20027.2	-11.709	7.155	-0.607	13.735	22.686	11.794	151.453
DBV _{50/61D}	20277.2	9.795	-8.466	2.585	13.202	10.771	21.814	157.681
PCB _{158B}	19027.1	0.178	-13.488	3.642	13.972	5.840	20.890	135.735
PCB _{82B}	18722.2	-12.949	-1.909	-5.485	14.192	22.722	6.855	172.776
PCB _{158D}	18026.1	0.918	-8.280	12.342	14.890	41.260	24.993	145.168
PCB _{82D}	17918.8	-7.792	-0.402	-12.680	14.888	18.046	42.402	165.018
MBV _{19A}	16594.4	-6.163	1.271	13.725	15.099	11.694	39.175	142.042
MBV _{19C}	16650.1	2.475	-3.723	-13.958	14.656	39.391	17.110	167.644

Table S10. Site energies (cm^{-1}), transition electric dipole moments (debies) and centers (\AA) for the PC645 using the QM/MMPol/ddCOSMO approach.

Chromophore	Site energy	μ_x	μ_y	μ_z				
DBV _{50/61B}	20057.1	-4.717	-12.853	1.505	13.774	21.341	3.420	27.810
DBV _{50/61D}	20404.7	6.600	11.536	0.155	13.291	15.233	-9.134	33.531
PCB _{158B}	18689.1	13.399	3.784	2.475	14.141	14.707	-15.912	14.238
PCB _{82B}	18500.4	4.272	-12.881	-4.598	14.329	28.416	6.079	46.202
PCB _{158D}	17584.9	8.585	3.561	11.541	14.818	5.773	16.729	20.137
PCB _{82D}	17560.7	0.712	-7.612	-12.624	14.759	-3.153	-6.211	44.469
MBV _{19A}	16878.3	1.891	-2.859	14.032	14.445	-2.688	-15.985	21.231
MBV _{19C}	16815.4	1.405	2.506	-14.329		17.209	20.479	41.849

Table S11. Total couplings (upper) and screening factors to the couplings (lower) for the PC577 using the MD-QM/MMPol approach. All values in (cm⁻¹).

s\V _{tot}	DBV _{50/61B}	DBV _{50/61D}	PCB _{20A}	PCB _{158B}	PCB _{82B}	PCB _{20C}	PCB _{158D}	PCB _{82D}
DBV _{50/61B}		12.75	20.96	24.02	-46.63	-6.24	-6.56	18.75
DBV _{50/61D}	0.1		-6.59	-7.17	18.80	19.70	22.89	-48.39
PCB _{20A}	0.6	1.2		101.14	18.17	1.65	-1.46	26.91
PCB _{158B}	0.6	0.6	0.6		20.32	-0.91	7.96	9.73
PCB _{82B}	0.6	0.6	0.8	0.7		28.61	7.07	0.15
PCB _{20C}	1.2	0.5	0.4		0.5		81.14	17.04
PCB _{158D}	0.6	0.7	-	0.7	1.0	0.6		19.89
PCB _{82D}	0.6	0.6	0.5	0.8	-	0.8	0.7	

Table S12. Total couplings (upper) and screening factors to the couplings (lower) for the PC612 using the MD-QM/MMPol approach. All values in (cm⁻¹).

s\V _{tot}	DBV _{50/61B}	DBV _{50/61D}	PCB _{20A}	PCB _{158B}	PCB _{82B}	PCB _{20C}	PCB _{158D}	PCB _{82D}
DBV _{50/61B}		-7.3	7.5	24.4	-43.8	-7.9	-3.2	17.0
DBV _{50/61D}	-0.3		-6.2	-3.1	17.1	16.3	25.4	-45.4
PCB _{20A}	0.6	1.1		129.2	13.4	-6.7	-4.3	6.5
PCB _{158B}	0.6	0.6	0.5		20.1	-1.4	7.2	5.5
PCB _{82B}	0.6	0.6	0.8	0.7		29.5	7.4	1.0
PCB _{20C}	1.1	0.6	1.2		0.6		114.5	16.8
PCB _{158D}	0.6	0.6	0.5	0.7	1.0	0.6		20.3
PCB _{82D}	0.6	0.6	0.6	1.0	-	0.8	0.7	

Table S13. Total couplings (upper) and screening factors to the couplings (lower) for the PC630 using the MD-QM/MMPol approach. All values in (cm⁻¹).

s\V _{tot}	DBV _{50/61B}	DBV _{50/61D}	PCB _{158B}	PCB _{82B}	PCB _{158D}	PCB _{82D}	MBV _{19A}	MBV _{19C}
DBV _{50/61B}		340.2	20.7	-44.3	20.4	19.1	15.7	-44.5
DBV _{50/61D}	0.6		14.9	22.1	21.5	-48.4	-43.5	10.7
PCB _{158B}	0.6	0.6		18.7	11.6	-12.4	119.3	1.9
PCB _{82B}	0.6	0.6	0.6		-12.8	-14.4	19.6	54.9
PCB _{158D}	0.5	0.5	0.6	0.6		18.7	1.9	113.4
PCB _{82D}	0.6	0.6	0.6	0.7			52.5	18.4
MBV _{19A}	0.6	0.6	0.6	0.7	-	0.6		-3.90
MBV _{19C}	0.6	0.6	-	0.5	0.6	0.7	0.7	

Table S14. Total couplings (upper) and screening factors to the couplings (lower) for the PC645 using the MD-QM/MMPol approach. All values in (cm⁻¹).

s\V _{tot}	DBV _{50/61B}	DBV _{50/61D}	PCB _{158B}	PCB _{82B}	PCB _{158D}	PCB _{82D}	MBV _{19A}	MBV _{19C}
DBV _{50/61B}		312.1	5.1	-21.0	12.7	14.2	14.5	-31.0
DBV _{50/61D}	0.5		17.5	20.5	16.4	-47.0	-39.1	13.4
PCB _{158B}	0.6	0.6		13.3	10.4	-13.4	111.8	1.6
PCB _{82B}	0.6	0.6	0.7		-14.4	-10.2	19.2	52.4
PCB _{158D}	0.5	0.5	0.6	0.6		16.9	1.2	111.9
PCB _{82D}	0.6	0.6	0.6	0.7	0.7		53.9	18.9
MBV _{19A}	0.7	0.6	0.5	0.7	-	0.5		-3.34
MBV _{19C}	0.6	0.6	-	0.5	0.6	0.7	-	

Table S15. Total couplings (upper) and screening factors to the couplings (lower) for the PC577 using the QM/MMPol/ddCOSMO approach. All values in (cm⁻¹).

s\V _{tot}	DBV _{50/61B}	DBV _{50/61D}	PCB _{20A}	PCB _{158B}	PCB _{82B}	PCB _{20C}	PCB _{158D}	PCB _{82D}
DBV _{50/61B}		-18.5	19.0	27.9	-52.1	-12.9	-0.7	20.9
DBV _{50/61D}	-1.0		-13.3	-1.5	20.6	18.6	28.1	-52.5
PCB _{20A}	0.5	1.1		75.0	16.8	-7.2	2.8	26.8
PCB _{158B}	0.6	0.4	0.5		14.0	2.7	4.9	6.3
PCB _{82B}	0.6	0.6	0.7	0.6		26.0	6.2	2.4
PCB _{20C}	1.2	0.5	1.7	0.9	0.6		78.5	16.6
PCB _{158D}	0.2	0.6	0.9	0.6	0.6	0.5		13.7
PCB _{82D}	0.6	0.6	0.6	0.6	0.4	0.7	0.6	

Table S16. Total couplings (upper) and screening factors to the couplings (lower) for the PC612 using the QM/MMPol/ddCOSMO approach. All values in (cm⁻¹).

s\V _{tot}	DBV _{50/61B}	DBV _{50/61D}	PCB _{20A}	PCB _{158B}	PCB _{82B}	PCB _{20C}	PCB _{158D}	PCB _{82D}
DBV _{50/61B}		-16.2	15.7	27.5	-52.1	-14.0	-1.3	19.0
DBV _{50/61D}	-0.8		-12.8	-0.8	19.0	15.0	28.4	-51.6
PCB _{20A}	0.5	1.2		78.9	16.7	-10.7	1.4	24.6
PCB _{158B}	0.6	0.3	0.5		13.7	1.9	4.8	4.4
PCB _{82B}	0.6	0.6	0.7	0.6		28.1	4.8	1.9
PCB _{20C}	1.1	0.5	1.5	1.1	0.6		74.6	17.6
PCB _{158D}	0.5	0.6	2.2	0.6	0.6	0.5		13.5
PCB _{82D}	0.6	0.6	0.6	0.6	0.3	0.7	0.6	

Table S17. Total couplings (upper) and screening factors to the couplings (lower) for the PC630 using the QM/MMPol/ddCOSMO approach. All values in (cm⁻¹).

s\V _{tot}	DBV _{50/61B}	DBV _{50/61D}	PCB _{158B}	PCB _{82B}	PCB _{158D}	PCB _{82D}	MBV _{19A}	MBV _{19C}
DBV _{50/61B}		212.7	25.8	-48.0	27.7	18.4	11.2	-46.6
DBV _{50/61D}	0.4		26.4	20.5	30.9	-51.3	-48.8	9.1
PCB _{158B}	0.6	0.5		11.9	8.7	-11.4	68.4	3.0
PCB _{82B}	0.5	0.5	0.5		-9.8	-9.2	17.7	56.4
PCB _{158D}	0.4	0.5	0.5	0.5		12.5	3.4	72.7
PCB _{82D}	0.5	0.5	0.5	0.6	0.5		55.8	16.0
MBV _{19A}	0.5	0.6	0.5	0.6	0.5	0.5		-4.8
MBV _{19C}	0.6	0.5	0.6	0.5	0.5	0.6	0.6	

Table S18. Total couplings (upper) and screening factors to the couplings (lower) for the PC645 using the QM/MMPol/ddCOSMO approach. All values in (cm⁻¹).

s\V _{tot}	DBV _{50/61B}	DBV _{50/61D}	PCB _{158B}	PCB _{82B}	PCB _{158D}	PCB _{82D}	MBV _{19A}	MBV _{19C}
DBV _{50/61B}		238.0	28.7	-46.9	27.4	18.7	13.4	-46.8
DBV _{50/61D}	0.4		24.8	19.9	30.8	-49.4	-50.4	8.2
PCB _{158B}	0.5	0.5		11.8	9.7	-10.1	76.9	2.9
PCB _{82B}	0.5	0.5	0.6		-10.1	-7.9	16.6	54.6
PCB _{158D}	0.4	0.5	0.5	0.5		12.1	3.5	79.7
PCB _{82D}	0.5	0.5	0.5	0.6	0.6		56.0	16.4
MBV _{19A}	0.6	0.6	0.5	0.6	0.6	0.5		-4.6
MBV _{19C}	0.6	0.5	0.6	0.5	0.5	0.6	0.7	

Supporting Information for

Single amino acid mutation controls hole transfer dynamics in DNA-methyltransferase HhaI complexes

Marina Corbella,^a Alexander A. Voityuk^{bc} and Carles Curutchet^a

^aDepartament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain ^bInstitució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain ^cInstitut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Campus Montilivi, 17071 Girona, Spain

Table of contents

Molecular dynamics simulations	2
Quantum-mechanical calculations	3
Estimation of HT rates	3
Kinetic Monte Carlo Algorithm	3
Structural deformation of the DNA stack	4
HT parameters	5
References	6
Figure S1. RMSD deviations	7
Figure S2. Stacking overlap	8
Figure S3. Rise base pair step parameter	8
Figure S4. Slide base pair step parameter	9
Figure S5. Shift base pair step parameter	9
Figure S6. Twist base pair step parameter	10
Figure S7. Roll base pair step parameter	10
Figure S8. Tilt base pair step parameter	11
Figure S9. Representation of the binding domain in the DNA-X system	11
Figure S10. Scheme of base pair step parameters	12
Figure S11. Distribution of HT pathways for DNA and DNA-X systems	13
Figure S12. Distribution of HT pathways for WT and Q237W systems	14
Table S1. Electronic couplings, free energy differences, and time constants for the forward/backward HT reactions	15

Molecular Dynamics Simulations

Four systems were studied (DNA, DNA-X, DNA-X bound to the WT and Q237W mutant enzymes). The starting DNA duplex structure was created with standard B-DNA fibre coordinates using the NAB module of the Amber12 suite of programs^[1] and adopting the sequence of the Ru-1 assembly described in Ref. ^[2]. This duplex was modified in order to delete a cytosine nucleobase and generate the abasic X site (DNA-X system). The unmodified duplex was also retained (DNA system) in order to assess the impact of the X gap in the DNA HT properties. The structure of the native *Hhal* enzyme was based on the X-ray crystal structure reported (Protein Data Bank ID code 1MHT) at 2.8 Å resolution,^[3] where the DNA duplex was substituted by the DNA-X system just described (WT system). In the Q237W mutant, the Gln237 was substituted by a tryptophan (Q237W system). Protonation states of all titratable residues were computed at neutral pH using the automated PROPKA server^[4], indicating a standard ionization state for all residues. Systems were solvated in a truncated octahedron box with TIP3P^[5] water molecules at 14 Å marginal radius and sodium ions were added to obtain neutral systems, using the Leap module of the Amber package^[1]. The protein and the nucleic acids were described using the ff99SB^{[6],[7]} and the parmbsc0^[8] AMBER force fields, respectively. In addition, we performed MD simulations of a single guanine nucleobase and a tryptophan in water solution, in order to calibrate the relative oxidation energies of both molecules.

The solvent and counterions of the 4 systems considered (DNA, DNA-X, WT and Q237W) were first minimized during 1000 steps, and then the full systems were further minimized for 2500 steps. In the Q237W system, the latter minimization step included constraints (500 kcal/mol·Å²) on the nucleobases and the inserted Trp237 in the binding domain (See Fig. 2). Then, the DNA and DNA-X systems were gradually heated up to 300 K with weak constraints in the nucleic acids (10 kcal/mol·Å²) at constant volume during 200 ps. Simulations were then extended for 200 ns (P = 1 bar, T = 300 K). The WT system was gradually thermalized up to 300 K at constant volume during 200 ps, applying distance restraints over the three hydrogen bonds involving Gln237 and G₃ and the hydrogen bond involving Gln237 and Ser87.^[9] Finally, 200 ns of unrestrained MD simulation (P = 1 bar, T = 300 K) were carried out. On the other hand, the Q237W system was gradually thermalized up to 300 K during 50 ps with constraints (10 kcal/mol·Å²) in the nucleobases and the Trp237 in the binding domain, followed with 150 ps performed by eliminating the constraints in the Trp237, but including distance restraints (20 kcal/mol·Å²) between the Trp237 and the G₂ and G₄ sites, and dihedral angle restraints (20 kcal/mol·Å²) in order to maintain the stacked orientation of Trp237. The simulation was then extended for 200 ns (P = 1 bar, T = 300 K), gradually releasing all restraints along the first 15 ns. Finally, the free guanine and tryptophan systems were initially energy minimized for 1000 steps, then thermalized up to 300 K at constant volume during 20 ps, and subsequently we extended the simulation for 50 ns (P = 1 bar, T = 300 K) for production purposes.

All runs were performed with Amber12^[1] using an integration time step of 2 fs in conjunction with SHAKE algorithm to restrain all bonds involving hydrogen, periodic boundary conditions, the Particle Mesh Ewald approach, and a nonbonded cutoff equal to 10 Å.

The mutual arrangement of neighboring base pairs in the different systems along the MD simulations were characterized by six base pair step parameters (BPSP, see Fig. S10), three translations (rise, shift and slide) and three rotations (twist, tilt and roll), as well as the stacking overlap between adjacent pairs, using the x3dna software package.^[10] The reference configuration for a B-DNA describes 0.0, -0.2 and 3.4 Å shift, slide and rise parameters, and 0°, -3° and 36° tilt, roll and twist angles, respectively.^[11] We could not determine BPSP parameters for the binding domain due to the missing cytosine base.

Quantum-mechanical calculations

Quantum-chemical calculations were performed using the semiempirical INDO/S method^[12,13] on 1000 structures extracted each 10 ps along the last 100 ns of the MD trajectories. The INDO/S method has been shown to provide a reliable description of electronic couplings of stacked nucleobases.^[14] In particular, quantum-chemical calculations were performed on a model system comprising the nucleobases in the binding domain (except the X abasic site) plus the Gln (wild-type) or Trp (Q237W mutant) residues inserted into the DNA pocket, as shown in Fig. 1, given that experiments indicate that HT transport along the DNA stack toward the binding site is not the rate-limiting step of the process.^[2] Such models were created by deleting sugar and backbone atoms, where the C1' atom of the sugar moiety was replaced by a hydrogen atom with the bond length adjusted to 1.08 Å. Electrostatic embedding effects exerted by the environment, including protein, DNA, solvent and counterions, were considered adopting the charges as defined in the force field used for the MD simulations.

For structures of the WT system, four electronic states of the lowest energy corresponding to hole states on the G sites were analyzed. In the Q237W mutant, a further state corresponding to the tryptophan radical cation was also included. These states represented by four or five highest-occupied molecular orbitals (HOMOs) of the neutral species (HOMO→HOMO-3/HOMO-4) are associated with each of the four guanine bases in the stack and the tryptophan. HT couplings between the states were estimated using the Fragment Charge Difference method (FCD)^[15].

$$V_{DA} = \frac{(E_2 - E_1)|\Delta q_{12}|}{\sqrt{(\Delta q_1 - \Delta q_2)^2 + 4\Delta q_{12}^2}} \quad (1)$$

For HT, the adiabatic splitting $E_2 - E_1$ can be estimated through the one-electron energies of the highest occupied molecular orbitals (HOMOs) of neutral systems. In Eq. (1), Δq_1 and Δq_2 are the difference of hole charges on the donor and acceptor sites in the adiabatic states of interest, Δq_{12} is the corresponding off-diagonal term. The calculation of these quantities as well as underlying approximations are described in detail in Ref.^[15] and^[16].

Estimation of HT rates

Hole transfer rates were estimated for all forward and reverse HT reactions shown in Fig. 3 using Marcus Theory:

$$k_{HT} = \frac{2\pi}{\hbar} |V|^2 \frac{1}{\sqrt{4\pi k_B T \lambda}} e^{-(\lambda + \Delta G^0)^2 / 4\lambda k_B T} \quad (2)$$

In Eq. (2), ΔG^0 is the driving force of the charge transfer reaction. We estimate this driving force from the energy difference between the donor/acceptor ionization potentials, λ is the reorganization energy, composed by the internal (λ_{int}) and the solvent (λ_s) terms. We assume a value of 0.7 eV in all cases.^[17] For charge transfer in biomolecules, λ is usually assumed to be in the range of 0.5 to 1.5 eV. For the simulation of each HT rate, we used the energies and squared couplings averaged over 1000 structures sampled along the MD trajectories.

Kinetic Monte Carlo Algorithm

We estimated the effective rates and the paths of the overall HT processes from G_1 to G_4 for each system adopting a Kinetic Monte Carlo algorithm.^[18] These two guanines are separated by a distance of ~12 Å. The rates estimated by Eq. (2) were used to generate a cumulative probability distribution function (CPD) for each possible step. This function is associated with a transition $i \rightarrow n$ according to Eq. (3), where $\sum_m k_{im}$ is the sum of transfer rates from the same origin (i).

$$CPD_{i(n)} = \frac{\sum_{p=1}^n k_{ip}}{\sum_m k_{im}} \quad (3)$$

The Monte Carlo hopping algorithm can be summarized in 5 steps: 1. Compute the CPD function for each allowed pairwise transfer; 2. Start at site i (G_1) and choose a random number $0 < r_1 \leq 1$ and hop to site j , where $CPD_{i(j)} \geq r_1 > 0$; 3. Compute the waiting time t_w at site i , generating a new random number $0 < r_2 \leq 1$ as $t_w = -(\ln(r_2)/\sum_m k_{im})$; 4. Choose a new random number and repeat steps 2 and 3 keeping track of the total accumulated waiting time; 5. Stop if the final site has been reached or if the total time exceeds a fixed value t_{\max} .

10^6 realizations were generated for each system, and for each one of them, the time taken to reach G_4 as well as the pathway followed by the hole were recorded.

Structural deformation of the DNA stack

In order to understand the impact of the structural deformation of the DNA helix and the insertion of the amino acid in the DNA stack in the HT dynamics, we have performed MD simulations on the DNA and DNA-X systems, as well as the DNA-X system bound to the WT and Q237W mutant *Hhal* enzymes. For the Q237W mutant, steric hindrance only allows for one possible orientation of the Trp side chain in the DNA pocket, as illustrated in Fig. 1. Such orientation is consistent with the Trp fulfilling the π -gap generated in the DNA stack upon binding, as previously hypothesized.^[2] Detailed analysis of the positional root mean square deviations (RMSD) show that fluctuations of the DNA backbone are larger than those of the protein backbone (see Fig. S1). However, fluctuations of the binding domain are somewhat reduced, especially in comparison with the DNA-X system, where fluctuations are enhanced in the abasic position. In addition, in the DNA-X system the abasic site causes a displacement of G_3 to fill the gap due to the abasic site, and the ejection of the upper cytosine residue – the base pair of G_4 – to allow a better stacking between guanines G_3 and G_4 .

Analysis of the variation of the stacking overlap and base pair step parameter (BPSP) near the binding domain clearly indicates a remarkable deformation of the standard DNA structure (see Fig. S2-S8); in the DNA-X we observe the largest differences, especially at the CA/TG₄ pair, as a result of the missing cytosine mentioned above, whereas in the DNA/protein complexes WT and Q237W such deformation is somewhat reduced. This analysis illustrates the strong impact of both protein binding and the presence of the X abasic site on the structural parameters of the double helix.

On the other hand, it was suggested that the Q237W mutant, in contrast to the WT protein, retains long-range oxidative damage at distal 5'-GG-3' sites because the inserted Trp fulfills the π -gap in the DNA stack by adopting a stacked configuration relative to the nucleobases. In order to investigate this point, we also analyzed the orientation and degree of insertion of the Gln and Trp amino acids into the DNA pocket in the WT and Q237W assemblies. The position of the amino acid residue inside the double strand impacts the HT energies and couplings that control HT dynamics. We define the degree of insertion of Gln (WT) and Trp (Q237W) within the DNA π -stack as the ratio between the distance among upper and lower stacked guanines geometric centers and the sum of the center-to-center distances among the inserted amino acid side chain and such guanines, $d_{G2-G4}/(d_{aa-G2} + d_{aa-G4})$. Thus, the degree of insertion is in the range 0-1, where 1 indicates complete insertion of the amino acid in the π -stack. In Fig. 4 we report the degree of insertion of Gln and Trp along the MD trajectory, as well as the insertion of G_3 . There is a competition between Trp and G_3 to fill the abasic site space in the Q237W system. Initially, Trp237 remains almost completely inserted in the π -gap, but after 60 ns both residues seem to reach a stacked and complementary arrangement that remains stable until the end of the simulation. In contrast, the insertion of the Gln side chain in the WT system remains stable during the whole simulation due to three hydrogen bonds established between G_3 , Gln and

Ser87, which appear to play a key role in maintaining the structure of the assembly during the catalytic process.^[9]

HT parameters

The deformation of the DNA structure, as well as the insertion of the Gln or Trp amino acid in the π -gap, modulates the energies of the radical cation states involved in the HT process as well as the electronic interactions between such states as predicted by our calculations. As shown in Fig. 4, in DNA all guanine sites have similar energies averaged over the MD simulation. In addition, we predict similar values of electronic couplings, ~6–8 meV, between the G₁ and G₂/G₃ sites, as well as between G₂/G₃ and G₄. The coupling among the G₂–G₃ sites, V₂₃, in contrast, is significantly enhanced amounting to ~30 meV, despite the apparent similarity to V₁₂ and V₃₄ interactions. This difference could arise from the fact that only the nucleobases in the binding domain are described at the quantum chemical level, therefore potential superexchange effects mediated by the surrounding bases could explain this enhancement. Alternatively, the difference could arise due to the fact that the outer GC pairs in the binding domain are stacked to AT pairs instead of GC ones.

In the DNA-X system containing the abasic site, most couplings remain similar, except V₂₄ and V₃₄, those between the final G₄ site and the intermediate G₂/G₃ sites. In particular, V₂₄ increases from ~7 to ~15 meV, whereas V₃₄ raises from ~8 to ~66 meV. This dramatic change arises from the displacement of G₃ in order to fill the gap due to the abasic site, and the ejection of the upper cytosine residue – the base pair of G₄ – to allow a better stacking between guanines G₃ and G₄, as discussed in the previous section (See Fig. S9). Such structural change, in addition, leads to a significant destabilization of the G₃ cation state, whereas the other guanines in the binding domain retain similar relative energies. These enhanced couplings among guanines in the DNA-X compared to the undamaged DNA significantly enhance the rate of HT through the binding domain. Moreover, the destabilization of the G₃ state does not slow down the HT process, given that holes can directly hop from G₂ to G₄.

When the DNA-X helix binds the wild-type *Hhal* protein, however, the insertion of the Gln amino acid in the π -gap precludes the movement of G₃ site leading to a better stacking with G₄ observed in the DNA-X system. This structural change has a strong impact on the electronic interactions among the guanines in the binding site. In contrast to the enhancements of V₂₄ and V₃₄ estimated in DNA-X, upon binding both interactions are dramatically attenuated leading to MD-averaged values of ~2 and ~4 meV, respectively. Moreover, the V₂₃ coupling is halved, passing from ~31 meV to ~12 meV in WT system. The energies of the guanine cation states are also significantly modified upon binding and insertion of the Gln side chain. In particular, the energy of the G₂ and G₄ guanine states are destabilized compared to the other sites, especially in the case of G₄. Whereas in the undamaged DNA the guanine cation states span a range of energies of 36 meV, a range that increases to 104 meV in the DNA-X, upon binding the energy asymmetry is enhanced due to the strong destabilization of G₄, leading to a range of 183 meV (see Fig. 4). The destabilization of the neighboring G 3' and G 5' sites to the inserted Gln residue, as well as the attenuated electronic interactions among the guanines, suggest an important slow down of the HT rate in the DNA-X upon binding to the wild-type protein, thus explaining the inhibition of long-range oxidative damage observed experimentally.

Interestingly, the energetic landscape in the binding site observed for the WT system is considerably changed in the Q237W mutant, leading to a situation closer to that observed for the unbound DNA-X system. Compared to the WT complex, in the Q237W assembly both G₂ and G₄ sites are stabilized due to the insertion of the stacked tryptophan, thus recovering the possibility to have fast HT toward the final G₄ site in the binding domain. Moreover, the mutation of the inserted Gln into a Trp leads to enhanced electronic interactions between most guanine sites. In particular, V₁₂ and V₂₃ interactions are doubled, whereas V₂₄ and V₃₄ experience twenty and four-fold increases, respectively. Such increases are mainly associated with a superexchange effect due to the presence of the aromatic amino acid tryptophan, whose p-orbitals mediate the exchange of electrons between guanine orbitals.

The changes in the guanine electronic properties suggest that the HT process in the mutant will be significantly faster compared to the WT enzyme, in accord with the experimental observations. In the Q237W mutant, there's a further actor coming into play, the Trp residue. Because of that one should obtain reliable energies of guanines and Trp sites in order to properly estimate the HT dynamics. To this end we have performed INDO/S calculations along MD trajectories of single guanine and tryptophan in water. These calculations predict the Trp state being higher in energy by 0.35 eV, whereas the experimental estimates derived from the one-electron oxidation potentials at neutral pH are 1.29 and 1.03 V for G and Trp, respectively.^[19] Therefore, we have applied a correction of -0.61 eV to the energy of the tryptophan cation state in the simulation of the Q237W complex.

References

- [1] D. A. Case, T. A. Darden, T. E. I. Cheatham, C. L. Simmerling, J. Wang, R. . Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, et al., *AMBER 12* **2012**, University of California, San Francisco.
- [2] H. A Wagenknecht, S. R. Rajski, M. Pascaly, E. D. Stemp, J. K. Barton, *J. Am. Chem. Soc.* **2001**, 123, 4400–7.
- [3] S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, *Cell* **1994**, 76, 357–369.
- [4] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, 7, 525–537.
- [5] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, 79, 926.
- [6] J. Wang, P. Cieplak, P. A. Kollman, *J. Comput. Chem.* **2000**, 21, 1049–1074.
- [7] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, 65, 712–25.
- [8] A. Pérez, F. Lankas, F. J. Luque, M. Orozco, *Nucleic Acids Res.* **2008**, 36, 2379–94.
- [9] S. Mi, D. Alonso, R. J. Roberts, *Nucleic Acids Res.* **1995**, 23, 620–7.
- [10] X.-J. Lu, *Nucleic Acids Res.* **2003**, 31, 5108–5121.
- [11] C. Curutchet, A. A. Voityuk, *Angew. Chemie - Int. Ed.* **2011**, 50, 1820–1822.
- [12] J. Ridley, M. Zerner, *Theor. Chim. Acta* **1973**, 32, 111–134.
- [13] A. A. Voityuk, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, 3, 515–527.
- [14] A. A. Voityuk, *Chem. Phys. Lett.* **2006**, 427, 177–180.
- [15] A. A. Voityuk, N. Rösch, *J. Chem. Phys.* **2002**, 117, 5607.
- [16] N. Rösch, A. A. Voityuk, *Top. Curr. Chem.* **2004**, 237, 37–72.
- [17] K. Siriwong, A.A. Voityuk, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, 2, 780–794.
- [18] J. Singh, E. R. Bittner, D. Beljonne, G. D. Scholes, *J. Chem. Phys.* **2009**, 131, 194905.
- [19] S. Steenken, S. V Jovanovic, M. Curie, C. Kn, *J. Am. Chem. Soc.* **1997**, 119, 617–618.

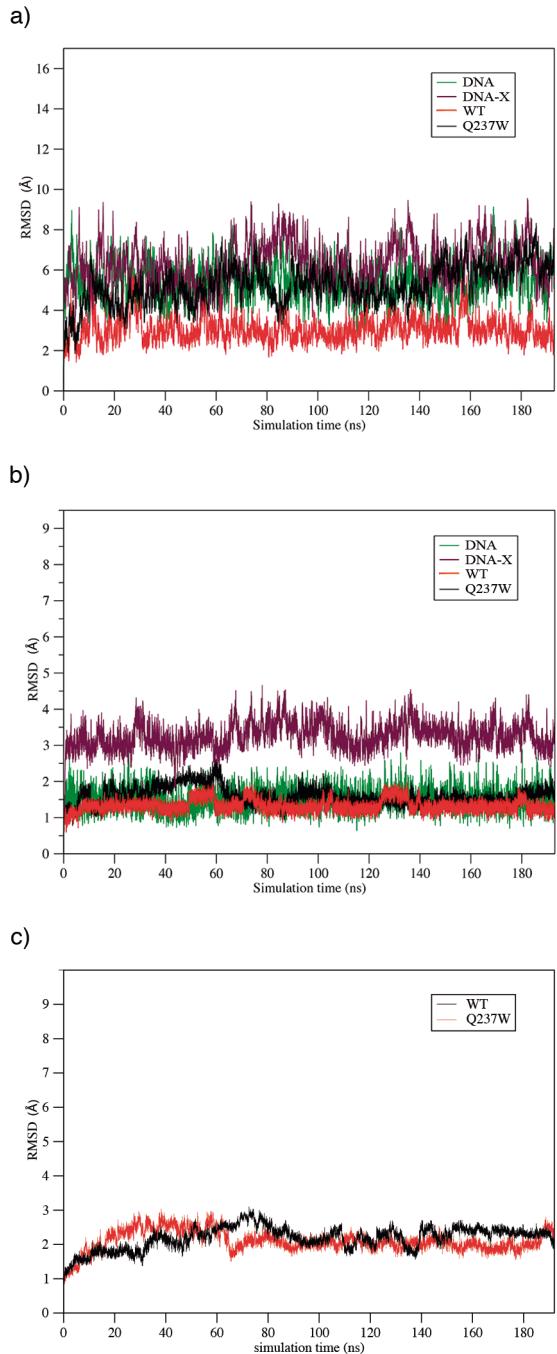


Figure S1. Positional root-mean square deviations of DNA, DNA-X, WT and Q237W systems corresponding to the a) DNA backbone, b) the DNA binding domain and c) the protein backbone.

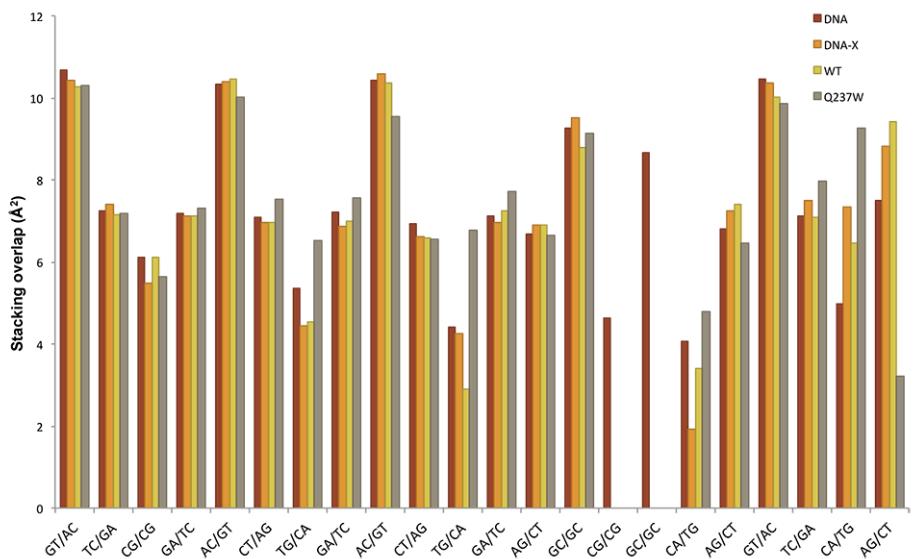


Figure S2. Stacking overlap between base pairs corresponding the DNA stack, the DNA-X stack with an abasic site, and the DNA-X stack complexed with the WT and Q237W mutant *HhaI* protein. Values are averaged along the MD classical trajectories.

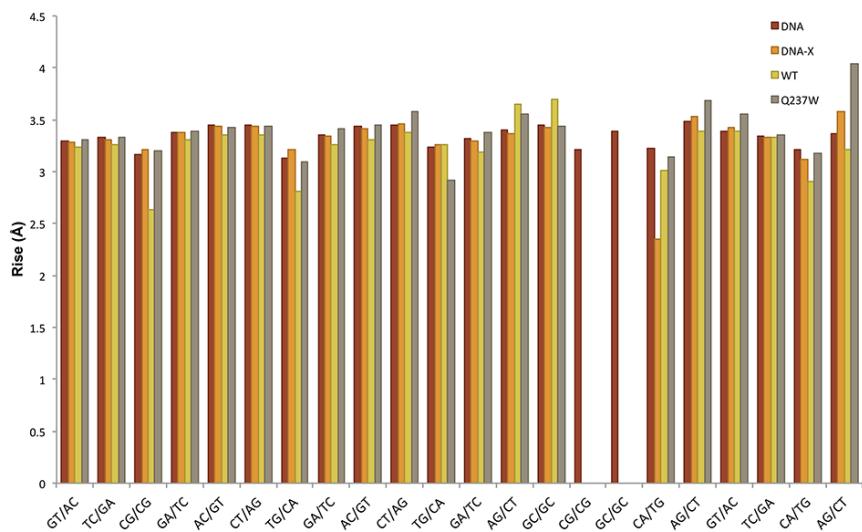


Figure S3. Rise base pair step parameter corresponding the DNA stack, the DNA-X stack with an abasic site, and the DNA-X stack complexed with the WT and Q237W mutant *HhaI* protein. Values are averaged along the MD classical trajectories.

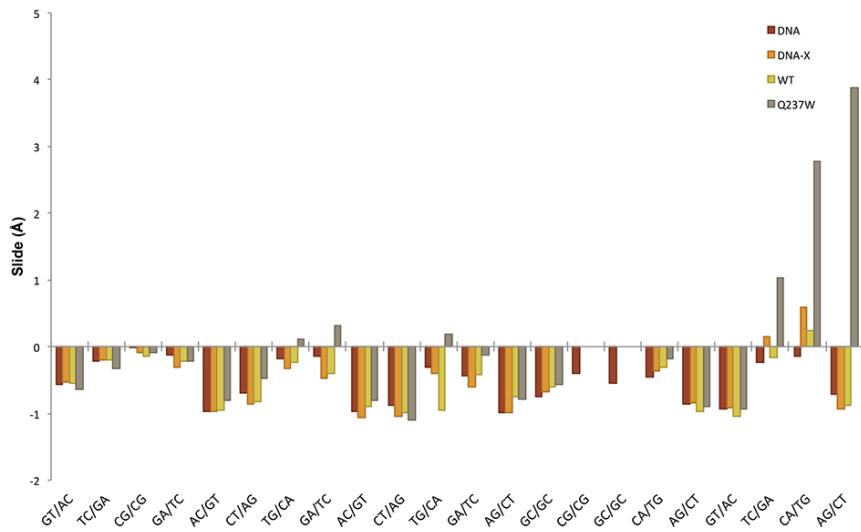


Figure S4. Slide base pair step parameter corresponding the DNA stack, the DNA-X stack with an abasic site, and the DNA-X stack complexed with the WT and Q237W mutant *HhaI* protein. Values are averaged along the MD classical trajectories.

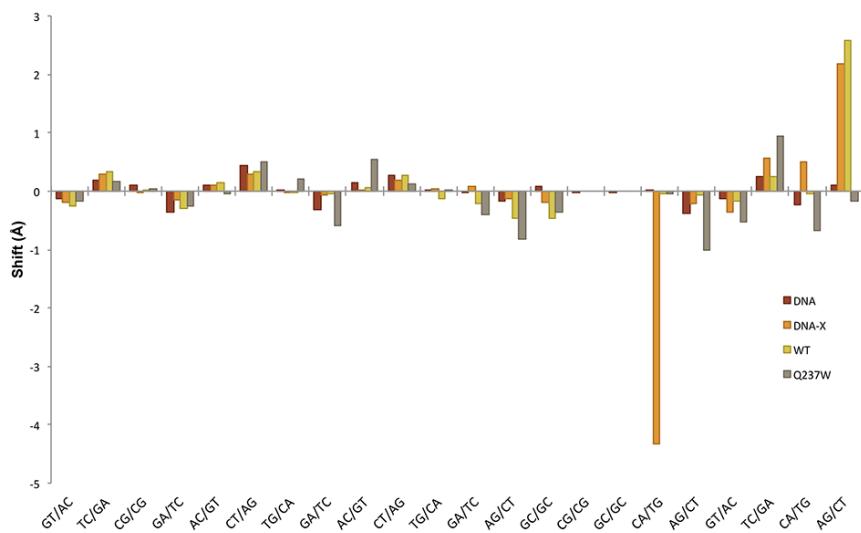


Figure S5. Shift base pair step parameter corresponding the DNA stack, the DNA-X stack with an abasic site, and the DNA-X stack complexed with the WT and Q237W mutant *HhaI* protein. Values are averaged along the MD classical trajectories.

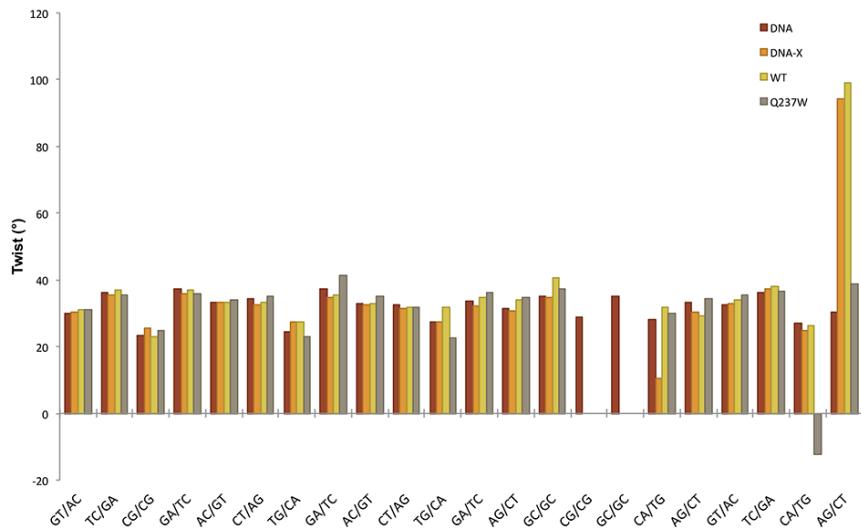


Figure S6. Twist base pair step parameter corresponding the DNA stack, the DNA-X stack with an abasic site, and the DNA-X stack complexed with the WT and Q237W mutant *HhaI* protein. Values are averaged along the MD classical trajectories.

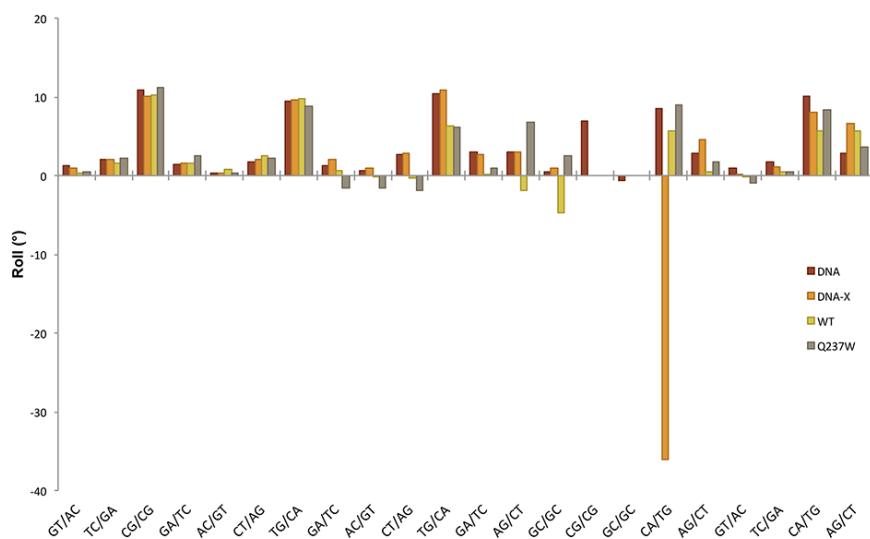


Figure S7. Roll base pair step parameter corresponding the DNA stack, the DNA-X stack with an abasic site, and the DNA-X stack complexed with the WT and Q237W mutant *HhaI* protein. Values are averaged along the MD classical trajectories.

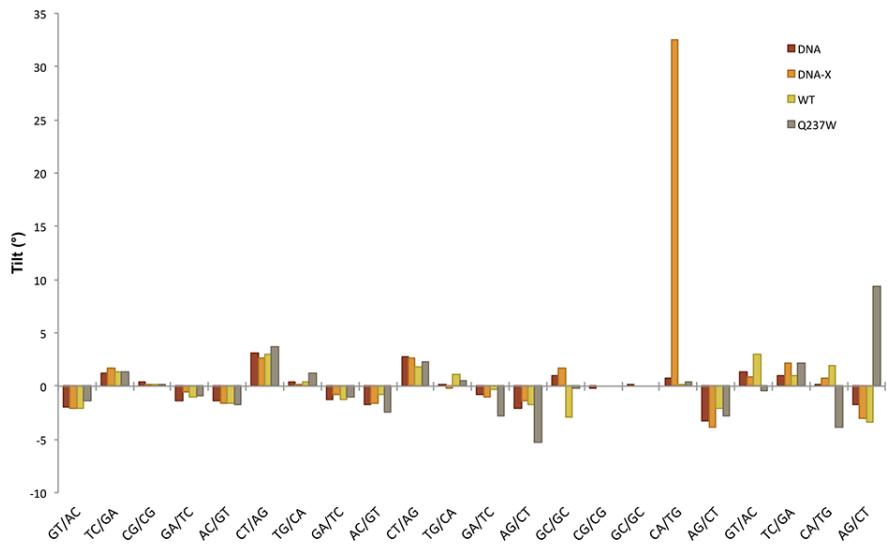


Figure S8. Tilt base pair step parameter corresponding the DNA stack, the DNA-X stack with an abasic site, and the DNA-X stack complexed with the WT and Q237W mutant *HhaI* protein. Values are averaged along the MD classical trajectories.

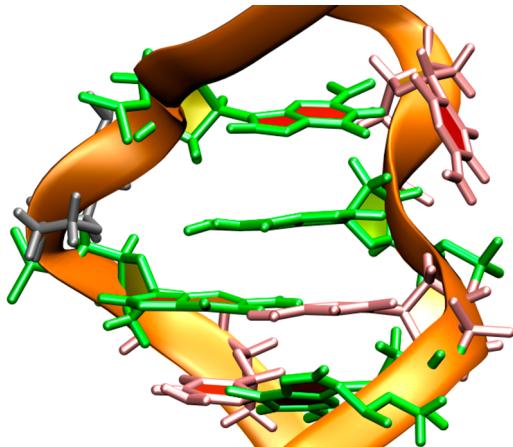


Figure S9. Structure of the binding domain in the DNA-X stack illustrating the displacement of G₃ in order to fill the gap due to the abasic site, and the ejection of the upper cytosine residue – the base pair of G₄ – to allow a better stacking between guanines G₃ and G₄.

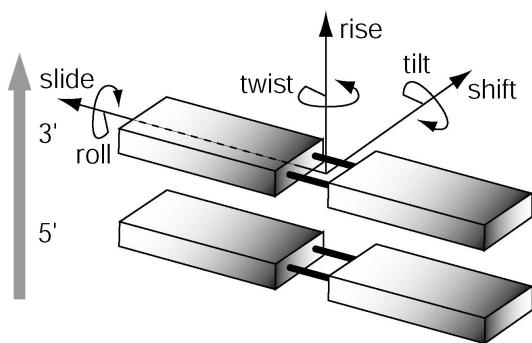


Figure S10. Parameters that define the arrangement of adjacent base pairs in a π stack.

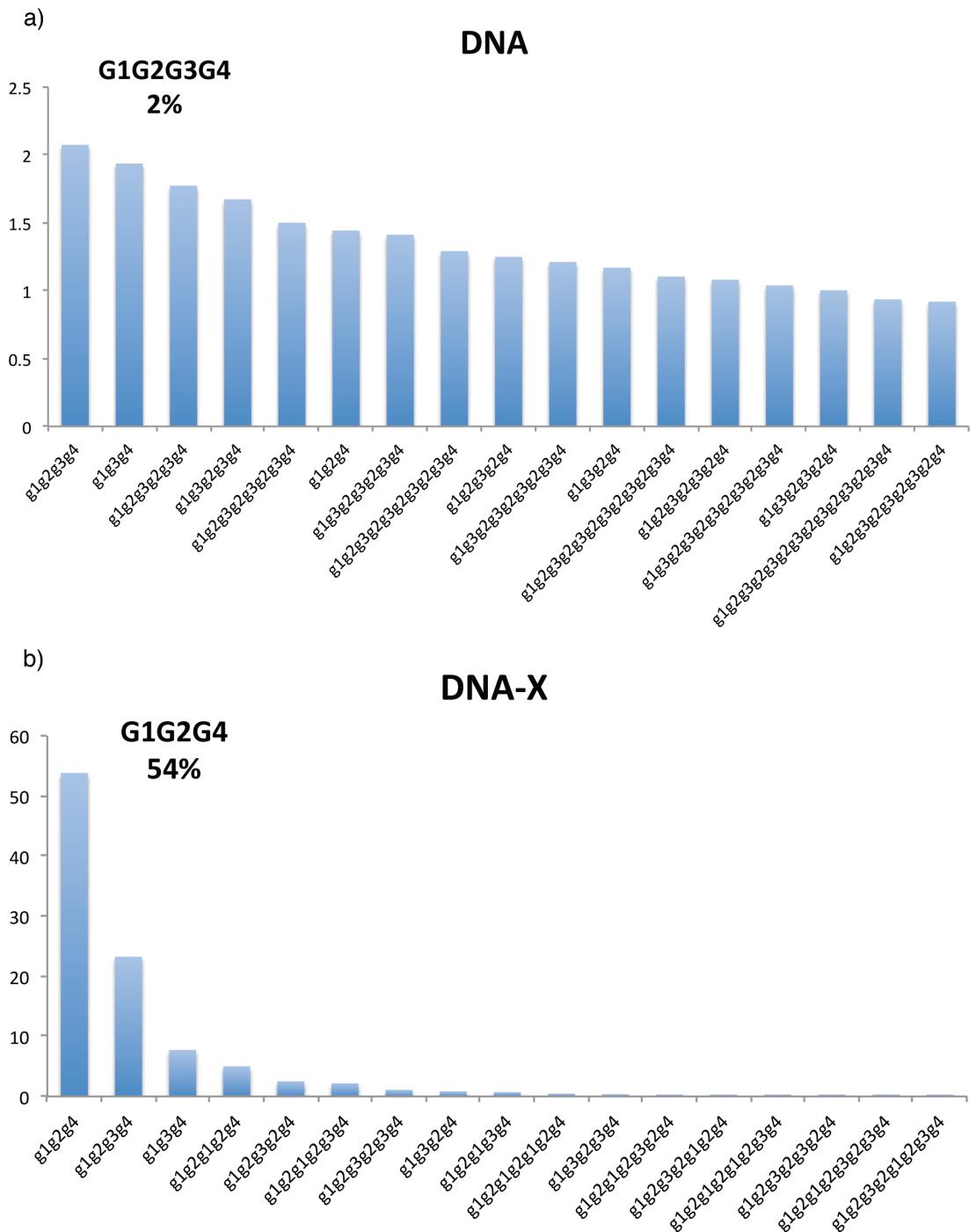


Figure S11. Distribution of preferred pathways obtained from the Kinetic Monte Carlo simulations corresponding to a) DNA and b) DNA-X.

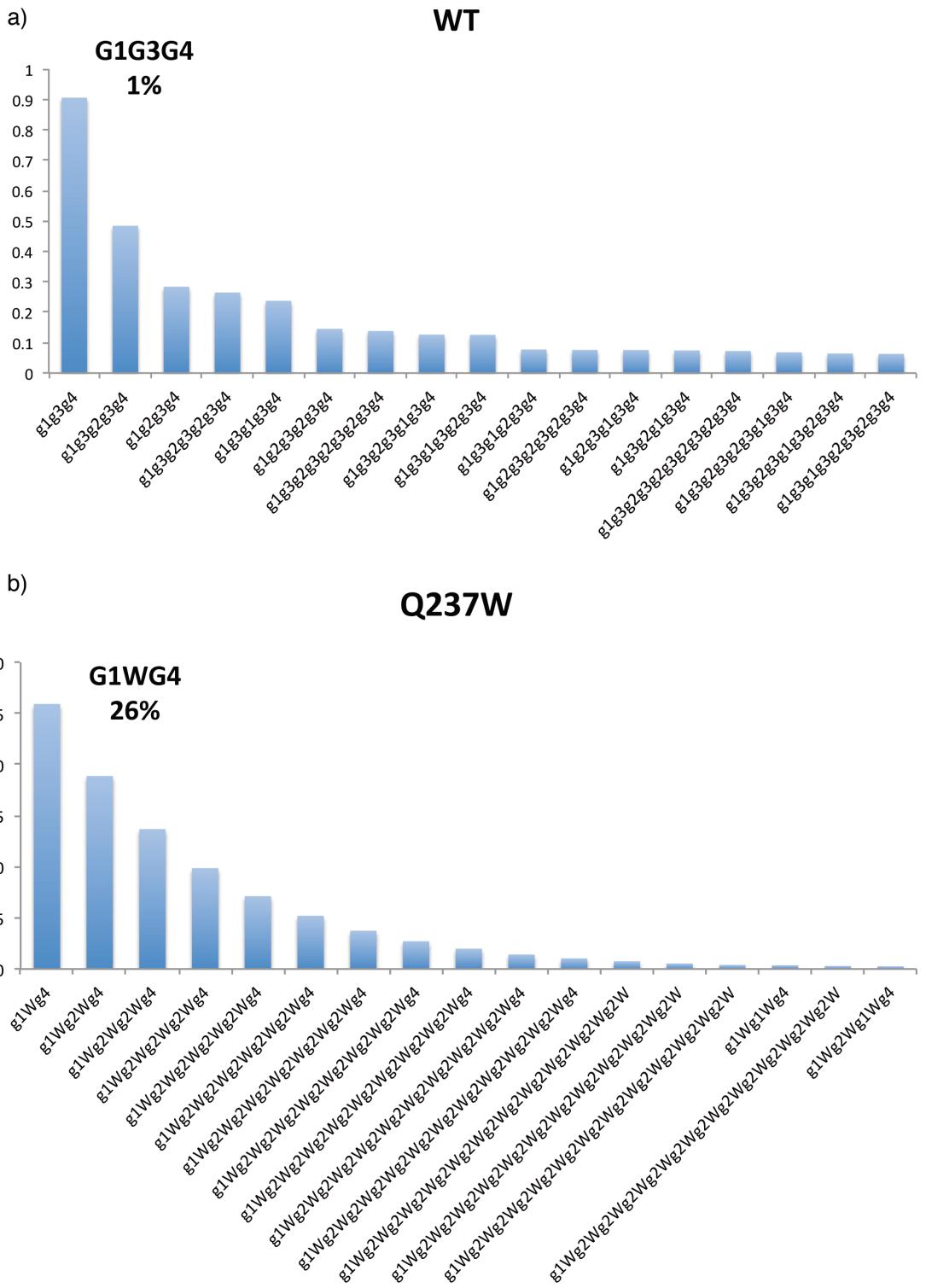


Figure S12. Distribution of preferred pathways obtained from the Kinetic Monte Carlo simulations corresponding to a) WT complex and b) Q237W complex.

Table S1. Electronic couplings, free energy differences, and time constants for the forward/backward HT reactions estimated for the DNA, DNA-X, WT and Q237W systems.

	$V^{[a]}$	$\Delta G^{[a]}$	$\tau_F^{[b]}$	$\tau_B^{[b]}$
DNA				
G1 - G2	8	4	0.85	0.71
G1 - G3	6	-14	0.98	1.70
G2 - G3	30	-19	3.4×10^{-2}	7.1×10^{-2}
G2 - G4	7	17	1.16	0.59
G3 - G4	8	36	1.56	0.38
DNA-x				
G1 - G2	8	-13	0.60	0.98
G1 - G3	6	80	5.90	0.26
G2 - G3	31	93	0.33	8.8×10^{-3}
G2 - G4	15	-10	0.17	0.26
G3 - G4	66	-104	1.6×10^{-3}	9.1×10^{-2}
WT				
G1 - G2	6	76	5.34	0.28
G1 - G3	4	-21	1.94	4.42
G2 - G3	12	-97	6.0×10^{-2}	2.45
G2 - G4	2	86	94.1	3.35
G3 - G4	4	183	124	0.10
Q237W				
G1 - G2	13	-27	0.16	0.44
G1 - G3	4	214	275	7.0×10^{-2}
G2 - G3	31	240	11.5	9.9×10^{-4}
G2 - G4	44	35	4.8×10^{-2}	1.2×10^{-2}
G3 - G4	16	-205	6.1×10^{-3}	17.7
G1 - W	20	-368 ^[c]	5.9×10^{-4}	984
G2 - W	96	-341 ^[c]	3.3×10^{-5}	19.1
G3 - W	17	-581 ^[c]	2.1×10^{-4}	1.4×10^6
W - G4	96	376 ^[c]	53.2	2.3×10^{-5}

[a] All values in meV. [b] Time constant for forward/backward HT reactions in ns adopting a reorganization energy value of $\lambda = 0.7$ eV. [c] Relative energy of the Trp237 site corrected by -0.61 eV (see text).