eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Proteins and DNA as Sources of Catalytic and Biological Activities: from *Huisgenases* to Chemotherapy

Doctoral Thesis

Mikel Odriozola Gimeno

Donostia/San Sebastián

2019

Tengo claro que después de estos cinco años sigo siendo la misma persona, con unas cuantas experiencias más, con algunas mejoras y con algunas cosas que habré empeorado, pero al fin y al cabo, la misma persona. Tengo mucha gente a la que agradecer su presencia y apoyo a lo largo de este periodo, no voy a entrar en nombres, no quiero ni olvidarme de nadie ni que sea un apartado para curiosos.

Me gustaría empezar por el inicio, dando gracias a las personas que me dieron la oportunidad de hacer la tesis en este grupo y a la gente que me ha supervisado durante todo este periodo. Del mismo modo, no me quiero olvidar de nadie del laboratorio/grupo, compañeros de vitrina y despacho, experimentales, computacionales, gente de estancia, estudiantes… Con algunos he compartido más momentos que con otros, pero sin duda todos han sido importantes para llevar el día a día. He tenido la suerte de poder hacer un trabajo multidisciplinar en el que hemos colaborado muchas personas, las cuales han aportado para que todo salga adelante. A su vez he coincidido con gente de muchos sitios y ámbitos en el Korta, tanto científicos como no. GRACIAS A TODOS.

During my PhD I had the chance to spend three months in the laboratory of Prof. Houk at UCLA. Without any doubt that experience was one of the best experiences of my life. Thus, I would like to thank all the people that I met during that period: supervisors, labmates, basketball coaches, roommates, friends… I am really aware about the importance of that stay in my PhD, it was a really enriching experience. THANKS TO EVERYBODY.

Bukatzeko euskeraz ekingo diot; argi daukat arlo zientifikoa ezinbestekoa dela doktoradutza burutzeko, baina bihotzez uste det askotan ahazten garela gainontzeko gauzetaz. Denbora askoan korrika ibiltzen gara, lanez gainezka eta ez det gezurrik esan nahi, asko ikasi det zalantza izpirik gabe, momentu on ugari izan ditut, jende asko ezagutu, experientzia onuragarri asko gehitu, baina dena ez da izan lorez beteriko bidea. Momentu txarrak ere izan dira, zalantza ugariko momentuak eta antsietate momentuak, batez ere azken urtean, azkenean, ez gera manual batekin jaio. Tesiak ordu asko eskatu izan dizkit, baina argi neukan eta orain argiago, laborategitik irtetzean inguruan jendea edukitzea eta beste jarduera batzuk egitea ezinbestekoa dela.

Egia esan balantze positiboa egiten det eta harro nago honetaz, baina balantze hori ez zen posible izango inguruko jende guztia gabe. Gainera ziur laborategitik atera naizen guztietan nere umorea ez zela hoberena eta jende horrek aguantau behar izan nau, bai irrifartsu eta bai serioago. Lehenik eta behin, FAMILIA, batez ere ama, aita eta anaia, beraiek izan dira egunero aguantatu nauetenak bai momentu onetan zein zailetan, maite zaituztet. Sin duda también al resto de la familia que me habéis apoyado en este proceso, aunque no lo hablemos agradezco mucho teneros. Bestalde, LAGUNAK (kuadrila eta gainontzekoak, lehendik zeudenak eta bidean eginikoak), denbora asko kendu diet hauei ere liburuxka hau idazteagatik, eskerrak ezagutzen nauten eta ulertzen duten, asko eskertzen dizuet zuen laguntza. Askotan, kalimotxo bat alde zaharrean hartzea edo donostitik bueltatxo bat ematea deskonetatzeko aukera ezin hobeak izan dira niretzat. Ze zortea dudan. Saskibaloiko Bostekoko taldekoei, disfrutatzen laguntzeagatik, titulu eta guzti. Entrenatu ditudan gazte guztiei, beren familiei eta klubeko guztiei, beraien laguntza eta nigan euki duten konfidantza guztiagatik. Selekziokoei, aukera ezin hobeak izan ditut eta deskonetatzeko ezin hobe. Apuntatu naizen aktibitate guztietakoei eta bost urte hauetan nerekin koinziditu dezuten guztioi. Momentu baxuak aparte gauza guzti hauek zoriontsu egin naute eta ziur urte batzuk barru atzera begiratzean irribarre bat aterako zaidala. Azken finean doktorea izango naiz orain, baina ze zortea dudan inguruan ditudan pertsonak edukitzeaz. MILA ESKER DENORI.

This Thesis Dissertation has been carried out in the Department of Organic Chemistry I of the Chemistry Faculty at the University of the Basque Country, in Donostia-San Sebastián, under the supervision of Dr. Iván Rivilla de la Cruz, Dr. Miquel Torrent Sucarrat and Prof. Fernando P. Cossío Mora.

Between September-December 2016, a short term research stay was done at the University of California, Los Angeles, Department of Chemistry and Biochemistry under the supervision of Prof. Kendall N. Houk and Dr. Marc García Borrás.

This Dissertation is divided into three main Chapters. Chapter 1 is a general introduction for the following Chapters 2 and 3. Firstly, in sections 1.1 and 1.2 an introduction focused on catalytic proteins, more specifically, pericyclases along with the employment of QM/MM methodology for the description of the activity of the catalytic proteins is provided. Subsequently, in section 1.3, the use of platinum(II) derivatives as antitumoral drugs is presented, emphasizing in their action mechanisms and the resistance strategies developed by the cancer cells. Chapter 2 describes the design and evaluation of a repeat protein as biocatalyst for (3+2) cycloadditions, presenting an enzyme with *Huisgenase* activity. To last, Chapter 3 discloses the design, synthesis and biological validation of a new family chemotherapeutic platinum(II) agents as potential drugs to overcome the cisplatin resistance mechanisms.

The numbering of the tables, figures, schemes and references of this dissertation is independent for each chapter. In Annex I and II, a figure containing the structures and nomenclature of the natural amino acids, as well as, a summary which includes all the structures of the novel platinum(II) complexes, respectively, are provided for an easier understanding of the main text.

LIST OF PUBLICATIONS

1.Microwave-Assisted Organocatalyzed Rearrangement of Propargyl Vinyl Ethers to Salicylaldehyde Derivatives: An Experimental and Theoretical Study. Tejedor, D.; Cotos, L.; Márquez-Arce, D.; Odriozola-Gimeno, M.; Torrent-Sucarrat, M.; Cossío, F. P.; García-Tellado, F. *Chem. Eur. J.* **2015**, 21, 18280-18289.

2.Discovering Biomolecules with Huisgenase Activity: Designed Repeat Proteins as Biocatalysts for (3+2) Cycloadditions. Rivilla, I.; Odriozola-Gimeno, M.; Aires, A.; Gimeno, A.; Jiménez-Barbero, J.; Torrent-Sucarrat, M; Cortajarena, A. L.; Cossío, F.P. *under revision*.

3.Imidazopyrimidine compounds as photosensitizer for photodynamic therapy. Lima, M. L. S. O.; Braga, C.B.; Becher, T. B.; Odriozola-Gimeno, M.; Torrent-Sucarrat, M.; Rivilla, I.; Cossío, F.P.; Marsaioli, A. J.; Ornelas, C. *Submitted.*

4.The π-stacking as the key to understand the photoredox catalysis. Odriozola-Gimeno, M.; Rivilla, I.; Porcarelli, L.; Freixa, Z.; Huix-Rotllant, M.; Cossío, F.P.; Torrent-Sucarrat, M. *Manuscript in preparation*.

*5.*Design, synthesis and validation of new chemotherapeutic agents against cisplatin resistance cancers. *Patent in preparation*.

PROTEINS AND DNA AS SOURCES OF CATALYTIC AND BIOLOGICAL ACTIVITIES: FROM *HUISGENASES* TO CHEMOTHERAPY

CHAPTER 2: DISCOVERING BIOMOLECULES WITH *HUISGENASE* ACTIVITY: DESIGNED REPEAT PROTEINS AS BIOCATALYSTS FOR (3+2) CYCLOADDITIONS

CHAPTER 3: DESIGN, SYNTHESIS AND BIOLOGICAL VALIDATION OF NEW CHEMOTHERAPEUTIC PLATINUM(II) AGENTS

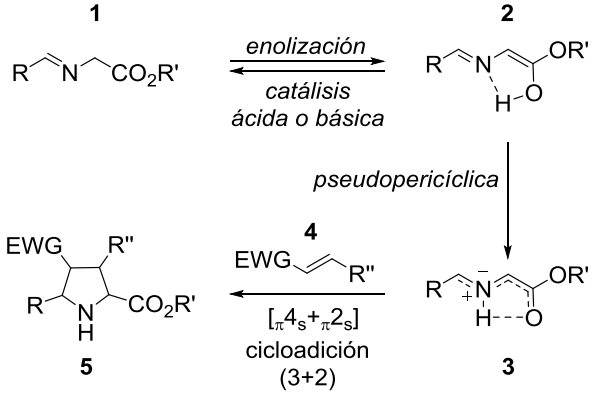
ANNEXES

Las proteínas son macromoléculas formadas por diferentes secuencias de aminoácidos, los cuales se unen mediante enlaces peptídicos. Estas unidades son capaces de realizar numerosas funciones biológicas, entre ellas se encuentra su habilidad para catalizar reacciones químicas. A este tipo de proteínas capaces de acelerar procesos químicos en sistemas biológicos se les denomina enzimas. Las enzimas son capaces de reducir las energías de activación de los procesos químicos que tienen lugar a nivel fisiológico, mediante la unión a los estados de transición y proporcionan un soporte consistente que facilita las transformaciones químicas. Existen diversas maneras de clasificar las enzimas, una de ellas consiste en dividirlas en función de la reacción que catalizan. Por ejemplo, las periciclasas son enzimas que catalizan reacciones pericíclicas.

Las reacciones pericíclicas, son aquellas que conllevan procesos de formación o ruptura de enlaces y que ocurren mediante mecanismos concertados en los que participan estados de transición cíclicos. De este modo, facilitan la formación de enlaces carbono-carbono de manera estereoselectiva y con control de la regioquímica. Pese a su abundante uso en la química sintética, son pocas las reacciones pericíclicas catalizadas por isomerasas, liasas y ligasas. Dentro del grupo de las periclicasas, las reacciones de cicloadición son las más abundantes y entre ellas, la actividad Diels-Alderasa es la más destacable. La reacción de Diels-Alder, también conocida como cicloadición (4+2), se refiere a la reacción entre un dieno conjugado y un dienófilo mediante la cual se obtienen anillos de seis unidades.

En cambio, si nos referimos a las cicloadiciones (3+2), también denominadas como reacciones de Huisgen o cicloadiciones 1,3-dipolares, con las que se forman ciclos de cinco unidades, no se ha descrito ninguna enzima que de forma inequívoca tenga esta actividad catalítica. Por ello, se propuso el nombre *Huisgenasa* para este tipo de enzimas. En el segundo capítulo de la presente Tesis Doctoral se describe el primer ejemplo en el que una proteína cataliza reacciones 1,3-dipolares entre iminas **1** y nitroalquenos π-deficientes **4**. En un primer paso de reacción, el cual puede ser llevado a cabo por la presencia de ácidos o bases, se generan los 1,3-dipolos **2**, los cuales, se transforman posteriormente a los NH-iluros de azometino **3**. Por último, estos

intermedios reaccionan con los nitroalquenos **4** para producir las prolinas **5** mediante un mecanismo concertado de tipo [$\pi 4_s + \pi 2_s$] (Esquema 1).

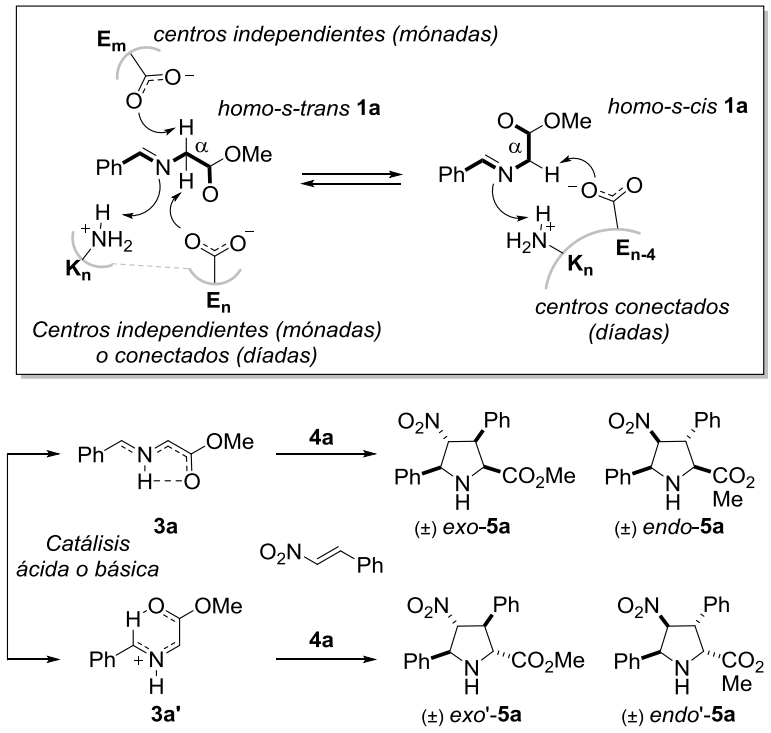


Esquema 1. Mecanismo para las cicloadiciones (3+2) entre los iluros de azometino **3** formado *in situ* y los nitroalquenos **4** mediante catálisis ácida o básica para formar las prolinas naturales **5**.

Teniendo en cuenta estas características, seleccionamos las proteínas CTPR (Consensus Tetratricopeptide Repeat) como buenas candidatas para catalizar este tipo de reacciones, debido a la presencia de varios residuos básicos y ácidos en su secuencia. La secuencia TPR consiste en 34 aminoácidos organizados en una estructura del tipo hélice-giro-hélice que suministra una gran estabilidad y que posibilita enlazar secuencias una tras otra para generar proteínas que contienen varias subunidades. La evaluación de su actividad catalítica fue realizada para dos series de estas proteínas, CTPRa y CTPR. Ambas difieren solamente en la secuencia del *loop* final del motivo TPR, en el cual la secuencia –DPRS de la serie CTPRa se transforma en –DPNN en la serie CTPR. A su vez, ambas series contienen siete residuos ácidos y tres o dos residuos básicos, en la CTPRa o CTPR, respectivamente, por subunidad.

Por ello, la capacidad catalítica de las proteínas CTPRa y CTPR fue evaluado en la reacción entre el glicinato de N-benzilideno **1a** y (*E*)-β-nitroestireno **4a** mediante agitación orbital a temperatura ambiente durante seis días y utilizando el THF como disolvente. Las proteínas originales de la serie CTPR que contienen una (CTPR1a) o tres (CTPR3a) repeticiones de la subunidad TPR, fueron capaces de catalizar la formación de los cuatro posibles isómeros de las pirrolidinas; *exo*-**5a**, *endo*-**5a**, *exo'*-**5a** y *endo'*-**5a**, de una forma proporcional y obteniendo un rendimiento global del 40% (Esquema 2). La obtención de estos cuatro isómeros no es habitual en este tipo de reacciones, siendo

pocos los casos descritos hasta la fecha. En cambio, las proteínas CTPR1 y CTPR3 solamente catalizaron la síntesis de los isómeros *exo*-**5a** y *endo*-**5a** con un rendimiento global del 20%.



Esquema 2. Mecanismo de formación de los iluros de azometino **3a** y **3a'** catalizado por las mónadas o díadas presentes en las proteínas CTPR. Reacción 1,3-dipolar entre los iluros de azometino **3a** y **3a'** parar producir las prolinas **5a** mediante el mecanismo concertado [$_{\pi}4_{s}+_{\pi}2_{s}$].

Por otro lado, se realizaron varias mutaciones en los residuos tanto ácidos como básicos de las proteínas para entender su implicación en el proceso catalítico. Todas las mutaciones realizadas en la serie CTPRa inhiben la síntesis de las prolinas *exo*'-**5a** y *endo*'-**5a**, además de reducir el rendimiento global. Por ello, propusimos que la formación de los cuatro cicloaductos que se obtenían para las secuencias originales de la CTPRa se basaban en la generación *in situ* de los iluros de azometino **3a** y **3a'**. Uno o dos residuos ácidos o básicos podían actuar como mónadas para catalizar la conversión de **1a** a los 1,3-dipolos **3a** y **3a'**. En otro mecanismo alternativo, estos intermedios estereoisoméricos podían ser formados mediante la catálisis de una díada catalítica, K_n/E_n-4, responsable de realizar los procesos de transferencia y abstracción de hidrógenos (Esquema 2). Una vez que estos intermedios reactivos eran formados, reaccionarían con los dipolarófilos **4a** para dar lugar a los productos deseados. Por un

lado, el 1,3-dipolo reaccionaría mediante un proceso concertado de tipo [$\pi 4_s + \pi 2_s$] para producir las prolinas *endo*-**5a** y *exo*-**5a**, mientras que por otro lado la reacción suprafacial entre dipolarófilo **4a** y el iluro de azometino **3a'** daría lugar a las prolinas no naturales *endo'*-**5a** y *exo'*-**5a**.

En las últimas décadas, la química computacional ha emergido como una gran herramienta para el análisis y comprensión de numerosos procesos químicos. En lo que se refiere a la evaluación catalítica de las proteínas, los métodos QM/MM han sido unos de los más empleados. Estos métodos combinan la precisión de la metodología de mecánica cuántica, QM, para el análisis de la región catalítica, con la mayor practicidad de los métodos de mecánica molecular, MM, con los que se estudia el resto de la proteína. Por ello, esta metodología fue empleada para validar nuestra hipótesis inicial.

En el primer caso, se estudió la actividad catalítica de la díada formada por los aminoácidos E22-K26, la cual permitiría la síntesis de los cuatro estereoisómeros. El estudio computacional mostró que ambas conformaciones de la imina, *homo-s-trans* and *homo-s-cis*, podían estabilizarse mediante la interacción con los residuos de la díada. Si nos referimos a la vía catalítica para la conformación *homo-s-trans*, en el primer paso, el residuo E22 abstrae uno de los dos hidrógenos del átomo C_α del ligando para posteriormente dar lugar a la formación del iluro de azometino **3a**. Hay que recordar que este intermedio es el precursor de los productos *endo*-**5a** y *exo*-**5a**. Por otro lado, el iluro de azometino **3a'**, precursor de los productos *endo'*-**5a** y *exo'*-**5a**, se obtendría mediante una ruta análoga para la conformación *homo-s-cis*. Este proceso se da mediante dos pasos, primero la imina **1a** es protonada, para proceder posteriormente a la abstracción de un hidrógeno de la posición C_α mediada por el residuo E22. De este modo, la díada catalítica podría catalizar la formación de ambos iluros de azometino **3a** y **3a'**, seguido de la reacción de cicloadición (3+2) que tiene como resultado la formación de los cuatro isómeros anteriormente descritos.

Otro mecanismo sintético posible es el que tiene lugar mediante la acción de residuos básicos o ácidos de forma individual, denominados como mónadas. Para el análisis de esta ruta sintética se eligieron los aminoácidos E2 y K13 como ejemplos de centros ácidos y básicos, respectivamente. Al contrario de lo observado en la catálisis

ACRONYMS AND ABBREVIATONS

A	Adenine
AcOEt	Ethyl acetate
AFM	Atomic Force Microscopy
ATP	Adenosine Triphosphate
BCL	B-cell Lymphoma
BRCA	Breast Cancer
CD	Circular Dichroism
CSP	Chemical Shift Perturbation
CTPR	Consensus Tetratricopeptide Repeat
CTR	Copper Transporter
COSY	Correlation Spectroscopy
d	Doublet (NMR)
DCM	Dichloromethane
deg	degrees
DIPEA	N,N-diisopropylethylamine
DFT	Density Functional Theory
DMATS	Dimethylallyl Triptophan Synthase
DMF	N,N-Dimethylformamide
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DPBS	Dulbecco's Phosphate-Buffered Saline
EC	Enzyme Commission
EDCI	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
eq.	Equivalent(s)
ERCC	Excision-Repair Cros-Complementing
ESI	Electrospray Ionization
EtOH	Ethanol
Et_3N	Triethylamine
EWG	Electron Withdrawing Group

FBS	Fetal Bovine Serum
FDA	Food and Drug Administration
FDC	Ferulic acid Decarboxylase
FITC	Fluorescein Isothiocyanate
FMN	Flavin Mononucleotide
FTIR	Fourier Transform Infrared Spectroscopy
G	Guanine
GSH	Glutathione
h	Hours
HER	Human Epidermal Growth Factor Receptor
Hex	Hexane
HOBt	1-hydroxybenzotriazol
HRMS	High Resolution Mass Spectra
HSQC	Heteronuclear Single Quantum Coherence
Hz	Hertz
IMOMM	Integrated Molecular Orbital + Molecular Mechanics
IPTG	Isopropyl β-d-Thiogalactoside
IR	Infrared Radiation
IUPAC	International Union of Pure and Applied Chemistry
J	Coupling constant
LB	Luria-Bertani
LUMO	Lowest Unoccupied Molecular Orbital
m	Multiplet (NMR)
MALDI-TOF	Matrix Assisted Laser Desorption Ionization–Time Of Flight
MAPK	Mitogen-Activated Protein Kinase
MS	Mass Spectrometry
MeOH	Methanol
min	Minute
mL	Mililitre
MM	Molecular Mechanics
MMR	DNA Mismatch Repair

MOPS	3-(N-morpholino)propanesulfonic acid
m_p	Melting point
MRE	Molar Residue Ellipticity
mW	Microwave
MW	Molecular Weight
MRP	Multidrug Resistance Protein
NAD	Nicotinamide Adenine Dinucleotide
NADH	Nicotinamide Adenine Dinucleotide Hydrogen
NADP	Nicotinamide Adenine Dinucleotide Phosphate
NADPH	Nicotinamide Adenine Dinucleotide Phosphate Hydrogen
NMR	Nuclear Magnetic Resonance
Nu	Nucleophile
OCT	Organic Cation Transporter
OD	Optical Density
ONIOM	Own N-layer Integrated Molecular Orbital Molecular Mechanics
P	Product
PBS	Phosphate-Buffered Saline
PES	Potential Energy Surfaces
PMS	Phenazine Methosulfate
PDB	Protein Data Bank
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
R	Reactant
RNA	Ribonucleic Acid
RPMI	Roswell Park Memorial Institute
r.t.	Room temperature
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEM	Standard Error of the Mean
SHx	Solvating helix
s	Singlet (NMR)

STD	Saturation Transfer Difference
t	Triplet (NMR)
TBTU	2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium tetrafluoroborate
TEA	Triethylamine
TEM	Transmission Electron Microscopy
THF	Tetrahydrofuran
TLC	Thin Layer Chromatography
TMS	Trimethylsilyl
TOF	Turnover frequency
TON	Turnover number
TPR	Tetratricopeptide Repeat
TROSY	Transverse Relaxation Optimized Spectroscopy
TS	Transition State
UDCA	Ursodeoxycholic Acid
UV	Ultraviolet
WHO	World Health Organization
WST-1	Water Souble Tetraziolium-1
wt	Wild type

Chapter 1:

Introduction

Abstract: In this first Chapter, some concepts that will result essential to understand the results displayed in Chapters 2 and 3 are introduced. Firstly, an approach to the catalytic proteins is presented (section 1.1), focusing mainly on the proteins that catalyze pericyclic reactions. Among them, a compilation of the available Diels-Alderase and *Huisgenase* proteins is offered. In the next section, a brief summary of the computational methods is reported, emphasizing on their application to analyze and predict the experimental data (section 1.2). More in detail, we have deepened into the QM/MM methods, their usefulness, drawbacks and their basis as a good strategy to describe catalytic reactions performed by proteins. In the last section, the use of platinum(II) derivatives as antitumoral drugs is presented (section 1.3). Additionally, we focus on the platinum resistance mechanisms developed by the carcinogenic cells, as well as, in the reaction mechanisms of cisplatin and nitrogen mustards.

1.1 CATALYTIC PROTEINS IN PERICYCLIC REACTIONS

Proteins are sequences of amino acids, which linked together form macromolecular complex structures that can perform numerous biological functions. The variation on the number of amino acids (20 natural residues are available) and how those are organized modify their functions; *i.e*, structural, transport, signalling, storage or enzymatic.[1] One of the first biological roles known for the proteins was the ability to perform chemical reactions, acting analogously to the chemical non-natural catalysts. These biological catalysts are named enzymes.[2] The amino acid sequence and their length, play an essential role in their activity, mainly because of their influence in the specific 3D rearrangement of the functional groups, known as protein folding.[3] Therefore, a vast number of protein foldings are reachable by altering the organization of the amino acids, affecting the structure as well as the dynamics necessary for the binding of the substrates, and consequently, the catalytic processes.[4]

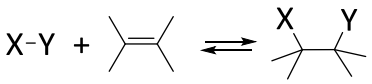
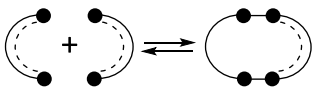
The catalytic activity of these proteins relies on lowering the activation energy of the reaction. Usually, this increase in the reaction rate is promoted by binding the transition state structure, which directly augments its concentration and enhances the reaction rate. Additionally, the catalytic proteins are able to convert a big number of substrate molecules per unit time.[5] These characteristics have aroused the interest of the field of green biotechnology in enzyme-catalyzed chemical reactions.

In this way, the understanding of the mechanisms involved in the catalysis and the study of their evolution[6] undergone by them have resulted to be crucial.[7] Moreover, once the involved processes were properly analyzed, the next step was the redesign of

1. Whitford, D., *Proteins Structure and Function*. Wiley & Sons: Chichester, England, 2005.
2. Copeland, R. A., *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis, Second Edition*. Wiley & Sons: New York, 2000.
3. Makhlynets, O. V.; Korendovych, I. V., *Nat. Chem.* **2016**, *8*, 823-824.
4. Rufo, C. M.; Moroz, Y. S.; Moroz, O. V.; Stöhr, J.; Smith, T. A.; Hu, X.; DeGrado, W. F.; Korendovych, I. V., *Nat. Chem.* **2014**, *6*, 303-309.
5. Marangoni, A. G., *Enzyme Kinetics: A Modern Approach*. Wiley&Sons: Hoboken, New Jersey, 2003.
6. Pál, C.; Papp, B.; Lercher, M. J., *Nat. Rev. Genet.* **2006**, *7*, 337-348.
7. Palmer, T.; Bonner, P. L., *Enzymes: Biochemistry, Biotechnology, Clinical Chemistry*. Woodhead Publishing: Cambridge, UK, 2001.

existing proteins,[8] aiming to find concrete or new reactivity.[9-11] These efforts have been very fruitful and the modification of the active catalytic sites or the binding site of the cofactors have resulted in incredible catalytic advances.[12] Among all the employed tools, computational developments have been essential for the evolution of the mentioned enzymes.[13,14] Depending on the chemical reaction they catalyze, enzymes are classified into six groups according to the Enzyme Commission (EC) of the IUPAC (Table 1.1).

Table 1.1. IUPAC EC classification of enzymes according to the reaction they catalyze.

EC Number	Enzyme Type	Catalyzed Reactions	
1	***Oxidoreductases***	Oxidation-reduction	$A^+ + B \longrightarrow A + B^+$ $A^- + B \longrightarrow A + B^-$
2	***Transferases***	Chemical group transfer	$A\text{-}X + B \longrightarrow A + B\text{-}X$ ($X = CH_3$, CH_3CO, NH_2)
3	***Hydrolases***	Hydrolytic bond cleavage	$A\text{-}B + H_2O \longrightarrow A\text{-}OH + H\text{-}B$
4	***Lyases***	Non-hydrolytic bond cleavage	
5	***Isomerases***	Isomerization	$A\text{-}B\text{-}C \longrightarrow A\text{-}C\text{-}B$
6	***Ligases***	Bond-forming reactions *e.g.* C-C bonds (6.4)	$A + B \rightleftharpoons A\text{-}B$ 

8. Moroz, Y. S.; Dunston, T. T.; Makhlynets, O. V.; Moroz, O. V.; Wu, Y.; Yoon, J. H.; Olsen, A. B.; McLaughlin, J. M.; Mack, K. L.; Gosavi, P. M.; Van Nuland, N. A. J.; Korendovych, I. V., *J. Am. Chem. Soc.* **2015**, *137*, 14905-14911.
9. Penning, T. M.; Jez, J. M., *Chem. Rev.* **2001**, *101*, 3027-3046.
10. Huang, P.-S.; Boyken, S. E.; Baker, D., *Nature* **2016**, *537*, 320-327.
11. Hilvert, D., *Annu. Rev. Biochem.* **2013**, *82*, 447-470.
12. Toscano, M. D.; Woycechowsky, K. J.; Hilvert, D., *Angew. Chem., Int. Ed.* **2007**, *46*, 3212-3236.
13. Kries, H.; Blomberg, R.; Hilvert, D., *Curr. Opin. Chem. Biol.* **2013**, *17*, 221-228.
14. Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N., *Angew. Chem., Int. Ed.* **2013**, *52*, 5700-5725.

Within the six categories, the enzymes are also divided into subclasses to which a second EC number is assigned. For instance, ligases that connect C-C bonds belong to the EC 6.4 group. Additionally, third and fourth EC numbers are assigned to each individual protein, providing a specific code for each independent enzyme. If we take into account the huge variety of functions performed by them, this classification is very helpful to provide a better understanding of their role.

1.1.1 Pericyclases

Pericyclases are the family of enzymes that catalyze pericyclic reactions. These reactions are concerted bond-forming and bond-breaking reactions that proceed through cyclic transition states without the need of reaction intermediates. This kind of reactions is one of the most powerful synthetic tools for the formation of regioselective and stereoselective formation of carbon-carbon bonds in organic chemistry. Among the pericyclic reactions, four main families have been described: sigmatropic rearrangements, electrocyclic reactions, cycloadditions, and group transfer reactions.[15,16] Despite their broad application in organic chemistry, few natural enzymes with this reactivity have been characterized.[17] Although pericyclic reactions catalyzed by isomerases, lyases and ligases are quite scarce,[18,19] some interesting pericyclic reactivities have been reported to date.[20]

Within the four classes of pericyclic reactions mentioned above, several enzymes that catalyze sigmatropic rearrangements have been described. For instance, the chorismate mutase catalyzes the suprafacial [3,3]-sigmatropic shift, Claisen rearrangement (Figure 1.1A), of chorismate to prephenate.[21,22] Other sigmatropic

15. Fleming, I., *Pericyclic Reactions*. Oxford University Press: Oxford, 1999.
16. Gill, G. B.; Willis, M. R., *Pericyclic Reactions*. Chapman & Hall Ltd.: London, England, 1974.
17. Jamieson, C. S.; Ohashi, M.; Liu, F.; Tang, Y.; Houk, K. N., *Nat. Prod. Rep.* **2019**, *36*, 698-713.
18. Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J., *Interdiscip. Sci.* **2010**, *2*, 115-131.
19. Ohashi, M.; Liu, F.; Hai, Y.; Chen, M.; Tang, M.-C.; Yang, Z.; Sato, M.; Watanabe, K.; Houk, K. N.; Tang, Y., *Nature* **2017**, *549*, 502-506.
20. Walsh, C. T.; Tang, Y., *Biochemistry* **2018**, *57*, 3087-3104.
21. Guo, H.; Cui, Q.; Lipscomb, W. N.; Karplus, M., *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9032-9037.
22. Andrews, P. R.; Smith, G. D.; Young, I. G., *Biochemistry* **1973**, *12*, 3492-3498.

rearrangements performed by pericyclases are the Cope rearrangements (Figure 1.1B). The suprafacial [3,3]-sigmatropic reaction has been found to be an intermediate step in the synthesis hapalindone/ambiguine,[23] which has also been confirmed to be necessary for the posterior synthesis of fischerindoles and welwitindoles.[24,25] Alternatively, a Cope rearrangement was suggested for the formation of 4-dimethylallyl tryptophan catalyzed by the 4-DMAT synthase.[26]

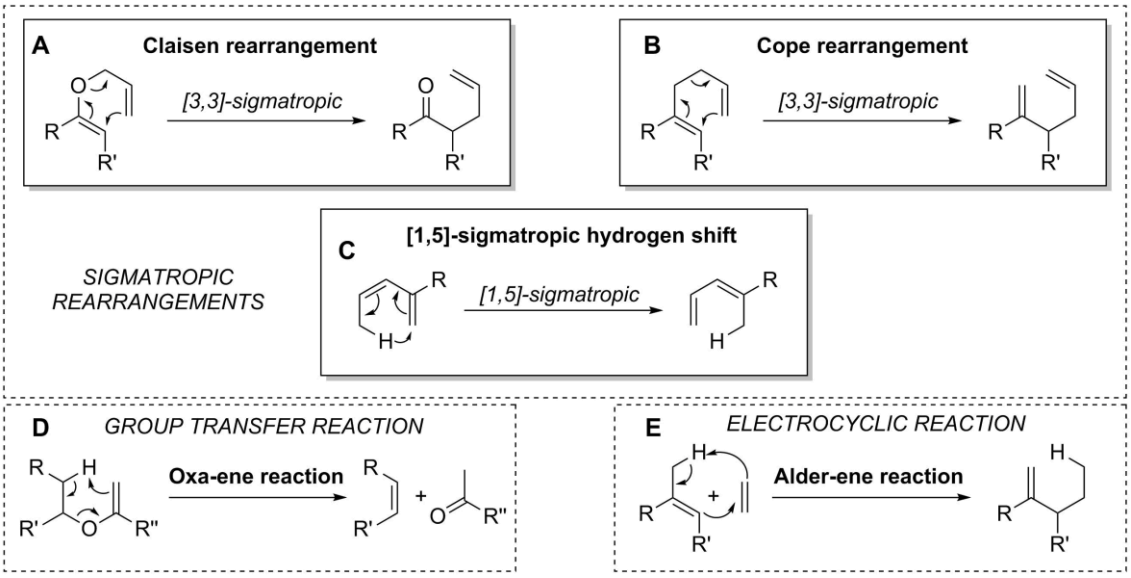


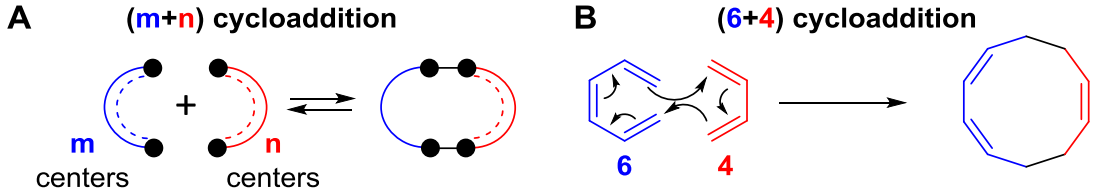
Figure 1.1. Five pericyclic reaction types catalyzed by enzymes. Sigmatropic rearrangements: **A**) Claisen rearrangement; **B**) Cope rearrangement and **C**) [1,5]-sigmatropic hydrogen shift. **D**) Oxa-ene group transfer reaction. **E**) Alder-ene electrocyclic reaction.

Similarly, a [1,5]-sigmatropic shift was described for the transformation of precorrin-8x into hydrogenobyrinic acid as an intermediate step in the biosynthesis of vitamin B_{12} catalyzed by precorrin-8x mutase (Figure 1.1C).[27,28] Within pericyclic lyases, the isochorismate pyruvate lyase has been reported as a catalyst for the conversion of isochorismate into salicylate and pyruvate through an oxa-ene group transfer reaction

23. Li, S.; Lowell, A. N.; Yu, F.; Raveh, A.; Newmister, S. A.; Bair, N.; Schaub, J. M.; Williams, R. M.; Sherman, D. H., *J. Am. Chem. Soc.* **2015**, *137*, 15366-15369.
24. Zhu, Q.; Liu, X., *Chem. Commun.* **2017**, *53*, 2826-2829.
25. Zhu, Q.; Liu, X., *Angew. Chem., Int. Ed.* **2017**, *56*, 9062-9066.
26. Luk, L. Y. P.; Qian, Q.; Tanner, M. E., *J. Am. Chem. Soc.* **2011**, *133*, 12342-12345.
27. Li, Y.; Alanine, A. I. D.; Vishwakarma, R. A.; Balachandran, S.; Leeper, F. J.; Battersby, A. R., *J. Chem. Soc., Chem. Commun.* **1994**, 2507-2508.
28. Shipman, L. W.; Li, D.; Roessner, C. A.; Scott, A. I.; Sacchettini, J. C., *Structure* **2001**, *9*, 587-596.

(Figure 1.1D).[29,30] Interestingly, an electrocyclic Alder-ene reaction was proposed in an intermediate step during the catalysis of crotonyl-CoA carboxylase/reductase (Figure 1.1E).[31]

The remaining subgroup of pericyclic reactions are the cycloaddition reactions. In the cycloaddition reactions, two-components containing π-systems react together to form two new σ-bonds. In this way, a new ring is formed through a concerted process including a reduction in the length of the conjugated system of the orbitals in each component.[32] Among the classes of pericyclic reactions, cycloadditions are the most abundant and useful ones, due to their ability to form many bonds in one step and to generate several stereogenic centers. These reactions are named as (m+n) cycloadditions, where m and n are the number of atoms of each component involved in the reactions; *e.g*, (4+2) cycloadditions (Diels-Alder reactions),[33] (3+2) cycloadditions (1,3-dipolar cycloadditions) and (6+4) cycloadditions (Scheme 1.1). The number of π-electrons and the suprafacial or antarafacial interaction is specified between square brackets. For instance, both concerted (4+2) and (3+2) cycloadditions occur through $[\pi 4_s + \pi 2_s]$ orbital topologies.



Scheme 1.1. A) General scheme for the cycloaddition reactions. **B**) Reaction example for the (6+4) cycloaddition, in which six and four atoms of the blue and red reactants, respectively, are involved.

29. DeClue, M. S.; Baldridge, K. K.; Künzler, D. E.; Kast, P.; Hilvert, D., *J. Am. Chem. Soc.* **2005**, *127*, 15002-15003.
30. Lamb, A. L., *Biochemistry* **2011**, *50*, 7476-7483.
31. Rosenthal, R. G.; Ebert, M.-O.; Kiefer, P.; Peter, D. M.; Vorholt, J. A.; Erb, T. J., *Nat. Chem. Biol.* **2013**, *10*, 50.
32. Kobayashi, S.; Jorgensen, K. A., *Cycloaddition Reactions in Organic Synthesis*. Wiley-VCH: Weinheim, Germany, 2001.
33. Diels, O.; Alder, K., *Liebigs Ann.* **1928**, *460*, 98-122.

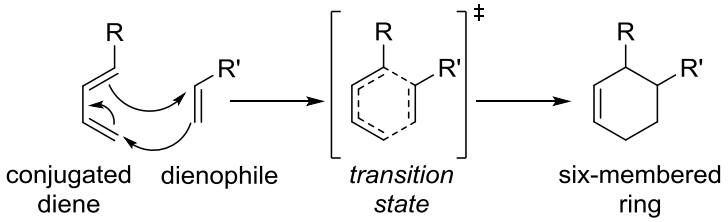
Concerning the enzyme catalyzed cycloaddition reactions the Diels-Alderase activity is the most abundant one.[34] Despite their broad application in chemistry no enzyme-assisted direct (3+2) cycloaddition, also known as Huisgen reaction,[35] has been reported so far.

By these means, in the next sections the biological most abundant cycloaddition, the Diels-Alder reaction, and the biologically unprecedented direct (3+2) cycloadditions are presented. Similarly to Diels-Alderases, our research group has suggested the name *Huisgenases* for this second family of enzymes.

1.1.2 Diels-Alderases

The (4+2) cycloaddition reaction between 1,3-dienes and alkenes was first described in 1928 by Otto Diels and Kurt Alder.[33] The conjugated 1,3-diene is the 4π component and the dienophile the 2π component. In this way, two new single bonds are formed in a concerted mechanism through a cyclic Transition State (TS) to yield cyclohexene rings (Scheme 1.2). This cyclization process has become one of the most common strategies for the regio- and stereocontrolled synthesis of six-membered rings.[36] Interestingly, the Diels-Alder reaction enables to generate up to four stereogenic centers in the same step. Moreover, carbon-carbon, carbon-heteroatom and heteroatom-heteroatom bonds can be formed with this methodology. By these means, it has become a versatile approach for the design of simple and complex natural compounds.[37] Given its synthetic relevance the discoverers of this reaction were awarded the Nobel Prize in Chemistry in 1950.

34. Klas, K.; Tsukamoto, S.; Sherman, D. H.; Williams, R. M., *J. Org. Chem.* **2015**, *80*, 11672-11685.
35. Huisgen, R., *Angew. Chem., Int. Ed.* **1963**, *2*, 565-598.
36. Fringuelli, F.; Taticchi, A., *The Diels Alder Reaction. Selected Practical Methods.* John Wiley & Sons: Chichester, England, 2002.
37. Nicolaou, K. C.; Snyder, S. A.; Montagnon, T.; Vassilikogiannakis, G., *Angew. Chem., Int. Ed.* **2002**, *41*, 1668-1698.



Scheme 1.2. General reaction scheme for the Diels-Alder reaction between a diene and a dienophile through a cyclic TS to yield six-membered rings.

Since its development in the 1920s, numerous works have been reported concerning synthetic, mechanistic and theoretical aspects.[38] Regarding the importance in the field of chemistry, the search for natural Diels-Alder activity aroused in the last decades.[39] In the pursuit of natural entities that catalyze Diels-Alder reactions, firstly, an antibody which performed a bimolecular (4+2) cycloaddition was characterized.[40,41] The efforts done in this research also resulted in the description of Diels Alder reactions that were catalyzed by RNA.[42]

Taking into account the efficient design of cyclohexene rings through this kind of reaction in organic chemistry, and along with the presence of numerous natural products that contain cyclohexene rings with defined stereochemical configuration, the existence of enzymes that catalyze Diels-Alder reactions was suggested.[43-45] The study of the biosynthetic process of these natural products finally resulted in the characterization of the first Diel-Alderase activity.[46] Ichichara and coworkers described

38. Franzen, J., *Kirk-Othmer Encyclopedia of Chemical Technology*. John Wiley & Sons: New Jersey, 2016.
39. Kim, H. J.; Ruszczycky, M. W.; Liu, H.-w., *Curr. Opin. Chem. Biol.* **2012**, *16*, 124-131.
40. Hilvert, D.; Hill, K. W.; Nared, K. D.; Auditor, M. T. M., *J. Am. Chem. Soc.* **1989**, *111*, 9261-9262.
41. Gouverneur, V.; Houk, K.; de Pascual-Teresa, B.; Beno, B.; Janda, K.; Lerner, R., *Science* **1993**, *262*, 204-208.
42. Tarasow, T. M.; Tarasow, S. L.; Eaton, B. E., *Nature* **1997**, *389*, 54-57.
43. Oikawa, H.; Tokiwano, T., *Nat. Prod. Rep.* **2004**, *21*, 321-352.
44. Stocking, E. M.; Williams, R. M., *Angew. Chem., Int. Ed.* **2003**, *42*, 3078-3115.
45. Kelly, W., *Org. Biomol. Chem.* **2009**, *6*, 4483-93.
46. Oikawa, H.; Katayama, K.; Suzuki, Y.; Ichihara, A., *J. Chem. Soc., Chem. Commun.* **1995**, 1321-1322.

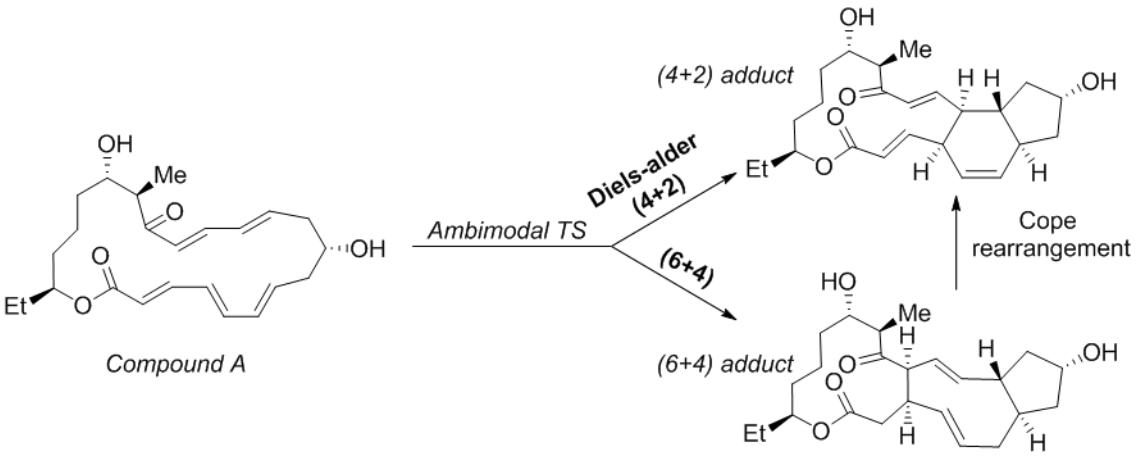
in 1995 the conversion of prosolanapyrone III into the exo adduct of solanopyrone A, although the Diels-Alderase, Solanapyrone Synthase, was described in later studies.[47]

It must be taken into consideration that (4+2) cyclases can undergo various catalytic mechanisms, for which the Diels-Alder reactivity is one possibility among others. Despite the instability of the substrates and the complexity of the involved multifunctional enzymes, numerous proteins that can potentially catalyze (4+2) cycloaddition reactions have been identified in the last years.[48,49] Nevertheless, in most of the cases in which the Diels-Alder activity have been proposed, the (4+2) cycloaddition reaction consisted of an intermediate step of the biosynthesis of secondary metabolites.[50,51] While the Diels-Alder reactions usually employed by synthetic chemists perform mainly intermolecular cycloadditions, the studied natural reactions occur through intramolecular processes late in the biosynthetic pathways. Interestingly, the effect of these enzymes on the reaction mechanism is performed in different manners, *e.g.*, providing a catalytic site that make the transition state energetically accessible or activating (LUMO lowering) the dienophile by hydrogen bonding.

More recently, the SpnF enzyme has been described as the first naturally selected enzyme that performs an unambiguous intramolecular Diels-Alderase activity.[52,53] The SpnF of *Escherichia coli* is an enzyme that participates in the biosynthesis of the natural compound Spynosin A. This enzyme catalyzes a transannular (4+2) cycloaddition with a rate acceleration of approximately 500-folds. Houk and coworkers have presented a possible ambimodal (6+4)/(4+2) reactivity for this synthetic

47. Kasahara , K.; Miyamoto, T.; Fujimoto, T.; Oguri, H.; Tokiwano , T.; Oikawa, H.; Ebizuka, Y.; Fujii, I., *ChemBioChem* **2010**, *11*, 1245-1252.
48. Zheng, Q.; Tian, Z.; Liu, W., *Curr. Opin. Chem. Biol.* **2016**, *31*, 95-102.
49. Ose, T.; Watanabe, K.; Mie, T.; Honma, M.; Watanabe, H.; Yao, M.; Oikawa, H.; Tanaka, I., *Nature* **2003**, *422*, 185-189.
50. Auclair, K.; Sutherland, A.; Kennedy, J.; Witter, D. J.; Van den Heever, J. P.; Hutchinson, C. R.; Vederas, J. C., *J. Am. Chem. Soc.* **2000**, *122*, 11519-11520.
51. Minami, A.; Oikawa, H., *J. Antibiot.* **2016**, *69*, 500-506.
52. Kim, H. J.; Ruszczycky, M. W.; Choi, S.-H.; Liu, Y.-N.; Liu, H.-W., *Nature* **2011**, *473*, 109-112.
53. Fage, C. D.; Isiorho, E. A.; Liu, Y.; Wagner, D. T.; Liu, H.-w.; Keatinge-Clay, A. T., *Nat. Chem. Biol.* **2015**, *11*, 256-258.

pathway of Spynosin A.[54,55] Interestingly, these (6+4) intermediates could be converted into (4+2) adducts through a Cope rearrangement. By these means, these authors considered the SpnF enzyme as the first example of a monofunctional Diels-Alder/(6+4)-ase (Scheme 1.3). More recently, a group of (6+4)/(4+2) pericyclases that catalyze the synthesis of streptoseomycin-type natural products through ambimodal transition states has also been characterized.[56,57]



Scheme 1.3. Diels-alder/(6+4)-ase activity of the SpnF enzyme for the transformation of Compound A into the (4+2) and (6+4) adducts through an ambimodal TS. Cope rearrangement that connects the (6+4) adduct with (4+2) adduct.[54,55]

The importance of the Diels-Alder reaction and the presence of few natural Diels-Alderases have triggered the design of rationally designed enzymes with this activity.[58] For instance, the absence of bimolecular natural Diels-Alder reactions has promoted the *de novo* computational design and experimental characterization of a novel Diels Alderase.[59]

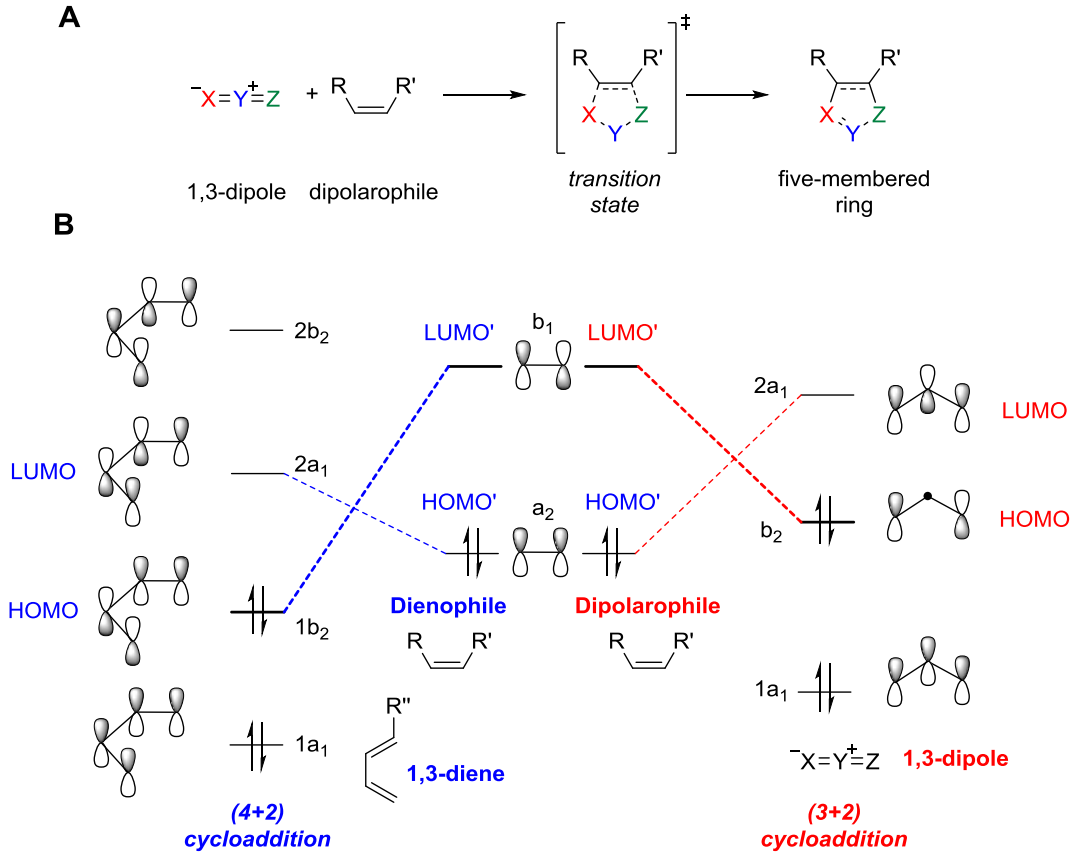
54. Patel, A.; Chen, Z.; Yang, Z.; Gutiérrez, O.; Liu, H.-w.; Houk, K. N.; Singleton, D. A., *J. Am. Chem. Soc.* **2016**, *138*, 3631-3634.
55. Yang, Z.; Yang, S.; Yu, P.; Li, Y.; Doubleday, C.; Park, J.; Patel, A.; Jeon, B.-S.; Russell, W. K.; Liu, H.-W.; Russell, D. H.; Houk, K. N., *Proc. Natl. Acad. Sci. U.S.A* **2018**, *115*, 848-E855.
56. Yu, P.; Patel, A.; Houk, K. N., *J. Am. Chem. Soc.* **2015**, *137*, 13518-13523.
57. Zhang, B.; Wang, K. B.; Wang, W.; Wang, X.; Liu, F.; Zhu, J.; Shi, J.; Li, L. Y.; Han, H.; Xu, K.; Qiao, H. Y.; Zhang, X.; Jiao, R. H.; Houk, K. N.; Liang, Y.; Tan, R. X.; Ge, H. M., *Nature* **2019**, *568*, 122-126.
58. Lichman, B. R.; O'Connor, S. E.; Kries, H., *Chem. Eur. J.* **2019**, *25*, 6864-6877.
59. Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St.Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D., *Science* **2010**, *329*, 309-313.

1.1.3 *Huisgenases*

In a paper recently sent to publication by our research group, the *Huisgenase*[60] term is proposed for proteins that catalyze reactions between 1,3-dipoles and π-deficient dipolarophiles. This kind of thermal pericyclic reactions was firstly described by Rolf Huisgen in 1963.[35] The Huisgen reaction is considered to be the analogous reaction of the Diels-Alder reaction for the formation of five-membered rings. Therefore, in the 1,3-dipolar cycloadditions, 2π electrons of the dipolarophile react with 4π electrons of the 1,3-dipole via a concerted mechanism to yield the five-membered rings (Scheme 1.4A).[61]

Additionally, this strategy has aroused as a good methodology for the synthesis of adducts containing new stereogenic centers. The X and Z atoms of the 1,3-dipole can be C, N or O, while the central atom Y is a nitrogen atom. Thus, different kind of bonds can be formed depending on the employed reactants; *e.g*, C-C, C-N or C-O. These reactions are also known as (3+2) cycloadditions and are electronically equivalent to Diels-Alder reactions. Generally, in normal Diels-Alder reactions, the main orbital interaction occurs between the HOMO of the diene and the LUMO' of the dienophile, whereas, in the 1,3-dipolar cycloadditions the interaction takes place between the HOMO of the 1,3-dipole and the LUMO' of the dipolarophile (Scheme 1.4B).[62,63]

60. Rivilla, I.; Odriozola-Gimeno, M.; Aires, A.; Gimeno, A.; Jiménez-Barbero, J.; Torrent-Sucarrat, M.; Cortajarena, A. L.; Cossío, F. P., *under revision.*
61. Huisgen, R., 1,3-Dipolar Cycloaddition Chemistry. Wiley, E. d., Ed. New York, 1984; pp 1-176.
62. Ess, D. H.; Houk, K. N., *J. Am. Chem. Soc.* **2008**, *130*, 10187-10198.
63. Gothelf, K. V.; Jørgensen, K. A., *Chem. Rev.* **1998**, *98*, 863-910.



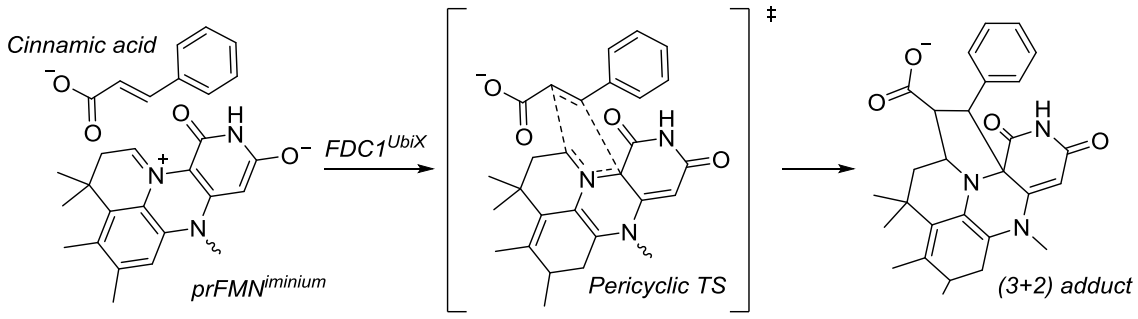
Scheme 1.4. A) General scheme for the 1,3-dipolar cycloaddition reaction between a 1,3-dipole and a dipolarophile through a pericyclic transition state to form five-membered rings. **B**) Molecular orbital diagrams for the (4+2) and (3+2) cycloadditions.

Since its discovery in the 1960s, the 1,3-dipolar reactions have been broadly employed by synthetic chemists.[64,65] Additionally, the structures of several natural products suggested the presence of (3+2) cycloadditions in the respective biosynthetic processes. According to this hypothesis, various natural products have been synthesized through this cycloaddition reaction.[66 - 68] Moreover, the presence of this kind of structures in natural products and the broad application of the Huisgen reaction have

64. Sankararaman, S., *Pericyclic Reactions-a Textbook: Reactions, Applications and Theory*. Wiley-VCH: Weinheim, Germany, 2005.
65. Padwa, A.; Pearson, W. H., *Synthetic Applications of 1,3-Dipolar Cycloaddition Chemistry Toward Heterocycles and Natural Products*. Wiley & Sons: Hoboken, New Jersey, 2002.
66. Painter, P. P.; Pemberton, R. P.; Wong, B. M.; Ho, K. C.; Tantillo, D. J., *J. Org. Chem.* **2014**, *79*, 432-435.
67. Krenske, E. H.; Patel, A.; Houk, K. N., *J. Am. Chem. Soc.* **2013**, *135*, 17638-17642.
68. Strych, S.; Journot, G.; Pemberton, R. P.; Wang, S. C.; Tantillo, D. J.; Trauner, D., *Angew. Chem., Int. Ed.* **2015**, *54*, 5079-5083.

motivated the search for natural entities that catalyze 1,3-dipolar cycloadditions. For instance, a (3+2) cycloaddition catalyzed by covalently modified DNA with a bimetallic complex has been recently reported by our group.[69]

In contrast to Diels-Alder reactions, unambiguous *Huisgenases* that generate carbon-carbon bonds have not been yet identified in living systems.[70] Interestingly, 1,3-dipolar reactions that form C-O bonds in intermediate steps of the biosynthetic processes have been described.[67,71] In 2015, Leys and coworkers, presented the first example of a retro (3+2) intermolecular cycloaddition reaction as an intermediate step in the enzymatic decarboxylation of α-β-unsaturated acids, catalyzed by the *Aspergillus niger* FDC^{UBIx} protein assisted by the prFMN cofactor (Scheme 1.5).[72] Additionally, in 2017, Marsh and coworkers characterized the similar reaction between the mentioned prFMN cofactor and 2-fluoro-2-nitrovinylbenzene catalyzed by the FDC enzyme.[73] More recently, a similar reactivity has been described in an intermediate step of the UbiD-related ferulic acid decarboxylase.[74]



Scheme 1.5. 1,3-dipolar cycloaddition reaction performed by the $FDC1^{UbiX}$ enzyme between the Cinnamic acid and the iminium species of the prFMN cofactor to form the final (3+2) adduct.[72]

69. Rivilla, I.; de Cózar, A.; Schäfer, T.; Hernandez, F. J.; Bittner, A. M.; Eleta-Lopez, A.; Aboudzadeh, A.; Santos, J. I.; Miranda, J. I.; Cossío, F. P., *Chem. Sci.* **2017**, *8*, 7038-7046.
70. Baunach, M.; Hertweck, C., *Angew. Chem., Int. Ed.* **2015**, *54*, 12550-12552.
71. Borowski, T.; de Marothy, S.; Broclawik, E.; Schofield, C. J.; Siegbahn, P. E. M., *Biochemistry* **2007**, *46*, 3682-3691.
72. Payne, K. A. P.; White, M. D.; Fisher, K.; Khara, B.; Bailey, S. S.; Parker, D.; Rattray, N. J. W.; Trivedi, D. K.; Goodacre, R.; Beveridge, R.; Barran, P.; Rigby, S. E. J.; Scrutton, N. S.; Hay, S.; Leys, D., *Nature* **2015**, *522*, 497.
73. Ferguson, K. L.; Eschweiler, J. D.; Ruotolo, B. T.; Marsh, E. N. G., *J. Am. Chem. Soc.* **2017**, *139*, 10972-10975.
74. Bailey, S. S.; Payne, K. A. P.; Saaret, A.; Marshall, S. A.; Gostimskaya, I.; Kosov, I.; Fisher, K.; Hay, S.; Leys, D., *Nat. Chem.* **2019**, *11*, 1049-1057.

1.2 COMPUTATIONAL CHEMISTRY: A USEFUL TOOL FOR THE ANALYSIS OF CHEMICAL PROBLEMS

Since 1960s,[75-77] the theoretical chemistry has undergone a dynamic boom, mainly due to the advances in the field of informatics, the development of new chemical tools, as well as, the optimization of the accuracy and efficiency of numerical calculations and simulations.[78-80] Therefore, all these improvements have broadened the spectrum of the applications of the computational methods and they have become great tools both for academic and industrial research.[81-83]

The set of techniques that compose the computational chemistry allows to solve a huge variety of chemical problems.[84] Among them, the employment of these methods prior to the synthetic process as predictive tools, and their use as a strategy to understand the chemical issues in a more accurate way, are the most commons applications.[85,86] By these means, theoretical chemistry is applied in the investigation of several chemical concepts:

-The study of the Potential Energy Surfaces (PES) allows to carry out optimizations of the chemical structures which permits to find their lowest energy

75. Cramer, C., *Essentials of Computational Chemistry: Theories and Models, 2nd Edition*. Chichester, England, 2004.
76. Bachrach, S. M., *Computational Organic Chemistry, 2nd Edition*. Hoboken, New Jersey, 2014.
77. Jensen, F., *Introduction to Computational Chemistry, 2nd Edition*. Chichester, England, 2007.
78. Lewars, E. G., *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics, 2nd Edition*. Springer: New York, 2011.
79. Thiel, W., *Angew. Chem., Int. Ed.* **2011**, *50*, 9216-9217.
80. Dykstra, C. E.; Frenking, G.; Kim, K.; Scuseria, G. E., *Theory and Applications of Computational Chemistry*. Eds. Elsevier: Oxford, UK, 2005.
81. Lipkowitz, K. B., *Chem. Rev.* **1998**, *98*, 1829-1874.
82. Richon, A. B., *Drug Discov. Today* **2008**, *13*, 659-664.
83. Houk, K. N.; Liu, F., *Acc. Chem. Res.* **2017**, *50*, 539-543.
84. Arrieta, A.; de la Torre, M. C.; de Cózar, A.; Sierra, M. A.; Cossío, F. P., *Synlett.* **2013**, *24*, 535-549.
85. Young, D., *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*. John Wiley & Sons, Inc.: New York, 2001.
86. Nguyen, Q. N. N.; Tantillo, D. J., *Chem. Asian J.* **2014**, *9*, 674-680.

geometries and to perform the characterization of their complete structure (lengths, angles, dihedrals...).[87]

-Analysis of the chemical reactivities, both to predict the reactive nature of the chemical reactants and to understand the chemical transformations previously studied experimentally.[88]

-Analysis of various chemical properties, for instance, prediction of spectra (IR, NMR, UV) or electron affinities and ionization potentials. Additionally, thermodynamic measurements can be easily done, and therefore, among other properties, equilibrium constants, rate constants or the reaction thermochemistry can be analized.[89 - 91] Furthermore, non-physically observable properties such as aromaticity can be also quantified.[92]

-Modelling of enzyme catalyzed reactions, prediction of the reactivity of the enzymes, development of new active sites, and generation of mutants from existing catalytic proteins.[93,94]

1.2.1 Brief classification of the computational methods

As it has been mentioned, the developments made in the last decades have resulted in numerous new computational tools.[95] Nevertheless, the selection of the proper methodology is fundamental in the pursuit of the correct description of a chemical problem. In this section, a brief introduction of the most common methodologies is presented, which are divided into six broad groups:

87. Schlegel, H. B., *J. Comput. Chem.* **2003**, *24*, 1514-1527.
88. Sperger, T.; Sanhueza, I. A.; Schoenebeck, F., *Acc. Chem. Res.* **2016**, *49*, 1311-1319.
89. Steinfeld, J. I.; Francisco, J. S.; Hase, W. L., *Chemical Kinetics and Dynamics 2nd Edition*. Prentice Hall: Upper Saddle River, New Jersey, 1999.
90. Laidler, K. J.; King, M. C., *J. Phys. Chem. A* **1983**, *87*, 2657-2664.
91. Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J., *J. Phys. Chem. A* **1996**, *100*, 12771-12800.
92. Schleyer, P. v. R.; Wu, J. I.; Cossío, F. P.; Fernández, I., *Chem. Soc. Rev.* **2014**, *43*, 4909-4921.
93. Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J., *Angew. Chem., Int. Ed.* **2006**, *45*, 6856-6859.
94. Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J., *Chem. Commun.* **2010**, *46*, 2354-2372.
95. Fernández, I.; Cossío, F. P., *Chem. Soc. Rev.* **2014**, *43*, 4906-4908.

a) **Molecular Mechanics (MM)**, also known as force field methods, use the classical laws of physics to describe the molecules by a simplified model, in which they are considered as a collection of balls (atoms) and springs (bonds). MM methods employ empirical force fields to reproduce the PES of the molecules and to find the bond lengths, dihedrals and angles for the structures of minimum energy. Therefore, these simplifications provide fast computational calculations, thus being useful for larger molecules. However, the most remarkable drawbacks are the need to generate new parameters for each specific situation and the impossibility to apply them in mechanisms where the electronic effects are predominant.[96,97]

b) **Semiempirical methods** are a group of Quantum Mechanics (QM) methods, which means that they predict geometries, energies and other chemical properties by solving the Schrödinger equation. These methods reduce their calculation time including many approximations to neglect the less important terms of the equations, and allow an easier solution of the integrals that are present on them. Therefore, they maintain the essential physics and they try to compensate for the limitations generated by the previous assumptions by including empirical parameters. Even though they provide a good tool for the qualitative study of the structures, the usefulness of these methods as a qualitative strategy for the characterization of the energies is directly influenced by the suitability of the employed parameters.[98]

c) ***Ab Initio* Methods** are a set of QM methods which do not include experimental data or parameters.[99] One of the main aims of *ab initio* calculations is to apply as few approximations and arbitrary decisions as possible. However, the Schrödinger equation cannot be solved in an exact way for molecules containing more than one electron, thus, a variable number of approximations are needed. The computational cost and the accuracy of the calculations vary depending on the number of included assumptions,

96. Rappe, A. K.; Casewit, C. L., *Molecular Mechanics Across Chemistry*. University Science Books: Sausalito, CA, 1997.
97. Williams, J. E.; Stand, P. J.; Schleyer, P. v. R., *Annu. Rev. Phys. Chem.* **2003**, *19*, 531-558.
98. Thiel, W., *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 145-157.
99. Hehre, W. J.; Radom, P. L.; Schleyer, P. V. R.; Pople, J., *Ab Initio Molecular Orbital Theory.* John Wiley & Sons: New York, 1986.

nevertheless, *ab initio* methods not only allow to calculate geometries and energies, but also permit the prediction of properties related with the electron distribution, *e.g.*, the polarity of the molecules. The main drawback of *ab initio* calculations is the higher amount of the needed computing resources, although the rapid developments are helping to solve these disadvantages.[100]

d) **Density Functional Theory** (**DFT**) **calculations** is a group of QM methods, which solve the Schrödinger equation in an alternative way. The main difficulty to solve a wave function is that it includes four variables, three spatial and one spin coordinates for each electron. This complexity is exponentially increased with the number of electrons, thus the DFT approach is focused on a physical observable as the electron density (a function that only depends on three spatial variables) instead of the wave function. For this approximation, functionals, mathematic entities related to a function, which connect the electron density with the energy were developed. By these means, the total electronic energy of the system is divided into four functionals of the electron density: kinetic energy, potential energy, electron-electron energy and exchange-correlation energy. Thus, this variation permits to perform faster calculations than *ab initio* methods. The DFT methods have become very popular in the last decades due to the confirmation of their accuracy in a broad spectrum of chemical problems when the appropriate functionals are selected.[101,102]

e) **Molecular Dynamics** are a series of methods that apply the laws of motions to molecules in order to solve chemical and biological problems. The application of these laws permits to obtain a trajectory, a sequence of frames containing the position and momentum of every atom at each particular time, which shows how molecules evolve over time. MD employ MM or QM, to generate the forces acting on the molecules and to calculate their motion, however, the employment of MM force fields permits the

100. Levine, I. N., *Quantum Chemistry, 5th Edition*. Prentice Hall: Englewood Cliffs, New Jersey, 2000.
101. Parr, R. G.; Yang, W., *Density Functional Theory of Atoms and Molecules*. Oxford University Press: New York, 1989.
102. Koch, W.; Holthausen, M. C., *A Chemist's Guide to Density Functional Theory, 2nd Edition*. Wiley-VCH: Weinheim, Germany, 2001.

calculation of bigger systems. MD can be applied in numerous fields, among them, the study of the interaction between the proteins and the ligands is one of the most commons.[103]

f) **Quantum Mechanics/Molecular Mechanics** (**QM/MM**) combine the previously described QM and MM methodologies providing a good tool for the evaluation of the catalysis performed by macromolecules, *e.g.*, organocatalysts and proteins. In this doctoral thesis, a catalytic reaction performed by the Consensus Tetratricoptide Repeat (CTPR) proteins is described employing QM/MM methods. Thus, in the next section, a deeper analysis of the QM/MM methodology is presented.

1.2.2 QM/MM ONIOM method: a tool for the description of catalytic proteins

The QM/MM methods combine the accuracy of the QM methodology with the higher speed of MM. This methodology was introduced in the early 1970s and has become of great interest to analyze local reactions in systems with a high amount of atoms.[104,105] The global molecular system is divided into two main parts; the chemically relevant area which is studied by the more expensive QM methods and the surrounding environment described by less accurate MM methods. Therefore, the total energy of the system ($E_{QM/MM}$) is divided into three main terms: the energy of the model analyzed by the QM methods (E_{QM}), the energy of the environment computed by means of MM (E_{MM}) and the energy of the interphase of the two methods (E_{QM-MM}) (Figure 1.2A).[106]

Therefore, one of the key points of this methodology is the treatment of the boundary between QM and MM phases when they are connected by covalent bonding. Different strategies have emerged to solve this problem: mechanical embedding, electronic embedding and polarizable embedding. Among them, the electronic embedding is the most broadly employed method, in which the atoms of the MM regions are allowed to polarize the QM region. Besides, the QM and MM regions belong,

103. Rapaport, D. C., *The Art of Molecular Dynamics Simulation, 2nd Edition*. Cambridge University Press: Cambridge, UK, 2004.
104. Honig, B.; Karplus, M., *Nature* **1971**, *229*, 558-560.
105. Warshel, A.; Levitt, M., *J. Mol. Biol.* **1976**, *103*, 227-249.
106. Friesner, R. A.; Guallar, V., *Annu. Rev. Phys. Chem.* **2005**, *56*, 389-427.

usually, to the same molecule and some covalent bonds need to be cut when both sections are divided. In this way, one or more electrons result unpaired in the QM part, which needs to be properly ended. The most common strategy to solve this problem is to include a link atom between both sections,[107] usually a hydrogen atom bonded to the QM frontier atom.[108]

However, this simplification contains some drawbacks that affect to the accuracy of QM/MM. One of the most controversial issues is the choice of the region that is treated by QM, which results essential to analyze if the selected part is the correct for an efficient study of the reaction mechanism. Another aspect to take into account is the way in which the boundary between the two regions is computed, which varies depending on the force field and the selected QM method. Besides, when the reaction environment is studied by MM, the conformational changes that are important for the reactivity are not studied in an accurate way.[109,110] Consequently, when QM/MM methods are selected for the study of reaction mechanisms, these drawbacks need to be taken into account. Nevertheless, with a correct selection of QM and MM methods the QM/MM methodology can provide a semiquantitative description of the reactions catalyzed by enzymes.[111,112]

107. Lin, H.; Truhlar, D. G., *Theor. Chem. Acc.* **2006**, *117*, 185-199.
108. Field, M. J.; Bash, P. A.; Karplus, M., *J. Comput. Chem.* **1990**, *11*, 700-733.
109. Hu, H.; Yang, W., *Theochem.* **2009**, *898*, 17-30.
110. Sousa, S. F.; Ribeiro, A. J. M.; Neves, R. P. P.; Brás, N. F.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J., *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7*, e1281.
111. Thiel, W. In *QM/MM Methodology: Fundamentals, Scope, and Limitations*, Multiscale simulation methods in molecular sciences, Jülich: NIC, J Grotendorst, N. A., S. Blgel, & D.Marx, Ed. Jülich: NIC, 2009.
112. Groenhof, G., Introduction to QM/MM Simulations. In *Biomolecular Simulations: Methods and Protocols*, Monticelli, L.; Salonen, E., Eds. Humana Press: Totowa, NJ, 2013; pp 43-66.

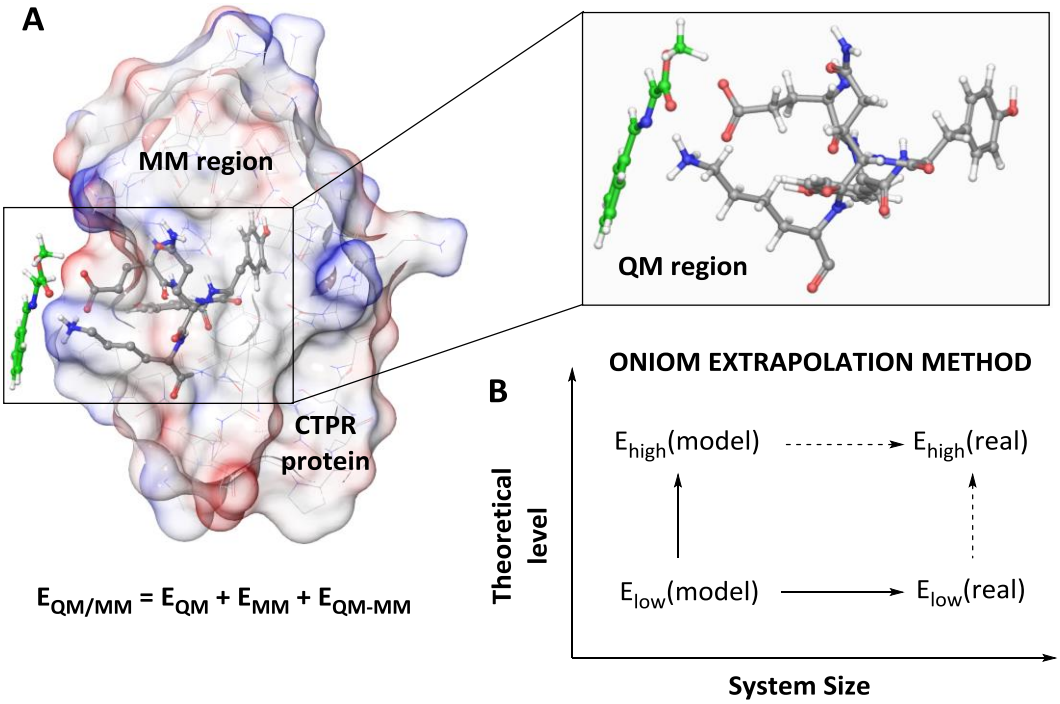


Figure 1.2. A) QM/MM model described for the CTPR protein presented in Chapter 2. The MM region is defined as surface and the QM region is presented with ball and sticks. **B**) Description of the ONIOM extrapolation method by the definition of high and low levels of energy for the model and the real systems.

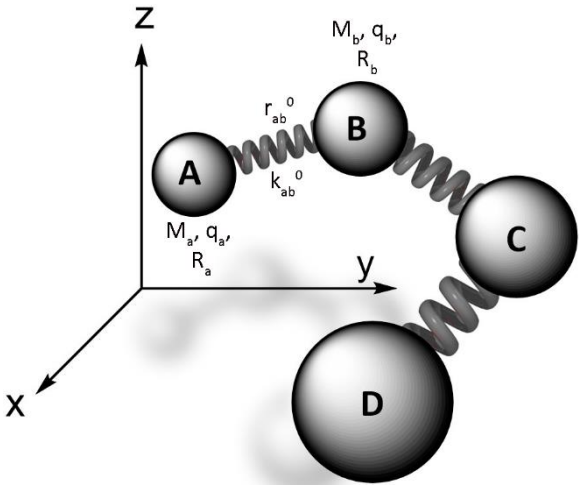
Among all the QM/MM strategies, ONIOM (Own N-layer Integrated molecular Orbital molecular Mechanics) methodology is one of the most popular in the literature. The new Integrated Molecular Orbital + Molecular Mechanics (IMOMM)[113] method developed in the 1990s is considered to be the first approach for the evolution of ONIOM method.[114] The previous methods were based on the aforementioned additive strategy, in which the total energy of the system was calculated by the sum of E_{QM}, E_{MM} and the energy of the interactions between both systems (E_{QM-MM}) (Figure 1.2A). In contrast, the ONIOM method uses an extrapolation method, calculating the model (A, the active site) in both high and low levels, whereas the total real system (AB) is calculated at the low level of theory (Figure 1.2B). Therefore, the global energy of the system is calculated summing the difference between the energy of the model system

113. Maseras, F.; Morokuma, K., *J. Comput. Chem.* **1995**, *16*, 1170-1179.
114. Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K., *J. Phys. Chem.* **1996**, *100*, 19357-19363.

in the high level, $E_{QM}(A)$, and the energy of the model system in the low level, $E_{MM}(A)$, to the energy of the whole real system at the low level, $E_{MM}(AB)$. In the next sections, a brief introduction of the employed MM and QM methods, more specifically DFT, is provided.[115,116]

1.2.2.1 MM methods employed in the QM/MM approach

MM methods employ the classical laws of physics for harmonic oscillators for the description of the molecules. In this manner, a molecule is defined as a group of atoms connected by springs (Figure 1.3), thus, the electrons are not explicitly treated in the interaction between atoms. The total MM energy is obtained by the sum of potential energy terms related to the stretching (E_r), bending (E_θ), torsion (E_ω) along with Coulombic (E_{coul}) and van der Waals interactions (Figure 1.3).



$$E = \sum E_r + \sum E_\theta + \sum E_\omega + \sum E_{coul} + \sum E_{vdw}$$

Figure 1.3. Description of MM model and definition of total MM energy.

The atoms and the interaction between them are described by some parameters: charges (q), radius (r and R), mass (M) or Hooke's constants (k). The set of parameters for various elements and bindings is gathered in what are defined as empirical force

115. Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K., *Chem. Rev.* **2015**, *115*, 5678-5796.
116. Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K., *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 327-350.

fields. They are constituted by three main elements: a group of atom types which defines their properties, a group of equations that define the PES of the molecules and additional parameters that suit the equations and atom types, obtained from experimental or higher level computational data. It is important to check if a previous parametrization process has been done for the selected force field and structure to obtain accurate results.

Numerous force fields have been described for MM (AMBER, CHARMM…), in our specific study the dreiding force field which is implemented for Gaussian and was developed by Mayo et al. has been employed.[117] This force field is one of the simplest generic force fields, however, it has been demonstrated that is a good tool to achieve accurate structures for biological samples. However, it is not so useful for the study of the breakage and formation of chemical bonds.

1.2.2.2 QM-DFT methods employed in the QM/MM approach

While MM methods are employed for the analysis of the protein environment, the reactivity of the protein is studied by means of more accurate QM methods, in our case, DFT methods. QM methods perform an explicit description of the electronic distribution of the molecules, considering the charge of the nuclei as well as the mass and number of electrons. These methods are based on the Schrödinger[118] equation to obtain the total energy of the system (E). Ψ (**r**, **R**), is the wave function for every nuclei and electrons which depends on the nuclear (R) and electronic positions (r), is essential to obtain the molecular properties; $\hat{H}$ is the Hamiltonian operator, a sum of the kinetic (T) and potential (V) operators. To solve this equation for systems containing more than one electron some approximations need to be done, *e.g.* Born Oppenheimer approximation.

$$\hat{H}\Psi\left(\vec{\mathbf{r}}\,;\vec{\mathbf{R}}\right) = E\Psi\left(\vec{\mathbf{r}}\,;\vec{\mathbf{R}}\right) \qquad (1.1)$$

The Density Functional Theory (DFT) methods take into account the difficulty to solve the Schrödinger equation and they include an approximation in which the physical

117. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., *J. Phys. Chem.* **1990**, *94*, 8897-8909.
118. Schrödinger, E., *Phys. Rev.* **1926**, *28*, 1049-1070.

observable electron density, $\rho(\vec{\boldsymbol{r}})$, is evaluated instead of the wave function. The electron density is only a function of the three spatial variables and therefore, provides less expensive and faster calculations. In this way, the energy and the properties of the molecules in its fundamental state are in function of the $\rho(\vec{\boldsymbol{r}})$, which varies in function of the total number of electrons.

$$\mathrm{E} = \mathrm{E}[\rho(\vec{\mathbf{r}})] \tag{1.2}$$

The energy is given by the combination of four functionals of the electron density and an additional term: the electronic kinetic energy, $T_S[\rho]$; the electron-nuclei attractive energy, $E_{ne}[\rho]$; the electron-electron repulsion, $J[\rho]$; the internuclear repulsion Coulomb energy, V_{NN}; and the exchange-correlation functional, $E_{xc}[\rho]$.[119] This last term has resulted to be the most difficult to determinate and some approximations need to be done in order to obtain accurate results: Local Density Approximation (LDA) and the Generalized Gradient approximation (GGA) are the most common ones.

$$\mathrm{E_{DFT}}[\rho] = \mathrm{T_S}[\rho] + \mathrm{E_{ne}}[\rho] + \mathrm{J}[\rho] + \mathrm{E_{xc}}[\rho] + \mathrm{V_{NN}} \tag{1.3}$$

In addition, to solve the problem of the exchange-correlation functional, hybrid functionals which combine DFT and Hartree-Forck (HF)[120,121] terms (HF approach states that the spatial distribution of each electron does not depend on the motion of the remaining electrons) have been developed. Among them, the B3LYP and M06-2X functionals are the most common ones.

The **B3LYP**[122] is a three parameter exchange-correlation hybrid functional, in which the exchange-correlation functional is defined by the following equation:

$$\mathrm{E_{XC}^{B3LYP}} = (1 - \mathrm{a_0})\mathrm{E_X^{LSDA}} + \mathrm{a_0}\mathrm{E_X^{HF}} + \mathrm{a_x}\mathrm{E_X^{B88}} + \mathrm{a_C}\mathrm{E_C^{LYP}} + (1 - \mathrm{a_C})\mathrm{E^{VWN}} \tag{1.4}$$

The values for the three parameters are a_0, a_x and a_c are, 0.20, 0.72 and 0.81, respectively. $\mathrm{E_X^{LSDA}}$ is the exchange functional, $\mathrm{E_X^{HF}}$ the exact HF exchange functional,

119. Cossío, F. P., Calculation of Kinetic Data Using Computational Methods. In *Rate Constant Calculation for Thermal Reactions: Methods and Applications*, John Wiley & Sons: Hoboken, New Jersey, 2011; pp 33–65.
120. Hartree, D. R., *Math. Proc. Camb. Philos. Soc* **1928**, *24*, 89-110.
121. Fock, V., *Z. Phys.* **1930**, *61*, 126-148.
122. Becke, A. D., *J. Chem. Phys.* **1993**, *98*, 1372-1377.

E_X^{B88} the Becke exchange functional,[123] E_C^{LYP} the correlation functional defined by Lee et al.[124] and E^{VWN} the functional described by Vosco[125] et al. Nevertheless, this functional has been reported no to be so accurate for noncovalent interactions and therefore, new functionals have been developed to solve this problems.

Among those new functionals the Minnesota functional **M06-2X**, [126] which facilitates the study of noncovalent interactions, proposed by Zhao and Truhlar is one of the most employed ones. For the better description of the noncovalent interactions, it depends on the spin density, reduced spin density gradient and spin kinetic energy density. In comparison with M06,[127] the M06-2X functional includes a 54% HF exchange instead of the 27% included in M06.

Once the desired functionals have been selected, the next step is the choice of the most suitable basis set, which is a set of functions developed as a linear combination of them for the good description of the molecular orbitals. In our particular case, the B3LYP method and 6-31G(d,p)[128] basis set have been used for the optimization of the structure, whereas single point calculations have been performed at the M06-2X/6-311+G(2d,2p) level. These basis sets are briefly described:

-6-31G(d,p): includes a split-valence double zeta basis set, in which the core electrons are represented with six gaussian functions whereas the valence electrons are represented with three and one gaussians. The (d,p) shows that “d” functions have been added to polarize the “p” functions of the heavy atoms, and “p” functions have been included in the “s” orbitals of the hydrogen atoms.

-6-311+G(2d,2p): includes a split valence triple zeta basis set, dividing the region of the valence electrons with three independents functions, one consisting on three gaussians and the other with only one. Additionally, “+” indicates that diffuse “s” and

123. Becke, A. D., *Phys. Rev. A* **1988**, *38*, 3098-3100.
124. Lee, C.; Yang, W.; Parr, R. G., *Phys. Rev. B* **1988**, *37*, 785-789.
125. Vosko, S. H.; Wilk, L.; Nusair, M., *Can. J. Phys.* **1980**, *58*, 1200-1211.
126. Zhao, Y.; Truhlar, D. G., *Acc. Chem. Res.* **2008**, *41*, 157-167.
127. Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.* **2008**, *120*, 215-241.
128. Hehre, W. J.; Ditchfield, R.; Pople, J. A., *J. Chem. Phys.* **1972**, *56*, 2257-2261.

"p" basis functions are added to each non H-atoms. Finally, two polarization functions are included both for the heavy and hydrogen atoms.

1.2.3 Computational characterization of reaction mechanisms

When the energy of a system is represented in terms of the position of the atoms the PES is obtained. The most relevant structures of the PES are the stationary points, *i.e*., geometries with an energy gradient value of zero. In this way, analyzing the stationary points in the PES, a clear picture of the reaction mechanism can be achieved (Figure 1.4).

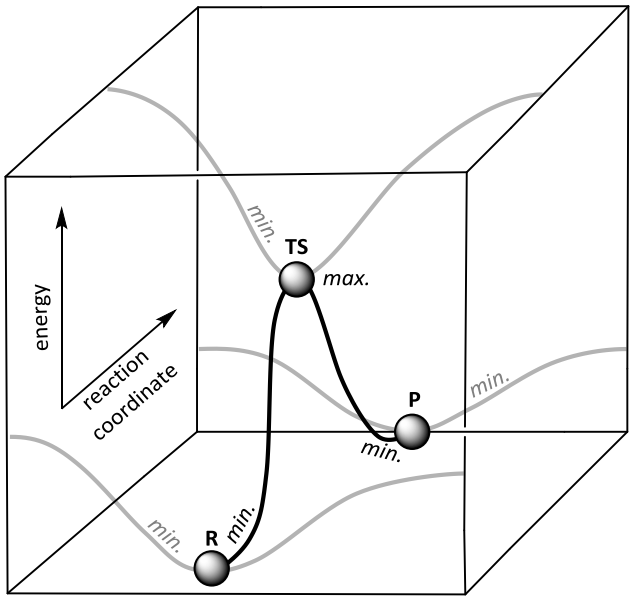


Figure 1.4. Simplified model of the transformation of the Reactant (R) to the Product (P) through its Transition Structure (TS). The energy vs the reaction coordinates is represented.

When Reactants (R), Products (P) or intermediates are considered, the structure results to be a minimum with respect to any direction of the reaction coordinates, thus, the second derivatives of the energy with respect to the nuclear coordinates must be positive. On the other hand, Transition Structures (TS) are also zero-gradient points, however, its second derivatives are positive in all the reaction coordinates except in one in which its value is negative. This negative value shows that the TS is a maximum in one reaction coordinate and it represents a saddle point. Additionally, the Intrinsic Reaction Coordinates (IRC) connect the structure of TS with the local minima of the reactant and products, being helpful to describe the complete reaction pathway.

1.3 PLATINUM(II) DERIVATIVES AND NITROGEN MUSTARDS IN CANCER THERAPY

Cancer is the term employed to name a collection of related diseases characterized by uncontrolled division, growth and spread of a group of cells. Additionally, these features result in the development of primary tumors that invade and damage the surrounding tissues. In healthy tissues, the growth and division of the cells are controlled, and the programmed cell death mechanisms are only activated when the cell viability is lost, giving rise to new cells. However, when carcinogenic processes are developed, cells become more abnormal and new ones are generated even if they are not needed. In this way, the malignant cells can spread through the blood or the lymph system and form novel tumors in other tissues; this process is known as metastasis.[129-131]

The origin of cancer is usually found in the genetic alterations of normal cells; moreover, this damage can be considerably affected by certain environmental factors.[132] These genetic changes vary from single mutations in DNA, to epigenetic changes (reversible covalent modifications performed in DNA or in related proteins)[133,134] or sequence deletions. Thus, each cancer has a singular combination of genetic changes and more than 100 classes of cancers are known depending on the affected organs or tissues.[135] The high degree of variability and their aggressiveness make cancer one of the most difficult diseases to treat. According to the World Health Organization (WHO), in 2018, 18.1 million of new cases aroused, resulting in 9.6 million deaths.[136]

129. Todd, A.; Grounwater, P. W.; Gill, J. H., *Anticancer Therapeutics: From Drug Discovery to Clinical Applications*. Wiley-Blackwell: Hoboken, New Jersey, 2017.
130. Link, W., *Principles of Cancer Treatment and Anticancer Drug Developmen*. Springer: Cham. Switzerland, 2019.
131. David, A. R.; Zimmerman, M. R., *Nat. Rev. Cancer* **2010**, *10*, 728-733.
132. Grivennikov, S. I.; Greten, F. R.; Karin, M., *Cell* **2010**, *140*, 883-899.
133. Sharma, S.; Kelly, T. K.; Jones, P. A., *Carcinogenesis* **2009**, *31*, 27-36.
134. Dawson, Mark A.; Kouzarides, T., *Cell* **2012**, *150*, 12-27.
135. Price, P.; Sikora, K., *Treatment of Cancer, 6th edition*. CRC Press: Boca Raton, Florida, 2014.
136. International Agency for Research on Cancer. World Health Organization. https://www.who.int/cancer/PRGlobocanFinal.pdf. (accessed Sept 23, 2019).

The high incidence of cancer has promoted the pursuit of an efficient treatment. Nevertheless, it is important to note that the survival rates are dependent on the type of cancer and the stage when they are detected. Currently, surgery, radiation therapy, chemotherapy, immunotherapy, targeted therapy, hormone therapy and stem cell transplant are the most broadly employed treatments. Among the different therapies, synthetic medicinal chemistry has emerged as an useful tool for the development of new anticancer drugs, thus, in the last decades, numerous new promising compounds have been designed.[137,138] Initially, the main target of cancer drugs was DNA, nevertheless, the design of new compounds has been focused on more selective targets, in this way, new targets have been included; *i.e*, RNA and proteins involved in cancer related metabolic pathways.[139-142]

The use of chemotherapy in cancer treatment began in the 1940s when the nitrogen mustards, previously developed for chemical warfare, were used to treat the non-Hodgkin's lymphoma.[143-145] The mechanism of these novel drugs was determined to be the alkylation of DNA on purine bases (adenine and guanine); thus, provoking DNA damage and a subsequent apoptotic cell death.[146] By these means, these first compounds give rise to the family of DNA alkylating agents, a kind of drugs that are capable to bind covalently an alkyl group to DNA and induced a cytotoxic effect. Additionally, platinum complexes are frequently included into this family of anticancer drugs due to the strong chemical bonds formed with the nucleic acids, nevertheless, in this case, the platinum is directly bound to the DNA.[147]

137. Bradbury, R., *Cancer*. Springer-Verlag: Berlin, Germany, 2007; Vol. 1.
138. Hait, W. N., *Nat. Rev. Drug Discov.* **2010**, *9*, 253-254.
139. Espinosa, E.; Zamora, P.; Feliu, J.; González Barón, M., *Cancer Treat. Rev.* **2003**, *29*, 515-523.
140. Atkins, J. H.; Gershell, L. J., *Nat. Rev. Cancer* **2002**, *2*, 645-646.
141. Wu, X.-Z., *Med. Hypotheses* **2006**, *66*, 883-887.
142. Neidle, S.; Thurston, D. E., *Nat. Rev. Cancer* **2005**, *5*, 285-296.
143. Papac, R. J., *Yale J. Biol. Med.* **2001**, *74*, 391-398.
144. Goodman, L. S.; Wintrobe, M. M.; Dameshek, W.; Goodman, M. J.; Gilman, A.; McLennan, M. T., *JAMA* **1946**, *132*, 126-132.
145. Chabner, B. A.; Roberts, T. G., *Nat. Rev. Cancer* **2005**, *5*, 65-72.
146. Cheung-Ong, K.; Giaever, G.; Nislow, C., *Chem. Biol.* **2013**, *20*, 648-659.
147. Avendaño, C.; Menénendez, C., *Medicinal Chemistry of Anticancer Drugs*. Elsevier B.V.: Oxford, England, 2008.

The alkylating agents and platinum complexes react directly with electron-rich atoms of biologic molecules to perform nucleophilic substitution reactions. More specifically, SN2 reactions have been described for the interaction between the DNA and nitrogen mustards or platinum(II) complexes.[148,149] Furthermore, these drugs mainly interact with the N-7 of the guanines, although a less selective interaction with the adenines has been characterized (Figure 1.5).[150]

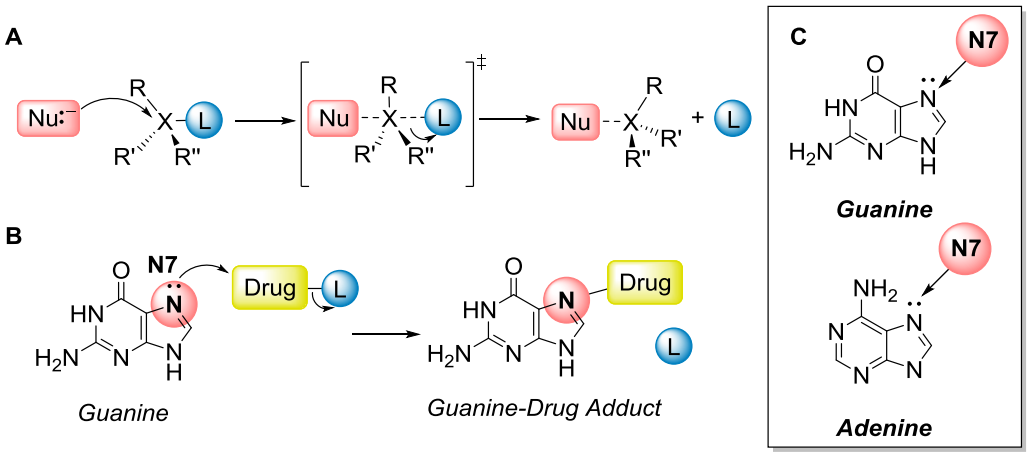


Figure 1.5. A) General reaction mechanism for SN2 reactions. **B**) SN2 reaction between the guanine and the anticancer drugs. **C**) Chemical structures of the purine nucleobases guanine and adenine.

In this doctoral thesis (Chapter 3) a novel family of platinum(II) complexes has been designed, thus in the next sections, we are going to focus on platinum(II) complexes as anticancer drugs. Additionally, a brief introduction into nitrogen mustards, the most broadly known DNA alkylating agents is going to be provided.

1.3.1 Nitrogen mustards in cancer therapy

As it has been mentioned before, the nitrogen mustards were the first compounds described as DNA alkylating agents. In World War I, the sulfur mustards were employed for chemical warfare and after the war, their evaluation as anticancer

148. Singh, R. K.; Kumar, S.; Prasad, D. N.; Bhardwaj, T. R., *Eur. J. Med. Chem.* **2018**, *151*, 401-433.
149. de Cózar, A.; Larrañaga, O.; Bickelhaupt, F. M.; San Sebastián, E.; Ortega-Carrasco, E.; Maréchal, J.-D.; Lledós, A.; Cossío, F. P., *ChemPhysChem* **2016**, *17*, 3932-3947.
150. Colvin, M. Alkylating Agents. In: Kufe, D.W., Weichselbaum, R.R., et al., editors. Holland-Frei Cancer Medicine. 6th edition. Hamilton (ON): BC Decker; 2003. Available from: https://www.ncbi.nlm.nih.gov/books/NBK12772/ (accessed 15 sept, 2019).

drug was studied.[151] The high toxicity of the sulfur mustards induced the modification of these compounds into nitrogen mustards, resulting in tumor suppression.[152] Since then, numerous nitrogen mustards derivatives have been designed trying to optimize their efficacy. These efforts have resulted in the approbation of some anticancer drugs; *e.g.*, chlorambucil, melphalan and ifosfamide (Figure 1.6A).[153] Although this Doctoral Thesis is mainly focused on the synthesis of platinum(II) complexes, in section 3.3.6 few nitrogen mustard derivatives are described.

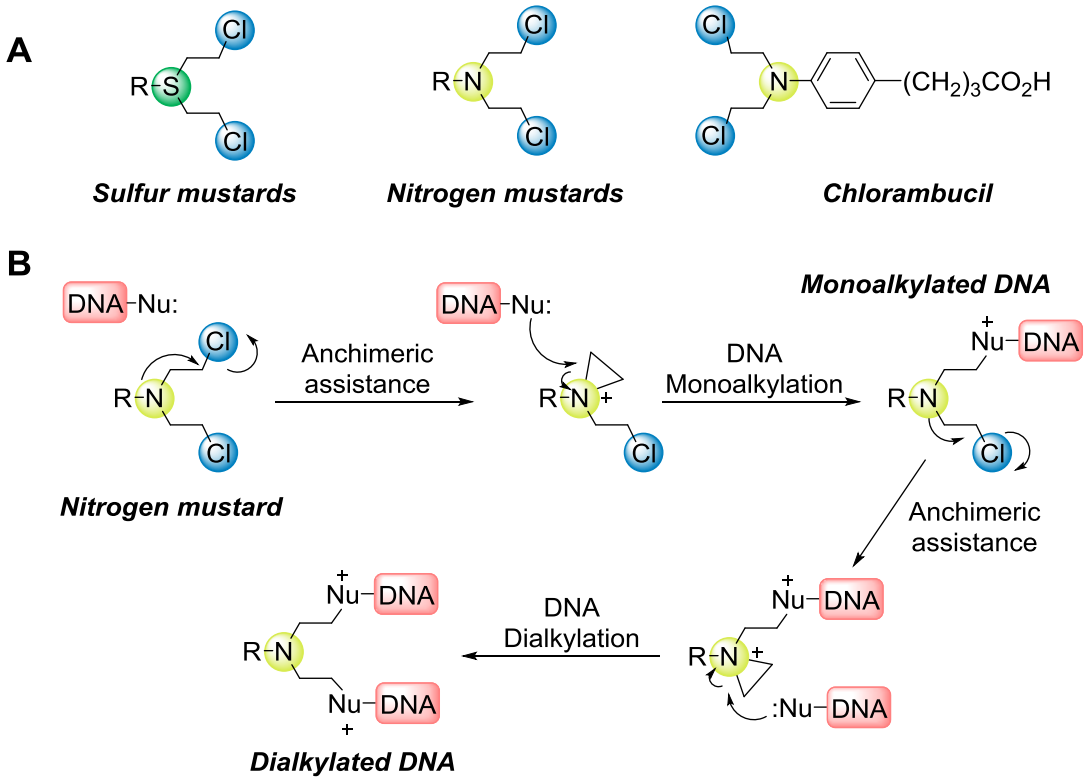


Figure 1.6. A) General structures for the sulfur mustards, nitrogen mustards and 2D chemical structure of one of the approved nitrogen mustards: Chlorambucil. **B**) Mechanism of DNA monoalkylation and dialkylation processes performed by the nitrogen mustards, which are enhanced by the anchimeric assistance of the nitrogen atom.

The nitrogen mustards provoke a cytotoxic effect and a subsequent cell death by mainly binding the N7 of the guanines of the DNA, besides, their interaction with the N3

151. Kohn, K. W., *Cancer Res.* **1996**, *56*, 5533-5546.
152. Gilman, A.; Philips, F. S., *Science* **1946**, *103*, 409-436.
153. Puyo, S.; Montaudon, D.; Pourquier, P., *Crit. Rev. Oncol. Hematol.* **2013**, *89*.

of adenines has also been observed.[154,155] Once the nitrogen mustard is inside the cells, its binding with the DNA occurs in a two-step process, in which the anchimeric assistance of the nitrogen atom is essential to increase its reactivity. Firstly, the central atom promotes an intramolecular SN2 cyclization to form the highly reactive aziridium cation, which, in a second step, undergoes a nucleophilic addition by the DNA, usually the N7 of the guanines, to form the monoalkylated adduct. In a subsequent step, this reaction is repeated to form the final dialkylated DNA adduct (Figure 1.6B).[156,157]

The dialkylation process of DNA induced by nitrogen mustards produces interstrand or instrastrand DNA adducts,[158-160] which affect directly to the replication or transcription processes of the DNA, promoting the cell death in a similar way to platinum complexes. Additionally, the combination of platinum complexes and nitrogen mustards has been proposed in order to evaluate the synergistic effect of their DNA binding activity.[161] However, the high reactivity promoted by the lone pair of the nitrogen produces a high toxicity.

1.3.2 Platinum(II) complexes in cancer therapy

The use of platinum(II) complexes in cancer therapy started in the 1960s with cisplatin which was described in 1845 by Peryone.[162] Initially, it was employed for the study of the effects of electric currents on cells.[163] Interestingly, they observed that cisplatin, cis-$PtCl_2(NH_3)_2$, was able to inhibit the cell division, and thus, it was administered to mice containing tumors, observing a tumor regression.[164] Since its

154. Povirk, L. F.; Shuker, D. E., *Mutat. Res-Rev. Genet.* **1994**, *318*, 205-226.
155. Mattes, W. B.; Hartley, J. A.; Kohn, K. W., *Nucleic Acids Res.* **1986**, *14*, 2971-2987.
156. Bonsaii, M.; Gholivand, K.; Abdi, K.; Valmoozi, A. A. E.; Khosravi, M., *Med. Chem. Commun.* **2016**, *7*, 2003-2015.
157. Larrañaga, O.; de Cózar, A.; Cossío, F. P., *ChemPhysChem* **2017**, *18*, 3390-3401.
158. Singh, R. K.; Kumar, S.; Prasad, D. N.; Bhardwaj, T. R., *Eur. J. Med. Chem.* **2018**, *151*, 401-433.
159. Bauer, G. B.; Povirk, L. F., *Nucleic Acids Res.* **1997**, *25*, 1211-1218.
160. Schärer, O. D., *ChemBioChem* **2005**, *6*, 27-32.
161. Karmakar, S.; Purkait, K.; Chatterjee, S.; Mukherjee, A., *Dalton Trans.* **2016**, *45*, 3599-3615.
162. Peyrone, M., *Justus Liebigs Ann. Chem* **1845**, *55*, 205-213.
163. Rosenberg, B.; Van Camp, L.; Krigas, T., *Nature* **1965**, *205*, 698-699.
164. Rosenberg, B.; Van Camp, L.; Trosko, J. E.; Mansour, V. H., *Nature* **1969**, *222*, 385-386.

approval by the Food and Drug Administration (FDA) in 1978, cisplatin has become one of the most employed chemotherapeutic agents.[165]

Cisplatin has shown great activity in different kind of tumors, however, its use presents certain drawbacks.[166] Among them, the high toxicity in humans[167] and the cisplatin resistance mechanisms (see section 1.3.3) developed by the tumoral cells.[168,169] This has resulted in the novel design and biological evaluation of numerous novel cisplatin-based derivatives.[170,171] Despite all the employed efforts, to date, only three compounds (cisplatin, carboplatin and oxaliplatin) have been approved by the FDA and the use of four additional complexes has been validated in some countries (Figure 1.7).

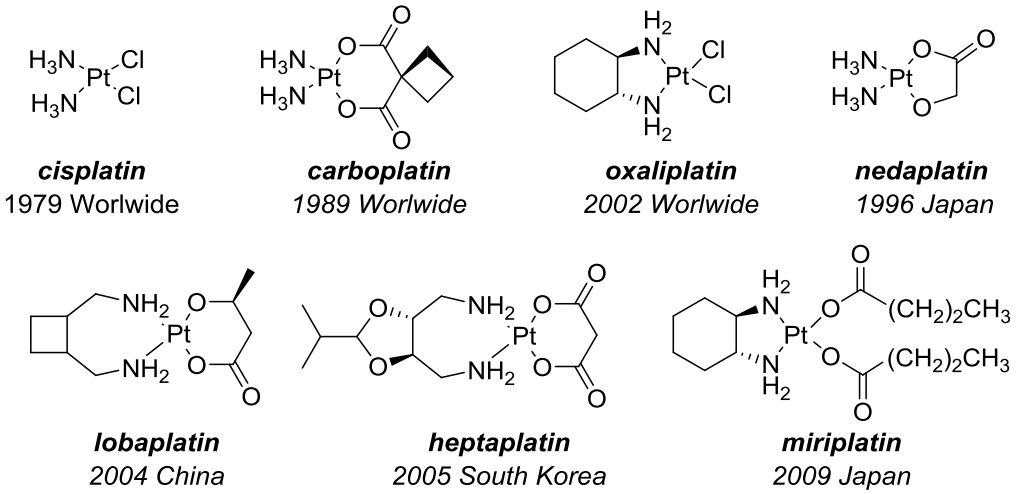


Figure 1.7. Approved platinum-based cancer drugs, their approbation year and countries.

Among all the approved drugs, cisplatin is the most broadly employed complex and it is been applied in diverse types of cancers: ovarian, lung, breast, carcinoma (cancers developed from epithelial cells), brain…[172] Cisplatin consists of a square-planar platinum complex [173] containing two ammonia groups and two labile chlorines

165. Barabas, K.; Milner, R.; Lurie, D.; Adin, C., *Vet. Comp. Oncol.* **2008**, *6*, 1-18.
166. Florea, A. M. B., D, *Cancers* **2011**, *3*, 1351-1371
167. Oun, R.; Moussa, Y. E.; Wheate, N. J., *Dalton Trans.* **2018**, *47*, 6645-6653.
168. Siddik, Z. H., *Oncogene* **2003**, *22*, 7265-7279.
169. Galluzzi, L.; Senovilla, L.; Vitale, I.; Michels, J.; Martins, I.; Kepp, O.; Castedo, M.; Kroemer, G., *Oncogene* **2011**, *31*, 1869-1883.
170. Kelland, L., *Nat. Rev. Cancer* **2007**, *7*, 573-584.
171. Johnstone, T. C.; Suntharalingam, K.; Lippard, S. J., *Chem. Rev.* **2016**, *116*, 3436-3486.
172. Dasari, S.; Bernard Tchounwou, P., *Eur. J. Pharmacol.* **2014**, *740*, 364-378.
173. Fenske, R. F.; Martin, D. S.; Ruedenberg, K., *Inorg. Chem.* **1962**, *1*, 441-452.

coordinated to the Pt(II) in the *cis* configuration. The high chloride concentration present in plasma prevents its hydrolysis and only when the complex is introduced inside the cells, the cisplatin reactive species are formed.[174,175] By these means, the first step for the interaction of cisplatin with DNA is the cellular uptake, which occurs through passive diffusion or several transport proteins.[176] Among this second group, the Copper Transporter-1 (CTR-1) and Organic Cation Transporters (OCT) have described as cisplatin transporters (Figure 1.8).

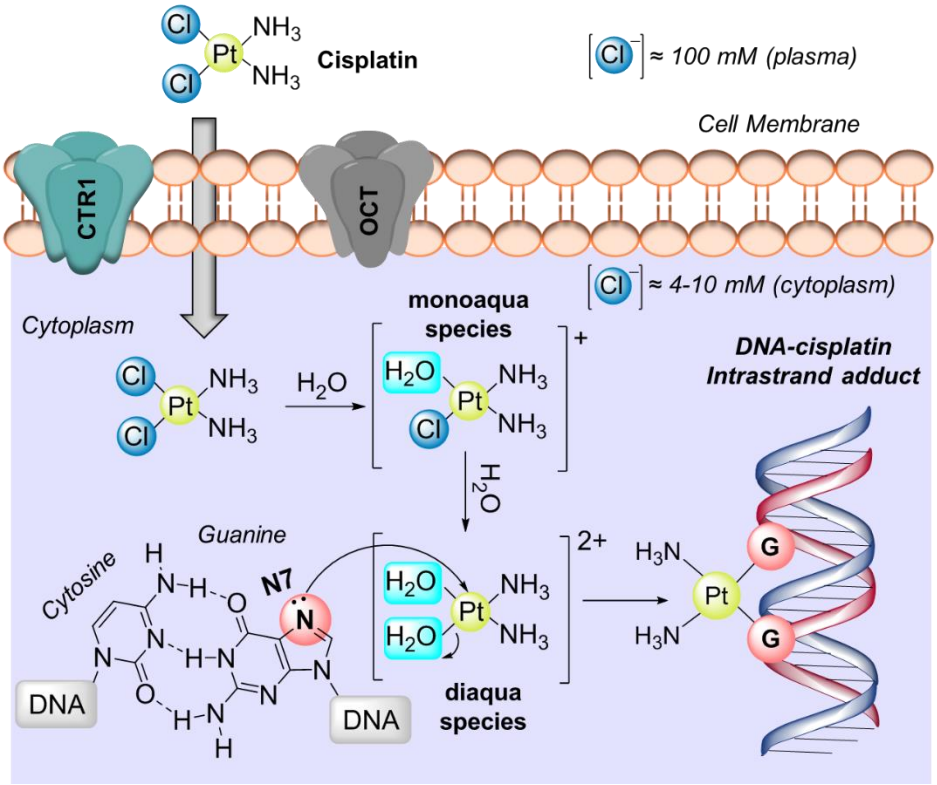


Figure 1.8. Cellular uptake of cisplatin and hydrolysis reaction of cisplatin into the reactive monoaqua, cis-$[Pt(NH_3)_2Cl(OH_2)]^+$, or diaqua, cis-$Pt(NH_3)_2(OH_2)_2]^{2+}$, species inside the cell. Nucleophilic attack of the N7 of the guanine to cisplatin to form, after two nucleophilic reactions, the most common DNA-cisplatin intrastrand adducts.

174. Davies, M. S.; Berners-Price, S. J.; Hambley, T. W., *Inorg. Chem.* **2000**, *39*, 5603-5613.
175. Kelland, L. R.; Farrell, N. P., *Platinum-Based Drugs in Cancer Therapy*. Humana Press: 2000.
176. Hall, M. D.; Okabe, M.; Shen, D.-W.; Liang, X.-J.; Gottesman, M. M., *Annu. Rev. Pharmacol. Toxicol.* **2008**, *48*, 495-535.

Once the cisplatin has entered the cell, the hydrolysis process begins.[177-179] Firstly, the platinum complex is hydrolyzed to lead to the intracellular formation of the more electrophilic monoaqua species, cis-$[Pt(NH_3)_2Cl(OH_2)]^+$, and in a subsequent step the diaqua species, cis-$Pt(NH_3)_2(OH_2)_2]^{2+}$, are formed (Figure 1.7).[180,181] Afterwards, the cisplatin interacts with the DNA, preferentially with the N7 of the guanine, although the interaction with the adenine has been also described.[182] In this way, the N7 present on the purine bases attacks the reactive cisplatin species to undergo the SN2 reaction, see Figure 1.5. Due to the presence of two electrophilic centers in its structure, the cisplatin has the ability to form either monofunctional or bifunctional adducts with the nucleic acids of the DNA.[183]

The formed adducts have been confirmed to occur mainly through intrastrand crosslinks, consisting on the formation of two novel bonds between the platinum metallic center and the N7-s of two guanines of the same DNA strand; thus, provoking a distortion of the DNA structure and a subsequent cell death (Figure 1.8).[184-187] However, among the intrastrand crosslinks, the binding process could occur in neighbor guanines to form 1,2-intrastrand or 1,3-crosslinks. Additionally, the bifunctional interaction with the DNA could also happen in different strands generating interstrand adducts.[188,189] Within the DNA-cisplatin adducts, the experimental binding studies have shown that

177. Song, I.-S.; Savaraj, N.; Siddik, Z. H.; Liu, P.; Wei, Y.; Wu, C. J.; Kuo, M. T., *Mol Cancer Ther.* **2004**, *3*, 1543-1549.
178. Yonezawa, A.; Masuda, S.; Yokoo, S.; Katsura, T.; Inui, K.-i., *J. Pharmacol. Exp. Ther.* **2006**, *319*, 879-886.
179. Ciarimboli, G.; Ludwig, T.; Lang, D.; Pavenstädt, H.; Koepsell, H.; Piechota, H.-J.; Haier, J.; Jaehde, U.; Zisowsky, J.; Schlatter, E., *Am. J. Pathol.* **2005**, *167*, 1477-1484.
180. Johnson, N. P.; Hoeschele, J. D.; Rahn, R. O., *Chem. Biol. Interact.* **1980**, *30*, 151-169.
181. Lippard, S. J., *Science* **1982**, *218*, 1075-1082.
182. Cohen, G. L.; Ledner, J. A.; Bauer, W. R.; Ushay, H. M.; Caravana, C.; Lippard, S. J., *J. Am. Chem. Soc.* **1980**, *102*, 2487-2488.
183. Fichtingerschepman, A. M. J.; Vanderveer, J. L.; Denhartog, J. H. J.; Lohman, P. H. M.; Reedijk, J., *Biochemistry* **1985**, *24*, 707-713.
184. Noll, D. M.; Mason, T. M.; Miller, P. S., *Chem. Rev.* **2006**, *106*, 277-301.
185. Jamieson, E. R.; Lippard, S. J., *Chem. Rev.* **1999**, *99*, 2467-2498.
186. Fichtinger-Schepman, A. M. J.; Van der Veer, J. L.; Den Hartog, J. H. J.; Lohman, P. H. M.; Reedijk, J., *Biochemistry* **1985**, *24*, 707-713.
187. Baik, M.-H.; Friesner, R. A.; Lippard, S. J., *J. Am. Chem. Soc.* **2003**, *125*, 14082-14092.
188. Sherman, S. E.; Lippard, S. J., *Chem. Rev.* **1987**, *87*, 1153-1181.
189. Wang, D.; Lippard, S. J., *Nat. Rev. Drug. Discov.* **2005**, *4*, 307-320.

1,2-intrastrand GpG (60-65%) and ApG 1,2-intrastrand (20-25%), the ones that involve adjacent bases, are the most commons. On the other hand, 1,3-intrastrand (2%), monofunctional adducts (2%) and 1,2/1,3-interstrand adducts (2%) result to be less susceptible (Figure 1.9).[170,185]

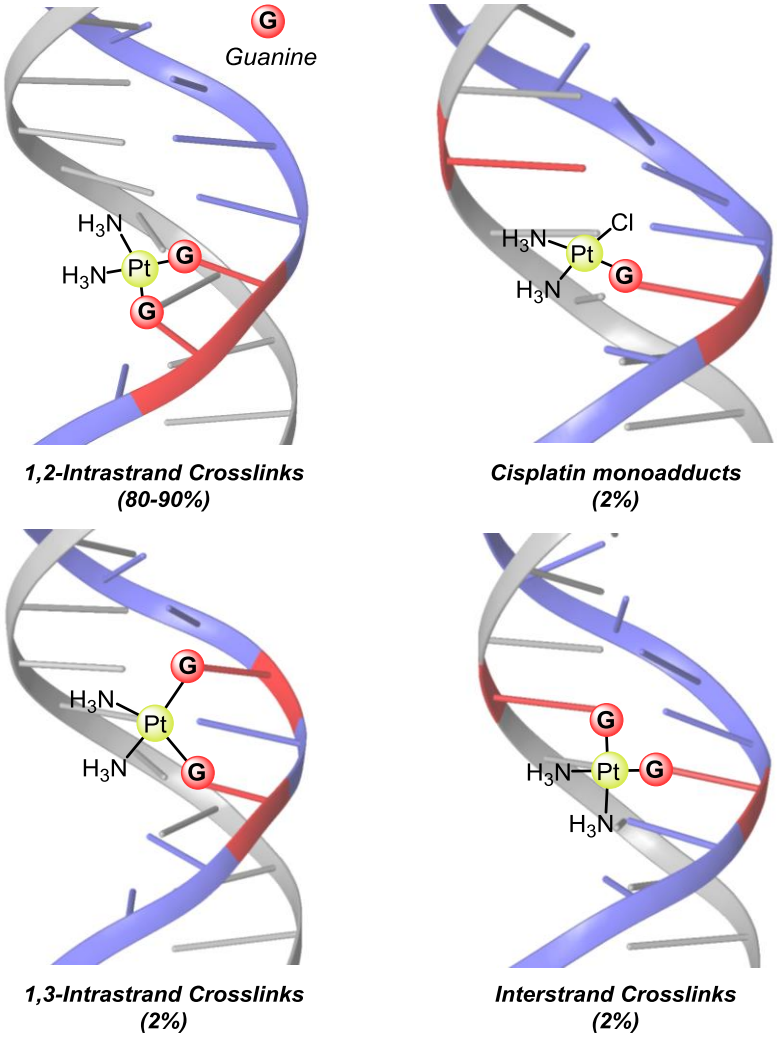


Figure 1.9. Principal DNA adducts formed by cisplatin: 1,2-intrastrand crosslinks, 1,3-intrastrand crosslinks, cisplatin monoadducts and interstrand crosslinks.

When the reactivity of cisplatin is analyzed, the described interactions between the cisplatin and some proteins must be taken into consideration. The presence of nucleophilic groups in some proteins, *e.g*, the sulfur atom of the cysteine residue, could provoke the SN2 reaction between the protein and cisplatin. This reactivity has been described for some proteins such as glutathione (GSH) or hemoglobin. Therefore, the

protein-cisplatin adducts would not allow the interaction with the DNA, reducing the effect of the drug.

Regarding the feasible cisplatin adducts, it is determined that the damage caused by cisplatin in DNA is mainly based on 1,2-intrastrand crosslinks, unfortunately, the majority of the tumor cells develop cisplatin resistance mechanisms and are able to repair this type of damage.[190] By these means, the main drawbacks of cisplatin therapy are considered to be the high toxicity of these compounds, due to the lack of selectivity, and the aforementioned developed resistance mechanisms. For these reasons, the search of cisplatin derivatives that could overcome these weaknesses is an extensive field of research.

1.3.3 Platinum drug resistance mechanisms

One of the main weaknesses of chemotherapy is the anticancer drug resistance mechanism developed by the carcinogenic cells, which has a direct impact on the efficacy of the employed compounds. For instance, cisplatin is one of the principal treatments in ovarian cancer; however, the A2780 ovarian cancer cell line has been proved to be able to develop drug resistance mechanisms after a continuous exposure to cisplatin.[191] These resistance mechanisms have been described in numerous diseases, in which the treatment results to be effective in an initial stage, but novel tumoral cells are developed afterwards.[192 - 195] These processes have provoked the need for the development of new drugs which could overcome this resistance mechanisms.[196-199] Kroemer and coworkers suggested the classification of the platinum resistance

190. Basu, A.; Krishnamurthy, S., *J. Nucleic Acids* **2010**, *2010*, 201367.
191. Parker, R. J.; Eastman, A.; Bostick-Bruton, F.; Reed, E., *J. Clin. Invest.* **1991**, *87*, 772-777.
192. Kelderman, S.; Schumacher, T. N. M.; Haanen, J. B. A. G., *Mol. Oncol.* **2014**, *8*, 1132-1139.
193. Galluzzi, L.; Vitale, I.; Michels, J.; Brenner, C.; Szabadkai, G.; Harel-Bellan, A.; Castedo, M.; Kroemer, G., *Cell Death Dis.* **2014**, *5*, e1257.
194. Martin, L. P.; Hamilton, T. C.; Schilder, R. J., *Clin. Cancer Res.* **2008**, *14*, 1291-1295.
195. Shen, D.-W.; Pouliot, L. M.; Hall, M. D.; Gottesman, M. M., *Pharmacol. Rev.* **2012**, *64*, 706-721.
196. Perez, R. P., *Eur. J. Cancer.* **1998**, *34*, 1535-1542.
197. Vasey, P. A., *Br. J. Cancer* **2003**, *89*, S23-S28.
198. Agarwal, R.; Kaye, S. B., *Nat. Rev. Cancer* **2003**, *3*, 502-516.
199. Shahzad, M. M. K.; Lopez-Berestein, G.; Sood, A. K., *Drug Resist Updat* **2009**, *12*, 148-152.

pathways in four main groups: *pre-trarget*, *on-target*, *post-target* and *off-target* resistance.[200] In this section, a brief summary of the four groups is going to be provided.

The first group of mechanisms is related with the processes that take place previous to the interaction of the drug with the DNA and are known as *pre-target resistance* mechanisms. Among them two main strategies have been developed by tumor cells, the reduction of intracellular accumulation of cisplatin and the increased sequestration of it by proteins with nucleophilic entities.

The decrease of the intracellular concentration of the drug is promoted by the reduction on the cisplatin uptake or by an increase of the efflux to the plasma. In order to decrease the uptake of the complexes, the mechanisms implicated in the transport of cisplatin inside the cells are altered; *i.e.*, CTR1 and OCT.[201-203] On the other hand, the augment of the export of cisplatin to plasma is mediated by Adenosine Triphosphatases (ATPases) and Multidrug Resistance Proteins (MRP), thus their expression levels are crucial to predict the sensitivity of cancers to cisplatin chemotherapy.[204-206] Besides, the binding of cisplatin with proteins that include nucleophilic species has also been described, resulting in an inactivation of the drug. In this way, the nucleophilic proteins act as cytoplasmic scavengers reducing the concentration of reactive cisplatin. GSH, methionine, metallothioneins and other proteins including a high concentration of

200. Galluzzi, L.; Senovilla, L.; Vitale, I.; Michels, J.; Martins, I.; Kepp, O.; Castedo, M.; Kroemer, G., *Oncogene* **2012**, *31*, 1869-1883.
201. Ishida, S.; Lee, J.; Thiele, D. J.; Herskowitz, I., *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14298-14302.
202. Song, I.-S.; Savaraj, N.; Siddik, Z. H.; Liu, P.; Wei, Y.; Wu, C. J.; Kuo, M. T., *Mol. Cancer Ther.* **2004**, *3*, 1543-1549.
203. Guttmann, S.; Chandhok, G.; Groba, S. R.; Niemietz, C.; Sauer, V.; Gomes, A.; Ciarimboli, G.; Karst, U.; Zibert, A.; Schmidt, H. H., *Oncotarget* **2017**, *9*, 743-754.
204. Nakayama, K.; Kanzaki, A.; Ogawa, K.; Miyazaki, K.; Neamati, N.; Takebayashi, Y., *Int. J. Cancer* **2002**, *101*, 488-495.
205. Aida, T.; Takebayashi, Y.; Shimizu, T.; Okamura, C.; Higasimoto, M.; Kanzaki, A.; Nakayama, K.; Terada, K.; Sugiyama, T.; Miyazaki, K.; Ito, K.; Takenoshita, S.; Yaegashi, N., *Gynecol. Oncol.* **2005**, *97*, 41-45.
206. Yamasaki, M.; Makino, T.; Masuzawa, T.; Kurokawa, Y.; Miyata, H.; Takiguchi, S.; Nakajima, K.; Fujiwara, Y.; Matsuura, N.; Mori, M.; Doki, Y., *Br. J. Cancer* **2011**, *104*, 707-713.

cysteine have been characterized for this kind of resistance mechanism (Figure 1.10).[207,208]

The second family of strategies developed by tumor cells are known as *on-target* resistance mechanisms and refer to the processes that happen once the cisplatin DNA adducts are formed. The cancer cells are able to avoid the apoptotic effect of the formed cisplatin-DNA binding by repairing the damaged adducts or by tolerating the unrepaired DNA lesions through the employment of a particular class of DNA polymerases, involved in the Translesion DNA synthesis (TLS).[209]

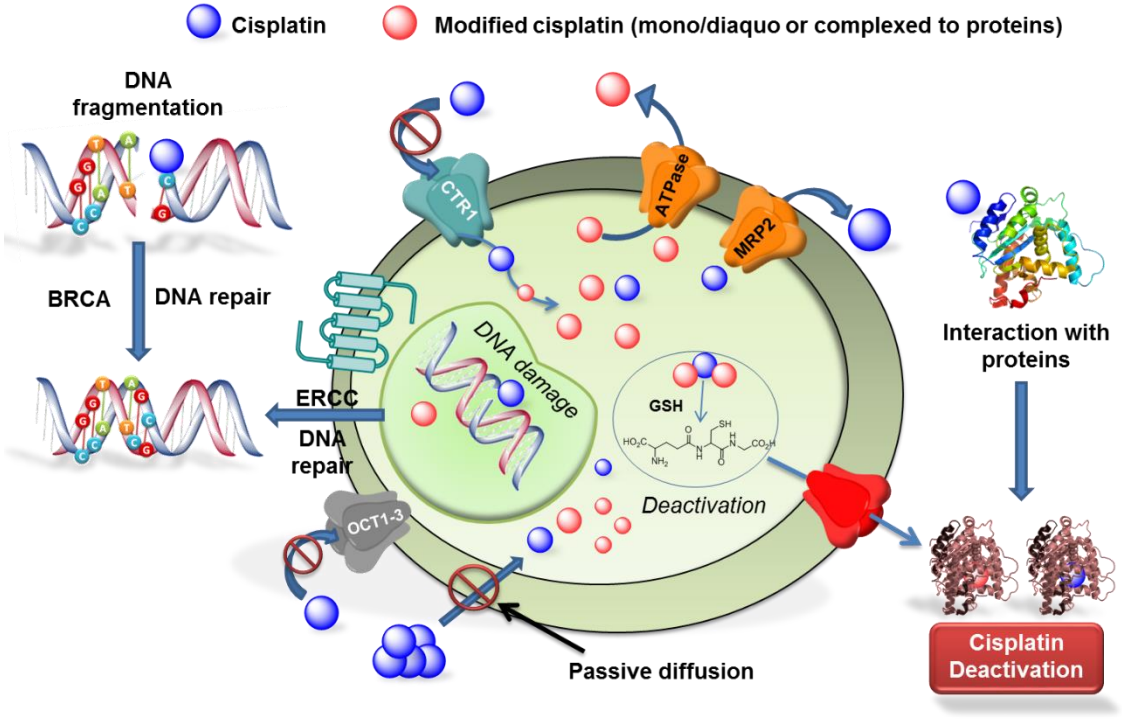


Figure 1.10. Summary of the main resistance mechanisms related to *pre-target* and *on-target* strategies.

Moreover, the increased concentration of Excision Repair Cross Complementing (ERCC), involved in the Nucleotide Excision Repair (NER) pathways, or Breast Cancer (BRCA) proteins, crucial in the Homologous Recombination (HR) pathways, would allow the repair of the DNA lesions, whereas the lower expression of the DNA Mismatch Repair

207. Casini, A.; Reedijk, J., *Chem. Sci.* **2012**, *3*, 3135-3144.
208. Zhang, C. X.; Chang, P. V.; Lippard, S. J., *J. Am. Chem. Soc.* **2004**, *126*, 6536-6537.
209. Shachar, S.; Ziv, O.; Avkin, S.; Adar, S.; Wittschieben, J.; Reißner, T.; Chaney, S.; Friedberg, E. C.; Wang, Z.; Carell, T.; Geacintov, N.; Livneh, Z., *EMBO J.* **2009**, *28*, 383-393.

(MMR) system is relevant for the suppression of the pro-apoptotic signals when DNA damages are detected (Figure 1.10).[194]

The third family of strategies are the *post-target* resistance processes related to the regulation of the cell death induced by the effect of cisplatin. Among these mechanisms, the alteration of the pro-apoptotic signaling pathways and the modification of the cell death executioner proteins are the most common strategies. For instance, the tumor-suppressor protein, p53, responsible for the control of DNA repair and apoptotic processes have been detected to be inactivated in cisplatin resistant cells.[210-212] Moreover, the functions of some other relevant proteins involved in the pro-apoptotic signaling cascades have been confirmed to be altered in those cells, *i.e.*, Caspases, Mitogen-Activated Protein Kinases (MAPKs) and B-cell Lymphona (BCL-2) proteins.[213-215]

The last family of strategies are the *off-target* resistance mechanisms, which are not directly engaged by cisplatin, however, they are able to reduce the lethal signals induced by cisplatin. One of the most studied mechanisms is the overexpression of the Human Epidermal Growth Factor Receptor (HER2), which is directly involved in the uncontrolled growth and division of the cells.[216,217]

In conclusion, numerous factors are involved in the development of platinum resistance mechanisms. Besides, the complete understanding of these strategies and

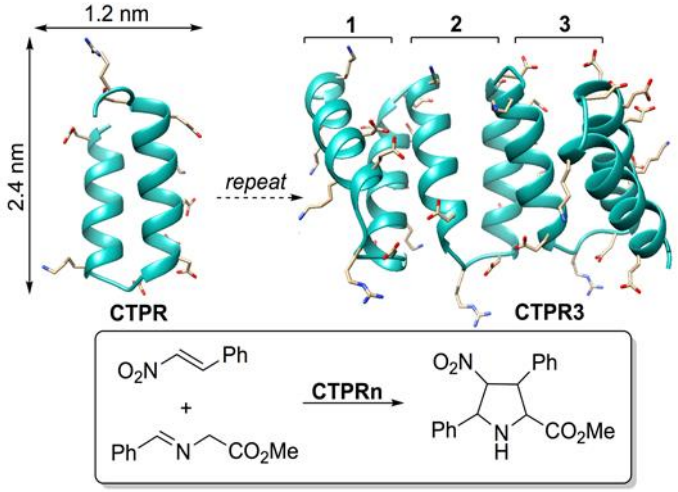
210. Perego, P.; Giarola, M.; Righetti, S. C.; Supino, R.; Caserini, C.; Delia, D.; Pierotti, M. A.; Miyashita, T.; Reed, J. C.; Zunino, F., *Cancer Res.* **1996**, *56*, 556-562.
211. Castedo, M.; Coquelle, A.; Vitale, I.; Vivet, S.; Mouhamad, S.; Viaud, S.; Zitvogel, L.; Kroemer, G., *Ann. N. Y. Acad. Sci.* **2006**, *1090*, 35-49.
212. Lin, X.; Howell, S. B., *Mol. Cancer Ther.* **2006**, *5*, 1239-1247.
213. Mueller, T.; Voigt, W.; Simon, H.; Fruehauf, A.; Bulankin, A.; Grothey, A.; Schmoll, H.-J., *Cancer Res.* **2003**, *63*, 513-521.
214. Cho, H. J.; Kim, J. K.; Kim, K. D.; Yoon, H. K.; Cho, M.-Y.; Park, Y. P.; Jeon, J. H.; Lee, E. S.; Byun, S.-S.; Lim, H. M.; Song, E. Y.; Lim, J.-S.; Yoon, D.-Y.; Lee, H. G.; Choe, Y.-K., *Cancer Lett.* **2006**, *237*, 56-66.
215. Brozovic, A.; Osmak, M., *Cancer Lett.* **2007**, *251*, 1-16.
216. Boone, J. J. M.; Bhosle, J.; Tilby, M. J.; Hartley, J. A.; Hochhauser, D., *Mol. Cancer Ther.* **2009**, *8*, 3015-3023.
217. Huang, D.; Duan, H.; Huang, H.; Tong, X.; Han, Y.; Ru, G.; Qu, L.; Shou, C.; Zhao, Z., *Sci. Rep.* **2016**, *6*, 20502.

the design of new drugs that are able to overcome the genesis of the resistance, are one of the major challenges of the anticancer therapy.

Chapter 2:

Discovering Biomolecules with *Huisgenase* Activity: Designed Repeat Proteins as Biocatalysts for (3+2) Cycloadditions

Abstract: Designed repeat proteins catalyze the 1,3-dipolar reaction between an imine, methyl N-benzylidene glycinate, and a π-deficient dipolarophile, (*E*)-β-nitrostyrene, in THF solution to form unnatural nitroproline esters, a reaction that no enzyme can catalyze. NMR studies and mutation experiments show that both acidic and basic residues can catalyze the reaction. The diastereocontrol of the reaction depends on the flexibility of the protein and on the number and location of the active lysine and glutamate residues, which can participate independently or forming dyads that promote the formation of unusual diastereomeric cycloadducts. QM/MM calculations permit to rationalize the origins of this *Huisgenase* activity and its diastereocontrol.



2.1 OBJECTIVES

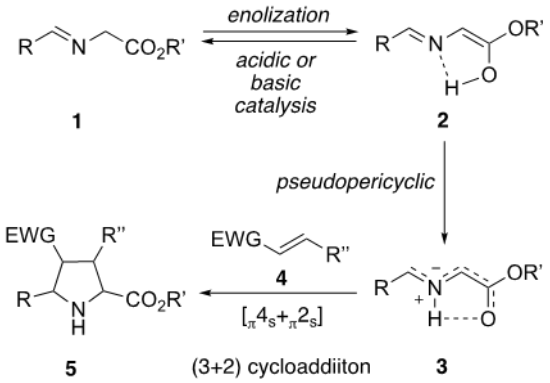
Previous to this work, proteins that catalyze the 1,3-dipolar reaction between imines and π-deficient dipolarophiles, the Huisgen[1] reaction, have not been described in living systems.[2] Taking into account the dilated experience of our group in (3+2) cycloadditions between azomethine ylides and nitrostyrenes,[3-6] we aim to describe the first *Huisgenase*. Therefore, in this Chapter, the consensus repeat proteins, which contain acidic and basic residues that could act as catalytic centers to perform these pericyclic reactions, would be evaluated as suitable metal-free biocatalysts. Several mutations are going to be performed to test the diastereoselectivity of these specific proteins. Moreover, QM/MM computational methods and NMR characterization will be employed to gain a better understanding of the catalytic activity of the CTPR proteins.

2.2 CATALYTIC DESIGN AND STRUCTURAL CONSIDERATIONS OF CTPR PROTEINS

2.2.1 CTPR proteins as suitable biocatalysts for (3+2) cycloadditions

According to a previous work,[4] the acidic or base catalysis is essential to promote (3+2) cycloadditions between imines **1** and π-deficient nitroalkenes **4** (Scheme 2.1). In a first step, the acidic or basic catalysts promote the enolization process that leads to the formation of the 1,3-dipoles **2** from the corresponding starting imines. Subsequently, via a pseudopericyclic process, the *NH*-azomethine ylides **3** are formed *in situ*.[5] The high reactivity of these intermediates produces the corresponding proline ester cycloadducts **5** via a concerted, although quite asynchronous [$_{\pi}4_{s}+_{\pi}2_{s}$] symmetry allowed mechanism (Scheme 2.1).

1. Huisgen, R., 1,3-Dipolar Cycloaddition Chemistry. Wiley, Ed. New York, 1984; Vol. 1, pp 1-176.
2. Baunach, M.; Hertweck, C., *Angew. Chem., Int. Ed.* **2015**, *54*, 12550-12552.
3. de Cózar, A.; Cossío, F. P., *Phys. Chem. Chem. Phys.* **2011**, *13*, 10858-10868.
4. Arrieta, A.; Otaegui, D.; Zubia, A.; Cossío, F. P.; Díaz-Ortiz, A.; de la Hoz, A.; Herrero, M. A.; Prieto, P.; Foces-Foces, C.; Pizarro, J. L.; Arriortua, M. I., *J. Org. Chem.* **2007**, *72*, 4313-4322.
5. Costa, P. R. R.; Sansano, J. M.; Cossío, U.; Barcellos, J. C. F.; Dias, A. G.; Nájera, C.; Arrieta, A.; de Cózar, A.; Cossío, F. P., *Eur. J. Org. Chem.* **2015**, 4689-4698.
6. Schleyer, P. v. R.; Wu, J. I.; Cossío, F. P.; Fernández, I., *Chem. Soc. Rev.* **2014**, *43*, 4909-4921.



Scheme 2.1. A general mechanism for (3+2) cycloadditions between *in situ* formed *NH*-azomethine ylides **3** and nitroalkenes **4** under acidic or basic catalysis to form unnatural proline esters **5**.

Consequently, in our pursuit for a biological candidate that potentially could catalyze the Huisgen reaction, we focus our search in available proteins that contain various acidic (aspartic acid and glutamic acid) and basic residues (arginine, histidine and lysine). Among all the possible candidates to become novel *Huisgenases,* proteins containing the Tetratricopeptide Repeat (TPR)[7,8] module were selected. This specific motif consists on a helix-turn-helix sequence of 34 amino acids that can be found in over 300 different proteins, moreover, this kind of repetitions have been associated with a wide range of biological functions.[9,10,11]

The alignment of individual TPR sequences allows the design of "*Consensus TPR*" (CTPR) sequences.[12] This design enables the construction of proteins with multiple repeated TPR motifs. Furthermore, their characteristic helix-turn-helix fold is well conserved when the number of repeats is modified. A noteworthy feature of the CTPR proteins is the high stability of their scaffold, which increases proportionally with the

7. Blatch, G. L.; Lassle, M., *Bioessays* **1999**, *21*, 932-939.
8. D'Andrea, L. D.; Regan, L., *Trends Biochem. Sci.* **2003**, *28*, 655-662.
9. Smith, D. F., *Cell Stress Chaperones* **2004**, *9*, 109-121.
10. Roberts, J. D.; Thapaliya, A.; Martínez-Lumbreras, S.; Krysztofinska, E. M.; Isaacson, R. L., *Front. Mol. Biosci* **2015**, *2*, 71.
11. Zhou, X.; Liao, H.; Chern, M.; Yin, J.; Chen, Y.; Wang, J.; Zhu, X.; Chen, Z.; Yuan, C.; Zhao, W.; Wang, J.; Li, W.; He, M.; Ma, B.; Wang, J.; Qin, P.; Chen, W.; Wang, Y.; Liu, J.; Qian, Y.; Wang, W.; Wu, X.; Li, P.; Zhu, L.; Li, S.; Ronald, P. C.; Chen, X., *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 3174-3179.
12. Main, E. R. G.; Xiong, Y.; Cocco, M. J.; D'Andrea, L.; Regan, L., *Structure* **2003**, *11*, 497-508.

number of tandem repeats.[13-15] Accordingly, these proteins provide robust hyper-stable structures that make possible to work in a wide range of temperatures and solvents.[16] Taking into account both requirements, the presence of acidic and basic residues that could act as catalytic centers and the robust properties of CTPR-based proteins, we considered them a good starting point for the screen and design of the first *Huisgenase*.

The evaluation of the catalytic activity of CTPR proteins was made with two different series of closely related proteins, namely CTPRa[13] and CTPR.[12] These two series, differ only in two residues located at the loop at the end of the repeat, where the final four residues change from –DPRS (CTPRa) to –DPNN (CTPR) (Figure 2.1).[14] These two series have been designed in models of one and three repeats; *i.e.* CTPR1, CTPR1a, CTPR3 and CTPR3a proteins. One of the most striking features of the CTPR sequences is the presence of diverse acidic (7 in CTPRa and CTPR series) and basic centers (3 in CTPRa and 2 in CTPR series) per repeat, which potentially can act as catalytic sites. Besides, the analysis of the crystal structure of the CTPR and CTPRa proteins shows a potential catalytic dyad, in which the acid-base amino acid pair formed by the E22 and K26 residues (Figure 2.1) could act cooperatively to perform the Huisgen reaction. Moreover, the sequences of the two series present glutamate (E), lysine (K), aspartate (D) and arginine (R) residues that lie away to each other and can behave as acidic and basic catalytic units or monads, *e.g.* see residues E2, K13, D31 and R33 in Figure 2.1B.

13. Kajander, T.; Cortajarena, A. L.; Main, E. R. G.; Mochrie, S. G. J.; Regan, L., *J. Am. Chem. Soc.* **2005**, *127*, 10188-10190.
14. Kajander, T.; Cortajarena, A. L.; Mochrie, S.; Regan, L., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 800-811.
15. Cortajarena, A. L.; Regan, L., *Protein Sci.* **2011**, *20*, 336-340.
16. Cortajarena, A. L.; Mochrie, S. G. J.; Regan, L., *Protein Sci.* **2011**, *20*, 1042-1047.

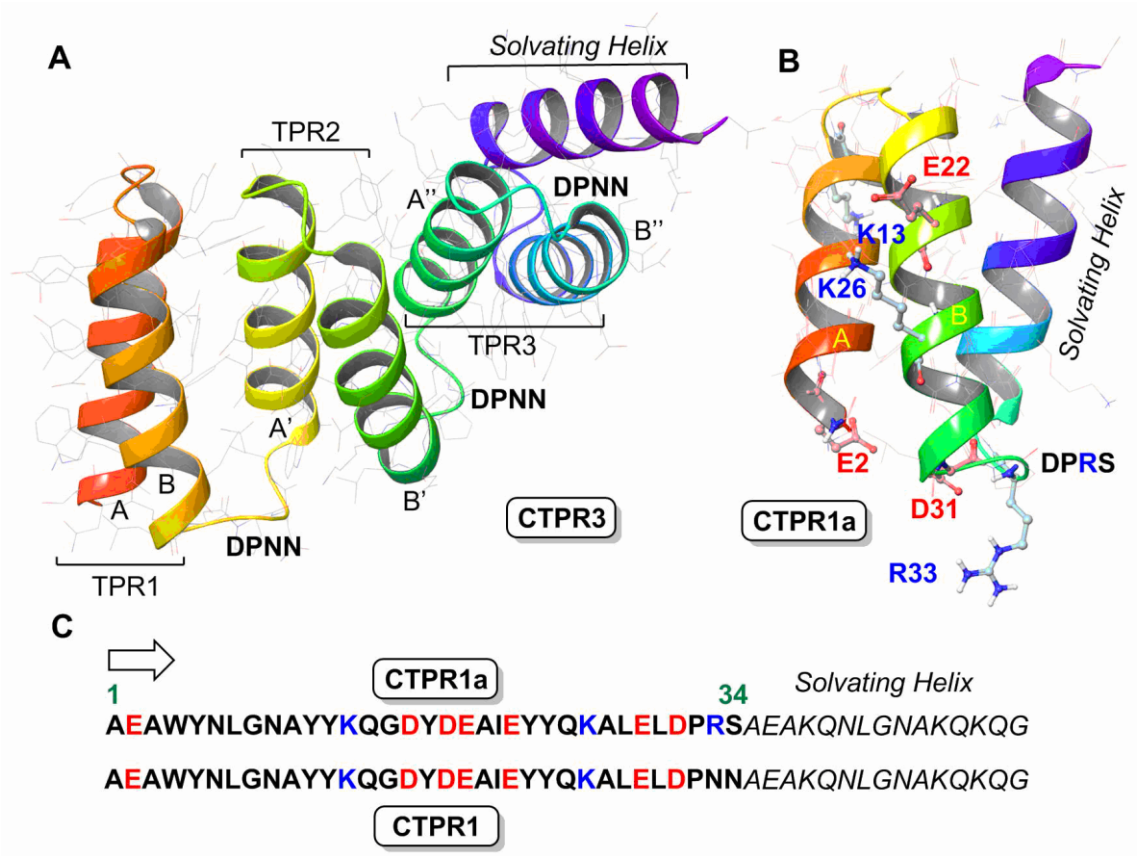


Figure 2.1. A) Model structure (pdb code: 1NA0) of CTPR3 protein showing three helical repeats (AB, TPR1-3) and one solvating helix (SHx). **B**) X-ray structure of CTPR1a protein. Acidic residues E2, E22 and D31 are highlighted in red. Basic residues K13, K26 and R33 are highlighted in blue. Structure generated from the X-ray diffraction analysis data performed on the supersymmetric homologue CTPR8a (pdb code: 2HYZ). **C**) Sequences for the one repeat CTPR1 and CTPR1a sequences including the solvating helix. The acidic and basic residues are highlighted in red and blue, respectively.

Within this context, in order to explore the potential of the promising CTPR proteins as biocatalysts of the reaction between azomethine ylides and 1,3-dipolarophiles, the aforementioned wild type (wt) CTPR1a, CTPR1, CPTR3a and CTPR3 were tested in (3+2) cycloadditions. In addition, to evaluate the catalytic role of the acidic and basic centers, mutants involving the substitution of these amino acids were developed. The design and synthesis of the CTPR proteins presented in this thesis was carried in the group of Prof. Aitziber López Cortajarena at CIC biomaGUNE. Some of the most relevant features of the designed proteins are described in the following section,

while the remaining information is reported in the experimental section and the Annex III.

2.2.2 Characterization of the CTPR variants and their stability under reaction conditions

With the purpose of evaluating the possible catalytic activity of the acidic and basic centers of the wt CTPR proteins, numerous mutations were performed (Table 2.1). Firstly, it is convenient to remind the difference in the sequence of CTPRa and CTPR series, where the –DPRS ending of CTPRa is converted into a DPNN sequence, thus substituting the potential active R33 basic residue for a N33. In the CTPR1a series, the acidic E2, E22 and D31 were mutated to alanine, whereas the basic center K26 was also modified to glutamine. Initially, we aim to introduce an *Ala* in all the mutants, although, some *Gln* were finally included to evaluate the relevance of a larger residue. On the other hand, for the CTPR1 series, all the present basic residues were mutated considering single mutations (K13A, K26Q and K26A) and a double mutation (K13A+K26A), which eliminates all the basic residues of the repeat. Furthermore, a mutant without basic residues was generated for the CTPR3 series, which combines the elimination of solvating helix with the K13A and K26A mutations.

Since small amounts of contaminant proteins might lead to catalytic activities not related to the designed CTPR variants, all the proteins mentioned above were purified and their identity and purity was confirmed by mass spectrometry. As shown in Table 2.1, all the proteins present a mass that matches their theoretical MW on the protein sequence. More details about the purification, the identification and the mass spectrometry spectra are explained in the experimental section and the Annex III.

Table 2.1. Mass spectrometry analysis of the purified CTPR variants and its mutations.

Protein	Calculated MW (Da)	Observed MW[a] (Da)
CTPR1a_wt	6018.54	6027.66
CTPR1a_E2A	5960.50	5966.27
CTPR1a_E22A	5960.50	5966.63
CTPR1a_K26Q	6018.84	6025.89
CTPR1a_D31A	5974.53	5979.38
CTPR1_wt	6329.79	6332.05
CTPR1_K13A	6272.69	6274.20
CTPR1_K26Q	6329.75	6339.79
CTPR1_K26A	6272.69	6275.65
CTPR1_K13A+K26A	6215.60	6215.65
CTPR3a_wt	14079.50	14081.10
CTPR3_wt	14360.34	14362.36
CTPR3_K13A+K26A-SHx[b]	11979.50	11536.29

[a]Determined by mass spectrometry. [b]-SHx denotes the absence of solvating helix. The four wild type (wt) series (CTPR1a, CTPR1, CTPR3a and CTPR3) are depicted in bold.

An important point to consider when mutations are performed, is the stability of the helical secondary structure of the proteins. In order to confirm that the α-helical domains were retained, the circular dichroism (CD) analysis was carried out (Figure 2.2A). Additionally, the evaluation of the conformational stability of the CTPR proteins in non-aqueous solvents, such as THF, was evaluated. The catalytic studies of the (3+2) cycloaddition catalyzed by the CTPRs were planned to be done in THF, due to the low solubility of the imine reactants in aqueous media, and therefore CD was accomplished for CTPR3 and CTPR3a proteins in this solvent. Upon dilution of CTPR protein from aqueous media into 95 % THF, it was observed a slight decrease in the CD signal at 222 nm that relies on the α-helical content of the proteins. After 24h under those conditions, the 80% of the CD signal was conserved, confirming that the α-helical secondary structure of both series was retained (Figure 2.2B). Furthermore, the resuspension of the sample after 24 hours results in the recovery of the CD signal, indicating that the proteins do not denature but slightly lose solubility under reaction conditions. Therefore, the analysis of the protein stability concludes that the (3+2) cycloadditions

could be carried in THF due to the preservation of the secondary structure of CTPR proteins under reactions conditions.

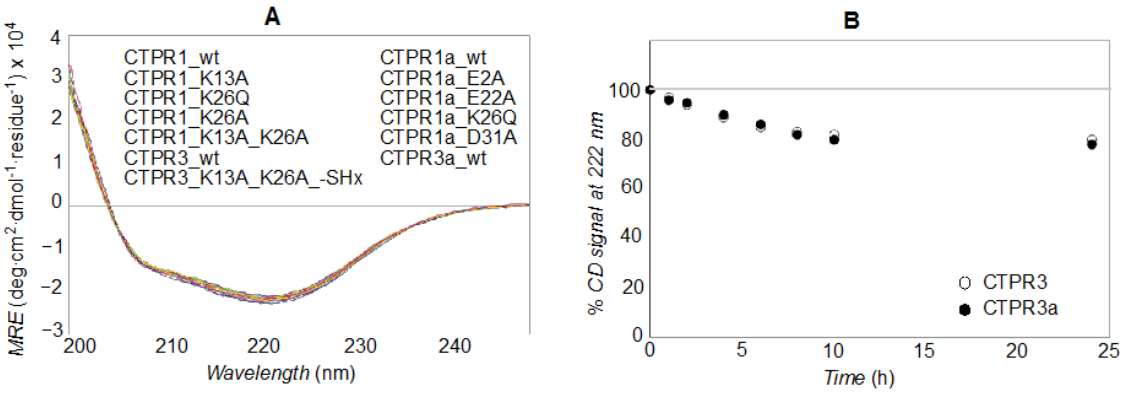
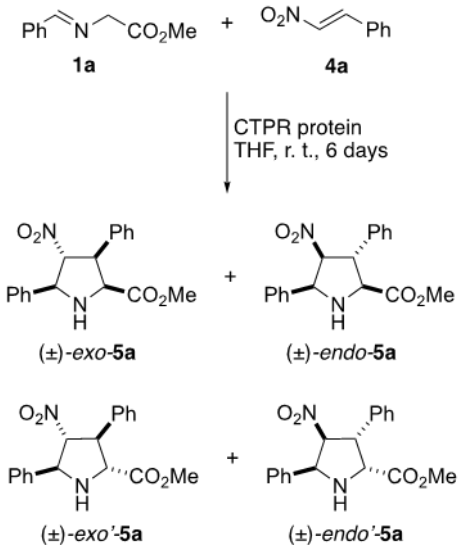


Figure 2.2. A) Circular dichroism spectra of the designed CTPR protein variants. MRE: Molar Residue Ellipticity. –SHx denotes the absence of the solvating helix. **B**) Stability of CTPR3a and CTPR3 proteins under reaction conditions. The percentage of CD signal at 222 nm that reports on the helical content of the proteins was monitored over time upon dilution of CTPR proteins in 95 % THF.

2.3 CATALYTIC *HUISGENASE* ACTIVITY OF CTPR PROTEINS

Once the rational analyses of the stability and the structure of the CTPR proteins were conducted, the following step was to test their potential *Huisgenase* activity. The (3+2) cycloadditions between stabilized azomethine ylides and dipolarophiles have been well studied both experimentally and computationally.[3] The employment of different chiral catalysts has resulted in achieving the formation of *endo* and *exo* cycloadducts.[3] Moreover, in a previous work,[4] the free thermal and microwave-assisted (3+2) cycloadditions between stabilized azomethine ylides and nitrostyrenes have been studied. When these reactions were carried out without solvent, both in thermal or microwave conditions, a mixture of *exo*, *endo* and *endo*' was obtained. Considering all these results, the reaction between methyl N-benzylidene **1a** and (*E*)-β-nitrostyrene **4a**, was selected to evaluate the catalytic activity of the designed CTPR proteins. The catalysis was performed at room temperature (r.t) for six days in THF, where the stability of the proteins was confirmed, employing a rotary mixer (Scheme 2.2).



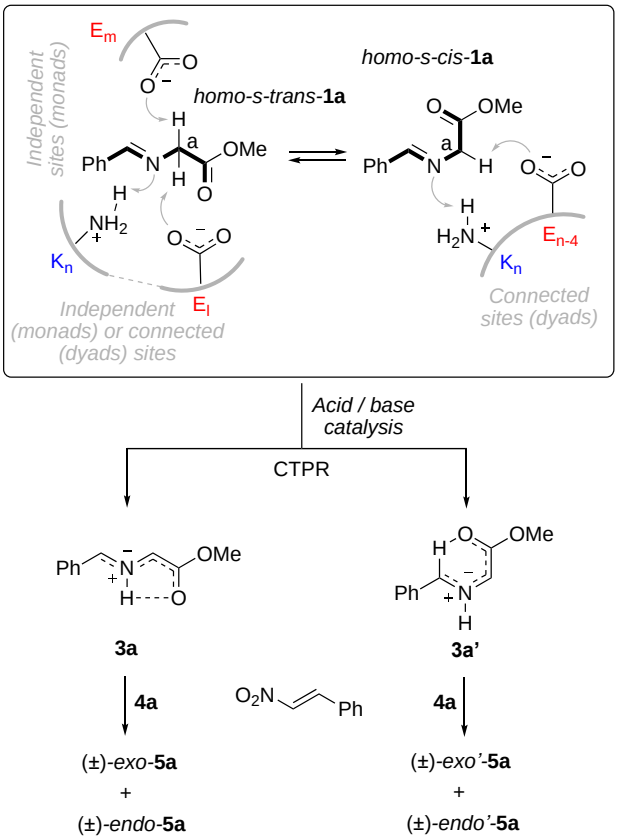
Scheme 2.2. (3+2) cycloaddition between *N*-benzylidene-glycinate **1a** and β-nitrostyrene **4a** catalyzed by CTPR proteins. The four possible racemic pairs of cycloadducts **5a** are indicated.

Firstly, the wt series were tested, resulting in a clear difference between CTPRa and CTPR series. CTPR1a_wt and CTPR3a_wt, catalyze the formation of the four possible pyrrolidine stereoisomers, *exo*-**5a**/*endo*-**5a**/*exo'*-**5a**/*endo'*-**5a**, in a virtually equimolar ratio with a global yield of 40% (see the experimental section and the Annex III for the structural characterization of the four diasteroisomeric cycloadducts and for the analysis of the reaction crude of the catalysis performed by CTPR3a_wt). Accordingly, the wild type series of CTPRa proteins were able to produce the *exo'*-**5a** and *endo'*-**5a** which are unusually formed in this kind of (3+2) cycloadditions. Conversely, CTPR1_wt and CTPR3_wt catalyze only the formation of *exo*-**5a** and *endo*-**5a** cycloadducts in an equimolar ratio for the two stereoisomers and a global yield of 20% (Table 2.2).

Taking into account these results, we hypothesized that the formation of the four cycloadducts observed for the CTPRa_wt catalysts is a consequence of the *in situ* generation of the two steroisomeric azomethine ylides **3a** and **3a**'. The rotation through the N-C_α single bond of the imine **1a**, results in an equilibrium between the *homo-s-cis* and *homo-s-trans* conformations (Scheme 2.3). One or two isolated acidic or basic residues could act as monads to promote a proton transfer/abstraction of one C_α-H bond of **1a** to generate the 1,3-dipoles **3a** and **3a'** in two independent processes. In addition, the two stereoisomeric azomethine ylides can be formed in alternative pathway, see

scheme 2.3, in which a K_n/E_n-4 catalytic dyad is the responsible to conduct the transfer/abstraction processes.

On the basis of the mechanism proposed in Scheme 2.3, there are two plausible and different routes that lead to the formation of *endo*-**5a**/*exo*-**5a** and/or *endo'*-**5a**/*exo'*-**5a** (3+2) cycloaddition products. On one hand, the reaction between the 1,3-dipole **3a** with the dipolarophile **4a** would produce the *endo*-**5a** and *exo*-**5a** cycloadducts via $[\pi 4_s+\pi 2_s]$ concerted, but most likely asynchronous mechanisms. On the other hand, the doubly suprafacial reaction of the azomethine ylide **3a'** with the β-nitrostyrene **4a**, results in the unnatural proline esters *endo'*-**5a** and *exo'*-**5a**. By these means, the four diastereoisomers can be theoretically obtained.



Scheme 2.3. Postulated formation of azomethine ylides **3a** and **3a'** from imine **1a** via acid/base catalysis from independent or connected active sites of CTPR proteins. 1,3-dipolar reactions of both azomethine ylides yield the *endo*, *exo*, *endo'* and *exo'* cycloadducts **5a** via $[\pi 4_s+\pi 2_s]$ mechanisms.

After the analysis of the wt series of CTPR proteins, all the generated mutants were tested in order to evaluate the initial hypothesis and to study the roles of the different acidic and basic centers. Every performed mutation result to inhibit the formation of the *exo'*-**5a** and *endo'*-**5a** (3+2) cycloadducts, promoting only the generation of *exo*-**5a** and *endo*-**5a** in an equimolar ratio. Moreover, the different mutations of the CTPR1a_wt, CTPR1_wt and the CTPR3_wt result in a considerable decrease on the chemical yields and the corresponding TON (Turnover number) and TOF (Turnover frequency) values see Table 2.2.

When acidic residues (E2, E22 and D31) were mutated in CTPRa series, the global yields were reduced close to 30 %, TON values close to 10^4 and TOF values of ca. $70h^{-1}$. On the contrary, the elimination of basic residues, K13 and K26, in both CTPR and CTPRa series, resulted in lower TON and TOF numbers and yields between 11-24%. Interestingly, the double mutation of K13 and K26 amino acids resulted in the lowest values in CTPR1 (11%) and CTPR3 (18%) series. These results suggest that lysine (K), basic residues, are more active in terms of *Huisgenase* activity than aspartate (D) and glutamate (E) residues. Another relevant feature of the catalysis with the mutated CTPR protein series was that any single or double mutation was unable to completely inhibit the catalytic activity of the protein. Therefore, this characteristic suggests that arginine (R), aspartate, glutamate and lysine residues are able to elicit independently the *Huisgenase* activity that leads to the final *exo*-**5a** and *endo*-**5a** products.

Table 2.2. Cycloaddition reactions between imine **1a** and β-nitrostyrene **4a** in the presence of proteins CTPR1a, CTPR1, CTPR3a and CTPR3 with diverse mutations.

Protein Mutation(s)	Yield[a] mmol ± 0.1·10^{-2} [%]	Stereoisomeric Ratio[b] Yield (mmol·10^{-2}) [%]				TON[c] x10^3	TOF[d] (h^{-1})
		endo-**5a**	*exo*-**5a**	*endo'*-**5a**	*exo'*-**5a**		
CTPR1a_wt[e]	**12.7 [41]**	**2.5 ± 0.2 [20]**	**3.6 ± 0.4 [28]**	**3.0 ± 0.3 [24]**	**3.6 ± 0.4 [28]**	**14.2**	**98.8**
CTPR1a_E2A	8.9 [29]	4.4 ± 0.6 [49]	4.6 ± 0.7 [51]	--	--	10.1	70.2
CTPR1a_E22A	9.6 [31]	4.3 ± 0.6 [45]	5.3 ± 0.9 [55]	--	--	10.9	75.4
CTPR1a_K26Q	7.4 [24]	3.3 ± 0.4 [44]	4.1 ± 0.6 [56]	--	--	8.4	58.2
CTPR1a_D31A	9.6 [31]	4.5 ± 0.7 [47]	5.1 ± 0.9 [53]	--	--	10.5	72.9
CTPR1_wt[e]	**6.2 [20]**	**2.6 ± 0.2 [41]**	**3.6 ± 0.5 [59]**	**--**	**--**	**6.8**	**47.1**
CTPR1_K26Q	5.5 [18]	2.5 ± 0.2 [45]	3.1 ± 0.3 [55]	--	--	6.3	43.5
CTPR1_K13A	4.0 [13]	1.7 ± 0.1 [42]	2.3 ± 0.2 [58]	--	--	4.4	30.3
CTPR1_K26A	4.3 [14]	1.9 ± 0.1 [44]	2.4 ± 0.2 [56]	--	--	4.9	33.8
CTPR1+K13A +K26A	3.4 [11]	1.5 ± 0.1 [43]	1.9 ± 0.1 [57]	--	--	3.9	27.0
CTPR3a_wt[e]	**12.4 [40]**	**3.1 ± 0.2 [25]**	**3.4 ± 0.4 [27]**	**3.1 ± 0.4 [25]**	**2.8 ± 0.3 [23]**	**13.6**	**94.8**
CTPR3_wt[e]	**7.4 [24]**	**3.8 ± 0.5 [51]**	**3.6 ± 0.5 [49]**	**--**	**--**	**8.4**	**58.6**
CTPR3+K13A +K26A-SHx[f]	5.6 [18]	2.7 ± 0.3 [48]	2.9 ± 0.3 [52]	--	--	6.1	42.6

[a]Combined yield of isolated cycloadducts. [b]Determined by column purification correspond to average values after three experiments. Errors were calculated from standard deviations. The limiting reactant was **1a** (0.31 mmol). [c]TON: turnover number. [d]TOF: turnover frequency. [e]wt: wild type protein without mutations. [f]-SHx: without the solvating helix. Data for wild type CTPR protein series (CTPR1a_wt, CTPR3a_wt, CTPR1_wt and CTPR3_wt) is represented in **bold**.[17]

17. Harris, D. C., *Quantitative Chemical Analysis, 8th ed*. W.H. Freeman: New York, 2010.

In order to assess the cooperative effect of acidic and basic residues on the *Huisgenase* activity of CTPRs and their mutants, we studied the reaction between **1a** and **4a** under the same reaction conditions and in the presence of amino acid derivatives or other enzymes (Figure 2.3).

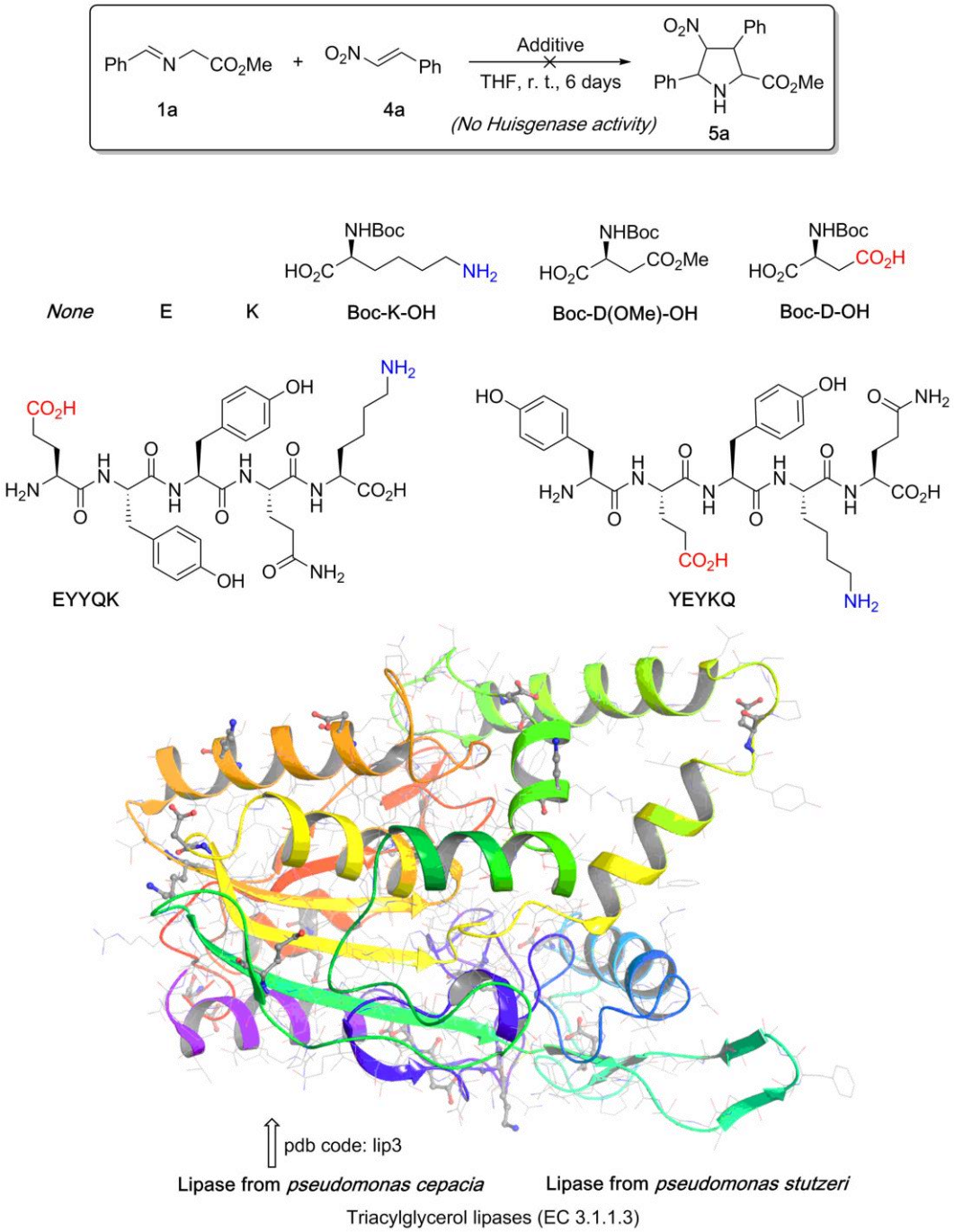


Figure 2.3. Amino acid derivatives, pentapeptides and triacylglycerol lipases tested in the (3+2) reaction between imine **1a** and nitrostyrene **4a**. Acidic and basic residues are highlighted in red and blue, respectively. The X-ray structure of the lipase from *Pseudomonas cepacia* is also provided. Aspartate, glutamate and lysine residues are highlighted in ball & stick representation.

In particular, lysine (K), glutamic acid (E), NH-Boc derivatives Boc-K-OH, Boc-D(OMe)-OH and Boc-D-OH were tested. Moreover, pentapeptides EYYQK and YEYKQ were also studied. The former pentapeptide possesses the same sequence present in the critical E22-Y23-Y24-Q25-K26 part of CTPR1a, whereas the second incorporates the E and K residues in a different order. Finally, we also studied two triacylglycerol lipases from *Pseudomonas cepacia* and *Pseudomonas stutzeri*, which incorporate α-helices with acidic and basic residues (Figure 2.3). Moreover, lipase TL from *Pseudomonas stutzeri* has been shown to be active in polar solvents, such as THF.[18,20] None of in THF under different concentrations (see experimental section) permitted to observe any noticeable *Huisgenase* activity. These results demonstrate that the whole structure of CTPR proteins is responsible for the observed catalytic activity, since isolated acidic or basic residues cannot, by themselves, promote the formation of cycloadducts **5a**.

As it is shown in Table 2.2, the formation of *endo'*-**5a** and *exo'*-**5a** was only achieved when the catalysis was carried out by the CTPRa series; CTPR1a and CTPR3a. One of the biggest differences between the CTPR and CTPRa series of proteins is the major stability, and therefore, smaller flexibility of the CTPR proteins in comparison with the CTPRa ones. In order to evaluate the stability of CTPR proteins, two main parameters need to be considered; the intrinsic stability of the repeated units (H) and the coupling between repeats (J). These descriptors can be calculated by the fit of the denaturation curves of both series to simple Ising models.[13] The evaluation of these parameters results in bigger stability for the CTPR repeat unit with regard to the CTPRa repeat unit, H values 4.95 and 3.66 kcal/mol, respectively. Additionally, thermal denaturation analyses were carried out to compare the stability between these two CTPR series.[12] The published results indicate that the CTPR3 protein presents a thermal denaturation midpoint 14.7˚C higher than the one for CTPR3a, 83.0˚C and 68.3˚C, respectively. All

18. Aoyagi, Y.; Agata, N.; Shibata, N.; Horiguchi, M.; Williams, R. M., *Tetrahedron Lett.* **2000**, *41*, 10159-10162.
19. Aoyagi, Y.; Saitoh, Y.; Ueno, T.; Horiguchi, M.; Takeya, K.; Williams, R. M., *J. Org. Chem.* **2003**, *68*, 6899-6904.
20. Hoyos, P.; Fernández, M.; Sinisterra, J. V.; Alcántara, A. R., *J. Org. Chem.* **2006**, *71*, 7632-7637.

these data confirm a clear destabilization of CTPRa vs CPTR series when proteins of the same size are evaluated.

The availability of the crystal structures for CTPR and CTPRa proteins, allows the employment of B factors[21] to assess the flexibility in the different regions of the proteins. Interestingly, considering the essential role in the formation of the *endo'*-**5a** and *exo'*-**5a** adducts, the flexibility of the catalytic dyad formed by the E22-K26 amino acid pair can be compared. The distribution of B-factors presented in Figure 2.4 for CTPR3 and CTPR3a shows overall higher values for CPTR3, indicating its higher flexibility. Furthermore, the side chains of the residues E22 and K26 also shows higher B-factor values for the CTPR3a in comparison with the CTPR3 protein.

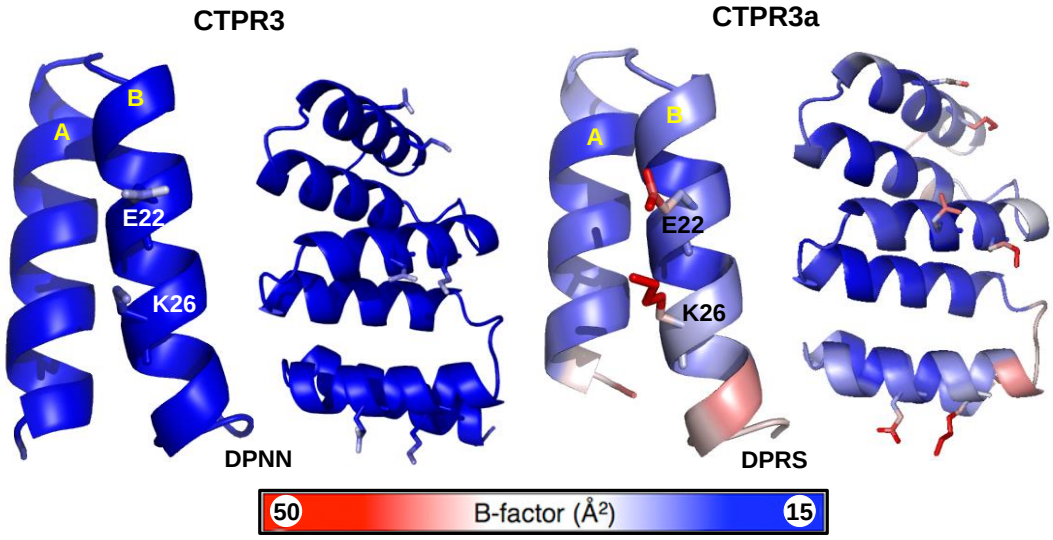


Figure 2.4. Structures of CTPR (PDB ID: 1NA0) and CTPRa (PDB ID: 2HYZ) showing on the left one repeat and on the right three repeat structures colored by their B-factors. Regions of lowest and highest B-factors from 15 to 50 Å² are shown in dark blue and red, respectively. The side chains of residues E22 and K26 are highlighted as sticks.

Since the side chains of the dyad pair in the CTPR3a protein show an increased flexibility, this could be crucial to access the conformations required to interact with imine **1** and, in turn, to generate different geometries for the 1,3-dipoles **3**. Contrarily, the less flexible side chains of CTPR3 protein would difficult the accessibility to the conformations necessary to perform the (3+2) cycloaddition reaction. In the following

21. Sun, Z. T.; Liu, Q.; Qu, G.; Feng, Y.; Reetz, M. T., *Chem. Rev.* **2019**, *119*, 1626-1665.

pages, all these hypotheses were tested by means of QM/MM simulations and NMR studies.

2.4 QM/MM STUDY OF THE *HUISGENASE* ACTIVITY OF CTPRa PROTEIN

In the previous section, based on the experimental results, two main routes to obtain the two reactive azomethine ylides **3a** and **3a'** were proposed. On one hand, the E_{n-4}/K_n amino acids present in the CTPR repeats would work cooperatively to catalyze the formation of the reactive 1,3-dipoles **3a** and **3a'** via a dyad pathway. On the other hand, isolated acidic or basic residues could also act as individual catalytic centers or monads to form the reactive azomethine ylides **3a**. Afterwards, the (3+2) cycloaddition with the dipolarophile **4a** should be carried out to yield the final 4-nitropolines esters **5a**.

With the purpose to evaluate the proposed reaction mechanisms, see Scheme 2.3, a QM/MM computational study of the reaction was performed. The (3+2) cycloaddition reaction was analyzed for the modeled structure CTPR1a using the ONIOM methodology at the M06-2X/6-311+G(2d,2p):dreiding//B3LYP/6-31G(d,p):dreiding level of theory (see Computational Methods section for further details). The QM part was formed by the imine **1a**, the nitroalkene **4a** and their corresponding derivatives, as well as by the key residues and their immediate neighbors, the remaining protein was included in the MM region.

2.4.1 Computational study of the reaction mechanism for the formation of azomethine ylides 3a and 3a' carried out by the catalytic dyad E22-K26

In our initial hypothesis, the main alternative to obtain the four possible stereoisomers of the (3+2) cycloaddition was the catalytic activity lead by the E22-K26 dyad. This specific active site was modelled defining a QM region where the two essential amino acids, E22 and K26, and the three connecting residues, Y23, Y24, Q25, were included (See Figures 2.5 and 2.6). The rest of the protein was included in the MM region.

As mentioned in the Scheme 3, the imine **1a** can adopt the *homo-s-trans* and *homo-s-cis* conformations, via rotation of the N-C_α single bond angle, which are the precursors of the azomethine ylides **3a** and **3a'**, respectively. Firstly, the imine **1a**, in the *homo-s-trans* conformation, approaches the catalytic center to form a pre-organized reactive complex, **RCa**. According to our calculations, **RCa** shows a hydrogen bond (1.50 Å) interaction between one hydrogen of the amino group in the K26 and an oxygen of the carboxylate group of E22. Additionally, these two residues interact directly with the imine **1a** via two weak interactions. On one hand, another hydrogen of the amino group of K26 residue interacts with the sp^2 hybridized oxygen atom of the ester group (2.02 Å), while the remaining oxygen of the E22 residue interacts with one hydrogen atom of the methylene group of **1a**. **RCa** is considered to be the initial point or the zero of the global reaction for the *homo-s-trans* conformation.

Once the initial complex is formed, the imine **1a** is transformed into the enolate **INTa** via the transition structure **TS1a** (Figure 2.5). This first transition step consists on a hydrogen abstraction process of the C_α atom by the acidic E22 residue, which contains a calculated activation barrier of 13.1 kcal/mol in terms of Gibbs free energy. Additionally, considering the distances involved in this hydrogen transfer (C----H--O, 1.47 Å and 1.19 Å, respectively), we can conclude that this endergonic process is associated with a late transition structure. Interestingly, the interactions that are presented in the enolate species **INTa** are considerably different to the ones in the **RCa**. Even though E22 and K26 residues continue interacting via a hydrogen bond, the distance between both has increased considerably from 1.50 to 1.86 Å. Moreover, the sp^2 hybridized oxygen atom of the ester group is now placed nearer to the amino group, to a distance of 1.49 Å, which facilitates the next step of the reaction.

Subsequently, this favorable disposition founded in **INTa** leads the enolization process to yield the enol **2a**. The process occurs via an essentially barrier-less step in which the saddle point **TS2a** corresponds to the proton transfer from the quaternized amino group of K26 to the enolate **INTa**. In this enol structure, the interaction between both residues is maintained (2.11 Å) and the interaction among the enol and the protein is conserved via a hydrogen bond of 1.52 Å. In a subsequent step, the reactive

azomethine ylide **3a**, which is the precursor of *endo*-**5a** and *exo*-**5a** cycloadducts, would be formed after a pseudopericyclic prototropy similar to that shown in Scheme 2.1.

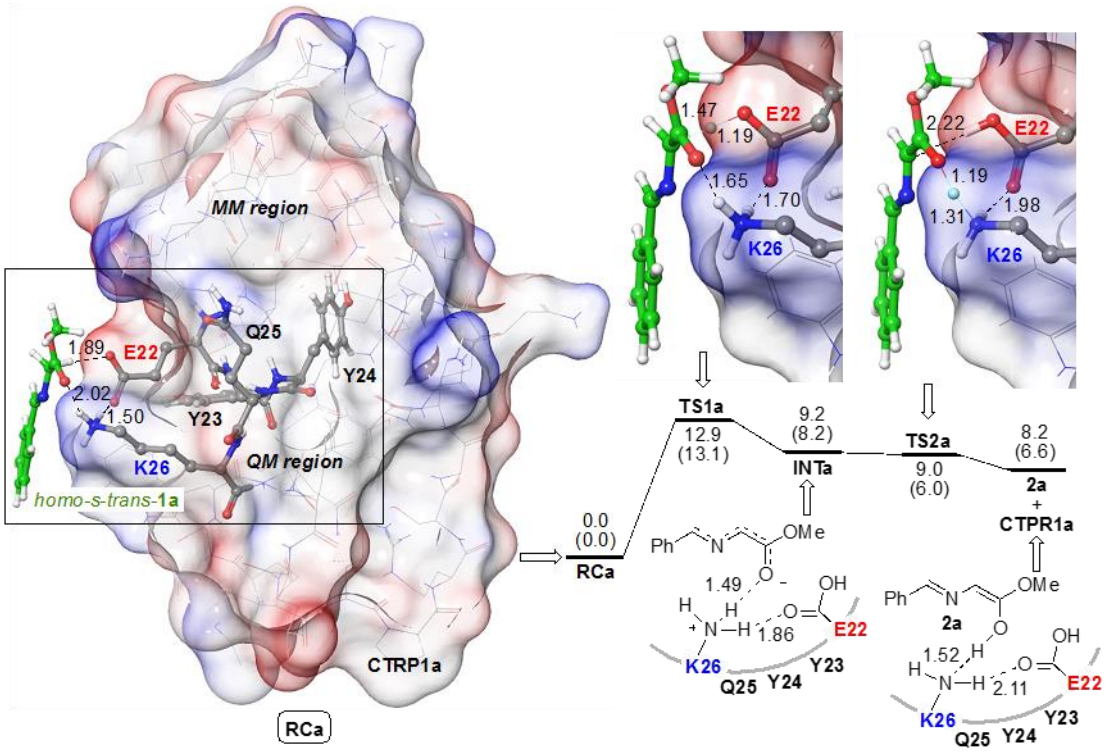


Figure 2.5. Calculated reaction profile for the formation of enol **2a** from imine **1a** in its *homo-s-trans* conformation and the neighborhood of the dyad formed by residues E22 and K26 of protein CTPR1a, calculated at the QM/MM ONIOM(M06-2X/6-311+G(2d,2p):dreiding)//ONIOM(B3LYP/6-31G(d,p):dreiding) level of theory. The QM region, defined by **1a** (highlighted in green) and E22-Y23-Y24-Q25-K26 residues (highlighted in gray), is represented in ball-and-stick mode. Bond distances are given in Å. Numbers on the stationary points (intermediates and transition structures) correspond to relative energies and Gibbs energies (at 298 K, in parentheses), in kcal/mol.

However, as it was mentioned in Scheme 2.3, the imine **1a** could also adopt an alternative *homo-s-cis* conformation via the rotation about the N-C_α single bond. Therefore, the dyad E22-K26 could interact likewise with this conformer to catalyze the formation of the azomethine ylide **3a'**, which is analyzed in the following paragraphs (Figure 2.6).

Firstly, the imine **1a**, in its *homo-s-cis* conformation, approaches the CTPR protein to form the reactive complex **RCb**. This initial complex consists on a hydrogen bond

between the amine group of K26 and one of the oxygens of the E22 (1.64 Å), an additional hydrogen bond between one of the hydrogens of the amino group in K26 and the N of the reactant (2.14 Å) and a third interaction between one of the hydrogens of the C_α and the remaining oxygen of E22 (2.18 Å). Accordingly, if we compare this initial reactant complex **RCb** with the previous **RCa**, we can conclude that the interactions between the *homo-s-trans* and *homo-s-cis* configurations of the imine **1a** and the active catalytic dyad are completely different. While the amino group of K26 interacts with the sp^2 hybridized oxygen atom from the ester group in **RCa**, the same residue is interacting with the N atom of the imine in **RCb**. In this way, the high flexibility of this protein (Figure 2.4) allows a noticeable distortion of the initial geometry of CTPR1a. As a result, the conformational change between both complexes, generates a reverse reactivity pattern which leads to the formation of the azomethine ylide **3a'** instead of the previous formed **3a**. Indeed, this new conformation, **RCb**, results to be 4.4 kcal/mol less energetic in terms of Gibbs energy in comparison with its *homo-s-trans* congener **RCa**.

The new orientation of the ammonium group of the K26 allows the N-protonation of the *homo-s-cis* **1a**, leading to the endergonic formation of the iminium cation **INTb**, which lies ca. 16kcal/mol above **RCb**. This hydrogen transfer step occurs via a transition state with an activation energy of 17.4 kcal/mol. The inspection of the C=N--H (1.17 Å) and H----NH_2-K26 (1.47 Å) distances of **TS1b**, shows a quite late transition structure. Besides, if we compare the existing interactions in **INTb** with the ones in **RCb**, we observe a shortening in the interaction between the hydrogen of the C_α position of the iminium cation and the oxygen of the E22 from 2.18 Å in **RCb** to 1.75 Å in **INTb**.

Thus, this iminium intermediate **INTb** can be deprotonated in the α position of the methylene group by the carboxylate group of E22. In this way, the iminium cation is transformed into the azomethine ylide **3a'** via **TS2b** with an activation energy of only 3.8 kcal/mol. In this case, the distances of **TS2b**, C_a---H (1.34 Å) and H---O-E22 (1.29 Å), show us a quite synchronous transition state structure (Figure 2.6). The formed **3a'** azomethine ylide interacts via a hydrogen bond with the protein (1.89 Å), generating a complex where E22 and K26 are also interacting through another hydrogen bond (2.03 Å).

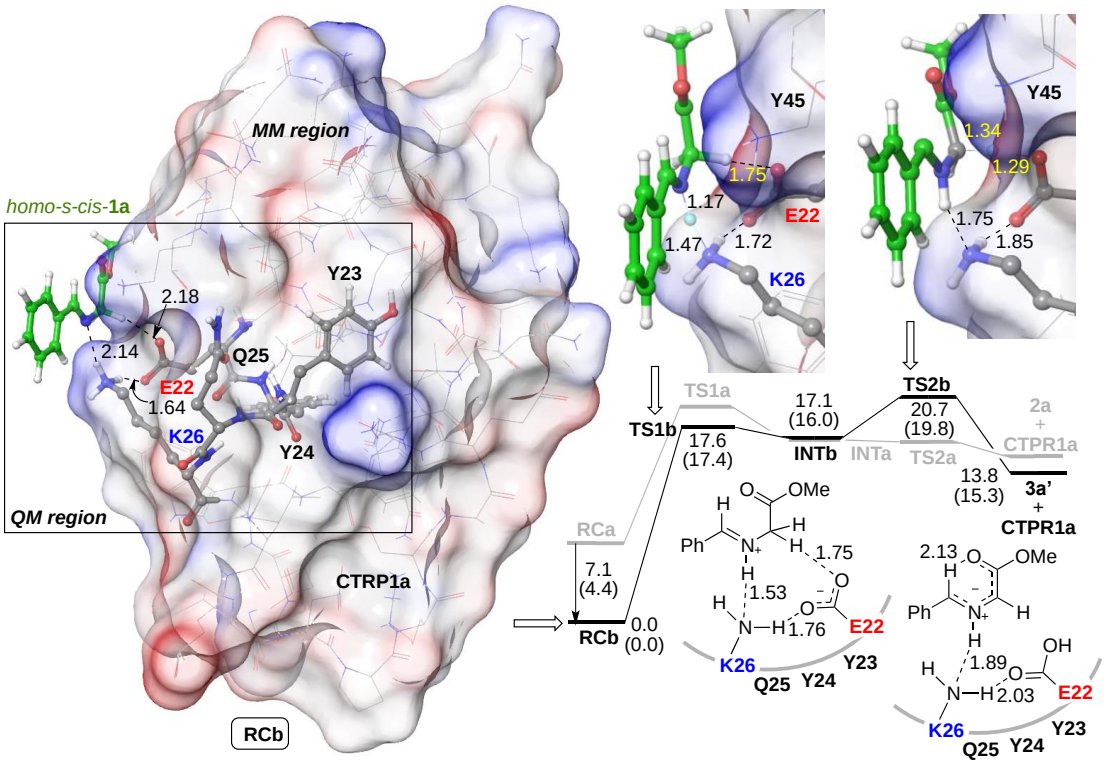


Figure 2.6. Calculated reaction profile for the formation of azomethine ylide **3a'** from imine **1a** in its *homo-s-cis* conformation and in the neighborhood of the dyad formed by residues E22 and K26 of protein CTPR1a. The reaction profile shown in Figure 2.5, in gray, is also included for comparison purposes. See the caption of Figure 2.5 for additional details.

The comparison of both pathways shown in Figure 2.6 confirms the energetic similarity of both transformations, which results compatible with the competitive formation of both azomethine ylides **3a** and **3a'** in the presence of CTPR1a.

2.4.2 Study of the (3+2) cycloaddition reaction carried out by the catalytic dyad E22-K26

As it was postulated in the Scheme 2.3, once the 1,3-dipoles **3a** and **3a'** were formed, these two reactive intermediates could promote the formation of *exo*-**5a** and *endo*-**5a** or *endo'*-**5a** and *exo'*-**5a**, respectively. It is worth to note that in the route of the *homo-s-trans* conformation of the imine **1a** the enol **2a** was obtained, which is transformed into the reactive azomethine ylide **3a** via a pseudopericyclic prototropy similar to the one shown in Scheme 2.1. Our calculations show that this complex

between **3a** and CTPR1a lies ca. 3 kcal/mol above enol **2a** in terms of Gibbs energy at 298K.

Consequently, the dipolarophyle, (*E*)-nitrostyrene **4a**, approaches the complexes formed between the CTPRa protein and the reactive 1,3-dipoles **3a** and **3a'** to initiate the (3+2) cycloadditions (Figure 2.7). Our calculations show that this final step occurs via concerted but quiet asynchronous **TS3-6**, as expected for other $[\pi 4_s + \pi 2_s]$ Woodward-Hoffmann topologies for thermally allowed (3+2) cycloadditions between azomethine ylides and nitroalkenes.[3,4]

With regard to the pathway that yields the *exo*-**5a** and *endo*-**5a** 4-nitropoline esters, the β-nitrostyrene **4a** reacts with the initial complex **3a**-CTPR1a via two competitive transition structures **TS3** and **TS4**. The formation of *endo*-**5a** product is mediated by the transition state, **TS3**, with a energy barrier of 3.5 kcal/mol, while the transition state for the (3+2) cycloaddition, **TS4**, which leads to the formation of exo-**5a** consist on a higher activation barrier of 6.8 kcal/mol. The evaluation of the distances for both **TS3** and **TS4** confirms the concerted but quite asynchronous nature of the transition structures, *i.e.* 2.09/2.65 Å and 2.08/2.65 Å, respectively.

In the case of the reaction between the **3a'**-CTPR1a complex and the dipolarophile **4a**, the transition structures **TS5** and **TS6** produce the final 4-nitropoline esters *exo'*-**5a** and *endo'*-**5a**. Therefore, the formation of *endo'*-**5a** takes place with an activation free energy of 5.2 kcal/mol in **TS5**, while the activation barrier in **TS6**, to yield *exo'*-**5a**, consists on 10.3 kcal/mol. Analogously to the pathways that yield *endo*-**5a** and exo-**5a**, the mechanism is concerted although quite asynchronous, in which the distances involved in the transition structures **TS5** and **TS6** are of 2.15/2.62 Å and 2.14/2.50 Å, respectively.

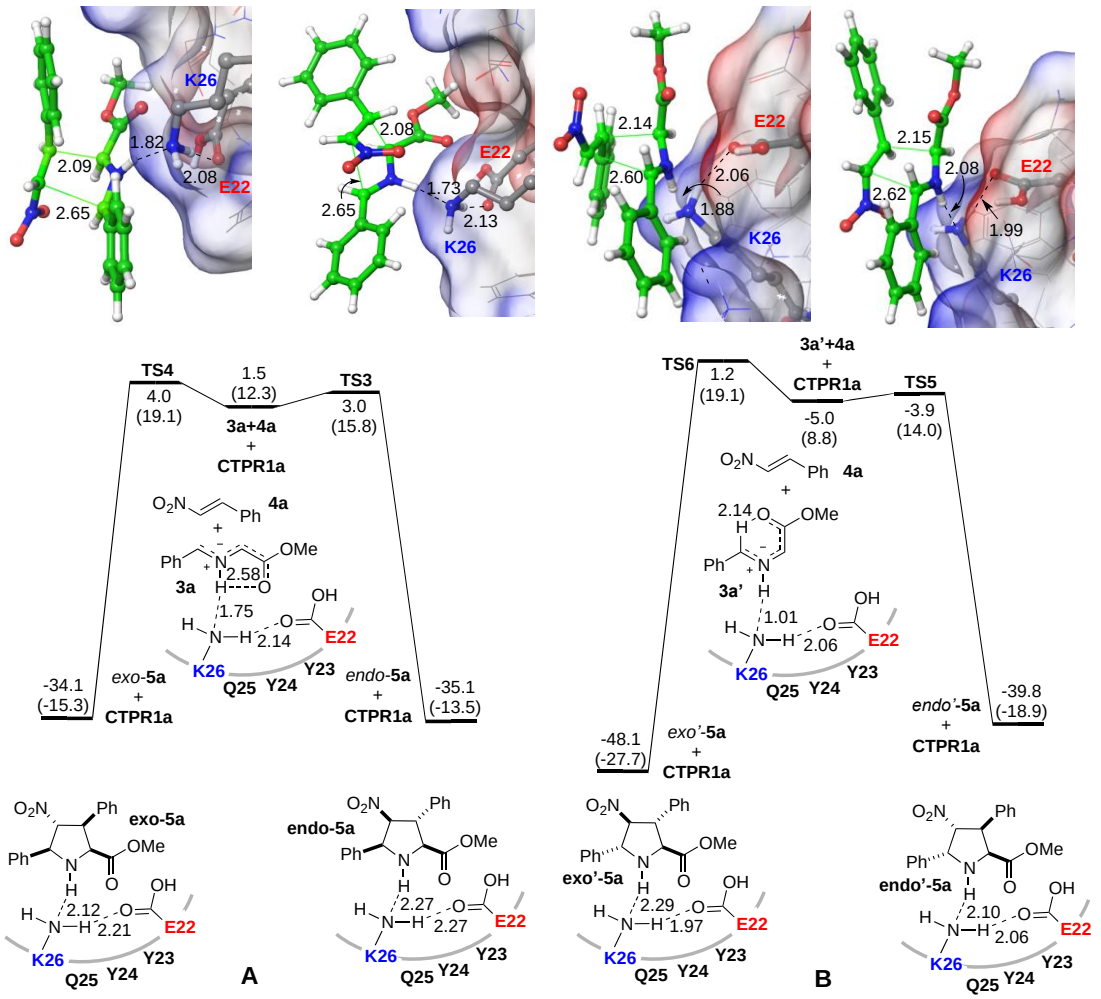


Figure 2.7. A) Calculated reaction profiles for the formation of unnatural proline esters *endo/exo*-**5a** and **B**) *endo'/exo'*-**5a** (**B**) from azomethine ylides **3a** and **3a'**, respectively, in the neighborhood of the dyad formed by residues E22 and K26 of protein CTPR1a. Values above and below the respective stationary points are the relative energies and Gibbs energies (in parentheses) with respect to reactive complexes **RCa** (Figure 2.5) and **RCb** (Figure 2.6).

In the four pathways presented in Figure 2.7, the complex between the CTPRa protein and the dipole or cycloadduct is maintained along the respective reaction coordinates through the interaction between the NH group and the nitrogen atom of K26. Additionally, the amino group of K26 and the neutral carboxyl group of E22 mantains a hydrogen bond that stabilizes the catalytic dyad. However, the interaction between the products and the CTPR1a protein results to be weaker than the one for the intermediates and transition structure. Therefore, the final cycloadducts can be released after the completion of the reactions and the protonation state of the E22/K26 dyad

would be recovered, thus re-initiating the catalytic cycle. Moreover, according to our calculations, the obtained values for the activation barriers make the conformational space energetically accessible for the reaction coordinates, in this manner, the catalytic dyad E22-K26 would be able to catalyze the formation of the four cycloadducts associated with the whole **1a + 2a** → **5a** process.

Consequently, this computational study is in good agreement with the experimental results and confirms the accessibility to the active dyad due to the flexibility of CTPR1a and the location of this active site in the surface of the protein. In addition, any mutation affecting the stabilization of this catalytic dyad in CTPR1a and CTPR3a series should result in the inhibition of the formation of *endo'*-**5a** and *exo'*-**5a** in the corresponding reaction mixtures. In contrary, as it is shown in the next section, single basic or acidic centers, are only able to catalyze the production of *endo*-**5a** and *exo*-**5a** cycloadducts.

2.4.3 Analysis of the reaction mechanism for the formation of azomethine ylides 3a catalyzed by acidic or basic monads

The CTPR1a protein contains six acidic centers (E2, D16, D18, E19, E29 and D31) and two basic residues (K13 and R33). Accordingly, these individual amino acids could act as catalytic monads of the Huisgen reaction. Thus we selected an acidic center, E2, and a basic center, K13, to analyze their potential individual catalytic activity. Initially, we hypothesized that the activity of those centers could produce both **3a** and **3a'**, however, our computational study concluded that only the **3a** intermediate could be obtained by the single monads, a result in line with our experimental results. When the computational work was carried out, the monads, both E2 and K13, result to be unable to stabilize the reactive complex between the CTPR1a protein and the *homo-s-cis* conformation of imine **1a**. When the reactant complex for *homo-s-cis* **1a** with the acidic or basic residues was optimized the equilibrium always turned into the *homo-s-trans* conformation, in this way, only the initial conformation needed for the 1-3-dipole **3a** was obtained.

The E2 active site was modelled defining a QM region in which the catalytic center and the neighbor residues A1 and A3 were included, while the rest of the protein was included in the MM region. Firstly, the *homo-s-trans* **1a** approaches the CTPR1a protein to generate the **RCc** reactant complex, in which a hydrogen bond is formed between one of the C_α-H protons of **1a** and the carboxyl group of E2 (Figure 2.8). This initial orientation allows the abstraction of one of the C_α-H protons by the acidic glutamate via a transition state, **TS1c**, with a Gibbs activation energy of ca. 13kcal/mol. In contrast to the abstraction process described for the dyad in Figure 2.5, this transition structure refers to an early transition state where the transferred proton is nearer from the C_α (1.11 Å) than from the oxygen of the E2 (1.59 Å). In this way, the enolate **INTc** is formed, being stabilized by an interaction between the nitrogen of the 1,3-dipole and the protonated E2.

In the subsequent step, the sp^2-hybridized nitrogen atom of the enolate imine intermediate **INTc** is protonated by the protonated carboxyl group of the glutamate residue in a barrierless process to yield the reactive azomethine ylide **3a**. The formation of this dipole, which directly interacts with the E2 residue via the NH group (1.54 Å), results to be a slightly endergonic process in terms of Gibbs free energy *c.a.* 3.3 kcal/mol. Consequently, this path described for the acidic monad is mechanistically similar to the one described in Figure 2.5 for the dyad, where the main difference was the double acceptor-donor role played by the *Glu* catalytic center (Figure 2.8).

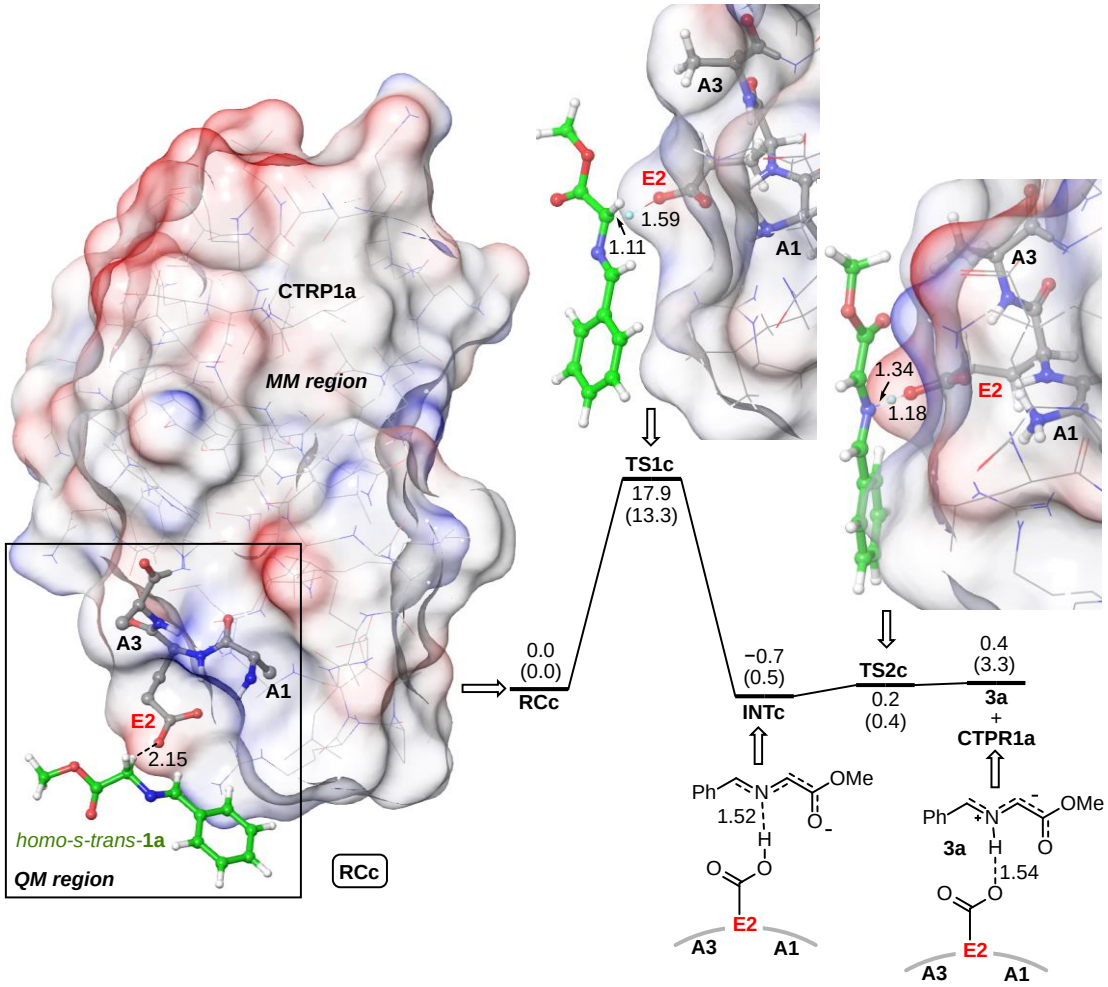


Figure 2.8. Calculated reaction profile for the formation of azomethine ylide **3a** from imine **1a** in its *homo-s-trans* conformation and in the neighborhood of the residue E2 of protein CTPR1a, forming the reactive complex **RCc**, calculated at the QM/MM ONIOM(M06-2X/6-311+G(2d,2p):dreiding)//ONIOM(B3LYP/6-31G(d,p):dreiding) level of theory. The QM region, defined by **1a** (highlighted in green) and A1-E2-A3 residues (highlighted in gray), is represented in ball-and-stick mode. Bond distances are given in Å. Values on the stationary points (intermediates and transition structures) correspond to relative energies and Gibbs energies (at 298 K, in parentheses), in kcal/mol, with respect to **RCc**.

An analogous computational study was carried out to understand the catalytic activity of the basic centers. In this case, the isolated K13 residue was selected for the modeling, defining as the QM region the respective lysine and the adjacent residues Y12 and Q14, whereas the rest of the protein was described with MM. The imine **1a** approaches the CPTR1a in its *homo-s-trans* conformation to generate the **RCd**, which is stabilized by the interaction between the ammonium group of the *Lys* residue and the

sp^2 hybridized nitrogen of the imine. Firstly, this nitrogen atom is protonated by the ammonium group of K13 through a process of an activation barrier of only 0.8 kcal/mol (Figure 2.9). Afterwards, the obtained **INTd**, in which the NH group of the ligand is interacting with the K13 residue, needs to be reoriented to the **INT'd**. In this way, one of the hydrogens of the C_α of the N-protonated iminium is directly interacting with the neutral amino group, creating a necessary conformation for the generation of the **3a** azomethine ylide.

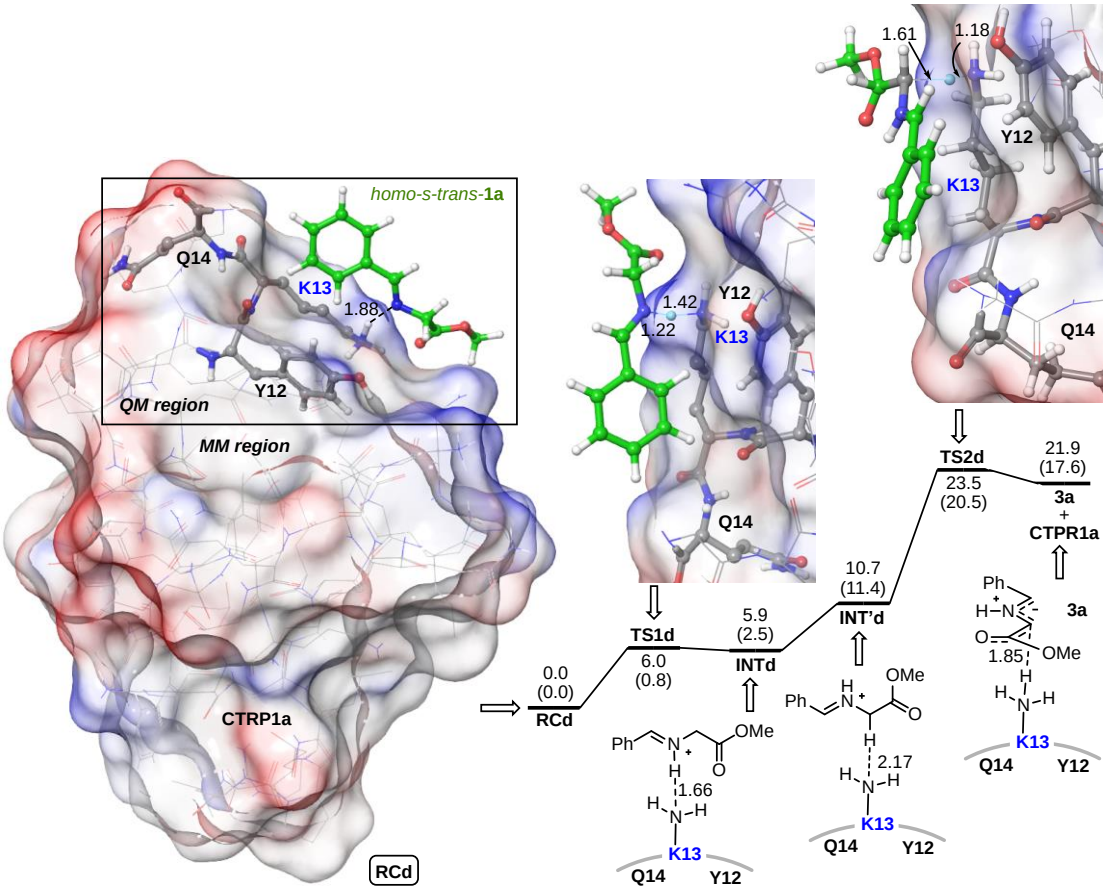


Figure 2.9. Calculated reaction profile for the formation of azomethine ylide **3a** from imine **1a** in its *homo-s-trans* conformation and in the neighborhood of the residue K13 of protein CTPR1a, forming the reactive complex **RCd**, calculated at the QM/MM ONIOM(M06-2X/6-311+G(2d,2p):dreiding)//ONIOM(B3LYP/6-31G(d,p):dreiding) level of theory. The QM region, defined by **1a** (highlighted in green) and Y12-K13-Q14 residues (highlighted in gray), is represented in ball-and-stick mode. Bond distances are given in Å. Values on the stationary points (intermediates and transition structures) correspond to relative energies and Gibbs energies (at 298 K, in parentheses), in kcal/mol, with respect to **RCd.**

Therefore, in the second step of this pathway, the neutral amino group of the catalytic lysine residues abstracts a hydrogen of the C_α position of the N-protonated iminium cation to yield the reactive azomethine ylide **3a** in an endergonic process, 17.6 kcal/mol. This second transition state, **TS2d**, results to be the rate limiting step of the reaction coordinate, consisting on an activation Gibbs energy of 9.1 kcal/mol. Accordingly, in the catalytic pathway for the basic monad, the K13 residue plays a double role, firstly, protonating the imine **1a** and secondly, abstracting a hydrogen from the C_α position.

In conclusion, isolated acidic or basic residues of CTPR proteins are able to catalyze by their own the formation of the 1,3-dipole **3a**, thus leading either *endo* or *exo* cycloadducts in the following (3+2) cycloaddition processes, analogously to the one depicted in Figure 2.7. These results are in good agreement with the experimental work, confirming that the *exo*' and *endo*' cycloadducts can only be obtained by the catalytic dyad and explaining the yield lowering of the *Huisgenase* activity when basis or acidic residues are removed (Table 2.2).

2.5 NMR STUDIES OF THE INTERACTION BETWEEN METHYL N-BENZYLIDENE GLYCINATE AND CTPR PROTEINS

In the last decades, the Saturation Transfer Difference NMR (STD-NMR) experiments and 2D ^{1}H-^{15}N correlation experiments have emerged as powerful tools to determine the main structural features of this kind of interactions.[22,23,24] Therefore, in order to gain more insight on the existing interactions in the studied catalysis, both STD-NMR and 2D ^{1}H-^{15}N correlation studies were carried out in collaboration with the group Professor Jesús Jiménez-Barbero at CIC bioGUNE.

22. Williams, M.; Davites, T., *Protein-Ligand Interactions: Methods and Applications*. 2nd ed.; Humana Press.: New York, 2013.
23. Meyer, B.; Peters, T., *Angew. Chem., Int. Ed.* **2003**, *42*, 864-890.
24. Fernández-Alonso, M. C.; Berbis, M. A.; Canales, A.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J., New Applications of High-resolution NMR in Drug Discovery and Development. In *New Applications of NMR in Drug Discovery and development*, Garrido, L.; Beckmann, N., Eds. Royal Society of Chemistry: 2013; pp 7-42.

In a first approach, the experimental conditions for the NMR studies (temperature, solvent etc.) were optimized. It is important to remind that the described (3+2) cycloadditions were always performed in THF-*d8* aiming to avoid the hydrolysis of the 1,3-dipole precursor **1a**. However, the optimization of the NMR studies suggested the employment of an aqueous phosphate buffered solution, which stabilizes the tertiary structure of the protein and allowed a better NMR analysis. The flexibility of the protein involves different motions that occur at different timescales, from ps to ms. We can consider that in the initial conditions, there are motions in the intermediate regime (ms or slower) in the chemical shift time scale that hamper the NMR analysis, due to the extreme broadening of the NMR cross peaks. Probably, this timescale is not relevant for the catalysis. The switch to phosphate buffer permits to eliminate these slower motions and to keep the intrinsic ps to ns motion in the protein backbone and side chains that allows the chemical shift perturbation and STD analysis.

Initially, the interaction of the **1a** imine with the selected larger CTPR3a and CTPR3 proteins was studied by means of STD-NMR experiments. These experiments are based on the fact that when the ligands are weakly bounded to the protein, an equilibrium between the bound and free ligand state is formed (Figure 2.10A). Taking this into account, the STD-NMR experiments involve two main experiments; on one hand, an *off resonance* ^{1}H NMR spectrum is recorded without protein saturation, secondly, an *on-resonance* experiment is done, in which the protein is selectively saturated by irradiating at a region where only the resonance of the protein is recorded. In this way, a difference spectrum could be obtained that only contains the signals of the ligand that have received the saturation (Figure 2.10B).[25] In our specific case, the studied protein would be the CTPR3 or CTPR3a protein and the ligand the imine **1a**.

25. Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J., *J. Chem. Educ.* **2011**, *88*, 990-994.

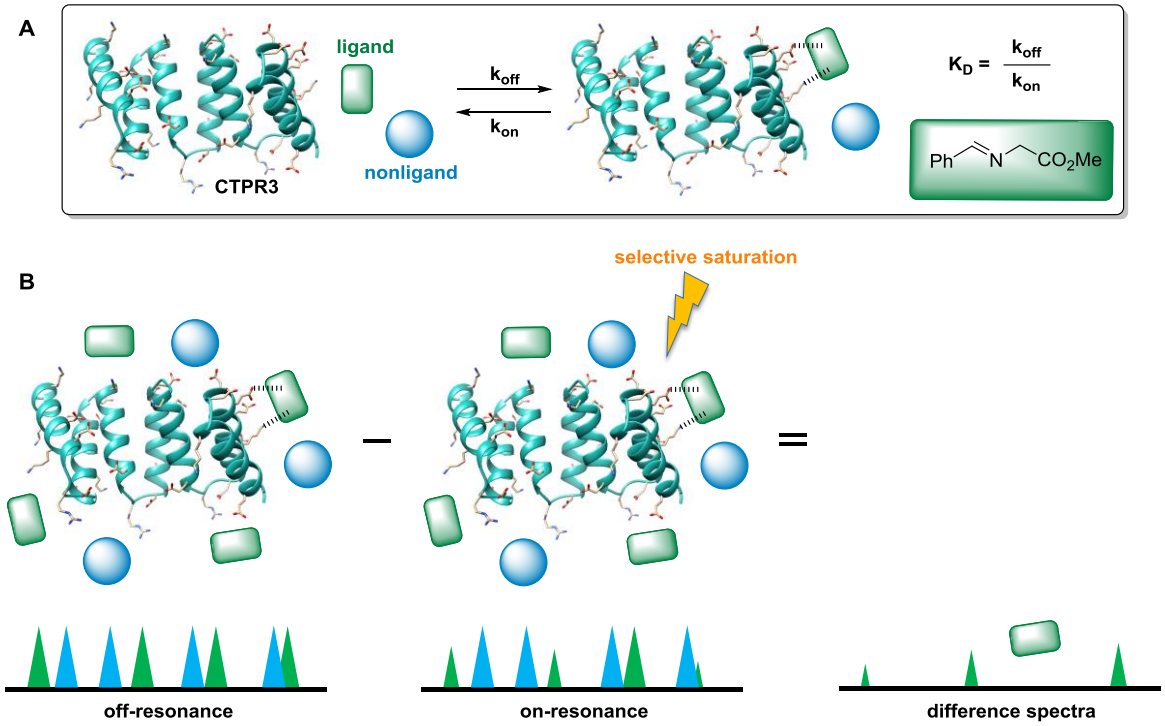


Figure 2.10. A) Scheme of the equilibrium between free and bound ligand **1a** and the CPTR3 protein. **B**) Scheme of the STD-NMR experiment, in which the *on-resonance* spectra is subtracted from the *off-resonance* spectra to obtain de saturation transfer difference spectra.

A 1:50 molar ratio mixtures of CTPR3a/CTPR3:**1a** were employed to perform the STD-NMR experiments. These analyses showed clear STD response for protons in both proteins CTPR3a/CTPR3, obtaining a similar spectrum in both cases. As it is depicted in Figure 2.11, STD signals were obtained for the benzylidene protons a-d, as well as for those of the glycinate moiety. In conclusion, these results confirm the close proximity between the ligand **1a** and the CTPR3 proteins. Additionally, during the experiments, the imine **1a** was hydrolyzed by the presence of a small amount of water in the media, to produce benzaldehyde. Even though the effect of this side reaction could be reduced by lowering the temperature to 278 K, STD signals for the aldehyde product were always obtained.

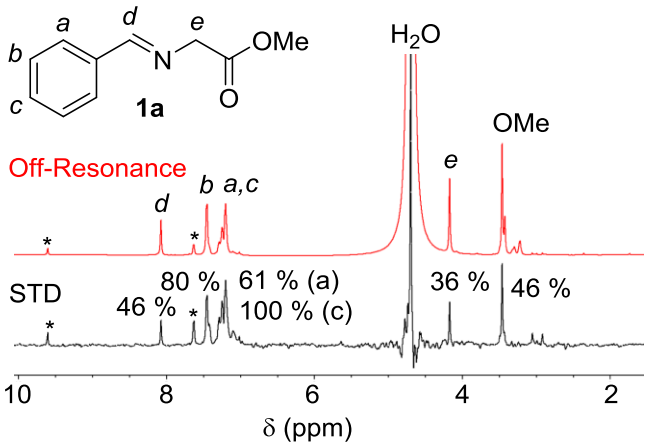


Figure 2.11. STD-NMR spectrum (*on-resonance* frequency δ 0.53 ppm) and ^{1}H-NMR reference spectrum (*off-resonance* frequency δ -25 ppm) of a sample containing 1.5 mM of **1a** and 30μM of CTPR3a at 278 K in D_2O (600 MHz). Signals marked with asterisks correspond to benzaldehyde, produced by partial hydrolysis of **1a**. The relative STD response, in %, of the different ^{1}H signals of **1a** is also given.

In a second approach, the interactions between the imine **1a** and the CTPR proteins were studied by means of ^{1}H-^{15}N TROSY NMR experiments. Firstly, the analysis was intended to be done with both CTPR3 and CTPR3a series, however, when the experiments were carried out with the CTPR3a proteins, more broadened peaks were obtained due to a conformational exchange process with an intermediate rate in the NMR chemical shift timescale. In addition, it is interesting to remember that the B-factor analyses (Figure 2.4) also confirm a very different behavior of the CTPR3a series. Thus, this protein was precluded for the direct analysis by ^{1}H-^{15}N TROSY NMR experiments. However, we must take into consideration that the associated structure for the TPR motifs is identical in both proteins, furthermore, the CTPR3 protein has been previously studied by NMR and the ^{1}H and ^{15}N resonance assignments.[26] In view of this data, the ^{1}H-^{15}N TROSY NMR experiments were performed with the more NMR-readable and characterized CTPR3 protein (Figure 2.1).

The chemical shift perturbation (CSP) induced by imine **1a** in the backbone was analyzed by ^{1}H-^{15}N TROSY of ^{15}N-labeled CTPR3 protein (Figure 2.12). Analogously to the

26. Main, E. R. G.; Stott, K.; Jackson, S. E.; Regan, L., *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5721-5726.

STD-NMR experiments, the decomposition of imine **1a** into benzaldehyde and glycine ester was also observed, although the temperature was reduced to 278K. For this reason, experiments by the addition of isolated glycine methyl ester and benzaldehyde were done as control experiments (Figure 2.13).

The initial experiments that were done with imine **1a** show a ppm shift of several ^{1}H-^{15}N signals of the CTPR3 protein, which allow us to determine the most perturbed amino acids (Figure 2.12A). Considering that the previous STD-NMR experiments confirm a notorious interaction between the proteins and the ligand, the chemical shift that was observed in the ^{1}H-^{15}N TROSY spectra for the CTPR3 protein in presence of **1a** is mostly attributed to the interaction between CTPR3 and the 1,3-dipole precursor. However, both imine **1a** and benzaldehyde contain an aromatic ring and a planar dipole with an electron-withdrawing nitrogen or oxygen atom able to act as a hydrogen bond acceptor which could bind to the protein in a similar way. Thus, some contribution of the benzaldehyde in the observed chemical shift could not be excluded.

Before introducing the discussion of the results of this section, we should note that the numbering of the residues in the NMR section includes the GAMDPGNS-N extension derived of the cloning that is showed in the experimental section. The study on the imine **1a** addition results on two mainly perturbed peaks, the positively charged K68 and K102 residues (according to the previous nomenclature the K26 of the second and third repetitions of CTPR3, respectively). These results are in good agreement with the chemical yields obtained for the mutants where these lysines were eliminated, due to the lowering on both yields and TON/TOF values. Therefore, these results suggest that these basic amino acids located on analogous positions of the B' and B'' helices of TPR2 and TRP3 repeats are good candidates to act as catalytic sites. Additionally, as it is shown in Figure 2.12C the K68 and K102 are located relatively close to E64 and E98, respectively, encoding a catalytic dyad (Figure 2.12B). Nevertheless, no other lysine residues were comparably perturbed.

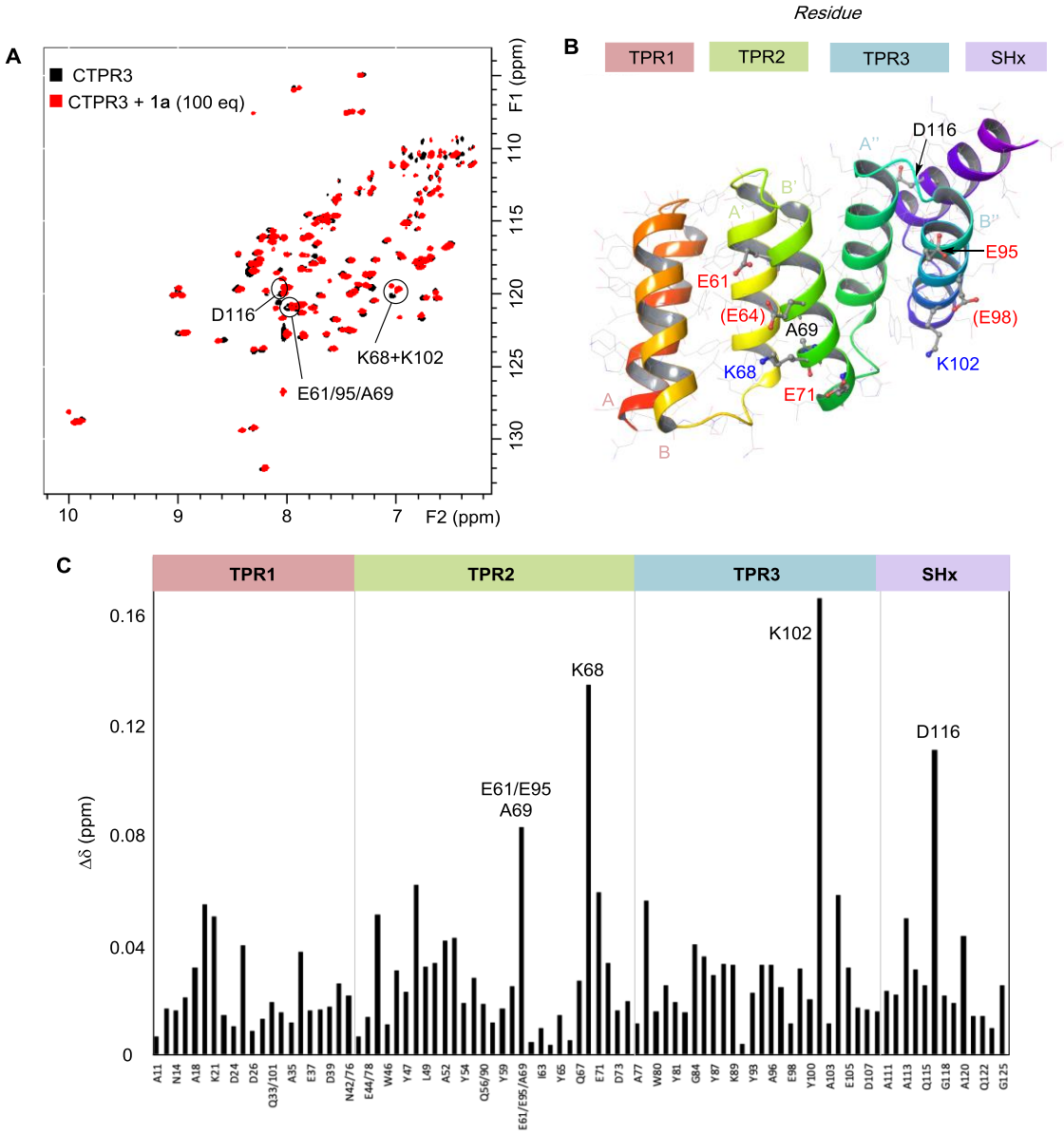


Figure 2.12. A) Overlay of the ^{1}H-^{15}N TROSY spectra of the CTPR3 protein for the apo form (in black) and upon addition of 100 eq. of **1a** (in red). **B**) Positions of the most perturbed residues shown in (**A**) on X-ray structure of CTPR3 (pdb code: 1NA0, see Figure 1A). Not significantly perturbed E64 and E98 residues, close to K68 and K102, respectively, are shown in parentheses. Acidic and basic residues are highlighted in red and blue, respectively. **C**) Average ^{1}H and ^{15}N chemical shift perturbation ($\Delta\delta=[(\Delta\delta^2_H+0.14\cdot\Delta\delta^2_N)/2]^{1/2}$, in ppm)[27] plotted for each amino acid of CTPR3 upon addition of **1a** (100 equiv.).

In addition, some negatively charged amino acids show chemical shift perturbations, suggesting the independent catalytic potential of single amino acids or

27. Williamson, M. P., *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1-16.

monads. Fundamentally, the E61 residue located on the B' helix of TPR2 and the E95 located on the equivalent B'' helix of TPR3 are the most perturbed ones, although their cross peaks could not be distinguished from A69. Moreover, CSP values greater than 2σ were observed for D116, Y19, N48, E71, A79 and L104 (Figure 2.12C). As mentioned before, control experiments for isolated glycine methyl ester and benzaldehyde were also performed. In those analyses, depicted in Figure 2.13, similar profiles were obtained upon benzaldehyde or imine **1a** addition, whereas the glycine methyl ester did not provide any noticeable effect.

These results suggest that B' and B'' helices of CTPR3 protein share potential catalytic sites involving both acidic and basic sites that would be able to perform the synthesis of the azomethine ylides derived from imine **1a**. Moreover, the conclusions obtained in these experiments could be probably extended to the analogue residues in CTPR1 series. In conclusion, these data are in good agreement with the activity results obtained in the catalysis and the computational study, thus, the ^{1}H-^{15}N TROSY and STD NMR experiments confirm the *Huisgenase* activity of CTPR protein series.

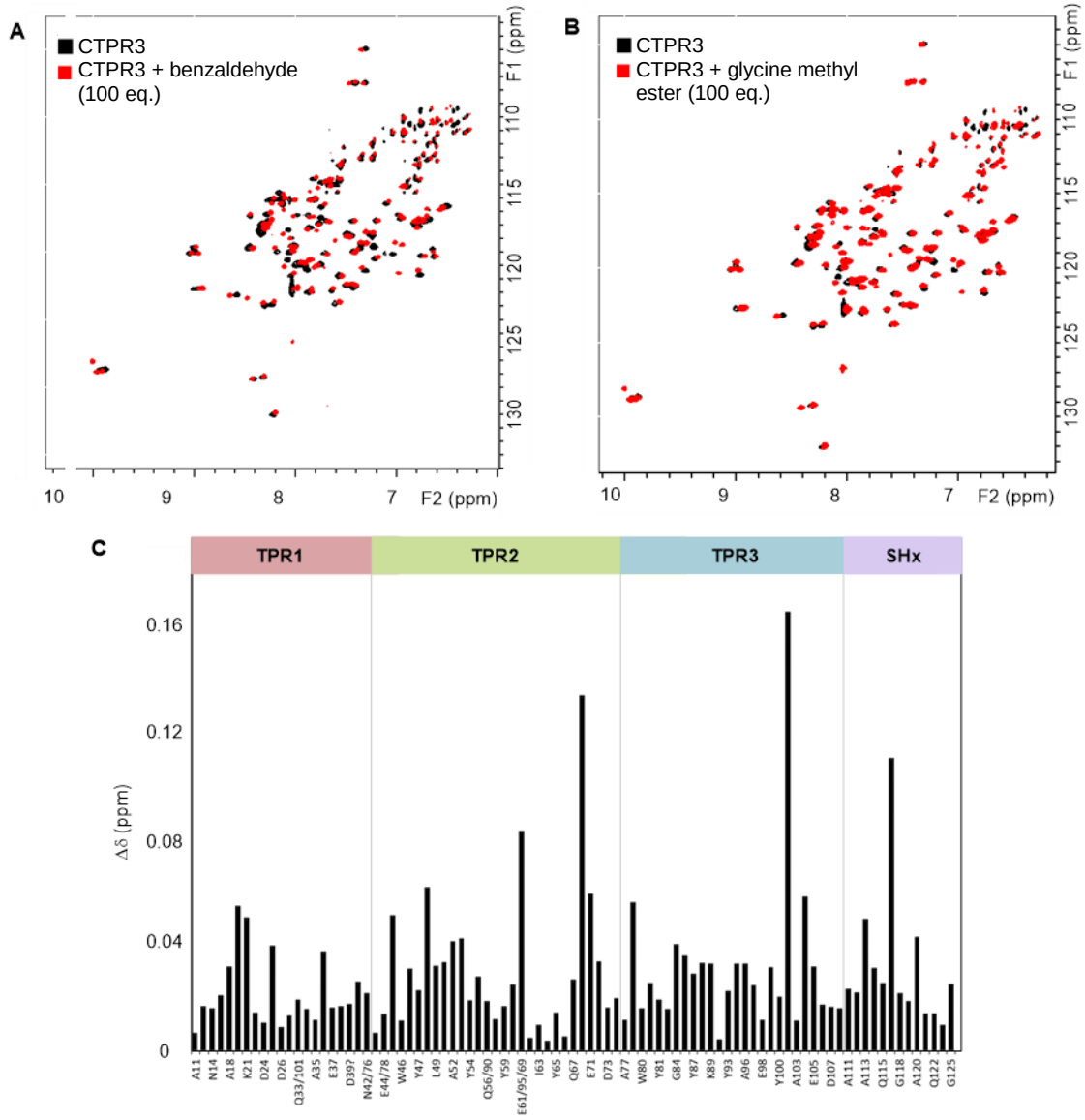


Figure 2.13. A) Overlay of the ^{1}H-^{15}N TROSY spectra of the CTPR3 for the apo form (in black) and red upon addition of benzaldehyde (100 equiv.). **B**) Overlay of the ^{1}H-^{15}N TROSY spectra of the CTPR3 for the apo form (in black) and red upon addition of glycine methyl ester. (100 equiv). **C**) Average ^{1}H and ^{15}N chemical shift perturbation ($\Delta\delta$ (ppm) = $[(\Delta\delta^2_H + 0.14\cdot\Delta\delta^2_N)/2]^{½}$) plotted for each amino acid of the CTPR3 upon addition of benzaldehyde (100 equiv.).

Additionally, we think that all the data obtained in this work can be a good starting point for a second generation of *Huisgenases*. The search for an optimized catalytic site or the rational design of more specific metal-containing sites would be helpful to achieve a diastereo- and enantioselective synthesis of unnatural proline amino acids. In this way, the repertoire of chemical reactions promoted by synthetic enzymes would be extended.

2.6 CONCLUSIONS

From the experimental and computational studies reported and discussed along this Chapter, the following conclusions can be drawn:

1. The employed consensus tetratricopeptide repeat (CTPR) proteins, both CTPR and CTPRa series with one or three repetitions, have shown a *Huisgenase* activity in THF solution catalyzing the (3+2) cycloaddition between the N-benzylidene glycinate and the (*E*)-β-nitrostyrene.
2. In the mentioned reaction, up to four racemic cycloadducts can be obtained in an equimolar ratio. Additionally, the diastereoselectivity of the reactions has been modified changing the flexibility of the CTPR variants. Therefore, the increased flexibility of the side chains in CTPRa series allows the formation of unusual diastereomeric cycloadducts, yielding the final four products in an equimolar ratio.
3. The modifications of several acidic and basic residues experiments confirm that this kind of residues result essential to catalyze the studied (3+2) cycloadditions. All these mutations generated in the protein, promote only the formation of *endo* and *exo* cycloadducts, consequently, inhibiting the production of *endo*' and *exo'*.
4. The carried QM/MM computational studies were able to rationalize the experimental findings. Firstly, the catalytic dyad E22-K26 is able to connect the *homo-s-cis* and *homo-s-trans* conformations of the starting imine with the corresponding (3+2) cycloadducts, forming the four stereoisomers. Secondly, the catalytic acidic or basic monads are able to catalyze the formation of the azomethine ylide **3a** and therefore, they only can produce the *exo*-**5a** and *endo*-**5a** cycloadducts.
5. The NMR experiments confirm the role of the CTPR protein in the *Huisgenase* activity, showing the direct interaction of the ligands with some of the crucial residues than can catalyze the studied (3+2) cycloadditions.

2.7 EXPERIMENTAL SECTION

General Remarks

Unless otherwise noted, reagents and substrates were purchased from commercial suppliers. Ligand **1a** was prepared following the previously described procedure.[28]

TLC was performed on 0.25mm silica gel 60 F254 aluminium plates and visualized with UV lamps. Column chromatography was carried out on columns of silica gel 60 (particle size 40-63 μm).

Infrared spectra were recorded on an Alpha-Bruker FTIR spectrometer with a single reflection ATR module. Wavenumbers are given in cm^{-1}.

High Resolution Mass Spectra (HRMS) analyses were carried out by SGIker services (Central Service of Alava and Bizkaia, University of the Basque Country) and performed on a LQ/QTOF, Agilent mass spectrometer using electrospray ionization (ESI) mode.

NMR spectra were recorded at 400 MHz for 1H NMR using $CDCl_3$ as solvent and TMS as internal standard. The data are reported as s=singlet, d=doublet, t=triplet and m=multiplet, coupling constant(s) (J) in Hz, integration.

Protein Design

Two types of consensus tetratricopeptide repeat sequences were used: consensus tetratricopeptide repeat module (CTPR) composed by 34 amino acids with the sequence AEAWYNLGNAYYKQGDYDEAIEYYQKALELDPNN,[12] and CTPRa consensus sequence composed by 34 amino acids with the sequence AEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRS.[13] Based on these modules, series of CTPR proteins were constructed by modular cloning.[29] CTPR proteins present an N terminal extension derived of the cloning (GAMDPGNS for CTPR series, and GAMGS for CTPRa series) and a solvating helix at the C-terminal end AEAKQNLGNAKQKQG. Different

28. Ayerbe, M.; Arrieta, A.; Cossío, F. P.; Linden, A., *J. Org. Chem.* **1998**, *63*, 1795-1805.
29. Kajander, T.; Cortajarena, A. L.; Regan, L., *Methods Mol. Biol.* **2006**, *340*, 151-170.

mutations were incorporated on CTPR and CTPRa repeats using QuickChange Site-Directed mutagenesis kit leading to the following CTPR and CTPRa modules: CTPRa_E2A, CTPRa_E22A, CTPRa_K26Q, CTPRa_D31A, CTPR_K13A, CTPR_K26A, CTPR_K26Q and CTPR_K13A_K26A.

In order to evaluate the role of the different potential catalytic residues, different proteins with 1 (CTPR1) or 3 repeats (CTPR3) of identical CTPR modules were constructed:

CTPR1a_wt:[GAMGS]AEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRSAEAKQNLGNAKQKQG.

CTPR1a_E2A:[GAMGS]A**A**AWYNLGNAYYKQGDYDEAIEYYQKALELDPRSAEAKQNLGNAKQKQG.

CTPR1a_E22A:[GAMGS]AEAWYNLGNAYYKQGDYDEAI**A**YYQKALELDPRSAEAKQNLGNAKQKQG.

CTPR1a_K26Q:[GAMGS]AEAWYNLGNAYYKQGDYDEAIEYYQ**Q**ALELDPRSAEAKQNLGNAKQKQG.

CTPR1a_D31A:[GAMGS]AEAWYNLGNAYYKQGDYDEAIEYYQKALEL**A**PRSAEAKQNLGNAKQKQG.

CTPR1_wt:[GAMDPGNS]AEAWYNLGNAYYKQGDYDEAIEYYQKALELDPNNAEAKQNLGNAKQKQG.

CTPR1_K13A:[GAMDPGNS]AEAWYNLGNAYY**A**QGDYDEAIEYYQKALELDPNNAEAKQNLGNAKQKQG.

CTPR1_K26A:[GAMDPGNS]AEAWYNLGNAYYKQGDYDEAIEYYQ**A**ALELDPNNAEAKQNLGNAKQKQG.

CTPR1_K26Q:[GAMDPGNS]AEAWYNLGNAYYKQGDYDEAIEYYQ**Q**ALELDPNNAEAKQNLGNAKQKQG.

CTPR1_K13A+K26A:[GAMDPGNS]AEAWYNLGNAYY**A**QGDYDEAIEYYQ**A**ALELDPNNAEAKQNLGNAKQKQG.

CTPR3a_wt:[GAMGS]AEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRSAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRSAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRSAEAKQNLGNAKQKQG.

CTPR3_wt:[GAMDPGNS]AEAWYNLGNAYYKQGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPNNAEAKQNLGNAKQKQG.

Finally, a **CTPR3_K13A+K26A** double mutant without solvating helix was generated:

[GNS]AEAWYNLGNAYY**A**QGDYDEAIEYYQ**A**ALELDPNNAEAWYNLGNAYYAQGDYDEAIEYYQAALELDPNNAEAWYNLGNAYYAQGDYDEAIEYYQAALELDPNN

Cloning and Molecular Biology

Genes encoding the CTPR and CTPRa proteins were constructed as previously described.[12,14] Genes encoding the CTPR mutants were obtained by Site-directed mutagenesis. All the genes were cloned into the pProEx-HTA vector for expression as his-tag fusion proteins. The gene encoding the CTPR3_K13A_K26A without solvating helix was synthesized by Biomatik Corporation. All the constructs were verified by DNA sequencing (Stab vida).

Protein Expression and Purification

Synthetic genes for the desired proteins cloned into pPROEx-HTA vector, coding for N-terminal $(His)_6$ tag and ampicillin resistance, were transformed into *Escherichia coli* C41 cells and cultured overnight at 37°C on agar plates. Overnight liquid cultures from single colonies were grown in 50 mL of Luria–Bertani (LB) media containing 100 µg mL^{-1} of ampicillin. 10 mL of saturated overnight cultures were then diluted into 1L of LB media supplemented with 100 µg mL^{-1} ampicillin. The cells were grown in an incubator-shaker (250 rpm) at 37°C until the optical density (OD_{600}) reached 0.6–0.8. Expression of CTPRs was induced with 0.6 mM isopropyl β-d-thiogalactoside (IPTG) followed by 5 h expression at 37°C. The cells were harvested by centrifugation at 5000 rpm for 20 min. The cell pellets were resuspended in lysis buffer consisting of 300 mM sodium chloride, 500 mM urea, 50 mM Tris pH 8.0. After 2 min. sonication at 30% power using a microtip

and Mison sonicator, lysed cells were centrifuged at 16 000 rpm for 30 min. and the protein supernatant was purified using standard Ni-NTA affinity purification protocols. The N-terminal hexahistidine tag was then cleaved from the CTPR proteins using TEV protease and a second Ni-NTA affinity column purification was performed in order to remove the his-tag and the TEV protease from the protein sample. Finally, the proteins were purified using FPLC over a Superdex 75 gel filtration column. As a final step, the aqueous solutions of CTPRs were dialyzed against 10 mM phosphate buffer pH 7.4 three times at 4°C using a dialysis membrane with molecular weight cutoff of 6-8 kDa. Protein concentration was measured by UV absorbance at 280 nm, using extinction coefficients at 280 nm calculated from amino acid sequence.

Expression and Purification of Isotopically Labeled Proteins

Synthetic genes for the desired proteins cloned into pPROEx-HTA vector, coding for N-terminal $(His)_6$ tag and ampicillin resistance, were transformed into *Escherichia coli* C41 cells and cultured overnight at 37°C on agar plates. Overnight liquid cultures from single colonies were grown in 50 mL of Luria–Bertani (LB) media containing 100 µg mL^{-1} of ampicillin. Then, 10 mL of overnight cultures were diluted into 1 L of M9 minimal medium that contains 1g/L $^{15}NH_4Cl$ supplemented with 100 µg mL^{-1} ampicillin. The cells were grown in an incubator-shaker (250 rpm) at 37°C until the optical density (OD_{600}) reached 0.6–0.8. Expression of CTPRs was induced with 0.6 mM isopropyl β-d-thiogalactoside (IPTG) followed by 18 h expression at 20°C. The cells were harvested by centrifugation at 5000 rpm for 20 min. The proteins are purified as described above for non-labeled proteins.

Protein characterization

a) Mass spectrometry analysis: MALDI-TOF mass spectrometry analysis of designed proteins was performed on a Maldi/TOF mass spectrometer UltrafleXtreme III (Bruker) with 3,5-Dimethoxy-4-hydroxycinnamic acid as a matrix.

b) SDS-PAGE analysis: The purified designed proteins with one repeat, CTPR1 and CTPR1a variants, were analyzed by SDS-PAGE gel electrophoresis using 4-20% gradient polyacrylamide gels with 1% SDS. The purified designed proteins with three repeats, CTPR3 and CTPR3a variants, were analyzed by SDS-PAGE gel electrophoresis using 15%

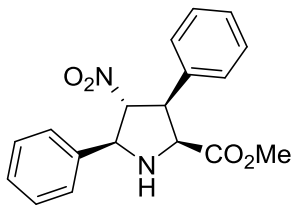
polyacrylamide gels with 1% SDS. For the staining of designed proteins, the gels were first washed with 40% ethanol and 7% acetic acid overnight to remove all SDS. They were then stained with Coomassie brilliant blue.

c) Circular dichroism: The protein secondary structure was examined by circular dichroism (CD) using a Jasco J-815 spectrometer. CD spectra were acquired at 20 μM protein concentration in 150 mM NaCl, 50 mM Tris pH 8.0 using 0.1 cm path length cuvette and a band-width of 1 nm at 1 nm increments and 10 second average time.

The protein stability under reaction conditions was evaluated measuring the protein secondary structure over the time by circular dichroism (CD) using a Jasco J-815 spectrometer. CD spectra were acquired at 20 μM protein concentration in THF/H_2O (95:5) using 0.1 cm pathlength cuvette using a band-width of 1 nm at 1 nm increments and 10 second average time.

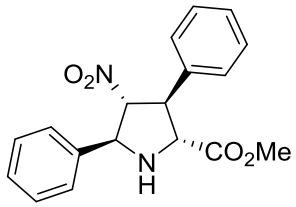
Catalysis Experiments: Cycloaddition of imine 1a with β-nitrostyrene 4a

A mixture of *N*-methyl benzylidene glycinate (**1a**, 55 mg, 0.31 mmol) and β-nitrostyrene (**4a,** 60 mg, 0.40 mmol), was dissolved in 2 mL of THF at room temperature. CTPR protein (0.3 mL, 30μM) in THF (2 ml) was added and the reaction mixture was stirred in an orbital stirrer at room temperature for 6 days at r.t. Then, the solvent was evaporated under vacuum. The resulting residue was dissolved in 20 mL of DCM and washed three times with brine. The organic layer was dried with $MgSO_4$, and the solvent was eliminated. In this manner, we are capable to obtain a yellow oil, which was purified by silica gel chromatography (Ethyl Acetate/Hexane 1:4) to give four stereoisomers, *exo´*-**5a**, *exo*-**5a**, *endo*-**5a** and *endo´*-**5a** (see Table 2.2 for the respective yields). All the experiments were carried out in triplicate and the crude reaction mixtures were analyzed by ^{1}H-NMR. These compounds were compared with samples already described in the literature.[4]

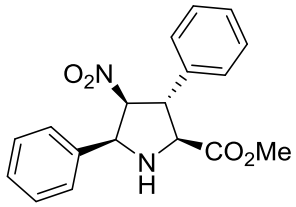


Methyl (2S,3S,4R,5S)-4-nitro-3,5-diphenylpyrrolidine-2-carboxylate (*exo-**5a***). Yellow solid. **m_p**= 114-116˚C. **FTIR** (neat, cm^{-1}) 3355, 1743, 1544, 1347. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.58–7.52 (m, 3H), 7.43–7.38 (m, 2H), 7.38–7.32 (m, 1H), 7.31–7.25

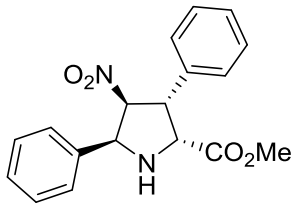
(m, 2H), 7.21 (d, *J* = 1.2 Hz, 1H), 5.19 (dd, *J* = 9.4, 6.9 Hz, 1H), 4.74 (t, *J* = 7.6 Hz, 2H), 4.48 (dd, *J* = 9.1, 4.7 Hz, 2H), 4.36 (dd, *J* = 9.7, 7.3 Hz, 1H), 3.26 (s, 3H), 2.71 (s, 1H).



Methyl (2R,3S,4R,5S)-4-nitro-3,5-diphenylpyrrolidine-2-carboxylate (*exo'-**5a***). Yellow solid. $\mathbf{m_p}$= 110-113˚C. **FTIR** (neat, cm^{-1}) 3356, 1744, 1544, 1346. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.50 (dd, *J* = 7.3, 1.7 Hz, 2H), 7.38 (ddd, *J* = 7.8, 6.5, 1.4 Hz, 2H), 7.33–7.26 (m, 3H), 7.21 (s, 1H), 5.13–5.07 (m, 1H), 5.04 (dd, *J* = 7.1, 4.0 Hz, 1H), 4.71–4.63 (m, 1H), 4.02 (dd, *J* = 9.0, 7.4 Hz, 1H), 3.72 (d, *J* = 1.6 Hz, 3H), 2.86 (s, 1H).



Methyl (2S,3R,4S,5S)-4-nitro-3,5-diphenylpyrrolidine-2-carboxylate (*endo-**5a***). Yellow solid. $\mathbf{m_p}$= 120-122˚C. **FTIR** (neat, cm^{-1}) 3314, 1735, 1545, 1370. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.44 (dd, *J* = 8.1, 6.5 Hz, 2H), 7.41–7.36 (m, 5H), 7.35–7.32 (m, 2H), 5.31 (dd, *J* = 6.5, 3.5 Hz, 1H), 4.95 (dd, *J* = 10.9, 6.5 Hz, 1H), 4.25 (dd, *J* = 7.4, 3.5 Hz, 1H), 4.19 (t, *J* = 8.1 Hz, 1H), 3.84 (s, 3H), 3.39 (t, *J* = 9.9 Hz, 1H).



Methyl (2R,3R,4S,5S)-4-nitro-3,5-diphenylpyrrolidine-2-carboxylate (*endo'-**5a***). Yellow solid. $\mathbf{m_p}$= 122-124˚C. **FTIR** (neat, cm^{-1}) 3314, 1732, 1543, 1370. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.52-7.25 (m, 10H), 5.64 (dd, 1H, J = 7.7Hz, J'= 6.4 Hz), 5.47 (d, 1H, J = 8.0 Hz), 4.69 (d, 1H, J = 8.2 Hz), 4.50 (dd, 1H, J = 11.3Hz, J'=7.8 Hz), 3.28 (s, 3H), 2.73 (s,1H).

Control (blank) experiments were performed in triplicate in the absence of any additive. Furthermore, different additives such as isolated α-amino acids, pentapeptides or enzymes were tested under reaction conditions identical to the previously described in order to assess their hypothetical *Huisgenase* activity. In particular, lysine (K) and glutamic acid (E), as well as N-protected monomeric amino acids Boc-K-OH, Boc-D(OMe)-OH and Boc-D-OH were assessed. Pentapeptides EYYQK (with a sequence identical to that of the catalytic dyads of CTPR proteins) and YEYKQ (possessing E and K residues in a different order with respect to the previously mentioned pentapeptide) were also tested. Three different concentrations of potential catalyst: 30 μM (identical to that used with CTPRn proteins), 0.57 μM and 0.25 μM. The two latter concentrations

were chosen to reproduce the number of acidic residues of the whole proteins with n=1 (53 AA, including the solvating helix) and n=3 (121 AA, also including one solvating helix), respectively. In addition, two triacylglycerol lipases from *Pseudomonas cepacia* and *Pseudomonas stutzeri* were tested under the same conditions and at a concentration of 30 µM. Both lipases incorporate α-helices possessing acidic and basic residue, and lipase TL from *P.stutzeri* tolerates THF water-free conditions.

Computational Methods

All stationary points on the potential surface have been fully optimized using the two layer integrated molecular orbital and molecular mechanism (ONIOM) procedure.[30,31,32,33] The quantum mechanics (QM) region was treated using the hybrid B3LYP functional[34,35] with the 6-31G(d,p) basis set,[36,37] whereas the dreiding force field[38] was employed for the molecular mechanics (MM) layer. Hydrogen link atoms were applied to treat the QM/MM boundary. Moreover, the electronic embedding scheme[39,40] was used to account for the polarizing effect of atomic point charges defined in the MM layer on the QM region. Harmonic analyses were calculated at this level of theory to verify the nature of the corresponding stationary points (minima or transition states), as well as to provide the zero-point vibrational energy (ZPE) and the thermodynamic contributions to the enthalpy and free energy for T=298 K. Energies of

30. Maseras, F.; Morokuma, K., *J. Comput. Chem.* **1995**, *16*, 1170-9.
31. Dapprich, S.; Komiromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J., *J. Mol. Struct.: THEOCHEM* **1999**, *461-462*, 1-21.
32. Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J., *J. Comput. Chem.* **2003**, *24*, 760-769.
33. Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K., *Chem. Rev.* **2015**, *115*, 5678-5796.
34. Becke, A. D., *J. Chem. Phys.* **1993**, *98*, 5648-52.
35. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., *J. Phys. Chem.* **1994**, *98*, 11623-11627.
36. Hehre, W. J.; Ditchfield, R.; Pople, J. A., *J. Chem. Phys.* **1972**, *56*, 2257-61.
37. Hehre, W. J.; Radom, L.; Schleyer, P. V. R.; Pople, J. A., *Ab Initio Molecular Orbital Theory.* Wiley: New York, 1986.
38. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III, *J. Phys. Chem.* **1990**, *94*, 8897-909.
39. Bakowies, D.; Thiel, W., *J. Phys. Chem.* **1996**, *100*, 10580-10594.
40. Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A., Jr.; Morokuma, K.; Frisch, M. J., *J. Chem. Theory Comput.* **2006**, *2*, 815-826.

the QM layer were refined using single-point M06-2X [41] calculations and the 6-311+G(2d,2p) basis set[36,37] at the optimized ONIOM(B3LYP/6-31G(d,p):dreiding) geometries. All geometries, frequencies and energies were calculated using the Gaussian 09 program package. [42] Some transition state structures present smooth potential energy surfaces and the default optimization criteria result difficult to be achieved. For these stationary points the convergence criteria have been reduced to a maximum step size of 0.01 au and a RMS force of 0.0017 au.

The CTPR3 pdb (PDB ID: 1NA0) structure has been used as the initial geometry. The CTPR1 protein with a solvating helix (total number of 49 amino acids) has been our computational model. The quantum regions were selected to capture the structural and electronic essential features of the acid-base reactions. In the structure of the protein, it only exists an acid-base amino acid pair, E22 and K26, with a short distance, *ca.* 3 Å, which allows the interaction with the imine and the formation of the azomethine ylide. In turn, the E22YYQK26 residues have been selected as the QM region for the catalytic dyad. Among the nine acid and six basic residues of the CTPR1 protein with the solvating helix, the E2 and K13, located on the surface of the first α-helix, were chosen to describe the catalysis of the monads. In these models, the QM regions include the selected acidic or basic centers with the preceding and subsequent residues, *i.e.* A1EA3 and Y12KQ14.

NMR Saturation Transfer Difference Experiments

The samples for STD experiments were prepared in phosphate-buffered saline (150 mM NaCl, 50 mM sodium phosphate pH = 6.8) using 1:50 protein/imine **1a** ratios with a protein concentration of 30 μM. The applied temperatures varied between 298

41. Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.* **2008**, *120*, 215-241.

42. *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.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, revision E.01; Gaussian, Inc.: Wallingford, CT, 2013*.

and 278 K. Temperatures were optimized to minimize imine-aldehyde conversion. In all cases, the on-resonance frequency was set at the aliphatic region (0.53 ppm) and the off-resonance frequency at -25 ppm. Protein saturation was achieved by using a series of 15 ms PC9 pulses with a total saturation time of the protein envelope signals of 2.5s with or without water suppression (WATERGATE schemes), using the Bruker AVANCE III 600 MHz and 800 (cryo) MHz spectrometers. A spin-lock filter (100 ms) was used to remove the NMR signals of the protein background.

2D 1H-15N TROSY experiments

For comparison with the sequence-specific assignments of the ^{1}H and ^{15}N resonances previously reported,[26] the ^{15}N-labeled protein samples (200-500 μM) were initially prepared in phosphate-buffered saline (150 mM NaCl, 50 mM sodium phosphate pH= 6.8) with the addition of 10% D_2O for field-frequency lock. Afterwards, buffer conditions and temperature were optimized to reduce imine-aldehyde conversion. Finally, chemical shift perturbations of ^{1}H and ^{15}N resonances of the protein (200-500 μM) in the presence of benzaldehyde (100 equiv.), glycine methyl ester (100 equiv.) and the imine **1a** (100 equiv.) were monitored at 278K using phosphate buffered solutions (150 mM NaCl, 50 mM sodium phosphate pH= 7.4). All two-dimensional ^{1}H-^{15}N Transverse Relaxation Optimized Spectroscopy (TROSY)-Heteronuclear Single Quantum Coherence (HSQC) spectra were recorded at 800 MHz on a Bruker AVANCE III equipped with TCI cryoprobe. The ^{1}H-^{15}N TROSY experiments were conducted with 256 t_1 and 2048 t_2 points with proton and nitrogen spectral widths of 11 and 40 ppm, respectively. Spectra were processed using Bruker TopSpin software and further analyzed using CCPNmr.[43] The ^{1}H-^{15}N composite chemical shift differences (Δδ) were calculated according to the following equation: $\Delta\delta\ (ppm) = [(\Delta\delta^2_H + 0.14\cdot\Delta\delta^2_N)/2]^{½}$.[27]

43. Skinner, S. P.; Fogh, R. H.; Boucher, W.; Ragan, T. J.; Mureddu, L. G.; Vuister, G. W., *J. Biomol. NMR* **2016**, *66*, 111-124.

Annexes

ChemPubSoc Europe

DOI: 10.1002/chem.201503171

CHEMISTRY
A European Journal
Full Paper

Domino Reactions

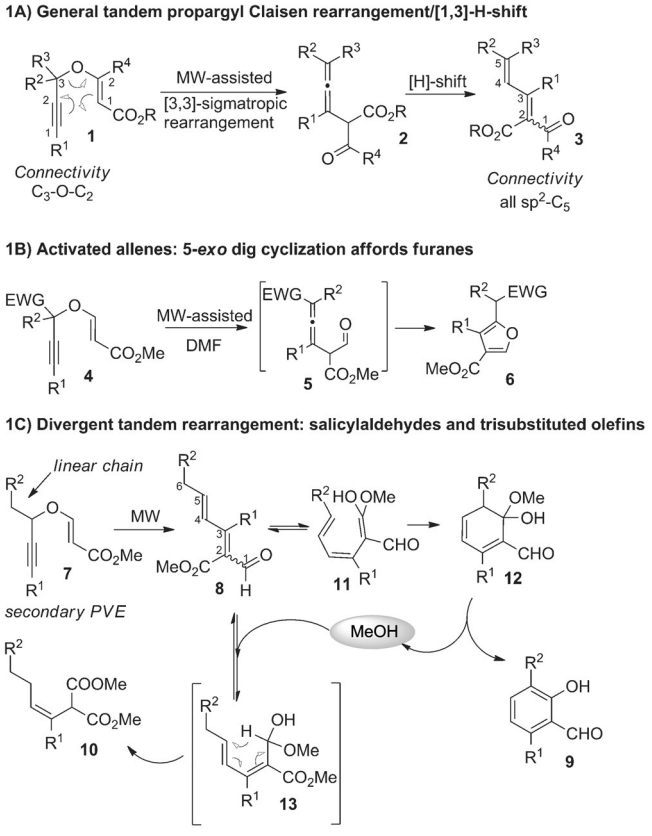
Microwave-Assisted Organocatalyzed Rearrangement of Propargyl Vinyl Ethers to Salicylaldehyde Derivatives: An Experimental and Theoretical Study

David Tejedor,*[a] Leandro Cotos,[a] Daniel Márquez-Arce,[a] Mikel Odriozola-Gimeno,[b] Miquel Torrent-Sucarrat,[b, c] Fernando P. Cossío,[b] and Fernando García-Tellado*[a]

Abstract: The microwave-assisted imidazole-catalyzed transformation of propargyl vinyl ethers (PVEs) into multisubstituted salicylaldehydes is described. The reaction is instrumentally simple, scalable, and tolerates a diverse degree of substitution at the propargylic position of the starting PVE. The generated salicylaldehyde motifs incorporate a broad range of topologies, spanning from simple aromatic monocycles to complex fused polycyclic systems. The reaction is highly regioselective and takes place under symmetry-breaking conditions. The preparative power of this reaction was demonstrated in the first total synthesis of morintrifolin B, a benzophenone metabolite isolated from the small tree *Morinda citrifolia* L. A DFT study of the reaction was performed with full agreement between calculated values and experimental results. The theoretically calculated values support a domino mechanism comprising a propargyl Claisen rearrangement, a [1,3]-H shift, a [1,7]-H shift (enolization), a 6π electrocyclization, and an aromatization reaction.

Introduction

Propargyl vinyl ethers (PVEs) **1** are a privileged group of small-size, densely functionalized, and readily accessible linear scaffolds. The main key to the chemical reactivity encoded in these structures is the [3,3]-sigmatropic rearrangement (propargyl Claisen rearrangement) shown in Scheme 1 A,[1] which takes place irreversibly and under thermodynamic control to generate the allene **2**.[2] Since the first, seminal thermally driven rearrangement achieved by Black and Landor in 1965,[3] which served to determine that propargylic systems could be accommodated into the Claisen rearrangement, a vast number of propargyl Claisen rearrangements have been successfully performed.[1] The main drawback of these rearrangements is the requirement for high temperatures, which forced the use



Scheme 1. The MW-assisted propargyl Claisen rearrangement of propargyl vinyl ethers **1**. EWG = electron-withdrawing group.

[a] *Dr. D. Tejedor, Dr. L. Cotos, D. Márquez-Arce, Dr. F. García-Tellado*
Departamento de Química Biológica y Biotecnología
Instituto de Productos Naturales y Agrobiología
Consejo Superior de Investigaciones Científicas
Avenida Astrofísico Francisco Sánchez 3
38206 La Laguna, Tenerife (Spain)
E-mail: fgarcia@ipna.csic.es
dtejedor@ipna.csic.es

[b] *M. Odriozola-Gimeno, Dr. M. Torrent-Sucarrat, Prof. F. P. Cossío*
Facultad de Química
Universidad del País Vasco-Euskal Herriko Unibertsitatea (UPV/EHU)
and Donostia International Physics Center (DIPC)
P. K. 1072, San Sebastian-Donostia (Spain)

[c] *Dr. M. Torrent-Sucarrat*
Ikerbasque, Basque Foundation for Science
María Díaz de Haro 3, 6°, 48013 Bilbao (Spain)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201503171.

of high-boiling solvents and harsh conditions. These experimental difficulties prevented their use in preparative organic synthesis, in sharp contrast to the counterpart involving allyl vinyl ethers.[4] However, this scenario dramatically changed in the last decade with the emergence of milder metal-catalyzed protocols[5] and the replacement of conventional heating with laboratory microwave (MW) equipment.[1] Under MW heating, allenes **2** rearrange to the more stable diene derivatives **3** through a pseudo-pericyclic [1,3]-H shift (Scheme 1 A). Overall, this tandem rearrangement transforms a C_3-O-C_2 linear structure, easily assembled through C–O bond-forming chemistry, into an all-sp^2-linear C_5 carbogenic block[6] armed with a reactive carbonyl group (aldehyde or ketone), an ester group, and a doubly conjugated diene. An exception to this outcome is represented by the activated allenes **5** (EWG = ester or secondary amide),[7] which directly rearrange to the furan derivatives **6** through a tandem enolization/O cyclization/H transfer process (Scheme 1 B).[8]

A particularly striking reactivity profile arises from monosubstituted PVEs **7** (R^3 = H, hereinafter referred to as secondary PVEs) bearing a linear chain at the propargylic position (Scheme 1 C). In these cases, the rearrangement affords a roughly equimolar mixture of salicylaldehyde **9** and trisubstituted olefin **10**.[9] The formation of salicylaldehyde **9** can be formally rationalized through a tandem enolization/6π electrocyclization/aromatization process with the net generation of a methanol molecule per molecule of salicylaldehyde. On the other hand, the stereoselective formation of the trisubstituted olefin **10** should be mediated by the formation of the hemiacetal **13** to launch an internal redox process involving a [1,5]-H shift. Overall, this manifold is a divergent chemical process that can produce two well-differentiated products from the same reactive intermediate and through two interconnected pathways (the product of one pathway launches the other). Remarkably, the manifold could be selectively funneled toward each of these two products by careful design of the experimental conditions. Thus, salicylaldehydes **9** could be cleanly obtained in 60–80% yield by MW irradiation of a xylene solution of PVEs **7** in the presence of 4 Å molecular sieves (MS) (300 mg per mole substrate).[9] Although initially we used 4 Å MS as additive, we later discovered that pyridine (200 mol%) could replace the MS, rendering the process homogenous and easier to perform.[10] It was assumed that the beneficial action of the two additives was related to a reduction of the activation barrier for the enolization process, otherwise the limiting step of the process. Although this is an uncommon reactivity for MS, there are some precedents for their use in enolization processes.[11] On the other hand, the MW irradiation of a methanolic solution of PVEs **7** exclusively afforded the olefin derivatives **10** in good to excellent yields (75–95%) and with complete stereoselectivity (the two alkyl chains are located in *trans* positions).[12]

The salicylaldehyde motif is an important building block for the preparation of numerous pharmacologically relevant coumarins,[13a] flavonoids,[13b] chromenes,[13c,d] catechols,[13e] and several mycotoxins,[13f] as well as chiral catalysts based on transition metal Schiff base complexes (salen catalysts).[14] Salicylaldehydes are commonly prepared by direct formylation of the corresponding phenol derivatives[15] under the classical Reimer–Tiemann conditions ($CHCl_3$/KOH)[15a] or the Duff procedure (hexamethylenetriamine/acetic acid /sulfuric acid).[15b,c] Interestingly, the 3,6-disubstitution pattern present in salicylaldehydes **9** cannot be easily accessed by using these standard protocols. Among other drawbacks, these reactions have regioselectivity problems when the *ortho* and *para* positions to the hydroxyl functionality are not conveniently blocked.[16]

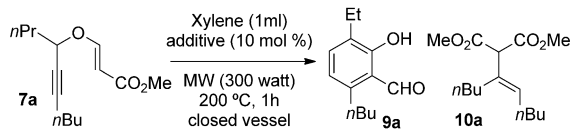
The efficiency and instrumental simplicity of this reaction together with the importance of the salicylaldehyde motif prompted us to undertake a systematic study of this reaction with the aim of transforming it into a general, robust, and preparative protocol for accessing salicylaldehyde-based structural motifs. The transformation of this reaction into a standard synthetic protocol with preparative value (academic and industrial) required implementation of a catalytic version incorporating the following: 1) the use of a cheap and easy handled catalyst (additive); 2) high tolerance to broad and diverse substitution patterns on the PVE; and 3) bench-friendly reaction processing. We report herein how these conditions could be met by using imidazole (10 mol%) as a basic catalyst and how this catalytic manifold could be used for the construction of salicylaldehyde motifs supported on a broad range of topologies, which spanned from simple aromatic monocycles to complex fused polycyclic systems. We also performed a theoretical study on this reaction that is in full agreement with the observed experimental results. Finally, the preparative power of this reaction has been demonstrated in the first total synthesis of morintrifolin B,[17] a benzophenone metabolite isolated from the small tree *Morinda citrifolia* L.

Results and Discussion

Catalytic synthesis of 3,6-disubstituted salicylaldehydes from secondary PVEs

Although pyridine had proved to be a convenient additive for this transformation,[10] the large excess required (200 mol%) precluded its use in large-scale preparative reactions. Guided by the current principles of sustainable chemistry,[18] we searched for new additives to be used as true catalysts. For this study, we chose the transformation of PVE **7a** into salicylaldehyde **9a** (Table 1). We had observed in earlier studies[9,10] that the substitution pattern of this PVE was one of the worst tolerated by this reaction with both 4 Å MS (43%)[9] and pyridine (53%).[10] Because the use of pyridine improved the efficiency of the reaction by 10%, we envisioned that this reaction could be a convenient benchmark for the assay of other basic additives better suited than pyridine to catalyze the required enolization of dienal **8** to intermediate **11** (Scheme 1 C). With this idea in mind, we assayed the bases shown in Table 1 (entries 1–15). Additionally, we assayed a set of common acids of moderate strength to see if they could also be suitable additives for this reaction (Table 1, entries 16–18). The reactions were performed under our previously established conditions [xylene (1 mL), MW (300 watt, 200 °C, and closed vessel), 1 h]

Table 1. Optimization of the additive for the MW-assisted synthesis of salicylaldehyde **9a**.

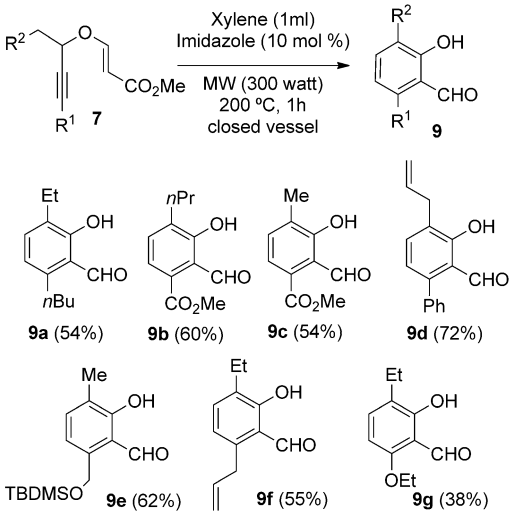


Entry	Additive[a]	pK_a[b]	**9a** + **10a** [%]	Ratio **9a**/**10a**[c]
1	DABCO	2.97	64	1.5
2	aniline	4.63	35	0.17
3	DMBA	4.68	36	0.16
4	pyridine[d]	5.2	87	1.6
5	imidazole	6.9	84	1.7
6	collidine	7.4	68	1.4
7	DMAP	9.2	78	2.1
8	DIPEA	10.8	64	1.4
9	quinuclidine	11	44	0.8
10	DBU	12	62	1.5
11	*t*BuOK	17	37	0.3
12	indole	21[e]	34	0.1
13	pyrrole	23[e]	34	0.1
14	urea	26.9[e]	42	0.7
15	TMP	37[e]	74	1.8
16	$C_6H_6CO_2H$	4.2	58	0.8
17	$C_5H_{10}NH_2^+AcO^-$	8.7	56	0.6
18	phenol	9.95	55	0.1
19	4 Å MS[f]		90	1

[a] Abbreviations: DMBA: 1,3-dimethylbarbituric acid; DMPA: 4-dimethylaminopyridine; DIPEA: *N*,*N*-diisopropylethylamine; DBU: 1,8-diazabicyclo[5.4.0]undec-7-ene; TMP: 2,2,6,6-tetramethylpiperidine. [b] In H_2O. [c] Determined by NMR spectroscopy. [d] 200 mol%. [e] In DMSO. [f] 300 mg mmol^{-1}.

with a catalytic amount (10 mol%) of additive. Imidazole and 4-(*N*,*N*-dimethylamino)pyridine proved to be the best additives (Table 1, entries 5 and 7), with the highest yields (84 and 78% combined yield) and **9a/10a** ratios of 1.7 and 2.1, respectively. The yields and ratios were in both cases similar to those obtained with an excess of pyridine (Table 1, entry 4) and substantially better than those obtained with 4 Å MS (90% combined yield, 1:1 ratio, Table 1, entry 19). Practical reasons (price and availability) made imidazole the preferred catalyst for this reaction. The use of acid additives had the reverse effect, generating the olefin **10a** as the main product, albeit with moderate efficiency (Table 1, entries 16–18).

The scope and efficiency of the catalytic manifold was first studied with the transformation of the secondary PVEs **7a**–**g** into the corresponding salicylaldehydes **9a**–**g** (Scheme 2).[19] In general, the reaction delivered the corresponding salicylaldehydes **9a**–**f** in moderate to good yields (54–72%) depending on the substitution pattern of the PVE. Esters, isolated double bonds, and protected hydroxyl groups were well tolerated. Interestingly, the PVE **7g** bearing an electron-rich alkyne in its structure could be also transformed into the corresponding hydroxylated salicylaldehyde **9g** (38%). In spite of this moderate-to-low yield, the importance of hydroxylated aromatic platforms[20] and the potential use of hydroxylated salicylaldehydes as convenient building blocks for accessing biologically relevant compounds[13,21,22] ensures a preparative value for this transformation. In addition, the easy and direct access to secondary PVEs containing the ethoxyacetylene motif (ethoxyacetylene is commercial available) allows for fast access to a broad range of 3-substituted 6-ethoxysalicylaldehyde platforms.



Scheme 2. Imidazole-catalyzed synthesis of salicylaldehydes **9** from secondary propargyl vinyl ethers **7**.

Catalytic synthesis of 3,4,6-trisubstituted salicylaldehydes from tertiary PVEs

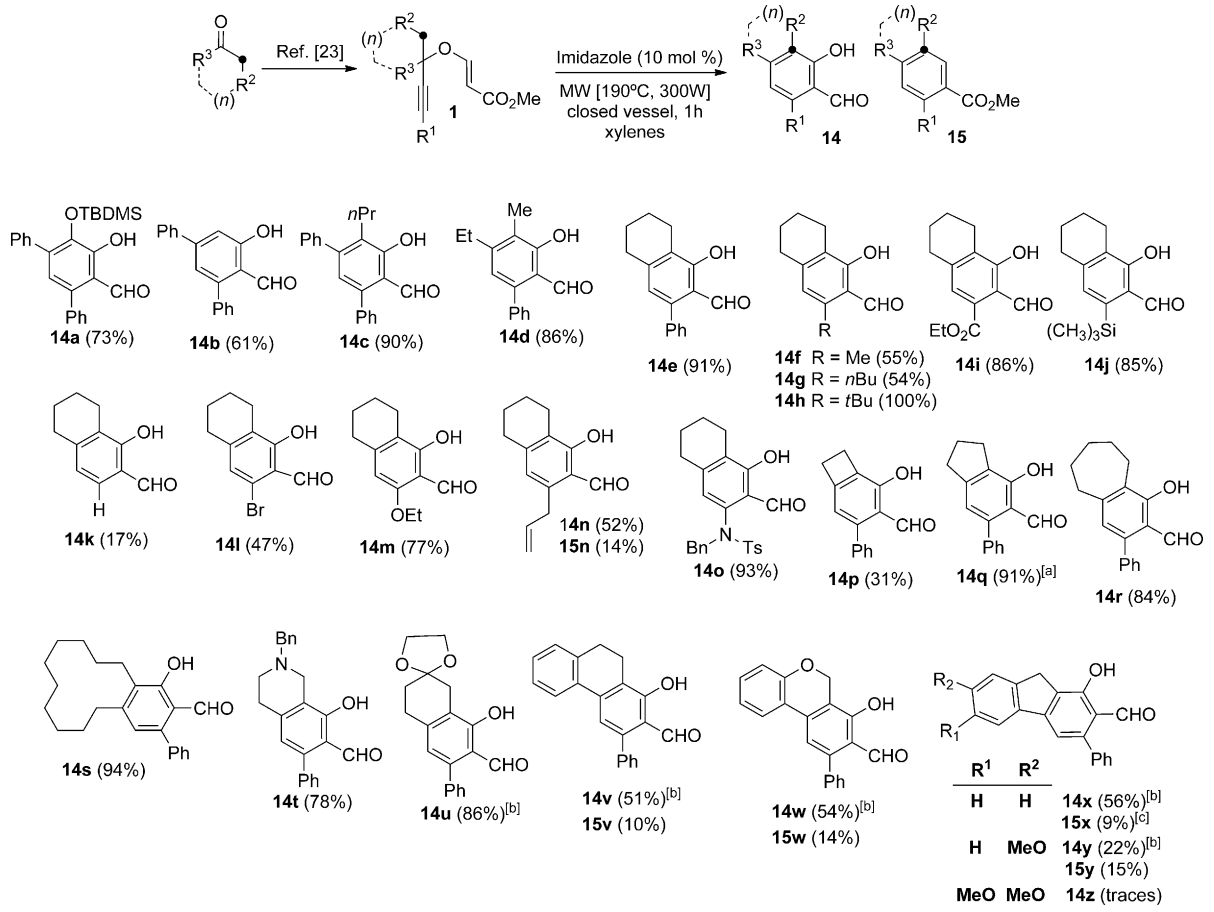
Once the catalytic version of the reaction had been standardized, we studied its extension to PVEs **1** bearing two substituents at the propargylic position (R^2 and $R^3 \neq H$, hereinafter referred to as tertiary PVEs). The main advantages associated with the incorporation of tertiary PVEs into the reaction manifold are related to: 1) skeletal complexity: the use of cyclic ketones (mono- or polycycles, simple or fused) should increase the power of the manifold to generate structural complexity and topological diversity; 2) reactivity: the presence of two substituents on the propargylic position of the tertiary PVE should favor both the propargyl Claisen rearrangement[1] and the enolization process (see Scheme 1C), which should translate into a net increase in the reaction efficiency. With these ideas in mind, we prepared an extensive set of tertiary PVEs **1**[23] incorporating a wide array of substituents and molecular topologies (Scheme 3). These PVEs were conveniently prepared from the corresponding tertiary propargyl alcohols and methyl propiolate according to our recently reported protocol.[24]

In general, tertiary PVEs proved to be excellent substrates for this catalytic reaction affording the corresponding aromatic derivatives in good average yield. The PVEs **1a**–**d**[23] derived from acyclic ketones delivered the corresponding aromatic derivatives **14a**–**d** in good average yield and incorporated varied substitution patterns at the aromatic ring, including alkyl, aryl, and oxygen-containing substituents. Remarkably, the last-named functional groups are introduced with high efficiency (**14a**, 73%) and in a chemo-differentiated manner (one as a free OH, the another as a protected OH group). This property enables orthogonal manipulation of each hydroxyl group at

the aromatic ring without taking special precautions or using special reagents. An interesting aspect of this reaction is the breakage of the symmetry observed when symmetrically disubstituted PVEs (ketones) are used ($CH_2R^2=R^3$). In these cases, the reaction places the R^3 substituent at the C^5 position of the ring, delivering the R^2 substituent (the sec-alkyl chain) at the corresponding C^6 position. This is a very important property and it allows the final differentiation of the otherwise identical propargylic substituents of the PVE, which increases the diversity-generating power of the reaction manifold. The derivative **14d** is a good example of this symmetry-breaking chemical differentiation ($R^3=$Et, $R^2=$Me).

The tertiary PVEs **1e**–**u**,[23] derived from symmetrical monocyclic ketones, were also convenient substrates for the reaction and provided a broad set of fused bicyclic salicylaldehyde derivatives featuring varied aromatic functionalization. Thus, the PVEs **1e**–**o**, armed with a cyclohexane motif, delivered the corresponding 1-hydroxy-5,6,7,8-tetrahydronaphthalene-2-carbaldehydes **14e**–**o** adorned with a wide array of substituents at the C^3 position. The substituents ranged from simple hydrogen, the worst case (**14k**, 17%), to heteroatoms including silicon (**14j**, 85%), bromine (**14l**, 47%), oxygen (**14m**, 77%), and nitrogen (**14o**, 93%). The low yield of **14k** is in accordance with our previous results using 4 Å MS[9] or pyridine[10] and confirms that substitution at the terminal alkyne position ($R^1\neq$H) is necessary to favor the formation of the enol intermediate **11** (see Scheme 1C). Thus, the high efficiency of this reaction when a bulky *tert*-butyl group is located at the terminal position of the triple bond in the parent PVE (**14h**, quantitative) is remarkable. The substitution of this group by less sterically demanding groups (*n*-Bu, Me) lowered the yield (54 and 55%, respectively). On the other hand, generation of the salicylaldehyde **14p** (31%) incorporating a fused cyclobutene ring is outstanding and reflects the potential of this catalytic reaction to generate molecular topologies not easily accessible by other methods.[25]

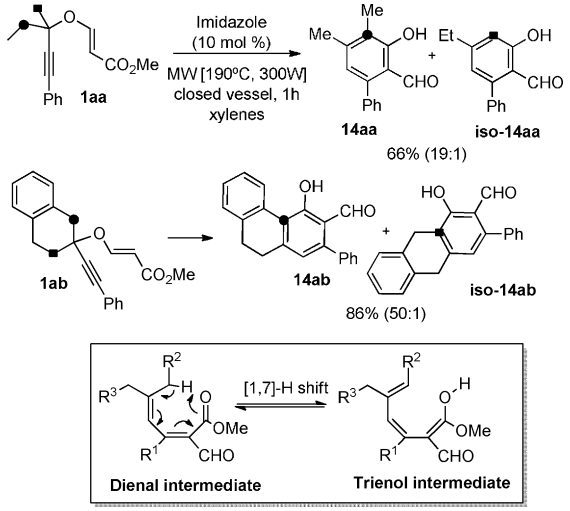
Other topologies based on the bicyclo[*n*.4.0] motive ($n=3$, 5, and 10) were explored. The reaction proved to be general with regard to the size of the non-aromatic ring (**14p**–**s**). An interesting case was offered by the derivative **14s** incorporating a C_{12} ring in its structure. These topologies incorporating sp^3-rich macrocycles are highly appreciated and underexploited structural motifs for drug discovery.[26] The reaction also allowed efficient access to C^6-substituted 8-hydroxy-1,2,3,4-tetra-



Scheme 3. Imidazole-catalyzed synthesis of salicylaldehydes **14** from tertiary propargyl vinyl ethers **1**. [a] The reaction was performed on a gram scale (1.1 g, 4.4 mmol). [b] The corresponding PVEs were not isolated. Yields refer to the two-step process: formation of the PVE from the corresponding propargyl alcohol and MW-assisted rearrangement. See the Experimental Section and the Supporting Information for details. [c] The yield was assigned by quantitative NMR integration of the crude reaction residue.

hydroisoquinoline-7-carbaldehydes (**14t**, 78%) and 1-hydroxy-5,6,7,8-tetrahydronaphthalene-2-carbaldehydes bearing a protected ketone at C^7 as a convenient handle for further ring functionalization (**14u**, 86%). Both structures are interesting scaffolds for medicinal chemistry.[27] Salicylaldehydes **14v–z** are examples of fused tricyclic structures with natural-product-like topologies.[28] Interestingly, in these cases, the formation of salicylaldehydes **14** is accompanied by formation of the corresponding benzoate derivatives **15**, which could stem from the 6π electrocyclization/aromatization processes of another diene **2** conformation arising from ketone–ester group interchange (see below). The subtle influence of the ketone skeleton on the selectivity of the reaction was apparent from the somehow surprising formation of the benzoate derivative **15n**, the only benzoate detected in the mono- and bicyclic series. An subtle electronic effect was also observed in the reactions involving fused bicyclic ketones. Whereas the formation of the 9*H*-fluorene derivatives **14x–z** showed a clear deterioration of the reaction efficiency with increasing electron richness of the aromatic ring (compare **14x** with **14y** and **14z**), the presence of an oxygen atom at the 9-position of the 9,10-dihydrophenantrene core was irrelevant in terms of efficiency, but not in terms of selectivity (compare **14v**/**15v** and **14w**/**15w**).

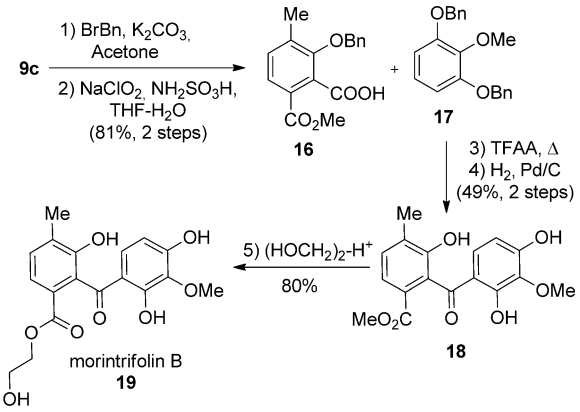
The reaction proved to be highly regioselective when asymmetrically substituted tertiary PVEs (ketones) were used (Scheme 4). Thus, PVE **1aa** bearing two different alkyl substituents at the propargylic position (R^2 = Et, R^3 = Me) delivered the salicylaldehyde **14aa** in 63% yield (19:1 isomeric ratio). The reaction manifold selectively incorporated the more substituted homopropargylic carbon atom into the final aromatic ring. Exquisite regioselectivity was also observed in the synthesis of the derivative **14ab** (86%, 50:1 isomeric ratio), showing a clear preference for the benzylic position over the alternative one. The observed regioselectivity could be explained by the relative stabilities of the isomeric trienol intermediates (see below; Scheme 4, inset), which in turn relies on the relative stability of the terminal double bonds.



Scheme 4. Regioselectivity in the imidazole-catalyzed reaction.

Application to the first total synthesis of morintrifolin B

The easy access to these multisubstituted salicylaldehydes scaffolds gave us the opportunity to use them as convenient starting materials for the synthesis of more complex structures embodying the parent aromatic ring. Among the possible structures, we chose benzophenones because of their chemical and biological relevance.[29] The family of natural benzophenones has more than 300 members with great structural diversity and bioactive properties including antifungal, anti-HIV, antimicrobial, antioxidant, and antiviral activity.[30] Morintrifolin B (**19**)[17] is a naturally occurring benzophenone isolated from the small tree *Morinda citrifolia* L. (Rubiaceae), commonly known as noni and widely distributed in southern Asia and the Pacific islands. The plant has found popular medicinal use for the treatment of asthma, bone fractures, cancer, cholecystitis, dysentery, lumbago, menstrual cramps, urinary difficulties, and many other ailments.[31] In spite of the pharmacological potential offered by the constituents of the plant, there are no reports to date on synthetic or biological studies on morintrifolin B. Because of our interest in this kind of compounds, we undertook a short synthesis of this benzophenone metabolite from the salicylic acid derivative **16** (Scheme 5), readily available from salicylaldehyde **9c** by simple phenol protection (BrBn, K_2CO_3, acetone, 93%) and aldehyde oxidation ($NaClO_2$, sulfamic acid, THF/H_2O, 87%). The trifluoroacetic anhydride-mediated reductive coupling with the 2-methoxy resorcinol derivative **17** assembled the benzophenone core of **19**, which was subjected to selective deprotection (H_2, Pd/C) to give the intermediate **18** (49% for two steps) featuring the same hydroxylation pattern as the target natural product. Finally, the acid-controlled transesterification of this intermediate with ethylene glycol afforded morintrifolin B (**19**) in 80% yield. The five-step synthesis delivered morintrifolin B in an overall yield of 32% from the corresponding propargyl vinyl ether **7c**. The

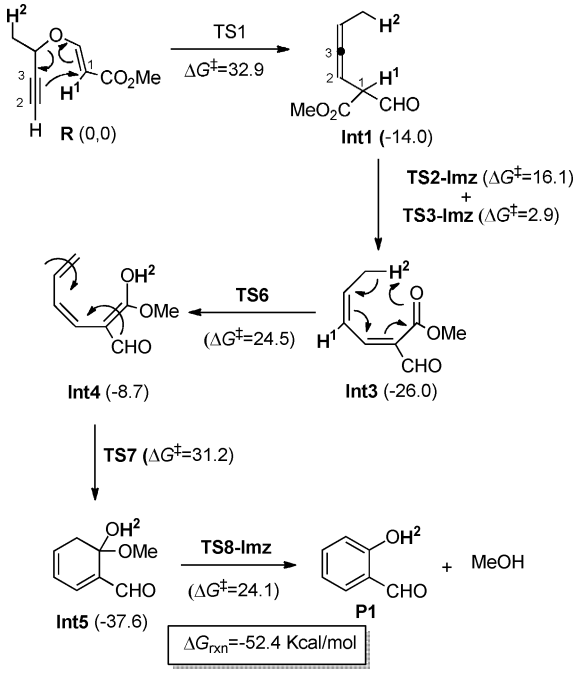


Scheme 5. Short synthesis of morintrifolin B.

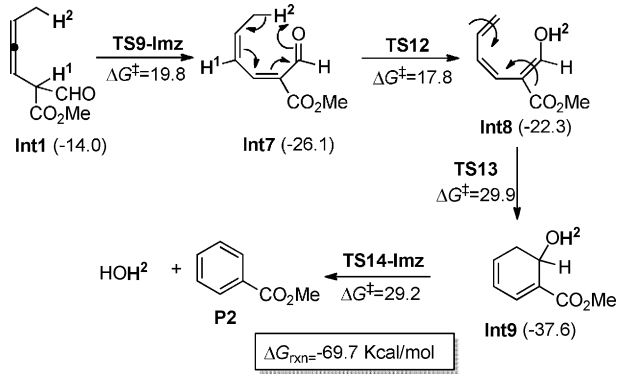
synthetic extension to other structurally related benzophenones and studies on their biological activity are in progress.

Computational studies on the reaction mechanism

To provide new insight into the reaction mechanism for the rearrangement of the propargyl vinyl ethers, we performed a computational DFT study using a simplified model denoted **R** in Scheme 6, where $R^1=R^2=R^4=H$ and $R^3=CH_3$. Some hydrogen-transfer processes (enolization and [1,3]-H shifts) show unrealistically[32] high free-energy barriers (ca. 55 kcal mol^{-1}). In these cases one imidazole molecule can act as a catalyst by taking advantage of its amphoteric character. The names of these stationary points end with "**-Imz**" to indicate the use of a molecule of imidazole as a suitable additive. All stationary points on the potential surface were fully optimized with the hybrid density functional B3LYP[33] and the 6-311+G(d,p) basis set.[34] Harmonic analyses were performed at this level of theory to verify the nature of the corresponding stationary points (minima or transition states), as well as to provide the zero-point vibrational energy and the thermodynamic contributions to the enthalpy and free energy for $T=298$ K. Moreover, intrinsic reaction coordinate[35] calculations were performed to ensure that the transition states connect the reactants and products belonging to the reaction coordinate under study. The final energies were obtained by performing single-point M06-2X[36] calculations with the 6-311+G(d,p) basis set at the optimized B3LYP geometries. The average difference between the B3LYP and M06-2X increments of energy was 3.1 kcal mol^{-1}. The raw data obtained with the B3LYP and M06-2X results are reported in the Supporting Information (see Tables S1–S5 in the Supporting Information) along with the Cartesian coordinates of each stationary point (see Table S6 in the Supporting Information). The values discussed in the text and the values reported in Schemes 6 and 7 are relative free energies evaluated at the M06-2X/6-311+G(d,p)//B3LYP/6-311+G(d,p) level of theory. All quantum chemistry calculations in this work were carried out with the Gaussian 09 program package.[37] The complete list of energies is reported in Tables S1–S5 of the Supporting Information. Exhaustive potential-energy surface scans with all the reaction mechanisms studied and selected geometrical parameters are shown in Figures S1–S15 of the Supporting Information. The transition states and intermediates of the studied domino rearrangement show several conformers with very similar energies (e.g., rotation about sp^3 carbon atoms and isomerization of double bonds).



Scheme 6. Reaction pathway for the domino process leading to the formation of salicylaldehyde derivative **P1**. Transition-state descriptors denoted **TS-Imz** refer to saddle points assisted by a molecule of imidazole (see text). Relative free energies are given in kcal mol^{-1}.



Scheme 7. Reaction mechanism for the domino process leading to the formation of methyl benzoate **P2**. Transition states including the **Imz** descriptor refer to saddle points assisted by a molecule of imidazole (see text). Relative free energies are given in kcal mol^{-1}.

The domino reaction starts with a thermally allowed [3,3]sigmatropic process of **R** to yield **Int1** (Scheme 6). This propargyl Claisen rearrangement shows an activation free energy barrier (**TS1**) of 32.9 kcal mol^{-1}, which is the highest activation energy along the alternative reaction paths (see below). As expected, **TS1** involves the cleavage and formation of a C–O and a C–C bond, respectively, as well as the conversion of a propargyl and an enol group into an allene and a carbonyl group, respectively (Figure 1 and Figure S1 of the Supporting Information). Allenyl intermediate **Int1** lies about 14 kcal mol^{-1} below **R**. Several conformations of **Int1** can be considered, and in this work we calculated six of them (**Int1a** to **Int1f**) with very similar relative energies (differences smaller than 2 kcal mol^{-1}, see the Supporting Information).

The next step of the domino process consists of conversion of allene **Int1** to unsaturated ester **Int3** (Scheme 6). This process formally corresponds to a thermal antarafacial [1,3] sigmatropic shift involving hydrogen atom H^1 and allyl scaffold C^1–C^3. The direct process is not geometrically favored, since the activation energy for this direct process is about 63 kcal mol^{-1}

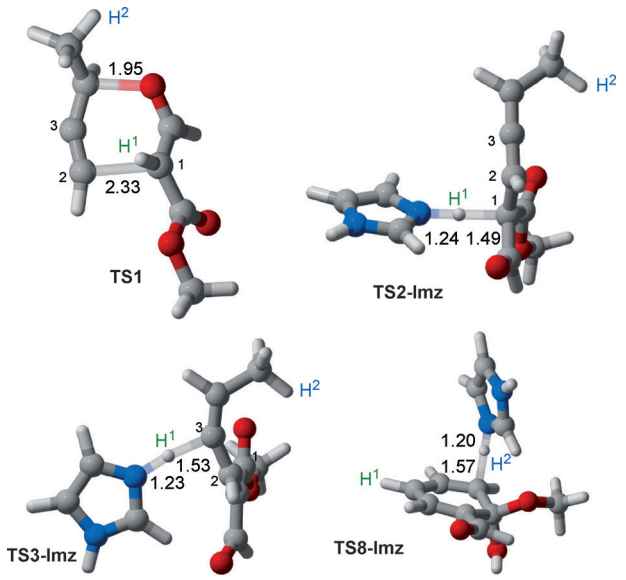


Figure 1. Fully optimized structures of saddle points **TS1** associated with [3,3] propargyl Claisen rearrangement, **TS2-Imz** and **TS3-Imz** corresponding to [1,3] hydrogen shift, and **TS8-Imz** associated with the aromatization process (see Scheme 6). Bond lengths are given in ångstrom.

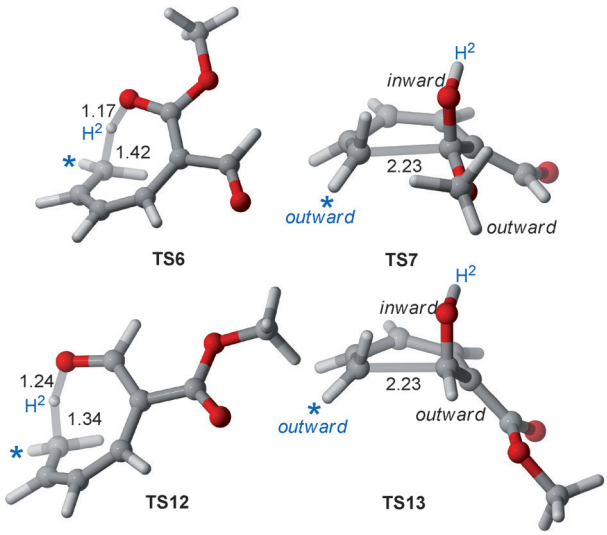


Figure 2. Fully optimized structures of saddle points **TS6**,**12**, associated with [1,7] sigmatropic shifts, and **TS7**,**13**, corresponding to [$\pi 6_d$] electrocyclizations (see Scheme 6 and Scheme 7). Bond lengths are given in ångstrom. Stars indicate the position of R^2 groups in more substituted transition structures (see Scheme 4).

(see the Supporting Information). However, our calculations indicate that this 1,3-prototropy is feasible under imidazole assistance. Thus, imidazole can remove proton $\mathbf{H^1}$ from **Int1** via **TS2-Imz** and transfer it to C^3 via **TS3-Imz** (see Figure 1 and the Supporting Information for additional details) with overall activation energy of about 20 kcal mol^{-1}, since the intermediate ion pair is almost isoenergetic with **TS3-Imz**. An alternative route involving the corresponding enol-allene intermediates **Int2** via tautomeric equilibria involving one molecule of imidazole can be also considered (see **TS4**, **TS5**, and **TS4-Imz** in Figures S3 and S6 of the Supporting Information).

Oxatriene **Int3** can be transformed into triene **Int4** via a thermally allowed antarafacial [1,7] sigmatropic shift. It is noteworthy that the geometry of saddle point **TS6** associated with this process permits an alternative pseudopericyclic process[38] involving one of the lone pairs of the sp^2-hybridized oxygen atom of **Int3** (Figure 2) to be excluded. The activation energy of this sigmatropic shift is lower than the activation energy associated with the first suprafacial [3,3] pericyclic reaction via **TS1** (see above).

From **Int4** the reaction proceeds through a thermal disrotatory electrocyclization to yield 1,3-cyclohexadiene intermediate **Int5** (Scheme 6). The features of saddle point **TS7** associated with this pericyclic step are those expected for a [$\pi 6_d$] process, in which the hydroxyl group $\mathbf{OH^2}$ formed in the previous step is necessarily inward (pseudoaxial) with respect to the C–C bond being formed (Figure 2). It results in destabilizing four-electron torquoelectronic interactions and steric congestion for this transition structure.[39] As a consequence, this step shows a relatively high activation energy, the second highest found in our calculations along the whole reaction coordinate. It is likely that this step and the first one associated with the [3,3] sigmatropic shift via **TS1** (see above) are responsible for the necessity of dielectric heating to complete these reactions. However, the effect of R^2 substituents in more complex substrates (see Scheme 4) is not critical, since these groups will always occupy the outward position in the corresponding [$\pi 6_d$] transition structures (see the stars in Figure 2).[40]

Finally, **Int5** can be subjected to an aromatization process to produce salicylaldehyde, **P1** with concomitant release of a methanol molecule (see Scheme 6, Figure 1 and Figures S4 and S7 of the Supporting Information). In this reaction step, the assistance of one molecule of imidazole is required to achieve a feasible activation energy ($\Delta G^{\ddagger} = 24.1$ kcal mol^{-1}, **TS8-Imz**). This final step is strongly exergonic and irreversible, as would be expected from the high resonance energy of the phenyl group generated in the course of the elimination/aromatization reaction.

An intriguing aspect of these reactions is the formation of methyl benzoates **15** in several polycyclic systems (see above, Scheme 3). Our computational model indicates that methyl benzoate (**P2** in Scheme 7) can be formed via oxatriene **Int7**, which differs with respect to **Int3** in the *E* configuration of the C^1–C^2 double bond. Despite this difference, **Int7** is almost isoenergetic with **Int3**, but its imidazole-assisted formation via **TS9-Imz** or (*Z*)-**Int6** (which in turn involves saddle points **TS10-Imz** and **TS11**, see Figures S9 and S12 of the Supporting Information) is less favored than in the previous case, which is in line with the minor formation of benzoates **15** in most cases (Scheme 3). Beyond **Int7**, the previously described antarafacial [1,7] sigmatropic shift leads to triene **Int8**, disrotatory electrocyclization of which leads in turn to cyclohexadienyl alcohol **Int9**. The relatively high energy of this pericyclic reaction is as expected for a transition structure involving an inward $\mathbf{OH^2}$ group (see the optimized structure of **TS13** in Figure 2). However, once again, R^2 groups in more substituted systems should occupy outward positions, which in general should not

jeopardize this reaction path. Finally, the imidazole-assisted elimination/aromatization step from **Int9** leads to methyl benzoate **P2** with an energy barrier significantly higher than those found for **P1** (see Schemes 6 and 7). This result is in nice agreement with the experimentally found formation of benzoates **15** as minor products, especially in the case of polycyclic systems.

Conclusion

We have developed a catalytic version of the MW-assisted formation of polysubstituted salicylaldehydes from propargyl vinyl ethers. The reaction manifold uses imidazole as the catalyst (10 mol%) to deliver an array of topologically diverse salicylaldehyde scaffolds spanning from simple aromatic monocycles to complex fused polycyclic systems. The reaction is scalable and instrumentally simple to perform, highly regioselective, and symmetry-disrupting: symmetrically substituted PVEs afforded asymmetrically (nonredundant) substituted salicylaldehydes. The preparative value of this transformation has been demonstrated in the five-step synthesis of the benzophenone-derived natural product morintrifolin B. A DFT study on a simplified model was performed. Calculations underpin a domino mechanism comprising a [3,3] propargyl Claisen rearrangement/[1,3]-hydrogen shift/[1,7]-hydrogen shift/6π electrocyclization/aromatization process. The use of imidazole lowers the energy of the two more difficult steps, that is, the 1,3-prototropic rearrangement and the final aromatization step, otherwise energetically very disfavored.

Experimental Section

General remarks

^{1}H NMR and ^{13}C NMR spectra of $CDCl_3$ solutions were recorded at 400 and 100 MHz or at 500 and 125 MHz (Bruker AC 200 and AMX2–500), respectively. MW reactions were conducted in sealed glass vessels (capacity 10 mL) with a CEM Discover MW reactor. FTIR spectra were measured on chloroform solutions with a PerkinElmer Spectrum BX FTIR spectrophotometer. Mass spectra (low-resolution EI/CI) were obtained with a Hewlett-Packard 5995 gas chromatograph/mass spectrometer. High-resolution mass spectra were recorded with a Micromass Autospec mass spectrometer. Microanalyses were performed with a Fisons Instruments EA 1108 carbon, hydrogen, and nitrogen analyzer. Analytical TLC was performed with E. Merck Brinkman UV-active silica gel (Kieselgel 60 F254) on aluminum plates. Flash column chromatography was carried out with E. Merck silica gel 60 (particle size less than 0.020 mm) and appropriate mixtures of ethyl acetate and hexanes, or ethyl acetate and dichloromethane as eluents. All reactions were performed in oven-dried glassware. All starting materials were obtained from commercial suppliers and used as received. PVEs **1v**, **1w**, **1x**, **1y**, and **1z** partially rearranged during the isolation and characterization process, so an alternative procedure was used to prepare **14v**, **14w**, **14x**, **14y**, and **14z** (see below).

Synthesis

Representative procedure for the MW-assisted synthesis of salicylaldehydes from the corresponding propargyl vinyl ethers: synthesis of 1-hydroxy-3-phenyl-5,6,7,8-tetrahydronaphthalene-2-carbaldehyde (14e): Propargyl vinyl ether **1e** (1.0 mmol) and imidazole (0.10 mmol) in dry xylene (1 mL) were placed in a MW closed vial and the solution was irradiated for 1 h in a single-mode MW oven (300 W, 190 °C). After removing the solvent at reduced pressure the products were purified by flash column chromatography (silica gel, *n*-hexane/EtOAc 95/5) to yield **14e** as an amorphous solid (229.6 mg, 91%). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 1.71–1.75 (m, 4H), 2.62–2.65 (m, 2H), 2.68–2.71 (m, 2H), 6.52 (s, 1H), 7.24–7.26 (m, 2H), 7.30–7.35 (m, 3H), 9.67 (s, 1H), 12.27 ppm (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 22.13, 22.25, 22.39, 30.5, 115.3, 122.1, 125.3, 127.9, 128.2 (2C), 130.0 (2C), 137.8, 143.6, 147.3, 161.2, 196.6 ppm; IR ($CHCl_3$): $\tilde{\nu}$ = 2939.6, 1635.7, 1617.2, 1559.6, 1405.3, 1366.7, 1299.4 cm^{-1}; LRMS (70 eV) *m/z* (%): 252 (100) [M^+], 251 (42), 234 (30), 233 (18), 223 (12), 165 (17); HRMS (EI-TOF): *m/z* calcd for $C_{17}H_{16}O_2$: 252.1150 [M]$^+$; found: 252.1144.

Representative telescoped procedure for the synthesis of salicylaldehydes 14v, 14w, 14x, 14y, and 14z from the corresponding tertiary propargylic alcohols: synthesis of 1-hydroxy-3-phenyl-9,10-dihydrophenanthrene-2-carbaldehyde (14v): The corresponding propargyl alcohol (1.0 mmol) and DABCO (0.10 mmol) were dissolved in hexane (1 mL); a small amount of CH_2Cl_2 was used if the alcohol did not dissolved well in hexane. Methyl propiolate was slowly added dropwise (1.5 mmol) and the reaction mixture was stirred for 5 min. The reaction mixture was filtered through a short column filled with silica gel with *n*-hexane/EtOAc (60/40). The solvent was evaporated off and the mixture was dissolved in xylenes (1 mL) and transferred to a MW closed vial. Imidazole (0.10 mmol) was added and the reaction mixture was irradiated for 1 h in a single-mode MW oven (300 W, 190 °C). After removing the solvent at reduced pressure the products were purified by flash column chromatography (silica gel, *n*-hexane/EtOAc 95/5) to give **14v** as an amorphous solid (153.2 mg; 51%). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 2.89–2.93 (m, 2H), 2.97–3.01 (m, 2H), 7.28–7.34 (m, 4H), 7.42–7.51 (m, 5H), 7.76–7.79 (m, 1H), 9.83 (s, 1H), 12.36 (s, 1H) ppm; ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ 19.8, 28.1, 116.7, 117.0, 124.4, 124.9, 127.0, 128.1, 128.39 (2C), 128.41, 129.2, 130.1 (2C), 133.1, 138.0, 138.6, 142.1, 145.4, 160.1, 196.6 ppm; IR ($CHCl_3$): $\tilde{\nu}$ = 2945.1, 2896.7, 2844.7, 1633.7, 1617.8, 1545.8, 1479.8, 1400.2, 1363.5, 1331.6, 1307.2, 1288.7, 1239.0 cm^{-1}; LRMS (70 eV): *m/z* (%): 300 (100) [M^+], 299 (41), 282 (27), 281 (32), 253 (16), 252 (21), 239 (12), 165 (15). HRMS (EI-TOF): *m/z* calcd for $C_{21}H_{16}O_2$: 300.1150 [M]$^+$; found: 300.1142. Further elution delivered pure methyl 3-phenyl-9,10-dihydrophenanthrene-2-carboxylate (**15v**) as an amorphous solid (31.4 mg; 10%). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 2.93 (br s, 4H), 3.64 (s, 3H), 7.24–7.32 (m, 3H), 7.35–7.44 (m, 5H), 7.33 (d, ^3J(H,H) = 7.1 Hz, 2H), 7.77 ppm (t, ^3J(H,H) = 7.1 Hz, 1H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 28.5, 28.8, 51.8, 124.2, 126.1, 127.10, 127.12, 127.8 (2C), 128.3, 128.41 (3C), 129.1, 129.7, 133.4, 136.2, 137.4, 137.9, 141.6, 141.7, 168.8 ppm; IR ($CHCl_3$): $\tilde{\nu}$ = 3028.8, 2951.1, 2841.6, 1716.4, 1603.0, 1436.2, 1305.8, 1252.4 cm^{-1}; LRMS (70 eV): *m/z* (%): 314 (57) [M^+], 283 (24), 252 (10), 239 (10), 170 (13), 169 (100), 141 (16), 115 (17), 91(20). HRMS (EI-TOF): *m/z* calcd for $C_{22}H_{18}O_2$: 314.1307 [M]$^+$; found: 314.1305.

Acknowledgements

This research was supported by the Spanish Ministerio de Economía y Competitividad (MINECO) and the European Regional Development Fund (CTQ2011-28417-C02-02 and CTQ2013-45415-P), the UPV/EHU (UFI11/22 QOSYC), and the

Basque Government (GV/EJ, grant IT-324-07). L.C. thanks Spanish MEC for a FPI grant and D.M.A. thanks CSIC for a JAE-intro grant. M.T.-S., M.O.-G., and F.P.C. thank the SGI/IZO- SGIker UPV/EHU and the DIPC for generous allocation of computational resources.

Keywords: domino reactions · microwave chemistry · organocatalysis · synthetic methods · total synthesis

[1] a) D. Tejedor, G. Méndez-Abt, L. Cotos, F. García-Tellado, *Chem. Soc. Rev.* **2013**, *42*, 458 – 471; b) Z.-B. Zhu, S. F. Kirsch, *Chem. Commun.* **2013**, *49*, 2272 – 2283; c) for other reactivity profile, see: D. Tejedor, D. González-Cruz, F. García-Tellado, J. J. Marrero-Tellado, M. L. Rodríguez, *J. Am. Chem. Soc.* **2004**, *126*, 8390 – 8391.

[2] S. Ma, *Chem. Rev.* **2005**, *105*, 2829 – 2871.

[3] D. K. Black, S. R. Landor, *J. Chem. Soc.* **1965**, 6784 – 6788.

[4] a) L. Claisen, *Chem. Ber.* **1912**, *45*, 3157 – 3167 ; b) *The Claisen Rearrangement*, (Eds.: M. Hiersemann, U. Nubbemeyer), Wiley-VCH, Weinheim, **2007**.

[5] a) For the first metal-catalyzed protocol, see: B. D. Sherry, F. D. Toste, *J. Am. Chem. Soc.* **2004**, *126*, 15978 – 15979; For other metal-catalyzed rearrangements, see: b) E. S. da Silva Júnior, N. Maria de Araujo, F. S. Emery, *J. Heterocycl. Chem.* **2015**, *52*, 518 – 521; c) A. Palisse, S. F. Kirsch, *Eur. J. Org. Chem.* **2014**, 7095 – 7098; d) D. V. Vidhani, J. W. Cran, M. E. Krafft, I. V. Alabugin, *Org. Biomol. Chem.* **2013**, *11*, 1624 – 1630; e) H. Huang, H. Jiang, H. Cao, J. Zhao, D. Shi, *Tetrahedron* **2012**, *68*, 3135 – 3144; f) T. Harschneck, S. F. Kirsch, *J. Org. Chem.* **2011**, *76*, 2145 – 2156; g) H. Wei, Y. Wang, B. Yue, P.-F. Xu, *Adv. Synth. Catal.* **2010**, *352*, 2450 – 2454; h) K. Tanaka, E. Okazaki, Y. Shibata, *J. Am. Chem. Soc.* **2009**, *131*, 10822 – 10823; i) J. T. Binder, S. F. Kirsch, *Org. Lett.* **2006**, *8*, 2151 – 2153; j) H. Menz, S. F. Kirsch, *Org. Lett.* **2006**, *8*, 4795 – 4797; k) B. D. Sherry, L. Maus, B. N. Laforteza, F. D. Toste, *J. Am. Chem. Soc.* **2006**, *128*, 8132 – 8133; l) M. H. Suhre, M. Reif, S. F. Kirsch, *Org. Lett.* **2005**, *7*, 3925 – 3927; for selected examples of Claisen rearrangements of *N*-propargyl amines, see: m) A. Saito, A.i Kanno, Y. Hanzawa, *Angew. Chem. Int. Ed.* **2007**, *46*, 3931 – 3933; *Angew. Chem.* **2007**, *119*, 4005 – 4007; n) A. Saito, S. Oda, H. Fukaya, Y. Hanzawa, *J. Org. Chem.* **2009**, *74*, 1517 – 1524.

[6] For selected examples of all-sp^2 C_5 building blocks, see: a) D. V. Vidhani, M. E. Krafft, I. V. Alabugin, *J. Org. Chem.* **2014**, *79*, 352 – 364; b) D. V. Vidhani, M. E. Krafft, I. V. Alabugin, *Org. Lett.* **2013**, *15*, 4462 – 4465; c) T. Sakaguchi, Y. Okuno, Y. Tsutsumi, H. Tsuchikawa, S. Katsumura, *Org. Lett.* **2011**, *13*, 4292 – 4295; d) C. D. Vanderwal, *J. Org. Chem.* **2011**, *76*, 9555 – 9567 and references cited therein.

[7] The parent PVEs armed with an ester or tertiary amide at the propargylic position are synthesized by the tertiary amine catalyzed ABB′ three-component reaction between a 1,2-ketoester (1,2-ketoamide) and methyl propiolate. D. Tejedor, A. Santos-Expósito, F. García-Tellado, *Chem. Eur. J.* **2007**, *13*, 1201 – 1209. For a tutorial review of ABB′ three-component reactions, see: D. Tejedor, F. García-Tellado, *Chem. Soc. Rev.* **2007**, *36*, 484 – 491.

[8] D. Tejedor, L. Cotos, F. García-Tellado, *Org. Lett.* **2011**, *13*, 4422 – 4425.

[9] D. Tejedor, G. Méndez-Abt, L. Cotos, M. A. Ramirez, F. García-Tellado, *Chem. Eur. J.* **2011**, *17*, 3318 – 3321.

[10] D. Tejedor, L. Cotos, F. García-Tellado, *J. Org. Chem.* **2013**, *78*, 8853 – 8858.

[11] a) M. Hasegawa, F. Ono, S. Kanemasa, *Tetrahedron Lett.* **2008**, *49*, 5220 – 5223; b) C. Palomo, R. Pazos, M. Oiarbide, J. M. Garcia, *Adv. Synth. Catal.* **2006**, *348*, 1161 – 1164; c) T. P. Yoon, E. N. Jacobsen, *Angew. Chem. Int. Ed.* **2005**, *44*, 466 – 468; *Angew. Chem.* **2005**, *117*, 470 – 472; d) L. M. Weinstock, S. Karady, F. E. Roberts, A. M. Hoinowski, G. S. Brenner, T. B. K. Lee, W. C. Lumma, M. Sletzinger, *Tetrahedron Lett.* **1975**, *16*, 3979 – 3982.

[12] D. Tejedor, G. Méndez-Abt, L. Cotos, F. García-Tellado, *Chem. Eur. J.* **2012**, *18*, 3468 – 3472.

[13] a) *Coumarins: Biology, Applications and Mode of Action*, (Eds.: R. O. Kennedy, R. D. Thornes), Wiley, New York, **1997**; b) *Flavonoids: Chemistry, Biochemistry and Applications*, (Eds.: O. M. Andersen, K. R. Markham), CRC, Boca Raton, **2006**; c) A. Gaspar, M. J. Matos, J. Garrido, E. Uriarte, F. Borges, *Chem. Rev.* **2014**, *114*, 4960 – 4992; d) R. Pratap, V. J. Ram, *Chem. Rev.* **2014**, *114*, 10476 – 10526; e) Salicylaldehydes rearrange to catechols under oxidative conditions (Dakin rearrangement). For a selected recent example, see: S. Chen, M. S. Hossain, F. W. Foss, *Org. Lett.* **2012**, *14*, 2806–2809 and references cited therein; f) S. Bräse, A. Encinas, J. Keck, C. F. Nising, *Chem. Rev.* **2009**, *109*, 3903 – 3990.

[14] For selected reviews, see: a) K. C. Gupta, A. K. Sutar, *Coord. Chem. Rev.* **2008**, *252*, 1420 – 1450; b) J. F. Larrow, E. N. Jacobsen, *Top. Organomet. Chem.* **2004**, *6*, 123 – 152; c) T. Katsuki, *Chem. Soc. Rev.* **2004**, *33*, 437 – 444.

[15] a) H. Wynberg, *Comp. Org. Syn.* **1991**, *2*, 769 – 775; b) J. C. Duff, E. J. Bills, *J. Chem. Soc.* **1934**, 1305 – 1308; c) For a representative example with preparative value: J. F. Larrow, E. N. Jacobsen, *Org. Synth.* **1998**, *75*, 1 – 11.

[16] G. Cozzi, *Chem. Soc. Rev.* **2004**, *33*, 410 – 421.

[17] Y. Deng, Y. W. Chin, H. Chai, W. J. Keller, A. D. Kinghorn, *J. Nat. Prod.* **2007**, *70*, 2049 – 2052.

[18] For a recent review, see: R. A. Sheldon, *Chem. Soc. Rev.* **2012**, *41*, 1437 – 1451.

[19] The scope of the synthesis of salicylaldehydes **9** from PVEs is further demonstrated in our initial communication.[9,10]

[20] a) D. A. Alonso, C. Nájera, I. M. Pastor, M. Yus, *Chem. Eur. J.* **2010**, *16*, 5274 – 5284; b) J. H. P. Tyman, in *Synthetic and Natural Phenols*, Elsevier, New York, **1996**; c) Z. Rappoport, in *The Chemistry of Phenols*, Wiley-VCH, Weinheim, **2003**.

[21] a) A. Estévez-Braun, A. G. González, *Nat. Prod. Rep.* **1997**, *14*, 465 – 475; b) F. Borges, F. Roleira, N. Milhazes, L. Santana, E. Uriarte, *Curr. Med. Chem.* **2005**, *12*, 887 – 916; c) A. Behrenswerth, N. Volz, J. Toräng, S. Hinz, S. Bräse, C. E. Müller, *Bioorg. Med. Chem.* **2009**, *17*, 2842 – 2851; d) E. Jijy, P. Prakash, M. Shimi, P. M. Pihko, N. Joseph, K. V. Radhakrishnan, *Chem. Commun.* **2013**, *49*, 7349 – 7351.

[22] For example, 5-hydroxylated coumarins (position C^5 in coumarins correlates with position C^6 in the parent salicylaldehyde) are endowed with potent nematicidal activity against the phytopathogenic nematode *Bursaphelenchus xylophilus*: K. Takaishi, M. Izumi, N. Baba, K. Kawazu, S. Nakajima, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5614 – 5617.

[23] See the Supporting Information for their structures and synthesis.

[24] D. Tejedor, S. J. Álvarez-Méndez, J. M. López-Soria, V. S. Martín, F. García-Tellado, *Eur. J. Org. Chem.* **2014**, 198 – 205.

[25] For selected reviews on cyclobutarenes and related compounds, see: a) A. K. Sadana, R. K. Saini, W. E. Billups, *Chem. Rev.* **2003**, *103*, 1539 – 1602; b) G. Mehta, S. Kotha, *Tetrahedron* **2001**, *57*, 625 – 659.

[26] a) E. Marsault, M. L. Peterson, *J. Med. Chem.* **2011**, *54*, 1961 – 2004; b) E. M. Driggers, S. P. Hale, J. Lee, N. K. Terrett, *Nat. Rev. Drug Discovery* **2008**, *7*, 608 – 624.

[27] a) C. M. Marson, *Chem. Soc. Rev.* **2011**, *40*, 5514 – 5533; b) Substituted tetrahydroisoquinolines as β-secretase inhibitors: L. A. Thompson, K. M. Boy, J. Shi, J. E. Macor, A. C. Good, L. R. Marcin, US 20080153868A1; c) Hydroxy-substituted 1,2,3,4-tetrahydroisoquinolines as inhibitors of phenylethanolamine *N*-methyltransferase (PNMT): D. J. Sall, G. L. Grunewald, *J. Med. Chem.* **1987**, *30*, 2208 – 2216.

[28] M. A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casaulta, A. Odermatt, P. Ertl, H. Waldmann, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17272 – 17277.

[29] S. Baggett, E. P. Mazzola, E. J. Kennelly, in *Studies in Natural Products Chemistry, Vol. 32*, Part L, (Ed.: Atta-ur-Rahman), Elsevier, Amsterdam, **2005**, pp. 721.

[30] S.-B. Wu, C. Long, E. Kennelly, *Nat. Prod. Rep.* **2014**, *31*, 1158 – 1117.

[31] O. Potterat, M. Hamburger, *Planta Med.* **2007**, *73*, 191 – 199.

[32] a) R. E. Plata, D. A. Singleton, *J. Am. Chem. Soc.* **2015**, *137*, 3811 – 3826; b) A. Winter, *Nat. Chem.* **2015**, *7*, 413 – 415.

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

[34] a) W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257 – 2261; b) W. J. Hehre, L. Radom, P. von R. Schleyer, J. A. Pople, in *Ab Initio Molecular Orbital Theory*, Wiley, New York, **1986**.

[35] a) K. Ishida, K. Morokuma, A. Kormornicki, *J. Chem. Phys.* **1977**, *66*, 2153 – 2156; b) C. Gonzalez, H. B. Schlegel, *J. Chem. Phys.* **1989**, *90*, 2154 – 2161; c) C. Gonzalez, H. B. Schlegel, *J. Phys. Chem.* **1990**, *94*, 5523 – 5527.

[36] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215 – 241.

[37] Guassian 2009, Revision B.01, 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. Haseg-awa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Per-alta, 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, J. M. 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. Moroku-ma, 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, D. J. Fox, Gaussian, Inc., Wallingford CT, **2010**.

[38] a) P. von R. Schleyer, J. I. Wu, F. P. Cossío, I. Fernández, *Chem. Soc. Rev.* **2014**, *43*, 4909 – 4921; b) D. Martin Birney, *Curr. Org. Chem.* **2010**, *14*, 1658 – 1668.

[39] a) J. D. Evanseck, B. E. Thomas IV, D. C. Spellmeyer, K. N. Houk, *J. Org. Chem.* **1995**, *60*, 7134 – 7141; b) V. A. Guner, K. N. Houk, I. W. Davies, *J. Org. Chem.* **2004**, *69*, 8024 – 8028; c) T.-Q. Yu, Y. Fu, L. Liu, Q.-X. Guo, *J. Org. Chem.* **2006**, *71*, 6157 – 6164.

[40] Another conceivable reaction mechanism involves the formation of a furan intermediate (see Scheme S1 and Figure S14 in the Supporting Information). This mechanism was discarded on the basis of its global free-energy barrier of 57.5 kcal mol^{-1} with respect to **R**.

Received: August 11, 2015
Published online on October 30, 2015

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Proteinak eta DNA Funtzio Katalitiko Zein Biologikoen Sustatzaile Gisa: *Huisgenasetatik* Kimioterapiara

Doktoretza Tesia

Mikel Odriozola Gimeno

Donostia/San Sebastián

2019

Tengo claro que después de estos cinco años sigo siendo la misma persona, con unas cuantas experiencias más, con algunas mejoras y con algunas cosas que habré empeorado, pero al fin y al cabo, la misma persona. Tengo mucha gente a la que agradecer su presencia y apoyo a lo largo de este periodo, no voy a entrar en nombres, no quiero ni olvidarme de nadie ni que sea ni un apartado para curiosos.

Me gustaría empezar por el inicio, dando gracias a las personas que me dieron la oportunidad de hacer la tesis en este laboratorio y a la gente que me ha supervisado durante todo este periodo. Del mismo modo, no me quiero olvidar de nadie del laboratorio/grupo, compañeros de vitrina y despacho, experimentales, computacionales, gente de estancia, estudiantes… Con algunos he compartido más momentos que con otros, pero sin duda todos han sido importantes para llevar el día a día. He tenido la suerte de poder hacer un trabajo multidisciplinar en el que hemos colaborado muchas personas, las cuales han aportado para que todo salga adelante. A su vez he coincidido con gente de muchos sitios y ámbitos en el Korta, tanto científicos como no. GRACIAS A TODOS.

During my PhD I had the chance to spend three months in the laboratory of Prof. Houk at UCLA. Without any doubt that experience was one of the best experiences of my life. Thus, I would like to thank all the people that I met during that period: supervisors, labmates, basketball coaches, roommates, friends… I am really aware about the importance of that stay in my PhD, It was a really enriching experience. THANKS TO EVERYBODY.

Bukatzeko euskeraz ekingo diot; argi daukat arlo zientifikoa ezinbestekoa dela doktoradutza burutzeko, baina bihotzez uste det askotan ahazten garela gainontzeko gauzetaz. Denbora askoan korrika ibiltzen gara, lanez gainezka eta ez det gezurrik esan nahi, asko ikasi det zalantza izpirik gabe, momentu on ugari izan ditut, jende asko ezagutu, experientzia onuragarri asko gehitu, baina dena ez da izan lorez beteriko bidea. Momentu txarrak ere izan dira, zalantza ugariko momentuak eta antsietate momentuak, batez ere azken urtean, azkenean, ez gera manual batekin jaio. Tesiak ordu asko eskatu izan dizkit, baina argi neukan eta orain argiago, laborategitik irtetzean inguruan jendea edukitzea eta beste jarduera batzuk egitea ezinbestekoa dela.

Egia esan balantze positiboa egiten det eta harro nago honetaz, baina balantze hori ez zen posible izango inguruko jende guztia gabe. Gainera ziur laborategitik atera naizen guztietan nere umorea ez zela hoberena eta jende horrek aguantau behar izan nau, bai irrifartsu eta bai serioago. Lehenik eta behin, FAMILIA, batez ere ama, aita eta anaia, beraiek izan dira egunero aguantatu nauetenak bai momentu onetan zein zailetan, maite zaituztet. Sin duda también al resto de la familia que me habéis apoyado en este proceso, aunque no lo diga agradezco mucho teneros. Bestalde, LAGUNAK (kuadrila eta gainontzekoak, lehendik zeudenak eta bidean eginikoak), denbora asko kendu diet hauei ere liburuxka hau idazteagatik, eskerrak ezagutzen nauten eta ulertzen duten, asko eskertzen dizuet zuen laguntza. Askotan, kalimotxo bat alde zaharrean hartzea edo donostitik bueltatxo bat ematea deskonetatzeko aukera ezin hobeak izan dira niretzat. Ze zortea dudan. Saskibaloiko Bostekoko taldekoei, disfrutatzen laguntzeagatik, titulu eta guzti. Entrenatu ditudan gazte guztiei, beren familiei eta klubeko guztiei, beraien laguntza eta nigan euki duten konfidantza guztiagatik. Selekziokoei, aukera ezin hobeak izan ditut eta deskonetatzeko ezin hobe. Apuntatu naizen aktibitate guztietakoei eta bost urte hauetan nerekin koinziditu dezuten guztioi. Momentu baxuak aparte gauza guzti hauek zoriontsu egin naute eta ziur urte batzuk barru atzera begiratzean irribarre bat aterako zaidala. Azken finean doktorea izango naiz orain, baina ze zortea dudan inguruan ditudan pertsonak edukitzeaz. MILA ESKER DENORI.

Doktoretza Tesi hau Kimika Fakultateko Kimika Organiko I Sailan burutu da, Donostian, Iván Rivilla de la Cruz eta Miquel Torrent Sucarrat Doktoreen gidaritzapean eta Fernando P. Cossío Mora Katedradunaren tutoretzarekin.

2016eko iraila eta abendu artean ikerketa egonaldi laburra burutu zen, Los Angeleseko UCLA unibertsitateko Kimika eta Biokimika Sailean Kendall N. Houk Katedradunaren eta Marc García Borrás doktorearen gidaritzapean.

Doktoretza Tesi hau hiru kapitulutan banatzen da. 1.Kapituluan hurrengo bi kapituluen ulermenerako lagungarria den sarrera orokor bat azaltzen da. Lehenik 1.1 eta 1.2 ataletan, erreakzio periziklikoak katalizatzen dituzten proteinetan eta entzimen aktibitatea aztertzeko erabilgarria den QM/MM metodologian sakontzen da batez ere. Ondoren, 1.3 atalean minbiziaren trantamenduan platino(II) konplexuen erabilera azaltzen da, gehienbat hauen erreakzio mekanismoak eta minbizi zelulek garatutako erresistentzia mekanismoak aztertuz. 2.Kapituluan, (3+2) zikloadizioak katalizatzen dituzten CTPR proteinen aktibitatea aurkezten da, hauen *Huisgenasa* aktibitatea erakutsiz. 3.Kapituluan, azkenik, platino(II) konplexuen familia berri baten diseinua, sintesia eta hauen minbiziaren aurkako erabileraren ebaluaketa biologikoa aurkezten da.

Taula, irudi, eskema eta erreferentzien zenbaketa independientea da Kapitulu bakoitzarentzat. Euskerazko atal honetan, ingelesez eskainitako hiru Kapituluak itzulita azaltzen dira, zati esperimental zein eranskinetan eskeintzen den informazioa ordea soilik ingeleseko atalean aurkitzen delarik.

ARGITALPEN ZERRENDA

1.Microwave-Assisted Organocatalyzed Rearrangement of Propargyl Vinyl Ethers to Salicylaldehyde Derivatives: An Experimental and Theoretical Study. Tejedor, D.; Cotos, L.; Márquez-Arce, D.; Odriozola-Gimeno, M.; Torrent-Sucarrat, M.; Cossío, F. P.; García-Tellado, F. *Chem. Eur. J.* **2015**, 21, 18280-18289.

2.Discovering Biomolecules with Huisgenase Activity: Designed Repeat Proteins as Biocatalysts for (3+2) Cycloadditions. Rivilla, I.; Odriozola-Gimeno, M.; Aires, A.; Gimeno, A.; Jiménez-Barbero, J.; Torrent-Sucarrat, M; Cortajarena, A. L.; Cossío, F.P. *under revision*.

3.Imidazopyrimidine compounds as photosensitizer for photodynamic therapy. Lima, M. L. S. O.; Braga, C.B.; Becher, T. B.; Odriozola-Gimeno, M.; Torrent-Sucarrat, M.; Rivilla, I.; Cossío, F.P.; Marsaioli, A. J.; Ornelas, C. *Submitted.*

4.The π-stacking as the key to understand the photoredox catalysis. Odriozola-Gimeno, M.; Rivilla, I.; Porcarelli, L.; Freixa, Z.; Huix-Rotllant, M.; Cossío, F.P.; Torrent-Sucarrat, M. *Manuscript in preparation*.

*5.*Design, synthesis and validation of new chemotherapeutic agents against cisplatin resistance cancers. *Patent in preparation*.

PROTEINAK ETA DNA FUNTZIO KATALITIKO ZEIN BIOLOGIKOEN SUSTATZAILE GISA: *HUISGENASETATIK* KIMIOTERAPIARA

2. KAPITULUA: *HUISGENASA* AKTIBITATEDUN BIOMOLEKULEN DISEINUA: CTPR PROTEINAK (3+2) ZIKLOADIZIOEN BIOKATALIZATZAILE GISA

3. KAPITULUA: PLATINO(II) KONPLEXU KIMIOTERAPEUTIKO BERRIEN DISEINU, SINTESI ETA ERABILPEN BIOLOGIKOA

AKRONIMO ETA LABURDURAK

A	Adenina
AcOEt	Etil azetatoa
AFM	Indar Atomikoaren Mikroskopia
ATP	Adenosina Trifosfatoa
BCL	B-zelulen Linfoma
BRCA	Bular Minbizia
CD	Dikroismo Zirkularra
CSP	Desplazamendu Kimikoaren Perturbazioa
CTPR	Kontsentsu Tetratrikopeptido Errepikapena
CTR	Kobre Garraitzailea
COSY	Korrelazio Espektroskopia
d	Dobletea (EMN)
DCM	Diklorometanoa
deg	graduak
DIPEA	N,N-diisopropiletil amina
DFT	Densitate Funtzionalaren Teoria
DMATS	Dimetilalil Triptofano Sintasa
DMF	N,N-Dimetil Formamida
DMSO	Dimetil Sulfoxidoa
DNA	Azido Desoxirribonukleiko
DPBS	Dulbecco's Phosphate-Buffered Saline
EC	Entzima Komisioa
EDCI	1-etil-3-(3-dimetilaminopropil)karbodiimida
EMN	Erresonantzia Magnetiko Nuklearra
eq.	Baliokide(ak)
ERCC	Excision-Repair Cross-Complementing
ESI	Elektroesprai Ionizazioa
EtOH	Etanola

Et_3N	Trietilamina
EWG	Talde elektroi erakarlea
FBS	Behi fetuko seruma
FDA	Elikagai eta Sendagaien Administrazioa
FDC	Azido Ferulikoaren Deskarboxilasa
FITC	Fluoreszeina Isotiozianatoa
FMN	Flabin-Mononukleotidoa
FTIR	Espektroskopia Infragorria
G	Guanina
GSH	Glutatioia
g.t	Giro tenperatura
h	Orduak
HER	Human Epidermal Growth Factor Receptor
Hex	Hexanoa
HOBt	1-Hidroxibenzotriazola
HRMS	Erresoluzio Handiko Masa Espektrometria
HSQC	Heteronuclear Single Quantum Coherence
Hz	Hertzio
IMOMM	Integraturiko Orbital Molekularra + Mekanika Molekularra
IPTG	Isopropil β-d-Tiogalaktosidoa
IR	Espektroskopia Infragorria
IUPAC	International Union of Pure and Applied Chemistry
J	Akoplamendu konstantea
LB	Luria-Bertani
LUMO	Lowest Unoccupied Molecular Orbital
m	Multipletea (EMN)
MALDI-TOF	Laserrez Lagundutako Matrize Desortzio Ionizazioa-Hegaldi Denbora
MAPK	Mitogenoz-Aktibaturiko Proteina Kinasa
MS	Masa espektrometria
MeOH	Metanola
min	Minutu

mL	Mililitro
MM	Mekanika Molekularra
MMR	DNA Base-Parekatze Okerreraren Konponketa
MOPS	Azido 3-(N-morfolino)propanosulfonikoa
MRE	Hondarren Eliptizate Molarra
mW	Mikrouhin
MRP	Multidrug Resistance Protein
NAD	Nikotinamida Adenina Dinukleotidoa
NADP	Nikotinamida Adenina Dinukleotido Fosfatoa
Nu	Nukleozale
OCT	Katioi Organikoen Garraiatzailea
OD	Dentsitate Optikoa
ONIOM	Own N-layer Integrated Molecular Orbital Molecular Mechanics
P	Produktua
PBS	Phosphate-Buffered Saline
PES	Energia Potentzialaren Gainazala
PM	Pisu Molekularra
PMS	Fenazina Metosulfatoa
PDB	Proteinen Datu-Bankua
QM	Mekanika Kuantikoa
QM/MM	Mekanika Kuantiko/Mekanika Molekular metodoak
R	Erreaktiboa
RNA	Azido Ribonukleikoa
RPMI	Roswell Park Memorial Institutua
SDS-PAGE	Sodio Dodezil Sulfato Poliakrilamida Gel Elektroforesia
SEM	Bateztesteko Errore Estandarra
SHx	Solbatzaio Helizea
s	Singletea (EMN)
STD	Saturazio Transferentzia Diferentzia

t	Tripletea (EMN)
TBTU	2-(1H-benzotriazol-1-il)-1,1,3,3-tetrametilaminio tetrafluoroboratoa
TEA	Trietilamina
TEM	Transmisio Mikroskopia Elektronikoa
THF	Tetrahidrofuranoa
TLC	Geruza meheko kromatografia
TMS	Trimetilsilila
TOF	Turnover frequency
TON	Turnover number
TPR	Tetratrikopeptido Errrepikapena
TROSY	Transverse Relaxation Optimized Spectroscopy
TS	Trantsizio-egoera
UDCA	Azido Ursodeoxikolikoa
WHO	Munduko Osasun Erakundea
WST-1	Uretan disolbagarria den Tetrazolio-1
wt	Jatorrizko tipo

1.Kapitulua:

Sarrera

Lehenengo kapitulu honetan, bigarren eta hirugarren kapituluan aurkezten diren emaitzen ulermenerako ezinbestekoak diren hainbat kontzeptu azaltzen dira. Lehenik, erreakzio perizikloak katalizatzen dituzten entzimen inguruko sarrera bat eskeintzen da, hauen artean, Diels-Alderasen eta *Huisgenasen* familietan sakontzen delarik. Hurrengo atalean, kimika teorikoak erabilitako metodo konputazionalei buruz hitz egiten da hauen aplikazioak azalduz eta zehazki, proteinek burututako katalisia aztertzeko baliagarriak diren QM/MM metodoen erabilera aztertuz. Azkenik, 1.3 atalean, minbiziaren aurkako platino(II) konplexuen erabilera aurkezten da, gehienbat, minbizi zelulek garatutako erresistentzian eta cisplatinoak zein mostaza nitrogenatuek DNA lotzeko duten mekanismoetan sakontzen delarik.

1.1 PROTEINEK KATALIZATUTAKO ERREAKZIO PERIZIKLIKOAK

Proteinak aminoazido desberdinez osaturiko sekuentziak dira. Aminoazidoek beraien artean loturak sortzeko gaitasuna dute, horrela, funtzio biologiko ugari burutzeko gai diren estruktura makromolekular konplexuak eratuz. Proteinek funtzio desberdinak bete ditzakete hauek osatzen dituzten aminoazido kopuruaren (20 dira naturan eskuragai daudenak) eta hauen antolamenduaren arabera, esaterako, garraio, entzimatiko, seinalizazio edota egitura funtzioak. [1] Hauen artean lehenetarikoa deskribatu zena erreakzio kimikoak katalizatzeko gaitasuna izan zen, katalizatzaile kimiko ez biologikoen antzeko funtzionamendua burutuz. Proteina hauek entzima izenaz ezagutzen dira.[2] Aminoazidoen sekuentziak eta honen luzera ezinbestekoak dira beraien aktibitate katalitikoa garatzeko, bi faktore hauek eragin zuzena baitute talde funtzionalen antolamenduan eta beraz, proteinen toleste prozesuan.[3] Aminoazidoen sekuentziak aldatuz proteina toleste ugari lortu daitezke, substratuek proteinekin duten elkarrekintzan eta beraz, prozesu katalitikoetan efektu zuzena izanik.[4]

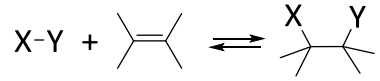
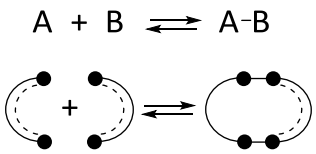
Proteina hauek prozesu kimikoen aktibazio energiak murriztuz burutzen dute katalisia. Gehienetan, erreakzio abiaduraren murrizpen hau proteinaren eta trantsizio-egoeraren arteko elkarrekintzaren bitartez gertatzen da. Modu honetan, trantsizio-egoera honen kontzentrazioa handitzen da, erreakzio abiaduraren areagotzea sustatuz. Horrez gain, entzimen ezaugarri garrantzitsuenetako bat denbora unitateko burutu ditzaketen transformazio kopurua da.[5] Ezaugarri hauek direla medio, azken urteetan entzimek katalizatutako erreakzioen azterketak interes handia sortu du arlo bioteknologikoan.

Hori horrela, prozesu katalitikoetan parte hartzen duten mekanismoen hausnarketa eta proteinek jasan dituzten prozesu ebolutiboen ulermena ezinbestekoa

1. Whitford, D., *Proteins Structure and Function*. Wiley & Sons: Chichester, England, 2005.
2. Copeland, R. A., *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis, Second Edition*. Wiley & Sons: New York, 2000.
3. Makhlynets, O. V.; Korendovych, I. V., *Nat. Chem.* **2016**, *8*, 823-824.
4. Rufo, C. M.; Moroz, Y. S.; Moroz, O. V.; Stöhr, J.; Smith, T. A.; Hu, X.; DeGrado, W. F.; Korendovych, I. V., *Nat. Chem.* **2014**, *6*, 303-309.
5. Marangoni, A. G., *Enzyme Kinetics: A Modern Approach*. Wiley&Sons: Hoboken, New Jersey, 2003.

bilakatu da.[6,7] Parte hartzen duten prozesuak aztertu ostean, hurrengo urratsa entzimen erreaktibitatearen eraldaketa izan zen, proteina berriak garatuz.[8-11] Honela, proteinen gune aktiboan zein kofaktoreen interakzio guneetan egindako aldaketek sekulako aurrerapenak ekarri dituzte. [12] Entzimen eboluzioan baliabide-konputazionalen garapena ezinbestekoa izan da.[13,14] Entzimak, katalizatzen dituzten erreakzioen arabera, sei multzo nagusitan banatzen dira IUPACen arabera (1.1 Taula).

1.1 Taula. IUPACek proposatutako entzimen sailkapena katalizatutako erreakzioaren arabera.

Familia Zenbakia	Entzima Mota	Katalizatutako Erreakzioa	
1	***Oxidoreduktasak***	Oxidakzio-Erredukzioak	$A^+ + B \longrightarrow A + B^+$ $A^- + B \longrightarrow A + B^-$
2	***Transferasak***	Transferentzia	$A\text{-}X + B \longrightarrow A + B\text{-}X$ $(X = CH_3, CH_3CO, NH_2)$
3	***Hidrolasak***	Hidrolisia	$A\text{-}B + H_2O \longrightarrow A\text{-}OH + H\text{-}B$
4	***Liasak***	Segmentazio ez-hidrolitikoa	
5	***Isomerasak***	Isomerizazioa	$A\text{-}B\text{-}C \longrightarrow A\text{-}C\text{-}B$
6	***Ligasak***	Loturen eraketa C-C loturak (6.4)	

Bestalde, sei multzo hauek beste hainbat azpitaldetan sailka daitezke bigarren EC zenbaki bat esleituz. Esaterako, C-C loturak burutzen dituzten ligasak EC 6.4 multzoaren

6. Pál, C.; Papp, B.; Lercher, M. J., *Nat. Rev. Genet.* **2006**, *7*, 337-348.
7. Palmer, T.; Bonner, P. L., *Enzymes: Biochemistry, Biotechnology, Clinical Chemistry*. Woodhead Publishing: Cambridge, UK, 2001.
8. Moroz, Y. S.; Dunston, T. T.; Makhlynets, O. V.; Moroz, O. V.; Wu, Y.; Yoon, J. H.; Olsen, A. B.; McLaughlin, J. M.; Mack, K. L.; Gosavi, P. M.; Van Nuland, N. A. J.; Korendovych, I. V., *J. Am. Chem. Soc.* **2015**, *137*, 14905-14911.
9. Penning, T. M.; Jez, J. M., *Chem. Rev.* **2001**, *101*, 3027-3046.
10. Huang, P.-S.; Boyken, S. E.; Baker, D., *Nature* **2016**, *537*, 320-327.
11. Hilvert, D., *Annu. Rev. Biochem.* **2013**, *82*, 447-470.
12. Toscano, M. D.; Woycechowsky, K. J.; Hilvert, D., *Angew. Chem., Int. Ed.* **2007**, *46*, 3212-3236.
13. Kries, H.; Blomberg, R.; Hilvert, D., *Curr. Opin. Chem. Biol.* **2013**, *17*, 221-228.
14. Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N., *Angew. Chem., Int. Ed.* **2013**, *52*, 5700-5725.

barnean sailkatzen dira. Era berean, hiru eta laugarren zenbakiak egokitzen zaizkie proteinei bakoitza kode espezifiko baten bidez izendatuz.

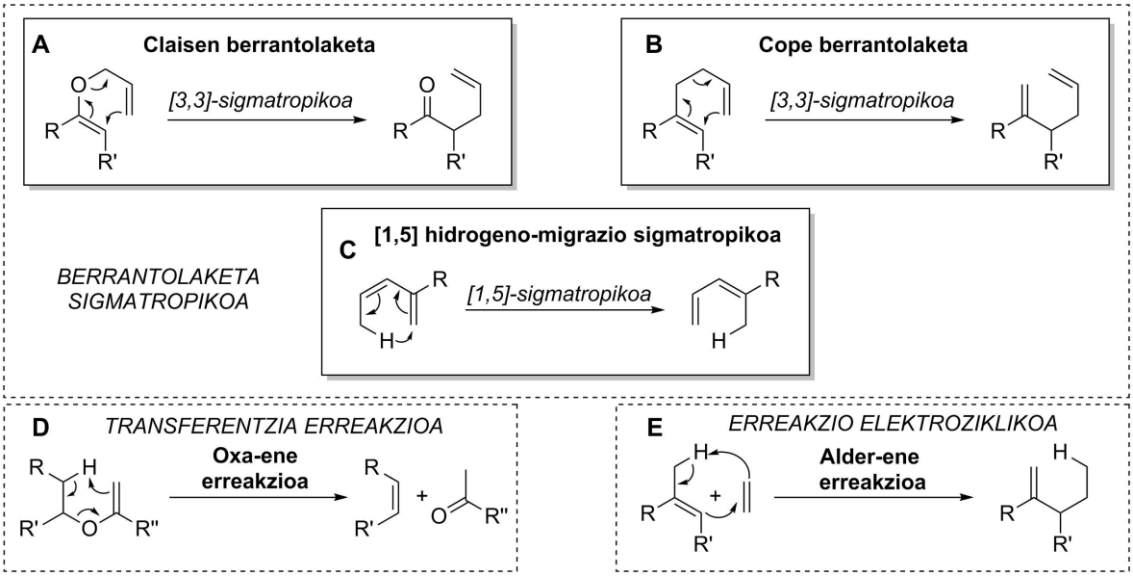
1.1.1 Periziklasak

Periziklasak, erreakzio periziklikoak katalizatzen dituzten entzimak dira. Erreakzio hauetan, urrats soil batean lotura berriak sortu edo hautsi egiten dira, trantsizio-egitura zikliko baten bitartez. Modu honetan, erreakzio periziklikoek C-C lotura kimiko regio- eta estereoselektiboen eraketa ahalbidetzen dute eta lau azpimultzo nagusitan sailkatu daitezke: berrantolaketa sigmatropikoak, erreakzio elektroziklikoak, zikloadizio erreakzioak eta transferentzia erreakzioak.[15,16] Kimika organikoaren arloan erabilpen zabala izan arren, gaur egun gutxi dira periziklasa aktibitatea duten entzimak.[17] Nahiz eta isomerasek, liasek eta ligasek katalizatutako erreakzio periziklikoak urriak izan,[18,19] badira erreaktibitate perizikloak burutzen duten hainbat entzima interesgarri.[20]

Aipatutako erreakzio periziklikoen lau azpimultzoak kontuan izanik, berrantolaketa sigmatropikoak katalizatzen dituzten hainbat proteina ezagunak dira. Korismato mutasak, hots, korismatotik prepenatorako [3,3]-berrantolaketa sigmatropiko suprafaziala, Claisenen berrantolaketa, katalizatzen du (1.1A Irudia).[21,22] Bestalde, [3,3]-erreakzio sigmatropiko suprafazial bat ezinbestekoa da hapalindol/ambiginaren sintesiaren bitarteko prozesuan.[23] Era berean, urrats hau nahitaezkotzat hartu izan da fischerindolen eta welwitindolen sintesian.[24,25] Azpimultzo

15. Fleming, I., *Pericyclic Reactions*. Oxford University Press: Oxford, 1999.
16. Gill, G. B.; Willis, M. R., *Pericyclic Reactions*. Chapman & Hall Ltd.: London, England, 1974.
17. Jamieson, C. S.; Ohashi, M.; Liu, F.; Tang, Y.; Houk, K. N., *Nat. Prod. Rep.* **2019**, *36*, 698-713.
18. Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J., *Interdiscip. Sci.* **2010**, *2*, 115-131.
19. Ohashi, M.; Liu, F.; Hai, Y.; Chen, M.; Tang, M.-C.; Yang, Z.; Sato, M.; Watanabe, K.; Houk, K. N.; Tang, Y., *Nature* **2017**, *549*, 502-506.
20. Walsh, C. T.; Tang, Y., *Biochemistry* **2018**, *57*, 3087-3104.
21. Guo, H.; Cui, Q.; Lipscomb, W. N.; Karplus, M., *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9032-9037.
22. Andrews, P. R.; Smith, G. D.; Young, I. G., *Biochemistry* **1973**, *12*, 3492-3498.
23. Li, S.; Lowell, A. N.; Yu, F.; Raveh, A.; Newmister, S. A.; Bair, N.; Schaub, J. M.; Williams, R. M.; Sherman, D. H., *J. Am. Chem. Soc.* **2015**, *137*, 15366-15369.
24. Zhu, Q.; Liu, X., *Chem. Commun.* **2017**, *53*, 2826-2829.
25. Zhu, Q.; Liu, X., *Angew. Chem., Int. Ed.* **2017**, *56*, 9062-9066.

honen barnean aurkitzen den beste berrantolaketa sigmatropiko mota bat, Cope berrantolaketa da (1.1B Irudia). Erreaktibitate mota hau, 4-DMAT sintasak katalizatutako 4-dimetilalil triptofanoaren sintesi prozesuan proposatua izan da.[26]



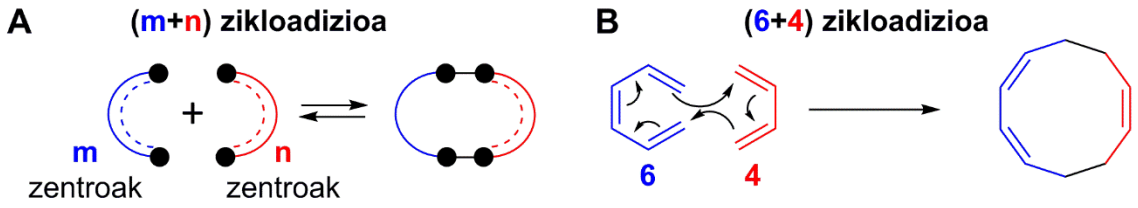
1.1. Irudia. Entzimek katalizatutako bost erreakzio perizikliko mota. Berrantolaketa sigmatropikoak: **A**) Claisen berrantolaketa; **B**) Cope berrantolaketa eta **C**) [1,5]-hidrogeno-migrazio sigmatropikoa. **D**) Oxa-ene erreakzioa. **E**) Alder-ene erreakzio elektroziklikoa.

Halaber, precorrin-8x mutasak precorrin-8x substratuaren azido hidrogenobirinikorako eraldaketa katalizatzen du. B_{12} bitaminaren biosintesiko artekaria den urrats honetan, [1,5]-hidrogeno-migrazio sigmatropikoa deskribatua izan da erreakzio mekanismotzat (1.1C Irudia).[27,28] Liasa periziklikoei dagokienez, isokorismato pirubato liasak, isokorismatoa salizilatora edota pirubatora transformatzen du, oxa-ene transferentzia erreakzio baten bidez (1.1D Irudia).[29,30] Horrez gain, Alder-ene

26. Luk, L. Y. P.; Qian, Q.; Tanner, M. E., *J. Am. Chem. Soc.* **2011**, *133*, 12342-12345.
27. Li, Y.; Alanine, A. I. D.; Vishwakarma, R. A.; Balachandran, S.; Leeper, F. J.; Battersby, A. R., *J. Chem. Soc., Chem. Commun.* **1994**, 2507-2508.
28. Shipman, L. W.; Li, D.; Roessner, C. A.; Scott, A. I.; Sacchettini, J. C., *Structure* **2001**, *9*, 587-596.
29. DeClue, M. S.; Baldridge, K. K.; Künzler, D. E.; Kast, P.; Hilvert, D., *J. Am. Chem. Soc.* **2005**, *127*, 15002-15003.
30. Lamb, A. L., *Biochemistry* **2011**, *50*, 7476-7483.

mekanismo elektroziklikoa proposatu da crotonyl-Coa karboxilasa/erreduktasak katalizatzen duen bide sintetikoaren bitarteko urrats batean (1.1E Irudia).[31]

Orain arte aipatutako erreakzioen artean, zikloadizio erreakzioak dira laugarren azpimultzoa. Zikloadizio erreakzioak, π-sistemak dituzten bi substraturen arteko erreakzioak dira, zeintzuetan bi σ-lotura berri eratzen diren. Bertan, egitura zikliko bat sortzen da prozesu kontzertatu baten bitartez.[32] Erreakzio perizikliko guztien artean, zikloadizio erreakzioak dira oparoenak eta erabilienak, batez ere, urrats bakan batean hainbat lotura eta zentro esterogeniko berri sortzeko ematen duten aukeragatik. Nomenklaturari dagokionez, (m+n) zikloadizioak bezala ezagutzen dira, zeinetan m eta n erreakzioan parte hartzen duten osagarri bakoitzeko atomo kantitatea diren, hala nola, (4+2) zikloadizioak (Diels-Alder erreakzioak),[33] (3+2) zikloadizioak (1,3-zikloadizio dipolarrak) edota (6+4) zikloadizioak (1.1. Eskema). Bestalde, parte hartzen duten π-elektroi kopurua zein prozesuen izaera antarafazial edota suprafaziala kako zuzenen artean adierazten da. Esaterako, (4+2) eta (3+2) zikloadizio kontzertatuak [$_{\pi}4_{s}+_{\pi}2_{s}$] orbital topologien bitartez gertatzen dira.



1.1 Eskema. A) Zikloadizio erreakzioen eskema orokorra. **B**) (6+4)-zikloadizioa, zeinetan sei eta lau atomok parte hartzen duten, erreaktibo urdinetik eta gorritik, hurrenez hurren.

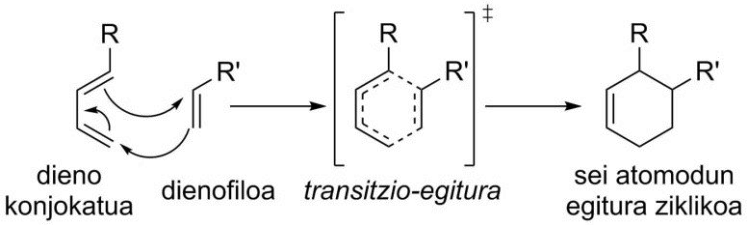
Entzimek katalizatutako zikloadizio erreakzioei bagagozkie, Diels-Alder aktibitate katalitikoa da ugariena.[34] Aldiz, kimika organikoan aplikazio anitz dituzten (3+2) zikloadizioak, Huisgenen erreakzioak bezala ere ezagutzen direnak,[35] katalizatzen dituzten entzimarik ez da deskribatu orain arte. Arrazoi hauek direla medio, datorren

31. Rosenthal, R. G.; Ebert, M.-O.; Kiefer, P.; Peter, D. M.; Vorholt, J. A.; Erb, T. J., *Nat. Chem. Biol.* **2013**, *10*, 50.
32. Kobayashi, S.; Jorgensen, K. A., *Cycloaddition Reactions in Organic Synthesis*. Wiley-VCH: Weinheim, Germany, 2001.
33. Diels, O.; Alder, K., *Liebigs Ann.* **1928**, *460*, 98-122.
34. Klas, K.; Tsukamoto, S.; Sherman, D. H.; Williams, R. M., *J. Org. Chem.* **2015**, *80*, 11672-11685.
35. Huisgen, R., *Angew. Chem., Int. Ed.* **1963**, *2*, 565-598.

ataletan, oparoak diren Diels-Alder erreakzioak eta aurrekari biologikorik ez duten (3+2) zikloadizioen nondik-norakoak azalduko dira.

1.1.2 Diels-Alderasak

1,3-dienoen eta alkenoen arteko (4+2) zikloadizio erreakzioa lehen aldiz 1928 urtean deskribatu zuten Otto Diels eta Kurt Alder kimikariek.[33] Kasu honetan, dieno konjokatua izango litzateke 4π osagaia eta dienofiloa, 2π osagaia. Bi substratu hauen arteko erreakzioaren bitartez, bi lotura bakun berri sortzen dira transitzio-egitura (TS) zikliko bat dela medio, ziklohexeno egiturak eratzeko (1.2 Eskema). Honenbestez, ziklazio prozesu hau sei atomoz osatutako eraztunak era regio- eta esterokontrolatu batean sortzeko estrategia erabilienetakoa bilakatu da, hainbat konposatu naturalen sintesian erabilia izanik.[36,37] Diels-Alder erreakzioak urrats bakan batean lau zentro estereogeniko berri sortzea eta are gehiago, karbono-karbono, karbono-heteroatomo zein heteroatomo-heteroatomo loturak sortzea ahalbidetzen du. Hortaz, kimika sintetikoan izan zuen eragin handia kontuan izanik, erreakzio honen aurkitzaileei Nobel Saria eman zieten 1950ean.



Scheme 1.2. Dieno konjokatu eta dienofilo baten arteko Diels-Alder erreakzioa.

1920ko hamarkadan erreakzio hau garatu zenetik, honen inguruan publikazio ugari argitaratu dira ikuspegi sintetiko, mekanistiko eta teorikoak azaldu nahian.[38] Kimikaren alorrean izandako garrantzia kontuan harturik, Diels-Alder aktibitatea duten entzimen bilaketak interes berezia izan du azken urteotan.[39] Bilatze honetan jarritako

36. Fringuelli, F.; Taticchi, A., *The Diels Alder Reaction. Selected Practical Methods.* John Wiley & Sons: Chichester, England, 2002.
37. Nicolaou, K. C.; Snyder, S. A.; Montagnon, T.; Vassilikogiannakis, G., *Angew. Chem., Int. Ed.* **2002**, *41*, 1668-1698.
38. Franzen, J., *Kirk-Othmer Encyclopedia of Chemical Technology*. John Wiley & Sons: New Jersey, 2016.
39. Kim, H. J.; Ruszczycky, M. W.; Liu, H.-w., *Curr. Opin. Chem. Biol.* **2012**, *16*, 124-131.

baliabide guztiek, (4+2) zikloadizioa katalizateko gai zen antigorputz bat deskribatzea ahalbidetu zuten.[40,41] Era berean, Diels-Alder erreakzioa katalizatzeko gai zen RNA sekuentzia bat deskribatua izan zen.[42]

Aipatu bezala, metodologia honen bitartez ziklohexeno eraztunak modu eraginkor batean lortu zitezkeenez, hainbat produktu naturalen egituran aurkitutako ziklohexeno taldeek, Diels-Alder aktibitatea burutzeko gai ziren entzimen presentzia iradoki zuten.[43-45] Hala izanik, prozesu biosintetiko guzti horien azterketak, lehen Diels-Alderasa entzimaren deskribapena ekarri zuen.[46] Ichichararen ikerketa taldeak 1995ean, prosolanapirona III substratuaren *exo*-solanopirona A-rako konbertsioa proposatu zuen, hala ere, Solanapirona Sintasa, lehen Diels-Alderasa, geroago deskribatua izan zen.[47]

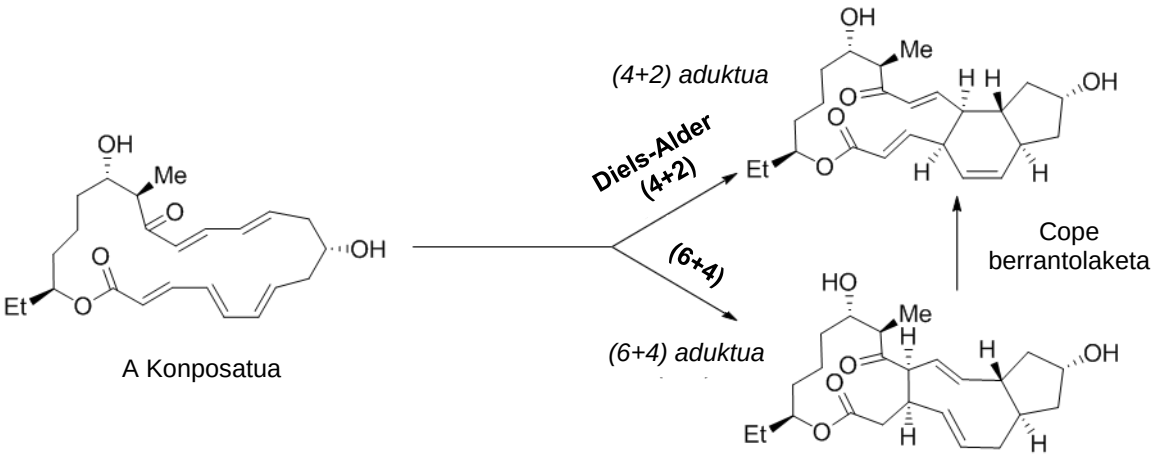
Analisi mekanistiko guzti hauek aztertzean kontuan izan beharreko faktore bat (4+2) ziklazio prozesuak mekanismo desberdinen bitartez gauzatu daitezkeela da, hauen artean, Diels-Alder erreaktibitatea aukera bat izanik. Era berean, erreakzio hauetan parte hartzen duten substratuen ezegonkortasunak eta entzimen konplexutasunak (4+2) zikloadizio mekanismoen azterketa zehatza zailtzen dutela aipatu beharra dago.[48,49] Diels-Alderasa entzima bezala proposatutako hautagai gehienetan, (4+2) zikloadizioa prozesu biosintetikoaren artekari baten-batean aurkeztua izan da soilik.[50,51] Kimikari sintetikoek erabilitako Diels-Alder erreakzioak prozesu intermolekularrak izan

40. Hilvert, D.; Hill, K. W.; Nared, K. D.; Auditor, M. T. M., *J. Am. Chem. Soc.* **1989**, *111*, 9261-9262.
41. Gouverneur, V.; Houk, K.; de Pascual-Teresa, B.; Beno, B.; Janda, K.; Lerner, R., *Science* **1993**, *262*, 204-208.
42. Tarasow, T. M.; Tarasow, S. L.; Eaton, B. E., *Nature* **1997**, *389*, 54-57.
43. Oikawa, H.; Tokiwano, T., *Nat. Prod. Rep.* **2004**, *21*, 321-352.
44. Stocking, E. M.; Williams, R. M., *Angew. Chem., Int. Ed.* **2003**, *42*, 3078-3115.
45. Kelly, W., *Org. Biomol. Chem.* **2009**, *6*, 4483-93.
46. Oikawa, H.; Katayama, K.; Suzuki, Y.; Ichihara, A., *J. Chem. Soc., Chem. Commun.* **1995**, 1321-1322.
47. Kasahara , K.; Miyamoto, T.; Fujimoto, T.; Oguri, H.; Tokiwano , T.; Oikawa, H.; Ebizuka, Y.; Fujii, I., *ChemBioChem* **2010**, *11*, 1245-1252.
48. Zheng, Q.; Tian, Z.; Liu, W., *Curr. Opin. Chem. Biol.* **2016**, *31*, 95-102.
49. Ose, T.; Watanabe, K.; Mie, T.; Honma, M.; Watanabe, H.; Yao, M.; Oikawa, H.; Tanaka, I., *Nature* **2003**, *422*, 185-189.
50. Auclair, K.; Sutherland, A.; Kennedy, J.; Witter, D. J.; Van den Heever, J. P.; Hutchinson, C. R.; Vederas, J. C., *J. Am. Chem. Soc.* **2000**, *122*, 11519-11520.
51. Minami, A.; Oikawa, H., *J. Antibiot.* **2016**, *69*, 500-506.

ohi dira, aldiz, deskribatutako prozesu biologiko hauetan erreakzio intramolekularrak dira erakusgai. Bestalde, entzima hauek erreakzio mekanismoan duten efektuari dagokionez, trantsizio-egoera errazten duen gune aktibo baten bitartez edota dienozalea hidrogeno lotura bitartez egonkortuz (LUMO murrizketa) lan egiten dute.

Aurkikuntza hauen ostean, *E. coli*-ko SpnF entzima deskribatu zen Diels-Alder aktibitate intramolekularra katalizatzen zuen entzima nabaritzat.[52,53] Proteina honek Spinosin A-ren biosintesian parte hartzen du, (4+2) zikloadizio transanularraren abiadura-konstantea, gutxi gorabehera, 500 aldiz azkartuz. Ikerketa berri batean, Houk eta bere lankideek (6+4)/(4+2) erreaktibitatea aurkeztu zuten Spinosin A-ren bide sintetikoan.[54,55] Gainera, lan horretan Cope berrantolaketa aurkeztua izan zen (6+4) bitartekarien konbertsiorako, (4+2) aduktuak emanez. Era honetan, SpnF entzima izan zen Diels-Alder/(6+4)-asa aktibitatea erakusten zuen lehen proteina (1.3. Eskema). Berriki, (6+4)/(4+2) periziklasen multzo bat deskribatuta izan da streptoseomicina-erako produktu naturalen sintesian.[56,57]

52. Kim, H. J.; Ruszczycky, M. W.; Choi, S.-H.; Liu, Y.-N.; Liu, H.-W., *Nature* **2011**, *473*, 109-112.
53. Fage, C. D.; Isiorho, E. A.; Liu, Y.; Wagner, D. T.; Liu, H.-w.; Keatinge-Clay, A. T., *Nat. Chem. Biol.* **2015**, *11*, 256-258.
54. Patel, A.; Chen, Z.; Yang, Z.; Gutiérrez, O.; Liu, H.-w.; Houk, K. N.; Singleton, D. A., *J. Am. Chem. Soc.* **2016**, *138*, 3631-3634.
55. Yang, Z.; Yang, S.; Yu, P.; Li, Y.; Doubleday, C.; Park, J.; Patel, A.; Jeon, B.-S.; Russell, W. K.; Liu, H.-W.; Russell, D. H.; Houk, K. N., *Proc. Natl. Acad. Sci. U.S.A* **2018**, *115*, 848-E855.
56. Yu, P.; Patel, A.; Houk, K. N., *J. Am. Chem. Soc.* **2015**, *137*, 13518-13523.
57. Zhang, B.; Wang, K. B.; Wang, W.; Wang, X.; Liu, F.; Zhu, J.; Shi, J.; Li, L. Y.; Han, H.; Xu, K.; Qiao, H. Y.; Zhang, X.; Jiao, R. H.; Houk, K. N.; Liang, Y.; Tan, R. X.; Ge, H. M., *Nature* **2019**, *568*, 122-126.



1.3. Eskema. SpnF entzimaren Diels-alder/(6+4)-asa aktibitatea A konposatua (4+2) eta (6+4) aduktuetan transformatuz. (6+4) aduktua, (4+2) aduktuarekin konektatzen duen Cope berrantolaketa.[54,55]

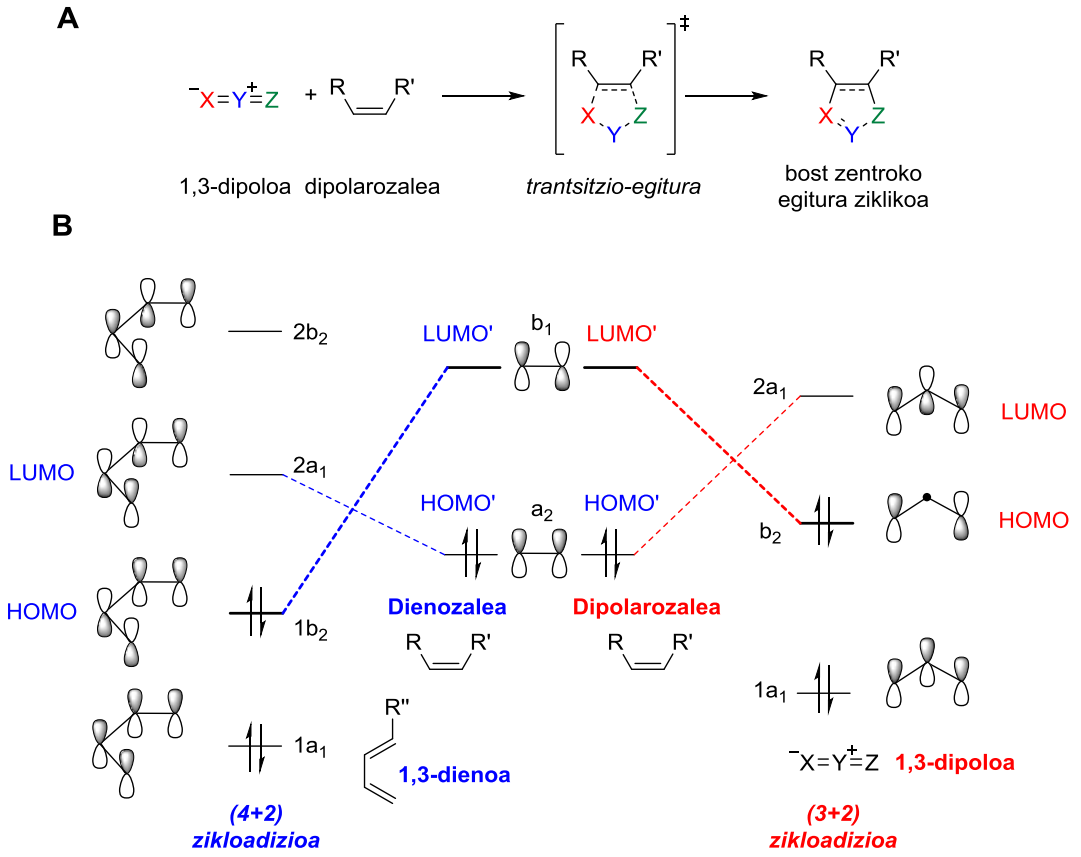
Aurkikuntza hauen ostean, Diels-Alder erreakzioen garrantziak, naturan aurkitutako Diels-Alderasa urriekin batera, Diels-Alder aktibitatea duten entzimen diseinu arrazionala bultzatu zuen. [58] Esaterako, Diels-Alderasa bimolekular baten bilakaera sustatu zen de novo modelizazio konputazionala eta lan experimentala bateratuz.[59]

1.1.3 *Huisgenasak*

Gure ikerketa taldeak, berriki argitaratzera bidalitako lan batean, *Huisgenasa*[60] izena proposatu du 1,3-dipoloen eta π-urritasuna duten dipolarozaleen arteko erreakzioa katalizatzen duten entzimak izendatzeko. Rolf Huisgenek 1963an deskribatu zuen erreakzio perizikliko termiko klase hau lehenengoz.[35] Huisgenen erreakzioa, Diels-Alder erreakzioaren analogotzat hartzen da, baina kasu honetan bost zentro dituzten eraztunak eratzeko erabiltzen delarik. Ala izanik, zikloadizio 1,3-dipolarra,

58. Lichman, B. R.; O'Connor, S. E.; Kries, H., *Chem. Eur. J.* **2019**, *25*, 6864-6877.
59. Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St.Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D., *Science* **2010**, *329*, 309-313.
60. Rivilla, I.; Odriozola-Gimeno, M.; Aires, A.; Gimeno, A.; Jiménez-Barbero, J.; Torrent-Sucarrat, M.; Cortajarena, A. L.; Cossío, F. P., *under revision.*

dipolarofiloaren 2π elektroiek 1,3-dipoloaren 4π elektroiekin erreakzionatzen dute bost unitateko egitura ziklikoak sortuz (1.4A Eskema).[61]



Scheme 1.4. A) Zikloadizio 1,3-dipolarraren eskema orokorra. **B**) (4+2) eta (3+2) zikloadizioen orbital molekularren konparaketa.

Bestalde, metodologia honen bitartez zentro estereogeniko berriak dituzten aduktuak lor daitezke, 1,3-dipoloaren X eta Z atomoak C, N edota O izan daitezkeela eta atomo zentrala berriz, Y, N izan ohi dela kontuan izanik. Honela, erabilitako substratuen arabera lotura mota desberdinak lortu daitezke, hauen artean, C-C, C-N edota C-O loturak. (3+2) zikloadizioen ezaugarri nagusienetako bat, Diels-Alder erreakzioekiko duten baliokidetasun elektronikoa da. Orokorrean, Diels-Alder erreakzioetan, interakzio nagusia dienoaren HOMO eta dienofiloaren LUMO' orbitalen artean gertatzen da, zikloadizio 1,3-dipolarretan, ostera, elkarrekintza 1,3-dipoloaren HOMO eta dipolarofiloaren LUMO' orbitalen artean ematen da (1.4.B Eskema).[62,63]

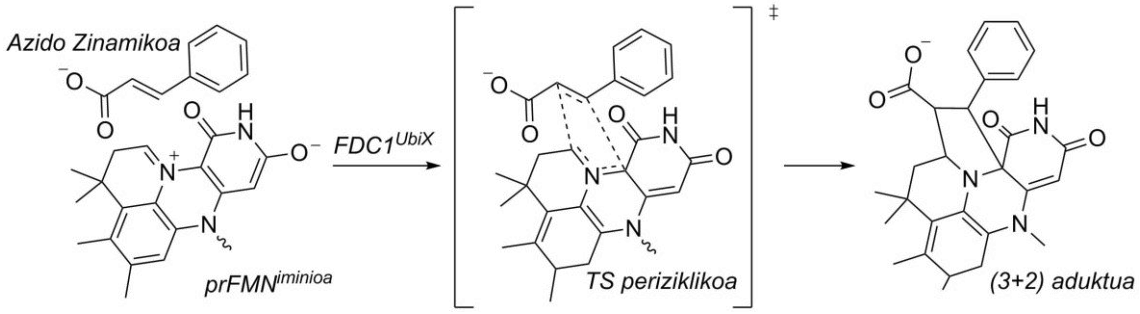
61. Huisgen, R., 1,3-Dipolar Cycloaddition Chemistry. Wiley, E. d., Ed. New York, 1984; pp 1-176.
62. Ess, D. H.; Houk, K. N., *J. Am. Chem. Soc.* **2008**, *130*, 10187-10198.
63. Gothelf, K. V.; Jørgensen, K. A., *Chem. Rev.* **1998**, *98*, 863-910.

60ko hamarkadan erreakzio hauen aurkikuntza eman zenetik, kimikari sintetikoek zikloadizio 1,3-dipolarrak aplikazio anitzetan erabili dituzte.[64,65] Horrez gain, zenbait produktu naturaletan halako egiturak aurkitu izan dira, (3+2) zikloadizioak katalizatzeko gai diren entzimen ikerketa biziagotuz.[66-68] Esaterako, modu kobalentean DNAri lotutako konplexu bimetaliko batek katalizaturiko (3+2) zikloadizio erreakzioa deskribatu du berriki gure ikerketa taldeak.[69]

Diels-Alder erreakzioekin alderatuta, C-C loturak eratzen dituzten eta *Huisgenasa* aktibitate soila duten entzimak ez dira oraindik ezagutzen.[70] Hala ere, badira hainbat kasu non, C-O loturak eratzen diren erreakzio 1,3-dipolarren bitartez prozesu biosintetikoen artekari gisa.[67, 71] Horrez gain, 2015ean Leysen-en ikerketa taldeak, lehenengo (3+2) retro-zikloadizio intermolekularra aurkeztu zuen; *Aspergillus niger*-en aurkitzen den FDC^{UBIx} proteinak prFMN kofaktorearen laguntzarekin α-β-azido asegabeen deskaboxilazio entzimatikoaren artekari batean hain zuzen ere (1.5. Eskema).[72] Ondoren, 2017an Marsh eta bere lankideek mekanismo berdintsu bat proposatu zuten, FDC proteinak katalizatutako prFMN kofaktorearen eta 2-fluor-2-nitrobinilbentzenoaren arteko erreakziorako.[73] Horrez gain, erreakzio honen analogo

64. Sankararaman, S., *Pericyclic Reactions-a Textbook: Reactions, Applications and Theory*. Wiley-VCH: Weinheim, Germany, 2005.
65. Padwa, A.; Pearson, W. H., *Synthetic Applications of 1,3-Dipolar Cycloaddition Chemistry Toward Heterocycles and Natural Products*. Wiley & Sons: Hoboken, New Jersey, 2002.
66. Painter, P. P.; Pemberton, R. P.; Wong, B. M.; Ho, K. C.; Tantillo, D. J., *J. Org. Chem.* **2014**, *79*, 432-435.
67. Krenske, E. H.; Patel, A.; Houk, K. N., *J. Am. Chem. Soc.* **2013**, *135*, 17638-17642.
68. Strych, S.; Journot, G.; Pemberton, R. P.; Wang, S. C.; Tantillo, D. J.; Trauner, D., *Angew. Chem., Int. Ed.* **2015**, *54*, 5079-5083.
69. Rivilla, I.; de Cózar, A.; Schäfer, T.; Hernandez, F. J.; Bittner, A. M.; Eleta-Lopez, A.; Aboudzadeh, A.; Santos, J. I.; Miranda, J. I.; Cossío, F. P., *Chem. Sci.* **2017**, *8*, 7038-7046.
70. Baunach, M.; Hertweck, C., *Angew. Chem., Int. Ed.* **2015**, *54*, 12550-12552.
71. Borowski, T.; de Marothy, S.; Broclawik, E.; Schofield, C. J.; Siegbahn, P. E. M., *Biochemistry* **2007**, *46*, 3682-3691.
72. Payne, K. A. P.; White, M. D.; Fisher, K.; Khara, B.; Bailey, S. S.; Parker, D.; Rattray, N. J. W.; Trivedi, D. K.; Goodacre, R.; Beveridge, R.; Barran, P.; Rigby, S. E. J.; Scrutton, N. S.; Hay, S.; Leys, D., *Nature* **2015**, *522*, 497.
73. Ferguson, K. L.; Eschweiler, J. D.; Ruotolo, B. T.; Marsh, E. N. G., *J. Am. Chem. Soc.* **2017**, *139*, 10972-10975.

bat UbiD-ko azido ferulikoaren deskarboxilasak burutzen duen katalisiaren etapa batean desrkribatua izan da.[74]



1.5. Eskema. FDC1^{UbiX} katalizatutako zikloadizio dipolarra.[72]

1.2 KIMIKA KONPUTAZIONALA: ARAZO KIMIKOEI ERANTZUNA EMATEKO BALIABIDE GISA

1960ko hamarkadaz geroztik,[75-77] kimika teorikoak sekulako hazkuntza jasan izan du, batez ere, arlo informatikoan, kimikoan, zein kalkulu numerikoen ebazpenean emandako aurrerakuntza guztiek bultzatuta.[78 - 80] Hobekuntza guzti hauek metodo konputazionalen aplikazioa sustatu dute, ikerketa industrial zein akademikoan indar handia hartuz.[81-83]

Kimika konputazionala osatzen duten teknika desberdinen bitartez galdera kimiko ugariri erantzuna emateko aukera zabaldu zen.[84] Honela, metodo hauek batez ere prozesu kimiko sintetikoak burutu baino lehen iragarpenak egiteko eta prozesu

74. Bailey, S. S.; Payne, K. A. P.; Saaret, A.; Marshall, S. A.; Gostimskaya, I.; Kosov, I.; Fisher, K.; Hay, S.; Leys, D., *Nat. Chem.* **2019**, *11*, 1049-1057.
75. Cramer, C., *Essentials of Computational Chemistry: Theories and Models, 2nd Edition*. Chichester, England, 2004.
76. Bachrach, S. M., *Computational Organic Chemistry, 2nd Edition*. Hoboken, New Jersey, 2014.
77. Jensen, F., *Introduction to Computational Chemistry, 2nd Edition*. Chichester, England, 2007.
78. Lewars, E. G., *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics, 2nd Edition*. Springer: New York, 2011.
79. Thiel, W., *Angew. Chem., Int. Ed.* **2011**, *50*, 9216-9217.
80. Dykstra, C. E.; Frenking, G.; Kim, K.; Scuseria, G. E., *Theory and Applications of Computational Chemistry*. Eds. Elsevier: Oxford, UK, 2005.
81. Lipkowitz, K. B., *Chem. Rev.* **1998**, *98*, 1829-1874.
82. Richon, A. B., *Drug Discov. Today* **2008**, *13*, 659-664.
83. Houk, K. N.; Liu, F., *Acc. Chem. Res.* **2017**, *50*, 539-543.
84. Arrieta, A.; de la Torre, M. C.; de Cózar, A.; Sierra, M. A.; Cossío, F. P., *Synlett.* **2013**, *24*, 535-549.

kimikoak modu sakon eta zehatzago batean aztertzeko erabiltzen dira.[85,86] Erabilpen hauek kontuan izanik, kimika teorikoa kontzeptu kimiko desberdinak aztertzeko erabili daiteke:

-Energia potentzialaren gainazalaren (PES) azterketarako. Modu honetan, molekulen optimizazioak burutu daitezke berauen energia baxueneko egiturak aurkitzeko eta estruktura horien karakterizazio osoa ezagutzeko (distantziak, angeluak, diedroak...).[87]

-Erreaktibitate kimikoen analisia alde batetik, honen nondik norakoak aurreikusiz eta bestetik, laborategian burututako erreakzioa ulertu ahal izateko.[88]

-Ezaugarri kimiko ezberdinen analisia, esaterako, espektro kimikoen (IG, EMN, UM) edota afinitate elektroniko zein ionizazio potentzialen iragarpena. Hauetaz gain, neurketa termodinamiko desberdinak kalkulatu daitezke, hauetatik propietate desberdinak neurtzeko: oreka-konstanteak, abiadura-konstanteak, erreakzioaren parametro termokimikoak…[89-91] Era berean, aromatizitatea bezalako zenbait behagai ez-fisiko kuantifikatu daitezke.[92]

-Entzimek katalizatutako erreakzioen modelizazioa, hauen erreaktibitatearen aurreikuspena eta entzima berrien diseinua.[93,94]

1.2.1 Metodo konputazionalen sailkapena

Aurretiaz aipatu moduan, azken hamarkadetan egindako aurrerapen ugariek metodo konputazional berrien garapena izan dute emaitzatzat.[95] Arazo kimiko zehatz

85. Young, D., *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*. John Wiley & Sons, Inc.: New York, 2001.
86. Nguyen, Q. N. N.; Tantillo, D. J., *Chem. Asian J.* **2014**, *9*, 674-680.
87. Schlegel, H. B., *J. Comput. Chem.* **2003**, *24*, 1514-1527.
88. Sperger, T.; Sanhueza, I. A.; Schoenebeck, F., *Acc. Chem. Res.* **2016**, *49*, 1311-1319.
89. Steinfeld, J. I.; Francisco, J. S.; Hase, W. L., *Chemical Kinetics and Dynamics 2nd Edition*. Prentice Hall: Upper Saddle River, New Jersey, 1999.
90. Laidler, K. J.; King, M. C., *J. Phys. Chem. A* **1983**, *87*, 2657-2664.
91. Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J., *J. Phys. Chem. A* **1996**, *100*, 12771-12800.
92. Schleyer, P. v. R.; Wu, J. I.; Cossío, F. P.; Fernández, I., *Chem. Soc. Rev.* **2014**, *43*, 4909-4921.
93. Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J., *Angew. Chem., Int. Ed.* **2006**, *45*, 6856-6859.
94. Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J., *Chem. Commun.* **2010**, *46*, 2354-2372.
95. Fernández, I.; Cossío, F. P., *Chem. Soc. Rev.* **2014**, *43*, 4906-4908.

baten azterketa burutu nahi bada, metodologia zuzenaren hautaketa ezinbestekoa da. Hau kontuan izanik, atal honetan metodologia arruntenen sailkapen bat eskaintzen da, hauek sei azpimultzotan banatzen direlarik:

a) **Mekanika Molekularra (MM)**, fisika klasikoak azaldutako arauak erabiltzen ditu molekulen eredu sinplifikatu bat sortzeko. Orokorrean, atomoak erradio jakin bateko esfera bezala deskribatzen dira eta beraien arteko loturak berriz, malgukiak balira bezala. MM metodoek indar-eremu enpirikoak erabiltzen dituzte molekulen PES-aren deskribapenerako eta energia minimoetan aurkitzen diren egituren loturak, diedroak eta angeluak kalkulatzeko. Modu honetan, hurbilketa hauek kalkuluen abiaduran efektu zuzena dute, sistema handiagoen azterketa ahalbidetuz. Sinplifikazio hauek badituzte hainbat desabantaila, horien artean, egoera espezifiko bakoitzean parametro berrien garapena eta efektu elektronikoak ezinbestekoak diren mekanismoetan erabiltzeko ezintasuna dira bereizgarrienak.[96,97]

b) **Metodo semienpirikoak**, metodo mekaniko kuantikoen (QM) familian sailkatzen dira eta beraz, geometriak, energiak, zein beste hainbat ezaugarri kimiko Schrödingerren ekuazioa ebatziz aurreikusten dituzte. Metodo hauetan zenbait sinplifikazio erabiltzen dira integralen ebazpena errazteko helburuarekin, modu honetan kalkuluen denbora murriztuz. Beraz, oinarrizko arau fisikoak mantentzen saiatzen dira eta burututako bakuntzeak konpentsatzeko ideiarekin hainbat parametro enpiriko eransten dituzte. Egitura kimikoen analisi kualitatibo egokirako erabilgarriak izan arren, aukeratutako parametroek eragin zuzena dute metodo semienpirikoen erabilera kuantitatiboan.[98]

c) ***Ab Initio* Metodoak** QM metodo multzo bat dira non ez diren ez parametroak ezta datu esperimentalik erabiltzen.[99] *Ab initio* kalkuluen helburu nagusienetako bat ahalik eta sinplifikazio eta hautazko erabaki gutxien erabiltzea da. Dena dela,

96. Rappe, A. K.; Casewit, C. L., *Molecular Mechanics Across Chemistry*. University Science Books: Sausalito, CA, 1997.
97. Williams, J. E.; Stand, P. J.; Schleyer, P. v. R., *Annu. Rev. Phys. Chem.* **2003**, *19*, 531-558.
98. Thiel, W., *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 145-157.
99. Hehre, W. J.; Radom, P. L.; Schleyer, P. V. R.; Pople, J., *Ab Initio Molecular Orbital Theory.* John Wiley & Sons: New York, 1986.

Schrödingerren ekuazioa ezin da modu zehatz batean ebatzi molekulek elektroi bat baino gehiago baldin badute eta hortaz, hainbat hurbilketa erabili behar dira. Beraz, kostu konputazionala eta kalkuluen zehaztasuna erabilitako sinplifikazioen araberakoak izango dira. Hala eta guztiz ere, *ab initio* metodoen bidez geometriak, energiak, zein elektroi distribuzioarekin zuzenki erlazionatuta dauden beste hainbat ezaugarri, polaritatea esaterako, kalkulatu daitezke. *Ab initio* metodoen desabantaila nagusia beharrezkoak diren baliabide kopurua da.[100]

d) **Dentsitate Funtzionalaren Teoriako** (**DFT**) **metodoak** QM metodoen azpimultzo bat dira eta beraz, beren oinarria Schrödingerren ekuazioaren ebazpenean dute. Ekuazio honetan aurkitzen den uhin funtzioa ebazteko zailtasun nagusia bertan aurkitzen diren lau aldagaiak dira; espazioko hiru koordenatuak eta spin koordenatu bat atomo bakoitzeko. Hortaz, uhin funtzioaren ebazpenaren konplexutasuna modu esponentzialean handituz doa sistemaren elektroi kopurua handitzearekin. DFT metodoek, Zailtasun honi irtenbidea aurkitu nahian uhin funtzioan oinarria izan ordez, behagai fisikoa den elektroi dentsitatean (soilik hiru aldagai espazialen funtsean dagoen funtzioa) dute oinarria. Hurbilketa hau aplikatu ahal izateko, energia elektroi dentsitatearen arabera aurkezten dituzten funtzionalak garatu ziren. Honela, *ab initio* metodoekiko abiadura irabazteko, sistemaren energia orokorra dentsitate elektronikoaren menpe dauden lau funtzionaletan banatua izan zen: energia zinetikoa, energia potentziala, elektroi-elektroi energia eta trukaketa-korrelazio energia. Azken urteotan DFT ikuspegiak garrantzia handia irabazi izan du, funtzional egokiak erabiltzen DFT metodoek erakutsi duten zehaztasunagatik batez ere.[101,102]

e) **Dinamika Molekularreko metodoek** (**MD**), Netwonen ekuazioak erabiltzen dituzten atomoen mugimendua deskribatzeko eta era honetan, hainbat galdera kimiko zein biologikoei erantzuna emateko. Metodo hauek, trajektoria bat, denbora unitate bakoitzean atomoko bakoitzak duen posizioa eta momentuaren informazioa lortzen da,

100. Levine, I. N., *Quantum Chemistry, 5th Edition*. Prentice Hall: Englewood Cliffs, New Jersey, 2000.
101. Parr, R. G.; Yang, W., *Density Functional Theory of Atoms and Molecules*. Oxford University Press: New York, 1989.
102. Koch, W.; Holthausen, M. C., *A Chemist's Guide to Density Functional Theory, 2nd Edition*. Wiley-VCH: Weinheim, Germany, 2001.

molekulen eboluzio dinamikoaren azterketa ahalbidetuz. Metodo hauek MM edota QM metodoak erabiltzen dituzte molekulen gain eragiten duten indarrak eta beraien mugimendua kalkulatzeko. MD-ren abantailak kontuak izanda, aplikazio ugari dituzte, esaterako, proteina eta ligandoen arteko elkarrekintzen azterketan.[103]

f) **Mekanika Kuantiko/Mekanika Molekular metodoek** (**QM/MM**) aurretiaz deskribatutako QM eta MM metodologiak batzen dituzte, sistema handien erreaktibitatea aztertu ahal izateko, proteinek burututako katalisia, esaterako. Doktoretza Tesi honetan, CTPR (Consensus Tetratricoptide Repeat) proteinen bidezko katalisia aztertu da QM/MM metodoak erabiliz, horregatik hurrengo atalean, QM/MM metodoen azterketa sakonago bat aurkeztuko da.

1.2.2 QM/MM ONIOM metodologia: proteinen aktibitate katalitikoa aztertzeko baliabidetzat

Aitzinean aipatu bezala, QM/MM metodoek QM metodologiaren zehaztasuna eta MM metodoen abiadura batzen dituzte. Metodologia hau lehenengoz 1970eko hamarkadan erabili zen eta orduz geroztik, atomo kopuru handiak dituzten sistemen erreaktibitatea aztertzeko oso baliagarria izan da.[104,105] Aztergai den sistema bi zati nagusitan banatu ohi da; alde batetik erreakzio kimikoa gertatzen den ingurua QM metodoen bitartez deskribatzen da eta bestetik, gainontzeko ingurua zehaztasun txikiagoa duten MM metodoen bitartez aztertzen da. Era honetan, sistemaren energia ($E_{QM/MM}$) hiru osagaietan banatzen da: QM sekzioaren energia (E_{QM}), MM inguruaren energia (E_{MM}) and bi sekzioen arteko interfaseari dagokion energia (E_{QM-MM}) (1.2.A Irudia).[106]

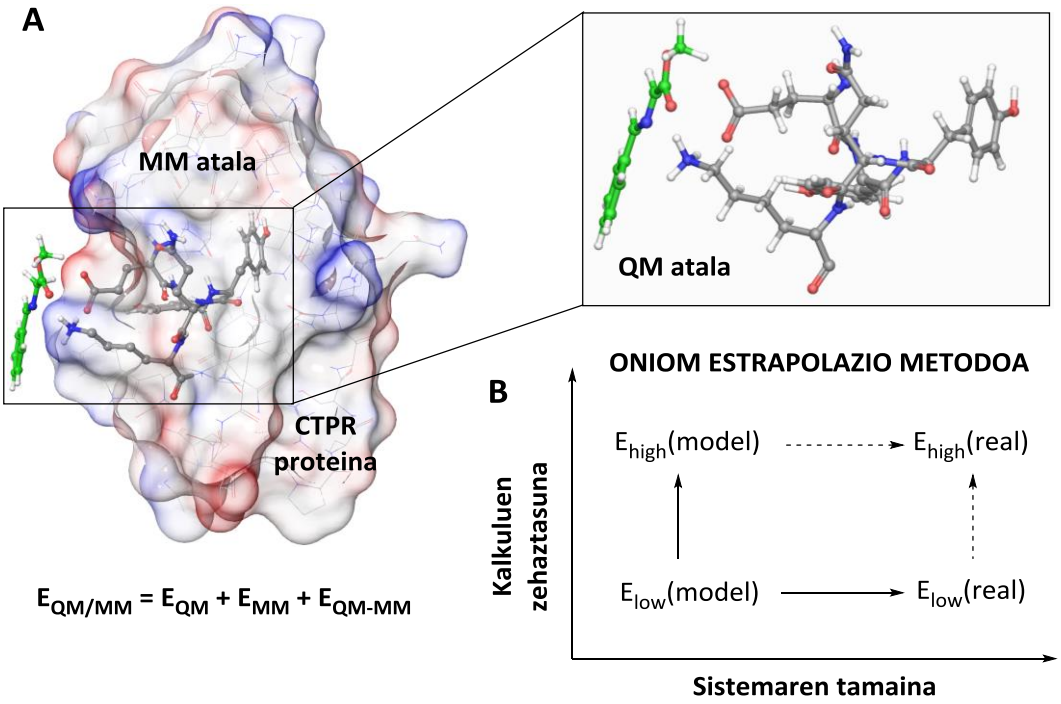
Metodologia honen gakoetako bat lotura kobalenteen bitartez konektatuta dauden QM eta MM sekzioen arteko interfasearen analisia da, zein modu desberdinetan aztertu daitekeen: txertaketa mekanikoa, txertaketa elektronikoa eta txertaketa polarizagarria. Hauen artean, txertaketa elektronikoa da metodo erabilienetakoa,

103. Rapaport, D. C., *The Art of Molecular Dynamics Simulation, 2nd Edition*. Cambridge University Press: Cambridge, UK, 2004.
104. Honig, B.; Karplus, M., *Nature* **1971**, *229*, 558-560.
105. Warshel, A.; Levitt, M., *J. Mol. Biol.* **1976**, *103*, 227-249.
106. Friesner, R. A.; Guallar, V., *Annu. Rev. Phys. Chem.* **2005**, *56*, 389-427.

zeinetan MM ataleko atomoek QM atala polarizatu dezaketen. Bestalde, orokorrean, QM eta MM atalak molekula berdinari dagozkie eta lotura kobalente baten bat moztu behar da biak banatzerakoan. Ondorioz, QM zatian elektroi bat edo gehiago parekatu gabe gelditzen direnez, hauek era egoki batean mugatzea beharrezkoa da. Helburu hau betetzeko estrategia erabiliena bi sekzioen artean atomo estekatzaile bat gehitzea da, gehienetan QM mugara loturiko hidrogeno atomo bat izanik.[107,108]

Hala eta guztiz ere, sinplifikazio honek QM/MM metodoen zehaztasuna baldintzatzen duten hainbat desabantaila ditu. Hauen artean eztabaida gehien zabaltzen dituenetako bat QM atalaren aukeraketa da, erreaktibitatearen analisi zehatz baterako ezinbestekoa baita. Aitzitik, QM eta MM zatiak konektatzen dituen mugaren analisiak ere eragin zuzena du kalkuluen egokitasunean eta kontuan izan behar da aukeratutako QM metodoa eta indar-eremuen arabera aldaketa esanguratsuak egon daitezkeela. Are gehiago, erreakzio ingurua MM metodologiaren bitartez aztertzen denez, erreaktibitatean efektu zuzena izan dezaketen aldaketa konformazionalen analisia ez da hain zehatza.[109,110] Beraz, QM/MM metodologia erreaktibitatea aztertzeko aukeratzen denean, alderdi negatibo hauek guztiak kontuan hartu behar dira. QM eta MM metodoen aukeraketa zuzenak, ordea, entzimek katalizatutako erreakzio mekanismoen analisi semi-kuantitatibo egokia du emaitzatzat.[111,112]

107. Lin, H.; Truhlar, D. G., *Theor. Chem. Acc.* **2006**, *117*, 185-199.
108. Field, M. J.; Bash, P. A.; Karplus, M., *J. Comput. Chem.* **1990**, *11*, 700-733.
109. Hu, H.; Yang, W., *Theochem.* **2009**, *898*, 17-30.
110. Sousa, S. F.; Ribeiro, A. J. M.; Neves, R. P. P.; Brás, N. F.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J., *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7*, e1281.
111. Thiel, W. In *QM/MM Methodology: Fundamentals, Scope, and Limitations*, Multiscale simulation methods in molecular sciences, Jülich: NIC, J Grotendorst, N. A., S. Blgel, & D.Marx, Ed. Jülich: NIC, 2009.
112. Groenhof, G., Introduction to QM/MM Simulations. In *Biomolecular Simulations: Methods and Protocols*, Monticelli, L.; Salonen, E., Eds. Humana Press: Totowa, NJ, 2013; pp 43-66.



1.2. Irudia. A) 2.Kapituluan aurkeztutako CTPR proteinarentzat aukeratutako QM/MM modeloa. MM atala gainazal moduan irudikatuta agertzen da eta QM atala "ball and sticks" moduan. **B**) ONIOM estrapolazio metodoaren azalpena, QM(high) eta MM(low) energi mailak definituz sistema modelo eta benetazko sistemarentzat.

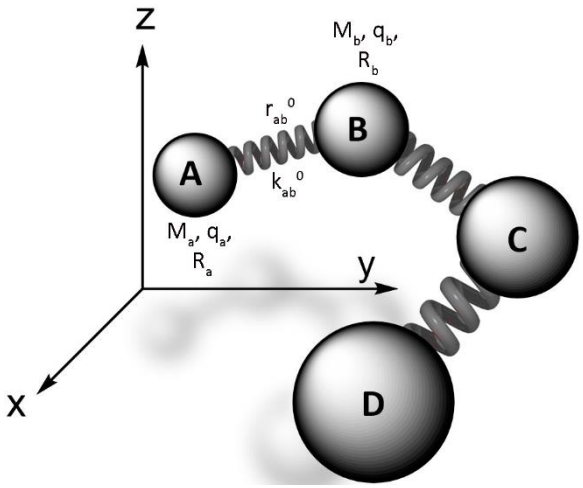
Gaur egun eskuragarri dauden QM/MM metodo guztien artean, ONIOM (Own N-layer Integrated molecular Orbital molecular Mechanics) metodologia da erabiliena eta 1990ko hamarkadan garatutako IMOMM (Integrated Molecular Orbital + Molecular Mechanics)[113] metodoa hartzen da ONIOM metodologiaren aitzindaritzat.[114] Ordura arte erabilitako metodoak estrategia aditiboan zuten oinarria, hau da, sistemaren energia E_{QM}, E_{MM} eta E_{QM-MM}, bi sistemen arteko mugari dagokion energiari batuz lortzen zuten (1.2.A Irudia). Berriz, ONIOM metodologia estrapolazio metodo baten oinarritzen da. Horretarako, ezarritako modeloaren (A, gune aktiboa) energia, alde batetik, QM zein MM metodoen bidez kalkulatzen da eta bestetik, benetazko sistema osoaren (AB) energia MM metodoaren bitartez lortzen da soilik (1.2.B Irudia). Sistemaren energia totala, ostera, QM metodo bidez kalkulatutako gune aktiboaren energiaren, E_{QM}(A), eta

113. Maseras, F.; Morokuma, K., *J. Comput. Chem.* **1995**, *16*, 1170-1179.
114. Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K., *J. Phys. Chem.* **1996**, *100*, 19357-19363.

sekzio berdinaren MM bidez kalkulatutako energiaren, $E_{MM}(A)$, arteko diferentziari, MM bidez kalkulatutako sistema osoaren energia, $E_{MM}(AB)$, batuz lortzen da. Hurrengo ataletan, MM eta QM metodoen sarrera labur bat eskeintzen da, bigarren multzo honetan batez ere DFT metodoen sakontzen delarik.[115,116]

1.2.2.1 QM/MM metodologian erabilitako MM metodoak

MM metodoek fisika klasikoak oszilatzaile harmonikoen deskribapenerako erabilitako arauak erabiltzen dituzte molekulen deskribapenerako. Honela, molekulak malgukiz konektaturiko esferen sekuentzia moduan irudikatzen dira eta beraz, elektroiak ez dira era esplizitu batean kontuan hartzen (1.3. Irudia).



$$E = \sum E_r + \sum E_\theta + \sum E_\omega + \sum E_{coul} + \sum E_{vdw}$$

1.3. **Irudia.** MM metodoetan erabilitako molekulen irudikapena eta energiaren definizioa.

MM metodoetan, bestalde, energia totala tentsio energia (E_r), tortsio energia (E_θ), flexio energia (E_ω), elkarrekintza elektroestatikoei dagokien energia Koulombioka (E_{coul}) eta van der Waalsen energien arteko batugai bezala azaltzen da (1.3. Irudia).

Atomoen eta beraien arteko elkarrekintzak hainbat parametroren bidez deskribatzen dira: karga (q), erradioa (r eta R), masa (M) eta Hooke-ren konstantea (k).

115. Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K., *Chem. Rev.* **2015**, *115*, 5678-5796.
116. Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K., *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 327-350.

Hau dela eta, atomo eta lotura desberdinen parametroak indar-eremu deiturikoetan biltzen dira. Era berean, indar-eremu hauek hiru osagai nagusi dituzte: beraien ezaugarriak definitutat dituzen atomo multzo bat da, molekulen PES-a deskribatzen dituzten hainbat ekuazio eta ekuazioen eta atomo moten arteko erlazioak ezartzen dituzten parametro gehigarri batzuk. Azken hauek, datu esperimentaletatik edo maila altuagoko datu konputazionaletatik lortzen dira gehienetan. Emaitzak lortu nahi izanez gero, ezinbestekoa da aztertu nahi diren egituren eta aukeratutako indar-eremuen aurretiaz egindako parametrizazioak eskuragai dauden ala ez aztertzea.

Gaur arte, indar-eremu desberdinak argitaratu dira (AMBER, CHARMM…) eta gure kasu espezifikoan, dreiding izeneko indar-eremua erabili izan da.[117] Hau, indar-eremuen artean sinpleenetakoa izanda ere, egitura biologikoen deskribapen zuzenerako baliagarria dela erakutsi da. Hala ere, ez da hain erabilgarria lotura kimikoen formakuntza eta haustura aztertu nahi bada.

1.2.2.2 QM/MM metodologian erabilitako QM-DFT metodoak

Esan bezala, gune katalitikoaren azterketa MM metodoen bitartez egiten da, ordea, proteinen erreaktibitatea aztertu nahi denean, ordea, QM metodo zehatzagoetara jo behar da, gure kasuan, DFT metodoak. MM metodoetan ez bezala, QM metodoen bitartez distribuzio elektronikoaren deskribapen esplizitu bat burutu daiteke, horretarako nukleoaren karga, masa eta elektroi kopurua kontuan hartuz. Metodoa hauek Schrödingerren [118] ekuazioan dute oinarria eta honen bitartez kalkulatzen dute sistemaren energia globala (E). Ψ (**r**, **R**), nukleo eta elektroi ororen uhin funtzioa da eta posizio nuklearren (R) eta elektronikoen (r) araberakoa da. $\hat{H}$ ordea operadore Hamiltoniarra da eta operadore zinetiko (T) zein potentzialen (V) batura moduan adierazten da. Sistema batek elektroi bat baino gehiago baldin baditu, Schrödingerren ekuazioa ebazteko orduan zenbait hurbilketa aplikatu behar dira, Born-Oppenheimerren hurbilketa, esaterako.

$$\hat{H}\Psi\left(\vec{\mathbf{r}}\,;\vec{\mathbf{R}}\right) = \mathrm{E}\Psi\left(\vec{\mathbf{r}}\,;\vec{\mathbf{R}}\right) \qquad (1.1)$$

117. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., *J. Phys. Chem.* **1990**, *94*, 8897-8909.
118. Schrödinger, E., *Phys. Rev.* **1926**, *28*, 1049-1070.

DFT metodoek Schrödingerren ekuazioa ebazteko zailtasun hau kontuan hartzen dute eta horregatik, behagai fisikoa den dentsitate elektronikoa erabiltzen dute, $\rho(\vec{r})$, uhin funtzioaren ordez. Dentsitate elektronikoa soilik hiru aldagai espazialen menpe dagoenez, kalkuluen kostua murrizten du. Beraz, DFT metodoetan energia eta molekulen ezaugarriak $\rho(\vec{r})$-aren arabera aurkezten dira, eta hau, era berean, elektroi zenbakiaren menpekoa da.

$$\mathrm{E} = \mathrm{E}[\rho(\vec{\mathbf{r}})] \tag{1.2}$$

Modu honetan, energia dentsitatearen menpe dauden hainbat funtzionalen konbinazio moduan ematen da: energia zinetiko elektronikoa, $T_S[\rho]$; nukleo-elektroi erakarpen energia, $E_{ne}[\rho]$; elektroi-elektroi aldarapena, $J[\rho]$; aldarapen elektronuklearra, V_{NN}; eta trukaketa-korrelazio funtzionala, $E_{xc}[\rho]$.[119] Azken termino hau da ebazteko zailena, hori dela eta, emaitza zehatza lortu nahian hainbat sinplifikazio burutzen dira, horien artean LDA (local density approximation) eta GCA (generalized gradient approximation) erabilienak izanik.

$$\mathrm{E_{DFT}}[\rho] = \mathrm{T_S}[\rho] + \mathrm{E_{ne}}[\rho] + \mathrm{J}[\rho] + \mathrm{E_{xc}}[\rho] + \mathrm{V_{NN}} \tag{1.3}$$

Trukaketa-korrelazio funtzionalaren arazoa ekiditeko beste aukera bat funtzional hibridoen erabilpena da. Hauetan, DFT eta Hartree-Fock (HF)[120, 121] metodoak (HF hurbilketan elektroi bakoitzaren antolamendu espazialak ez du beste elektroien mugimenduarekiko menpekotasunik) erabiltzen dira, hots, B3LYP eta M06-2X funtzionalak.

B3LYP[122] funtzionala hiru parametroz osaturiko trukaketa-korrelazio funtzional hibrido bat da, zeinetan, trukaketa-korrelazio energia ondorengo ekuazioaren bitartez lortzen den:

$$\mathrm{E_{XC}^{B3LYP}} = (1 - \mathrm{a_0})\mathrm{E_X^{LSDA}} + \mathrm{a_0}\mathrm{E_X^{HF}} + \mathrm{a_x}\mathrm{E_X^{B88}} + \mathrm{a_C}\mathrm{E_C^{LYP}} + (1 - \mathrm{a_C})\mathrm{E^{VWN}} \tag{1.4}$$

119. Cossío, F. P., Calculation of Kinetic Data Using Computational Methods. In *Rate Constant Calculation for Thermal Reactions: Methods and Applications*, John Wiley & Sons: Hoboken, New Jersey, 2011; pp 33–65.
120. Hartree, D. R., *Math. Proc. Camb. Philos. Soc* **1928**, *24*, 89-110.
121. Fock, V., *Z. Phys.* **1930**, *61*, 126-148.
122. Becke, A. D., *J. Chem. Phys.* **1993**, *98*, 1372-1377.

Hiru parametroen balioak hurrengoak dira: a_0, a_x eta a_c, 0.20, 0.72 eta 0.81, hurrenez hurren. E_X^{LSDA} trukaketa funtzionala, E_X^{HF} HF trukaketa funtzionala, E_X^{B88} Becke-ren trukaketa funtzionala,[123] E_C^{LYP} Lee-k eta kideek definituriko korrelazio funtzionala[124] eta E^{VWN} Vosco[125] eta kideek deskribatutako funtzionalei dagozkie. Funtzional hau hainbat arazo kimikoen deskribapen zehatzerako erabili bada ere, interakzio ez-kobalenteen deskribapenerako ez da hain zuzena eta ondorioz, berriki, beste funtzional gehigarri batzuk garatu dira.

Funtzional hauen artean, elkarrekintza ez-kobalenteen analisia errazten duenThrular eta Zhaok garatutako **M06-2X**[126] Minnesota funtzionala aurkitzen da. M06[127] funtzionalarekin alderatuz, M06-2X funtzionalak HF trukaketa funtzionalaren %54 erabiltzen du eta M06-k ordea, %27-a soilik.

Sistema bat aztertu nahi denean erabiliko den funtzionala aukeratu ostean, hurrengo urratsa base atomikoen aukeraketa da. Base atomiko hauek hainbat funtzio desberdinez osatuta daude eta orbital molekularren deskribapena ahalbidetzen dute. 2. kapituluan azaltzen den erreaktibitatea aztertzeko, B3LYP funtzionala eta 6-31G(d,p)[128] base atomikoak erabili ziren sistemen optimizaziorako. Horrez gain, kalkulu puntualak ere burutu ziren M06-2X/6-311+G(2d,2p) maila zehatzagoa erabiliz. Jarraian, erabilitako base atomikoak laburki deskribatzen dira:

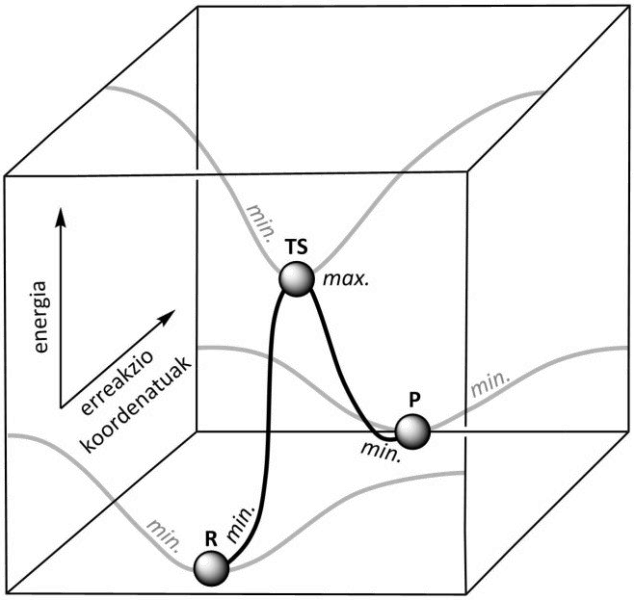
-6-31G(d,p): z bikoitzeko balentzia zatituaren eskema erabiltzen du. Alde batetik *core* elektroiak sei gaussianen bitartez adierazten dira eta bestetik, balentzia elektroiak hiru gaussianekin banatzen dira barne eremurako eta gaussiana batekin kanpo eremurako. (d,p)-k adierazten duen moduan, atomo astunen “p” orbitalak polarizatzeko “d” funtzioak erabili dira eta hidrogenoen “s” orbitalak polarizatzeko ordea, “p” funtzioak erabili dira.

123. Becke, A. D., *Phys. Rev. A* **1988**, *38*, 3098-3100.
124. Lee, C.; Yang, W.; Parr, R. G., *Phys. Rev. B* **1988**, *37*, 785-789.
125. Vosko, S. H.; Wilk, L.; Nusair, M., *Can. J. Phys.* **1980**, *58*, 1200-1211.
126. Zhao, Y.; Truhlar, D. G., *Acc. Chem. Res.* **2008**, *41*, 157-167.
127. Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.* **2008**, *120*, 215-241.
128. Hehre, W. J.; Ditchfield, R.; Pople, J. A., *J. Chem. Phys.* **1972**, *56*, 2257-2261.

-6-311+G(2d,2p): z hirukoitzeko balentzia zatituaren eskema erabiltzen du, balentzia elektroiak hiru funtziorekin deskribatzen ditu, horietako bat, hiru gaussianaz osatuta eta beste biak gaussiana bakarraz. Horrez gain, "+"-k adierazten duen moduan "s" eta "p" funtzio difusoak gehitu zaizkio hidrogenoak ez diren atomoei. Azkenik, bi polarizazio funtzio gehitu zaizkie atomo pisutsu zein hidrogeno atomoei.

1.2.3 Erreakzio mekanismoen analisi konputazionala

Sistema baten energia atomoen posizioaren arabera irudikatzen denean PES-a lortzen da. PES-aren puntu esanguratsuenak puntu geldikorrak dira; energia gradiente balio nuluak dituzten geometriak hain zuzen ere. Era honetan, PES-ean aurkitzen diren puntu geldikorrak aztertuz erreakzio mekanismoaren nondik-norakoak ulertu daitezke (1.4. Irudia).



1.4. **Irudia.** Erreaktibotik (R) Produkturako (P) transformazioa trantsizio-egituraren (TS) bitartez. Energiaren irudikapena erreakzio koordenatuen aurrean.

Erreaktiboak (**R**), produktuak (**P**) zein artekariak aztertzerakoan, geometria hauek erreakzio koordenatu guztietan minimo izaera izango dute eta ondorioz, hauen bigarren deribatuak balio positiboa izango du norabide guztietan. TS-ak kontuan hartzen direnean berriz, bigarren deribatuen balioak positiboak izango dira erreakzio koordenatu guztietan batean izan ezik, non balio negatiboa izango duen. Balio negatibo honek TS–aren geometria erreakzio koordenatu batean maximo bat dela esan nahi du.

Horrez gain, IRC-ek (Intrinsic Reaction Coordinates) erreakzioaren profila ulertzen laguntzen dute, trantsizio-egituraren geometriak, minimo lokalak diren erreaktibo eta produktuekin konektatzen baitituzte.

1.3 PLATINO(II) KONPLEXUAK ETA MOSTAZA NITROGENATUAK MINBIZIAREN TRATAMENDUAN

Minbizia, zelula multzo baten kontrol gabeko hazkuntza, hedapen eta zatiketa ezaugarritzat dituzten zenbait gaixotasun biltzeko erabiltzen den terminoa da. Gainera, aipatutako ezaugarri guzti hauek, inguruko ehunei erasotzeko eta kaltetzeko gaitasuna duten tumore primarioen garapenean dute emaitza. Ehun osasuntsuetan, hazkuntza eta zatiketa zelularra era kontrolatuan ematen dira eta kaltetuak ikusten direnean, programatutako zelula-heriotz mekanismoak aktibatzen dira. Prozesu kantzerigenoak garatzen direnean berriz, zelulek ezohiko garapen prozesuak izaten dituzte eta beharrik gabe ugaltzen jarraitzen dute. Modu honetan, zelula kaltegarriak gorputzean zehar zabaldu daitezke odolaren edo sistema linfatikoaren bitartez ehun desberdinetan tumore berriak garatuz. Prozesu honi metastasia deritzo.[129-131]

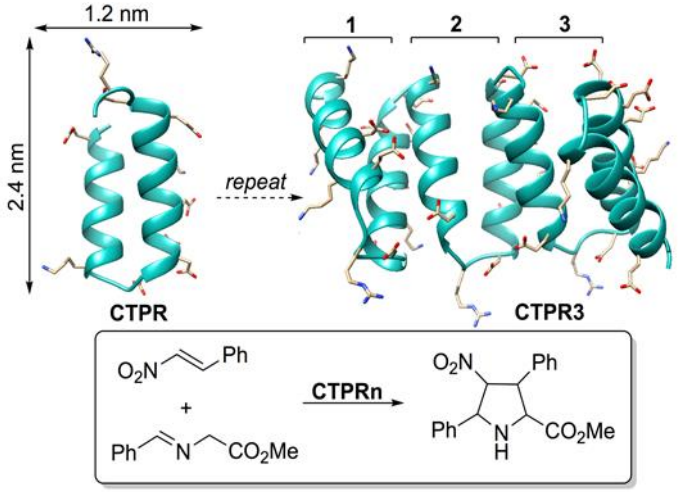
Minbiziaren sorrerak, orokorrean, zelula osasuntsuetan gertatzen diren alterazio genetikoetan du jatorria. Gehienetan, prozesu hauen sortzean hainbat inguruneko-faktoreek efektu zuzena izan ohi dute gainera.[132] Aldaera genetiko desberdinak aurki daitezke; DNA mutazio bakanak, aldaketa epigenetikoak (DNAn zein hainbat proteinetan ematen diren aldaketa kobalente itzulgarriak) [133, 134] edota sekuentzia delezioak esaterako. Aldakortasun hauek direla eta, minbizi mota bakoitzak bere aldaketa genetikoen konbinazio espezifikoa du eta gaur egun, kaltetutako organo edo ehunaren arabera 100 minbizi klase baino gehiago ezagutzen dira. [135] Gaixotasun hauen

129. Todd, A.; Grounwater, P. W.; Gill, J. H., *Anticancer Therapeutics: From Drug Discovery to Clinical Applications*. Wiley-Blackwell: Hoboken, New Jersey, 2017.
130. Link, W., *Principles of Cancer Treatment and Anticancer Drug Developmen*. Springer: Cham. Switzerland, 2019.
131. David, A. R.; Zimmerman, M. R., *Nat. Rev. Cancer* **2010**, *10*, 728-733.
132. Grivennikov, S. I.; Greten, F. R.; Karin, M., *Cell* **2010**, *140*, 883-899.
133. Sharma, S.; Kelly, T. K.; Jones, P. A., *Carcinogenesis* **2009**, *31*, 27-36.
134. Dawson, Mark A.; Kouzarides, T., *Cell* **2012**, *150*, 12-27.
135. Price, P.; Sikora, K., *Treatment of Cancer, 6th edition*. CRC Press: Boca Raton, Florida, 2014.

2.Kapitulua:

Huisgenasa Aktibitatedun Biomolekulen Diseinua: CTPR Proteinak (3+2) Zikloadizioen Biokatalizatzaile Gisa

CTPR proteinak iminen eta dipolarofiloen arteko erreakzio 1,3-dipolarrak katalizatzeko gai direla aurkezten da, orain arte entzima naturalen bitartez lortzea ezinezkoa zen katalisia emanez. EMN bidezko analisiak zein proteinan burututako mutazio desberdinen CTPR-tan aurkitzen diren gune azido zein basikoek katalisian duten garrantzia balioztatu zuten. Bestalde, proteinaren malgutasuna eta gune azido zein basikoen posizioek eragin zuzena dute erreakzioaren diastereokontrolean, monada eta diada katalikoen egituran eragiten baitute. Era berean, *Huisgenasa* aktibitatea ematen duen mekanismoa QM/MM bidezko analisi konputazionalaren bitartez azaltzen da.



2.1 HELBURUAK

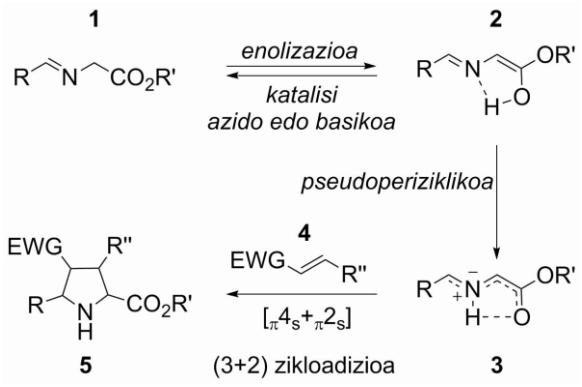
Orain arte ez da deskribatu, iminen eta dipolarofilo π-defizienteen arteko zikloadizio 1,3-dipolarra edota Huisgenen[1] erreakzioa, katalizatzen duten proteinarik.[2] Gure ikerketa taldeak azometino iluroen eta nitroestirenoen arteko (3+2) zikloadizioen inguruan duen esperientzia zabala kontuan harturik, *Huisgenasa* aktibitate esplizitoa duen lehen proteina karakterizatu nahi da.[3-6] Hau dela eta, kapitulu honetan gune katalitiko bezala parte hartu dezaketen hainbat aminoazido azido zein basikoak dituzten CTPR (Consensus Tetratricopeptide Protein) proteinen aktibitatea aztertuko da. Era berean, proteina hauen gain zenbait mutazio burutuko dira berauen diasteroselektibitatearen ulermenerako. Bestalde, CTPR proteinen erreakzio mekanismoa ulertu nahian, QM/MM metodo konputazionalak eta EMN metodoak erabiliko dira.

2.2 CTPR PROTEINEN EGITURA ETA HAUEK BURUTUTAKO KATALISIAREN DISEINUA

2.2.1 CTPR proteinak (3+2) zikloadizioen biokatalizatzaile izateko aukera

Aurretiaz argitaratutako hainbat lanen arabera,[4] **1** imina eta **4** nitroalkeno π-defizienteen arteko (3+2) zikloadizioak katalizatzeko katalisi azido edo basikoaren eragina ezinbestekoa da (2.1 Eskema). Lehenik, katalizatzaile azido edo basikoek iminen enolizazio prozesua sustatzen dute, era honetan **2** 1,3-dipoloak lortuz. Jarraian, prozesu pseudoperizikliko baten bitartez NH-azometino iluroak *in situ* eratzen dira.[5] Artekari hauek oso espezie erreaktiboak izanik, kontzertatua baina era berean asinkronoa den [$\pi 4_s + \pi 2_s$] mekanismoaren bitartez, **5** ester prolinikoak eratzeko gaitasuna dute (2.1. Eskema).

1. Huisgen, R., 1,3-Dipolar Cycloaddition Chemistry. Wiley, Ed. New York, 1984; Vol. 1, pp 1-176.
2. Baunach, M.; Hertweck, C., *Angew. Chem., Int. Ed.* **2015**, *54*, 12550-12552.
3. de Cózar, A.; Cossío, F. P., *Phys. Chem. Chem. Phys.* **2011**, *13*, 10858-10868.
4. Arrieta, A.; Otaegui, D.; Zubia, A.; Cossío, F. P.; Díaz-Ortiz, A.; de la Hoz, A.; Herrero, M. A.; Prieto, P.; Foces-Foces, C.; Pizarro, J. L.; Arriortua, M. I., *J. Org. Chem.* **2007**, *72*, 4313-4322.
5. Costa, P. R. R.; Sansano, J. M.; Cossío, U.; Barcellos, J. C. F.; Dias, A. G.; Nájera, C.; Arrieta, A.; de Cózar, A.; Cossío, F. P., *Eur. J. Org. Chem.* **2015**, 4689-4698.
6. Schleyer, P. v. R.; Wu, J. I.; Cossío, F. P.; Fernández, I., *Chem. Soc. Rev.* **2014**, *43*, 4909-4921.



2.1. Eskema. Katalisi azido edo basikoaren bitartez *in situ* eratzen diren **3** NH-azometino iluroen eta **4** nitroalkenoen arteko (3+2) zikloadizioaren mekanismo orokorra.

Ondorioz, Huisgenen erreakzioa katalizatu zezakeen proteina baten bilaketan, aminoazido basiko (arginina, histidina eta lisina) zein azido (azido glutamikoa eta aspartikoa) ugari zituzten proteinak ezinbestekoa ziren. Aukera guztien artean, tetratrikopeptido errepikapen (TPR) modulua zituzten proteinak aukeratuak izan ziren.[7,8] Gutxi gorabehera, 300 proteina desberdinetan aurkitu daitekeen sekuentzia hau, 34 aminoazidoz osaturiko helize-bira-helize motiboan datza eta funtzio biologiko ugari burutzeko gai dela konfirmatu da.[9-11]

Horrez gain, TPR sekuentzien lerrokatzeak “kontsentsu-TPR” (CTPR) sekuentzien diseinua ahalbidetzen du, modu honetan TPR egitura errepikatu anitz dituzten proteinak lortuz.[12] Bestalde, proteina hauen ezaugarri garrantzitsuenetako bat estruktura honen egonkortasun altua da, zeina errepikapen kopuruarekin batera handituz doan.[13 - 15]

7. Blatch, G. L.; Lassle, M., *Bioessays* **1999**, *21*, 932-939.
8. D'Andrea, L. D.; Regan, L., *Trends Biochem. Sci.* **2003**, *28*, 655-662.
9. Smith, D. F., *Cell Stress Chaperon.* **2004**, *9*, 109-121.
10. Roberts, J. D.; Thapaliya, A.; Martínez-Lumbreras, S.; Krysztofinska, E. M.; Isaacson, R. L., *Front. Mol. Biosci* **2015**, *2*, 71.
11. Zhou, X.; Liao, H.; Chern, M.; Yin, J.; Chen, Y.; Wang, J.; Zhu, X.; Chen, Z.; Yuan, C.; Zhao, W.; Wang, J.; Li, W.; He, M.; Ma, B.; Wang, J.; Qin, P.; Chen, W.; Wang, Y.; Liu, J.; Qian, Y.; Wang, W.; Wu, X.; Li, P.; Zhu, L.; Li, S.; Ronald, P. C.; Chen, X., *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 3174-3179.
12. Main, E. R. G.; Xiong, Y.; Cocco, M. J.; D'Andrea, L.; Regan, L., *Structure* **2003**, *11*, 497-508.
13. Kajander, T.; Cortajarena, A. L.; Main, E. R. G.; Mochrie, S. G. J.; Regan, L., *J. Am. Chem. Soc.* **2005**, *127*, 10188-10190.
14. Kajander, T.; Cortajarena, A. L.; Mochrie, S.; Regan, L., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 800-811.
15. Cortajarena, A. L.; Regan, L., *Protein Sci.* **2011**, *20*, 336-340.

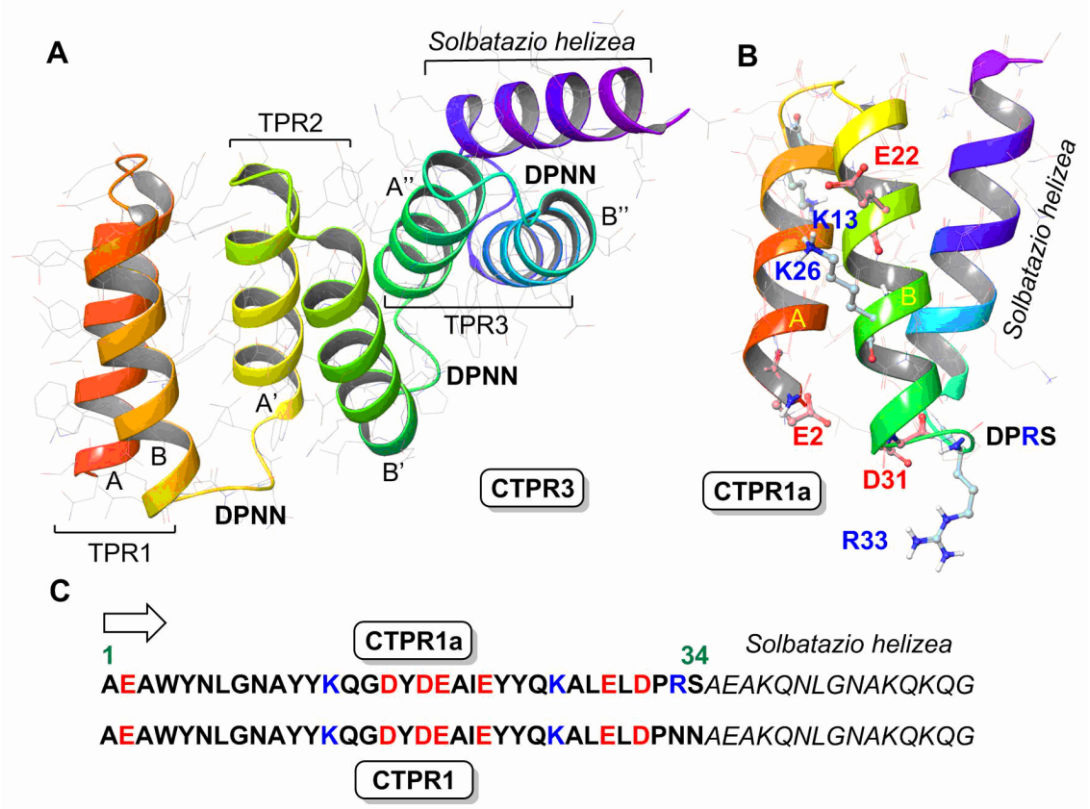
Ondorioz, proteina hauen diseinuaren bitartez tenperatura eta disolbatzaile desberdinetan erabili daitezkeen estruktura sendoak lortu daitezke. [16] Beraz, alsekuentzia hauetan aurkitzen diren gune azido zein basikoak eta estruktura hauen egonkortasuna kontuan hartuta, lehen *Huisgenasa* aktibitatearen deskribapenerako abiapuntu ezin hobea zirela proposatu genuen.

CTPR proteinen aktibitatearen azterketa bi proteina familia desberdinekin burutu zen, CTPRa[13] eta CTPR,[12] hain zuzen ere. Bi serie hauen arteko desberdintasuna sekuentziako azken hondarretan aurkitzen da, non CTPRa multzoa aurkitzen den –DPRS sekuentzia –DPNN sekuentzia izatera aldatzen den CTPR klasean (2.1. Irudia).[14] Bestalde, erreaktibitatearen ebaluaziorako bat eta hiru errepikapeneko sekuentziak erabili ziren, hau da, CTPR1, CTPR1a, CTPR3 eta CTPR3a proteinak. Entzima hauen ezaugarri nabarmenetako bat, errepikapen bakoitzean dauden gune azido (7 CTPRa eta CTPR sekuentzietan) zein basikoen (3 CTPRa-n eta 2 CTPR-n) presentzia da. Aitzitik, CTPR eta CTPRa proteinen kristal egituren azterketak, gune katalitiko izateko potentziala duen E22 eta K26 aminoazidoek eratutako diada katalitikoen presentzia erakutsi zuen (2.1 Irudia). Honetaz gain, monada katalitiko bezala jarduteko aukera duten hainbat glutamato (E), lisina (K), aspartato (D) edota arginina (R) hondar aurkitu daitezke, horien artean, E2, K13, D31 eta R33 (2.1B Irudia).

Ondorioz, CTPR proteinek azometino iluroen eta 1,3-dipolarofiloen arteko erreakzioa katalizatzeko gaitasuna, jatorrizko CTPR1a, CTRP1, CTPR3a eta CTPR3 proteinak erabiliz aztertu zen. Gainera, gune azido eta basikoen funtzioa ulertu ahal izateko, jatorrizko sekuentzien hainbat mutante garatu ziren. Atal honetan azaltzen diren CTPR proteinen diseinu eta sintesia Aitziber López Cortajarena Ikerbasque Research Proffessor-aren CIC biomaGUNE-ko laborategian burutu ziren. Proteina hauen deskribapenari buruzko ezaugarri nagusienak hurrengo atalean azaltzen dira eta

16. Cortajarena, A. L.; Mochrie, S. G. J.; Regan, L., *Protein Sci.* **2011**, *20*, 1042-1047.

gainontzeko karakterizazioa ingeleseko ataleko zati esperimentalean eta III. Eranskinean aurkitzen direlarik.



2.1. Irudia. A) CTPR3 proteinaren egitura (pdb kodea: 1NA0) non hiru errepikapenak (AB, TPR1-3) eta solbatazio helizea (SHx) irudikatuta dauden. **B**) X-izpien bidezko difrakzioaren bitartez lortutako CTPR1a proteinaren egitura. E2, E22 eta D31, hondar azidoak gorriz eta K13, K26 eta R33 hondar basikoak urdinez nabarmenduta agertzen dira. Egitura hau CTPR8a (pdb kodea: 2HYZ) homologo supersimetrikoa baliatuz garatu da. **C**) CTPR1 eta CTPR1a proteinen sekuentziak. Gune azido eta basikoak, gorriz eta urdinez, hurrenez hurren, azaltzen dira.

2.2.2 CTPR proteinen karakterizazioa eta erreakzio baldintzetan duten egonkortasunaren analisia

Aipatu bezala, CTPR proteinen funtzio katalitikoaren ebaluaziorako hainbat mutazio burutu ziren (2.1. Taula). Lehenik eta behin, CTPR seriean, CTPR serieko R33-S34 aminoazidoak, N33-N34 aminoazidoengatik ordezkatzen direla kontuan izan behar da. CTPR1a seriean aurkitzen diren E2, E22 eta D31 aminoazidoak alaninen ordez aldatu ziren eta K26-a berriz, glutamina batengatik. Hasierako ideia, mutazio guztietan alaninak gehitzea bazen ere, aminoazido luzeago baten efektua aztertzeko glutamina baten bat ere gehitu zen. CTPR1 multzoan gune basiko guztien banakako mutazioak (K13A, K26Q

eta K26A) zein mutazio bikoitz bat (K13A+K26A) ere burutu ziren, bigarren kasu honetan gune basiko guztiak ezabatzen zirelarik. Antzeko metodologia erabiliaz, CTPR3 multzoan gune basikorik ez zuen mutantea garatu zen, honetarako, K13A eta K26A mutazioa burutuz eta solbatazio helizea ezabatuz.

Proteinekin katalisia burutu nahi bada, beste proteinen ezpurutasun urriek aktibitatea eraldatu dezaketela kontuan izan behar da. Hau dela eta, proteina guztien purifikazioa eta masa espektrometria bidezko karakterizazio burutu zen. 2.1. Taulan, proteina guztien pisu molekularra azaltzen da, karakterizazioaren gainontzeko informazioa ingeleseko ataleko III.eranskinean zein atal esperimentalean deskribatzen direlarik.

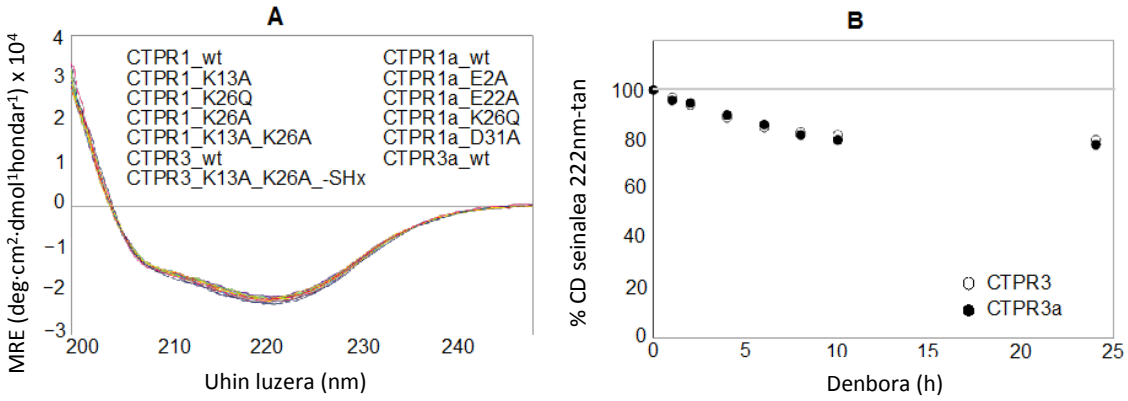
2.1. Taula. Purifikatutako jatorrizko CTPR proteinen zein mutanteen masa espektrometria.

Proteina	Kalkulatutako PM (Da)	PM esperimentala[a] (Da)
CTPR1a_wt	6018.54	6027.66
CTPR1a_E2A	5960.50	5966.27
CTPR1a_E22A	5960.50	5966.63
CTPR1a_K26Q	6018.84	6025.89
CTPR1a_D31A	5974.53	5979.38
CTPR1_wt	6329.79	6332.05
CTPR1_K13A	6272.69	6274.20
CTPR1_K26Q	6329.75	6339.79
CTPR1_K26A	6272.69	6275.65
CTPR1_K13A+K26A	6215.60	6215.65
CTPR3a_wt	14079.50	14081.10
CTPR3_wt	14360.34	14362.36
CTPR3_K13A+K26A-SHx[b]	11979.50	11536.29

[a]Masa espektrometriaz neurtuta. [b]-SHx solbatazio helizea. Jatorrizko proteina serieak (CTPR1a, CTPR1, CTPR3a eta CTPR3) letra lodiz irudikatzen dira.

Jatorrizko proteina batean mutazioak burutzen direnean egitura helikoidalaren egonkortasuna kontuan izan beharreko faktoreetako bat da. Mutazioak burutzean, proteinen egitura α-helikoidala mantentzen zela baieztatzeko helburuarekin dikroismo zirkular (CD) bidezko analisia burutu zen (2.2A Irudia). Horrez gain, CTPR proteinen

egonkortasun konformazionala aztertu zen disolbatzaile desberdinetan. Iminek uretan duten solubilitate baxua dela eta, entzima hauek katalizatutako (3+2) zikloadizio erreakzioa THF-n burutzea beharrezkoa zen, hortaz, CTPR3 eta CTPR3a proteinen CD analisia ere disolbatzaile honetan burutu zen. 24 ordu igaro ondoren CD seinalearen %80 kontserbatzen zela ikusi zen eta beraz, egitura α-helikoidala bi serieetan mantentzen zelarik (2.2B Irudia). Aitzitik, 24 ordu ostean proteinaren suspentsioa berriz ere burutzean CD seinalearen berreskuratzea ikusi zen, hortaz, CD seinalearen txikitzea solubilitatearen galera dela eta gertatzen da eta ez proteinak desnaturalizatzen direlako. Modu honetan, egitura sekundarioaren kontserbazioa ikusita, (3+2) zikloadizioak THF–n burutu daitezkeela ondorioztatu daiteke.

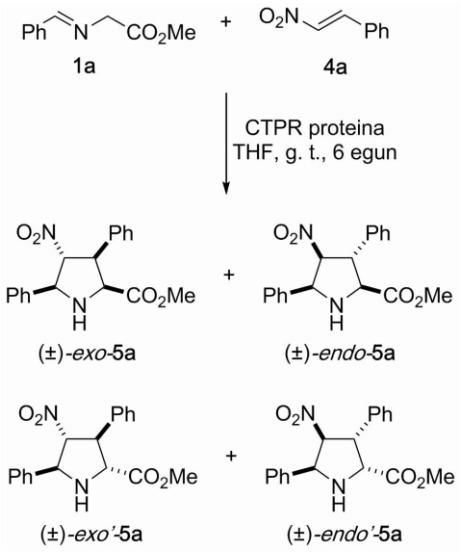


2.2. Irudia. A) CTPR proteinen dikroismo zirkularra (CD). MRE: Hondarren eliptizitate molarra. –SHx solbatazio helizearen gabezia adierazten du. **B**) Erreakzio baldintzetan CTPR3 eta CTPR3a proteinen egonkortasuna.

2.3 CTPR PROTEINEK BURUTUTAKO *HUISGENASA* AKTIBITATE KATALITIKOA

CTPR proteinen egitura eta egonkortasun analisiak burutu ostean, berauen *Huisgenasa* aktibitatearen ebaluazioa izan zen hurrengo urratsa. Azometino iluroen eta dipolarozaleen arteko (3+2) zikloadizioak sakonki aztertu izan dituzte kimikari esperimental zein konputazionalek eta katalizatzaile kiralen erabilerak nagusiki *endo* eta *exo* zikloaduktuen eraketa ekarri izan du.[3] Berriki argitaratutako ikerketa batean,[4] azometino iluro eta nitroestirenoen arteko (3+2) zikloadizioa bai baldintza termikoetan zein mikrouhinak erabiliz deskribatu da. Horrez gain, erreakzio hauek disolbatzaile gabe burutu zirenean *exo*, *endo* eta *endo*' produktuak lortzen ziren. Emaitza hauek kontuan harturik, **1a** N-benzilideno-glizinatoaren eta **4a** (*E*)-β-nitroestirenoaren arteko

erreakzioa aukeratu zen CTPR proteinen aktibitate entzimatikoa aztertzeko. Katalisia THF-n burutu zen, sei egunetan zehar giro tenperaturan (g.t.) irabiagailu orbital bat erabiliz (2.2. Eskema).



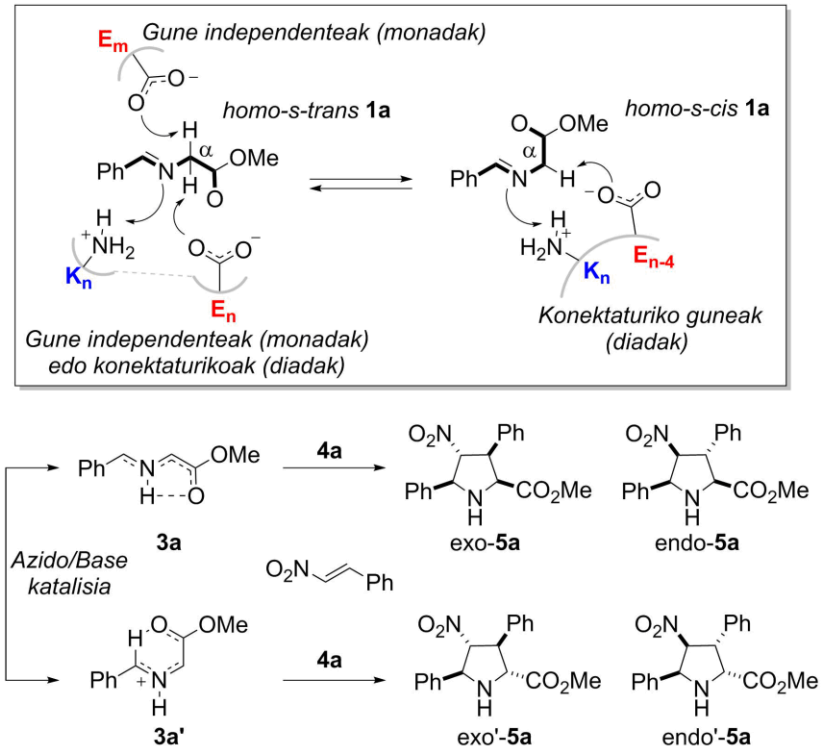
2.2. Eskema. CTPR proteinek katalizatutako **1a** *N*-benzilideno-glizinatoaren eta **4a** β-nitrostirenoaren arteko (3+2) zikloadizioa.

Lehenik eta behin, jatorrizko (wt) proteinak aztertu ziren, CTPRa eta CTPR sekuentzien artean desberdintasun adierazgarria lortuz. Errepikapen bateko zein hiru errepikapenez osaturiko CTPRa proteinek, CTPR1a_wt eta CTPR3a_wt, posible diren lau pirrolidina esteroisomeroen, *exo*-**5a**/*endo*-**5a**/*exo'*-**5a**/*endo'*-**5a**, sintesia katalizatzen dute 1:1 proportzioan eta %40-eko etekinarekin (lau zikloaduktu diasteroisomerikoen karakterizazioa ingelesez idatzitako sekzioko zati esperimentalean zein III.eranskinean azaltzen da). Beraz, CTPRa sekuentziek (3+2) zikloadizio urritan lortzen diren *exo'*-**5a** eta *endo'*-**5a** prolinen eraketa ere katalizaten dute. Bestalde, CTPR1_wt eta CTPR3_wt entzimek soilik *exo*-**5a** eta *endo*-**5a** zikloaduktuen formazioa katalizatzen dute %20-eko etekin globalarekin (2.2. Taula).

Emaitza hauek kontuan izanda CTPRa_wt entzimek burututako lau produktuen katalisia esteroisomeroak diren **3a** zein **3a**' azometino iluroen *in situ* eraketak sustatzen zuela proposatu genuen. Hortaz, **1a** iminaren N-C_α lotura sinplearen errotazioak *homo-s-cis* eta *homo-s-trans* konformazioen arteko oreka egoera bat sortzen du (2.3. Eskema). Honela, banaka edo binaka aurkitzen diren gune azido edota basikoen monada bidezko

katalisia burutu dezakete C_α-H loturako protoi abstrakzio/transferentzia prozesua sustatuz eta **3a** zein **3a'** dipoloen eraketa emanez bi prozesu independenteen bitartez. Horrez gain, azometino iluro hauek, K_n/E_n-4 diada katalitikoek bultzatutako abstrakzio/transferentzia prozesuen bitartez lortu daitezke (2.3. Eskema).

2.3. Eskeman azaltzen dira *endo*-**5a**/*exo*-**5a** eta/edo *endo'*-**5a**/*exo'*-**5a** (3+2) zikloadizio produktuen katalisia eman dezaketen proposatutako bi mekanismoak. Alde batetik, **3a** 1,3-dipoloaren eta **4a** dipolarofiloaren arteko erreakzioak [$\pi4_s$+$\pi2_s$] mekanismo kontzertatu asinkrono baten bitartez *endo*-**5a** eta *exo*-**5a** zikloaduktuen eraketa eman dezake. Bestalde, **3a'** azometino iluroaren eta **4a** β-nitroestirenoaren arteko erreakzio suprafazialak, *endo'*-**5a** eta *exo'*-**5a** prolina ez-naturalen sintesia ematen du. Beraz, prozesu hauen bitartez lau diasteroisomeroak eratu daitezke.



2.3. Eskema. Katalisi azido zein basiko bidezko **3a** eta **3a'** azometino iluroen eraketa ematen duten proposatutako mekanismoak. [$\pi4_s$+$\pi2_s$] mekanismoen bitartez emandako bi azometino iluroen eta **4a** nitroestirenoaren arteko zikloadizio 1,3-dipolarrak.

CTPR proteinen wt serieen aktibitate katalitikoa aztertu ondoren, gune azido zein basikoak modifikatuz diseinatutako mutanteak ebaluatu ziren. Burututako mutazio orok *exo'*-**5a** eta *endo'*-**5a** (3+2) zikloaduktuen eraketa inhibitzen zuen eta hortaz, bakarrik

exo-**5a** eta *endo*-**5a** prolinak bakarrik sortzen zituzten 1:1 proportzioan. Horrez gain, CTPR1a_wt, CTPR1_wt eta CTPR3_wt proteinen mutazioek etekina zein TON (konbertsio zenbakia) eta TOF (konbertsio frekuentzia) balioen murrizpena eragiten zuten (2.2. Taula).

CTPRa serieko gune azidoak (E2, E22 eta D31) eraldatzen zirenean, etekin orokorra %30-era murrizten zen, TON eta TOF balioak ordea, 10^4 eta $70h^{-1}$ balioetara jaisten ziren, hurrenez hurren. CTPR zein CTPRa proteinen gune basikoak, K13 eta K26, mutatzen zirenean lorturiko TON eta TOF balioak baxuagoak ziren eta errendimenduak %11-24 artekoak. Era berean, K13 eta K26 posizioetan eginiko mutazioek etekin baxuenak eman zituzten CTPR1 (11%) zein CTPR3 (18%) proteina multzoetan. Emaitza hauek kontuan hartuz, lisinak (K), gune basikoak hain zuzen ere, aspartato (D) eta glutamato (E) aminoazidoak baino aktiboagoak direla ondorioztatu daiteke. Bestalde, CTPR proteinetan eginiko mutazioen bidez *Huisgenasa* aktibitatea ezin izan zen guztiz inhibitu. Beraz, emaitza hauek, arginina (R), aspartato, glutamato zein lisinek independenteki *exo*-**5a** eta *endo*-**5a** aduktuak *Huisgenasa* aktibitatearen bitartez eratzeko gai direla iradokitzen dute.

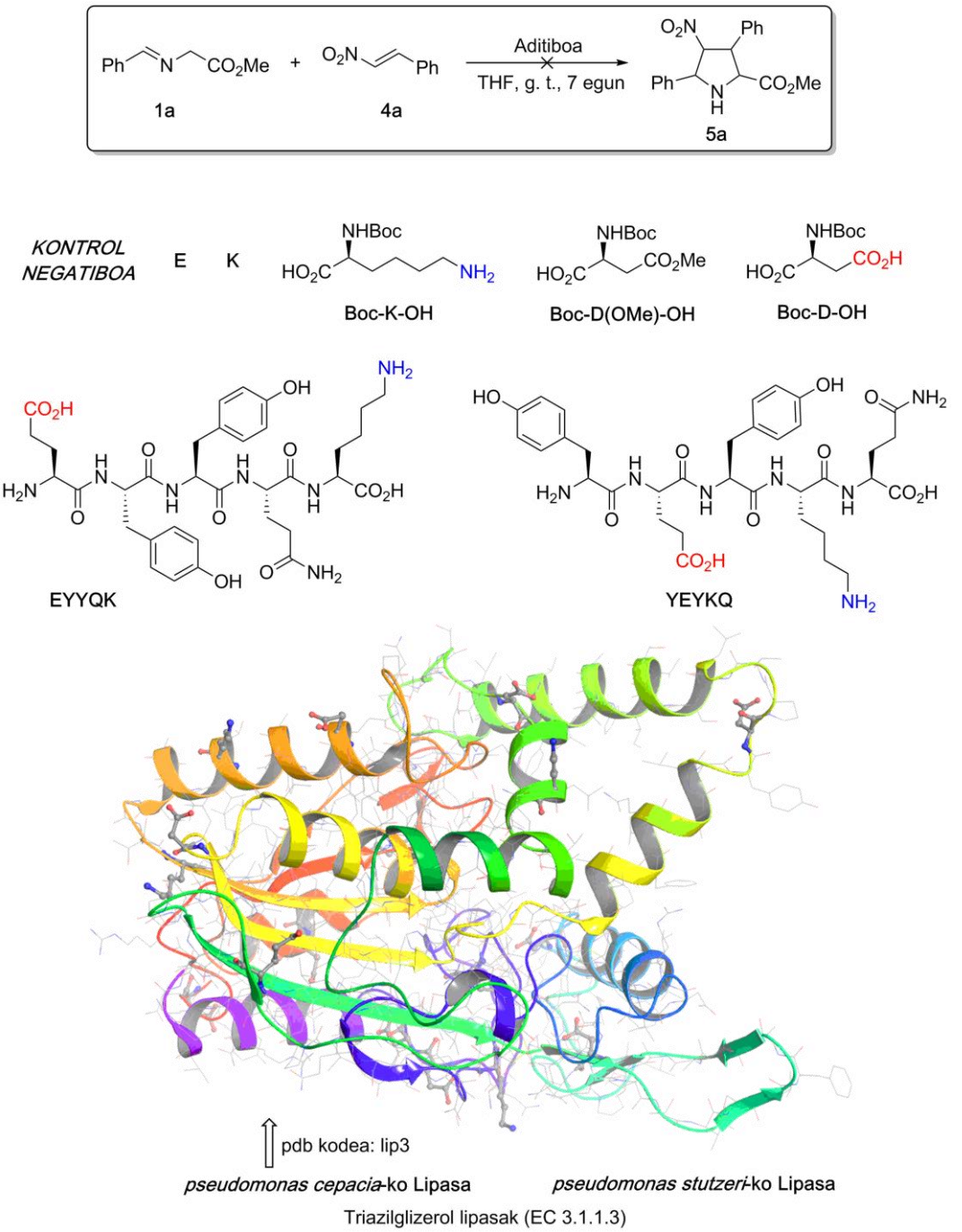
2.2. Taula. CTPRa/CTPR proteina desberdinek katalizatutako **1a** imina eta **4a** β-nitroestirenoaren arteko erreakzioa.

Proteinen Mutazioa(k)	Etekina [a] mmol ± 0.1·10^{-2} [%]	Estereoisomeroen Proportzioa[b] Yield (mmol·10^{-2}) [%]				TON[c] x10^3	TOF[d] (h^{-1})
		endo-**5a**	*exo*-**5a**	*endo'*-**5a**	*exo'*-**5a**		
CTPR1a_wt[e]	**12.7 [41]**	**2.5 ± 0.2 [20]**	**3.6 ± 0.4 [28]**	**3.0 ± 0.3 [24]**	**3.6 ± 0.4 [28]**	**14.2**	**98.8**
CTPR1a_E2A	8.9 [29]	4.4 ± 0.6 [49]	4.6 ± 0.7 [51]	--	--	10.1	70.2
CTPR1a_E22A	9.6 [31]	4.3 ± 0.6 [45]	5.3 ± 0.9 [55]	--	--	10.9	75.4
CTPR1a_K26Q	7.4 [24]	3.3 ± 0.4 [44]	4.1 ± 0.6 [56]	--	--	8.4	58.2
CTPR1a_D31A	9.6 [31]	4.5 ± 0.7 [47]	5.1 ± 0.9 [53]	--	--	10.5	72.9
CTPR1_wt[e]	**6.2 [20]**	**2.6 ± 0.2 [41]**	**3.6 ± 0.5 [59]**	**--**	**--**	**6.8**	**47.1**
CTPR1_K26Q	5.5 [18]	2.5 ± 0.2 [45]	3.1 ± 0.3 [55]	--	--	6.3	43.5
CTPR1_K13A	4.0 [13]	1.7 ± 0.1 [42]	2.3 ± 0.2 [58]	--	--	4.4	30.3
CTPR1_K26A	4.3 [14]	1.9 ± 0.1 [44]	2.4 ± 0.2 [56]	--	--	4.9	33.8
CTPR1+K13A +K26A	3.4 [11]	1.5 ± 0.1 [43]	1.9 ± 0.1 [57]	--	--	3.9	27.0
CTPR3a_wt[e]	**12.4 [40]**	**3.1 ± 0.2 [25]**	**3.4 ± 0.4 [27]**	**3.1 ± 0.4 [25]**	**2.8 ± 0.3 [23]**	**13.6**	**94.8**
CTPR3_wt[e]	**7.4 [24]**	**3.8 ± 0.5 [51]**	**3.6 ± 0.5 [49]**	**--**	**--**	**8.4**	**58.6**
CTPR3+K13A +K26A-SHx[f]	5.6 [18]	2.7 ± 0.3 [48]	2.9 ± 0.3 [52]	--	--	6.1	42.6

[a]Isolaturiko zikloaduktuen etekinen konbinazioz lorturiko etekina. [b]Purifikazio prozesuaren ostean hiru experimentu desberdinen batazbestekoari dagokio. Erroreak desbideratze estandarraren bitartez kalkulatzen dira. **1a** da erreaktio mugatzailea (0.31 mmol). [c]TON: konbertsio zenbakia. [d]TOF: konbertsio frekuentzia. [e]wt: mutazio gabeko jatorrizko sekuentzia. [f]-SHx: solbatazio helizea. Jatorrizko proteinak (wt) letra lodiz irudikatzen dira.[17]

17. Harris, D. C., *Quantitative Chemical Analysis, 8th ed*. W.H. Freeman: New York, 2010.

Zentro azido eta basikoek CTPR proteinen eta hauen mutanteek sustatutako *Huisgenasa* aktibitatean duten eragina aztertzeko, aurrez deskribatutako erreakzioa beste amino azido eratorrien, peptidoen zein entzimen presentzian ere burutu zen (2.3. Irudia).



2.3. Irudia. **1a** imina eta **4a** nitroestirenoaren arteko (3+2) zikloadizioan ebaluatutako aminoazidoen deribatuak, pentapeptidoak eta triazilglizerol lipasak.

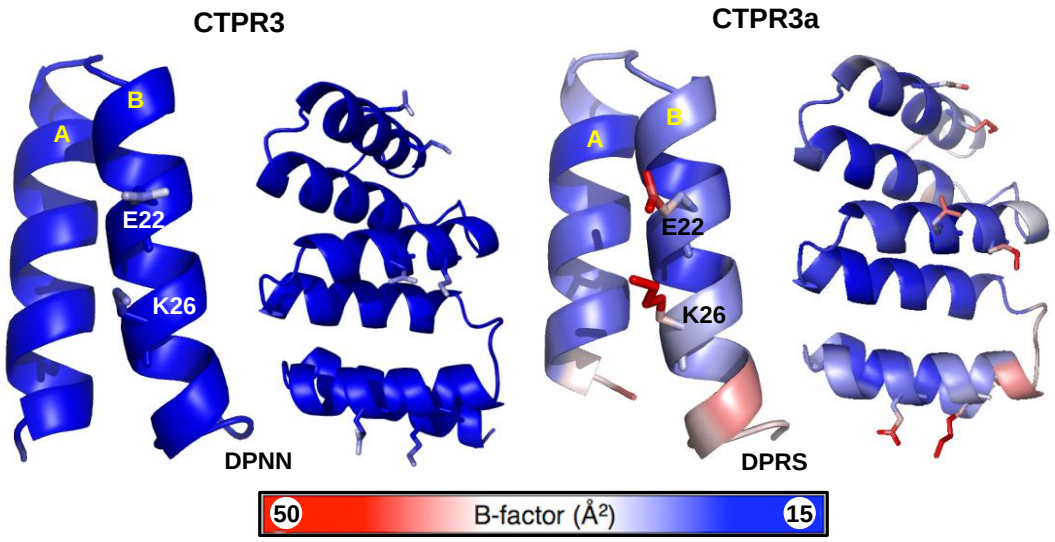
Zehazki, lisina (K), azido glutamikoa (E) zein NH-Boc eratorriak; Boc-K-OH, Boc-D(OMe)-OH eta Boc-D-OH ebaluatuak izan ziren. Hauetaz gain, EYYQK eta YEYKQ pentapeptidoak ere aztertuak izan ziren. Bukatzeko, *Pseudomonas cepacia*-ko eta *Pseudomonas stutzeri*-ko THF-n aktiboak diren bi triazilglikol lipasa ere ebaluatu ziren (2.3. Irudia).[18-20] Egindako azterketa guzti hauetan ez zen *Huisgenasa* aktibitaterik lortu, modu honetan CTPR proteinen estrukturak **5a** zikloaduktuen sintesian duen garrantzia konfirmatzen zelarik (2.3 Irudia).

2.2. Taulan azaltzen den moduan, *endo'*-**5a** eta *exo'*-**5a** aduktuak soilik CTPRa proteinen, CTPR1a eta CTPR3a, katalisiaren bitartez lortu ziren. CTPR eta CTPRa proteinen arteko diferentzia nagusienetako bat, CTPR proteinen egonkortasun handiagoa eta beraz, malgutasun txikiagoa da. Proteinen egonkortasun hau ebaluatzeko bi parametro interesgarri, errepikatutako unitateen egonkortasun intrintsekoa (H) eta errepikapenen arteko akoplamendua (J) dira.[13] Bi parametro hauen analisiak, CTPR proteinaren egonkortasun handiagoa baieztatu zuten, 4.95 eta 3.66 kcal/mol H balioak lortuz CTPR eta CTPRa proteinentzat hurrenez hurren. Gainera, desnaturalizazio termikoaren analisiak ere burutu ziren beraien arteko egonkortasuna alderatzeko.[12] Aurrez argitaratutako lanek, CTPR3 proteinak 14.7˚C altuagoa den desnaturalizazio puntua duela erakutsi zuten, 83.0˚C eta 68.3˚C, hurrez hurren. Ondorioz, CTPR3a proteinaren desegonkortasun handiagoa erakutsi zen.

CTPR eta CTPRa proteinen kristal egiturak eskuragarri izateak, proteinen atal desberdinen malgutasunak B faktoreen[21] erabileraren bitartez ebaluatzea ahalbidetu zuen. Are gehiago, *endo'*-**5a** zein *exo'*-**5a** aduktuen eraketan E22-K26 diada katalitikoak zuen garrantzia kontuan izanik, honen malgutasuna aztertzea ere posible izan zen. 2.4. Irudian erakutsitako B-faktoreen banaketan ikusten den moduan, CTPR3a proteinak, orokorrean, CTPR3-ak baino balio handiagoak ditu, honen malgutasun handiagoa

18. Aoyagi, Y.; Agata, N.; Shibata, N.; Horiguchi, M.; Williams, R. M., *Tetrahedron Lett.* **2000**, *41*, 10159-10162.
19. Aoyagi, Y.; Saitoh, Y.; Ueno, T.; Horiguchi, M.; Takeya, K.; Williams, R. M., *J. Org. Chem.* **2003**, *68*, 6899-6904.
20. Hoyos, P.; Fernández, M.; Sinisterra, J. V.; Alcántara, A. R., *J. Org. Chem.* **2006**, *71*, 7632-7637.
21. Sun, Z. T.; Liu, Q.; Qu, G.; Feng, Y.; Reetz, M. T., *Chem. Rev.* **2019**, *119*, 1626-1665.

konfirmatuz. E22-K26 diadari dagokionez, CTPR3a proteinarentzat lortutako balioak ere handiagoak dira.



2.4. Irudia. CTPR (PDB ID: 1NA0) eta CTPRa (PDB ID: 2HYZ) proteinen egiturak. Kasu bakoitzean, ezker aldean errepikapen bakarreko sekuentzia eta eskuinean hiru errepikapenekoa erakusten dira. B-faktorearen balioak 15 Å²-tik 50 Å²-era erakusten dira, urdinetik gorrirako erabilitako eskala baten bitartez. E22 eta K26 "sticks" irudikapena erabiliz azaltzen dira.

Hortaz, CTPR3a proteinan aurkitzen diren E22 eta K26 aminoazidoek osatutako diada katalitikoak erakusten duen malgutasun handiagoak, **1** iminak proteinarekin duen interakzioan eta beraz, **3** 1,3-dipoloak eratzeko behar diren konformazioetan eragin zuzena izan dezake. Aitzitik, CTPR3 proteinak erakusten duen malgutasun murritzagoak (3+2) zikloadizioa katalizatzeko behar diren konformazioak lortzea zaildu dezake. Hurrengo ataletan, aurkeztutako hipotesien QM/MM metodo eta EMN-ren bidezko azterketa azaltzen da.

2.4 QM/MM BIDEZKO CTPRa PROTEINAREN *HUISGENASA* AKTIBITATEAREN AZTERKETA

Aurreko atalean, emaitza esperimentaletan oinarrituta, **3a** eta **3a'** azometino iluroak lortzeko bi mekanismo desberdin proposatu dira. Alde batetik, CTPR sekuentzian aurkitzen diren E_{n-4}/K_n aminoazidoek era kooperatibo batean lan egin dezakete, hau da, diada katalitiko moduan, **3a** zein **3a'** 1,3-dipoloen katalisia emateko. Bestetik, banaka aurkitzen diren gune azido zein basikoek monada kalitiko moduan lan egin dezakete **3a**

azometino iluroa eratuz. **3a** zein **3a'** artekariak eratu ondoren, hauen eta **4a** dipolarofiloen arteko (3+2) zikloadizioaren bidez, **5a** 4-nitropolina esterrak lortzen dira.

Proposatutako mekanismo hauek ebaluatzeko, ikus 2.3. Eskema, erreakzioaren QM/MM bidezko analisia burutu zen. Hortaz, CTPR1a proteinak katalizatutako (3+2) zikloadizio erreakzioa ONIOM metodologia bitartez eta M06-2X/6-311+G(2d,2p):dreiding//B3LYP/6-31G(d,p):dreiding kalkulu maila erabiliz aztertu zen (ingeleseko "Computational Methods" atalean erabilitako metodologiaren nondik-norakoak azaltzen dira). QM atala **1a** iminak, **4a** nitroalkenoak eta katalisian parte hartzen duten aminoazido nagusiek eratzen dute eta proteinaren gainontzeko ingurunea MM metodologia bidez aztertu da.

2.4.1 E22-K26 diada katalitikoaren bitartez burututako 3a eta 3a' azometino iluroen eraketa mekanismoaren azterketa

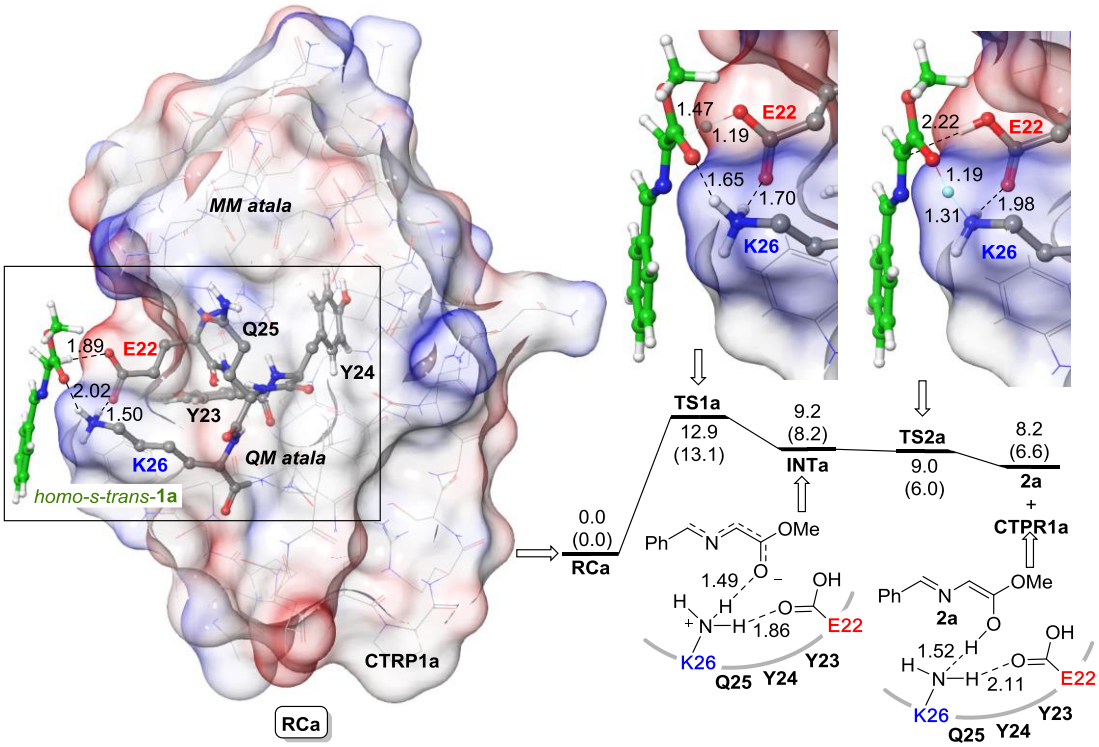
Proposaturiko lehenengo hipotesian, E22-K26 diada katalitikoak bultzatutako (3+2) zikloadizioaren bitartez, lau estereoisomeroak posibleak lortu zitezkeela azaltzen zen. Gune aktiboaren deskribapenerako, diada osatzen duten E22 eta K26 posizioak eta hauek konektatzen dituzten Y23, Y24 eta Q25 aminoazidoak QM atalean sartu ziren, gainontzeko proteina MM metodoen bitartez deskribatu zelarik (2.5. eta 2.6. Irudiak).

2.3. Eskeman azaltzen zen moduan, **1a** iminak, **3a** eta **3a'** azometino iluroen aurrekariak diren *homo-s-trans* zein *homo-s-cis*, hurrenez hurren, konformazioak hartu ditzake N-C_α loturaren inguruko angeluaren errotazioaren bitartez. Lehenik eta behin, **1a** *homo-s-trans* konformazioa duen imina gune katalitikora hurbiltzen da honekin elkarrekintza bat sortuz **RCa** konplexua eratzeko. Gure kalkuluen arabera, **RCa**-k K26-aren amino taldearen eta E22-aren karboxilatoaren oxigeno baten arteko hidrogeno-lotura bat erakusten du (1.50 Å). Horrez gain, **1a** iminak bi elkarrekintza ahulen bitartez bi aminoazido nagusiekin interakzionatzen du, hasierako egitura hau egonkortuz. Alde batetik, K26-aren amino taldeko beste hidrogeno batek iminaren ester taldean aurkitzen den eta sp^2 hibridazioa duen oxigenoarekin elkarrekiten du (2.02 Å) eta bestetik, E22 posizioko beste oxigenoak **1a**-ren metileno taldeko hidrogeno batekin elkarrekintza bat

eratzen du. **RCa** konplexu hau, *homo-s-trans* konformazioa duen iminaren mekanismoaren hasiera puntu bezala hartzen da.

Hasierako konplexua eratu ostean, **1a** imina **INTa** enolatora eraldatzen da **TS1a** trantsizio-egituraren bitartez (2.5. Irudia). Lehenengo urrats honetan E22-ak iminaren C_α posizioko hidrogeno baten abstrakzioa burutzen du, Gibbsen energia askeari dagokionez 13.1 kcal/mol-eko aktibazio energia duen prozesu batean. Hidrogenoaren transferentzia honetan parte hartzen duten distantziak kontuan hartuz, (C----H--O, 1.47 Å eta 1.19 Å, hurrenez hurren), erreakzio hau transitzio-egoera berantiar baten bidez gertatzen dela esan daiteke. Bestalde, enolato honetan aurkitzen diren elkarrekintzak desberdinak dira **RCa**-rekin alderatuz. E22 eta K26 hondarren arteko hidrogeno lotura mantentzen bada ere, honen distantzia nabarmenki handiagoa da, 1.50-tik 1.86-ra Å. sp^2 hibridazioa duen oxigeno atomoa ordea, amino taldetik gertuago aurkitzen da, 1,49 Å-eko distantziara, honek erreakzioaren hurrengo urratsa errazten duelarik.

INTa-k duen aldeko disposizio honek, beraz, **2a** enola eratzeko enolizazio prozesua ahalbidetzen du. **TS2a** egituraren bitartez ematen den prozesu hau K26-an aurkitzen den amino taldeko protoi baten **INTa**-rako transferentziari dagokio. Enol egitura honetan, bi aminoazidoen arteko elkarrekintza mantentzen da (2.11 Å) eta aldiz, erreaktiboak proteinarekin interakzionatzen du hidrogeno lotura baten bitartez (1.52 Å). Ondorengo urrats batean, 2.1 eskeman adierazitiko prototropia pseudoperizikliko baten bitartez *endo*-**5a** eta *exo*-**5a** zikloaduktuen aurrekaria den **3a** azometino iluroa sortuko litzateke.



2.5. **Irudia.** CTPR1a proteinak katalizatutako *homo-s-trans* konformazioan aurkitzen den **1a** iminaren **2a** enolerako eraldaketa erreakzioaren mekanismoaren analisia QM/MM ONIOM(M06-2X/6-311+G(2d,2p):dreiding)//ONIOM(B3LYP/6-31G(d,p):dreiding) metodologia erabiliaz. QM atala **1a** (berdez) eta E22-Y23-Y24-Q25-K26 (grisez) aminoazidoen bitartez osatuta dago eta "ball-and-stick" ereduaren bitartez irudikatzen da. Distantziak Å-etan ematen dira. Puntu geldikorretan (artekariak eta trantsizio-egoerak) adierazitako balioak energia erlatiboei eta Gibbs-en energia askeei 298 K-tan (parentesi artean) dagozkie, kcal/mol-etan adierazita.

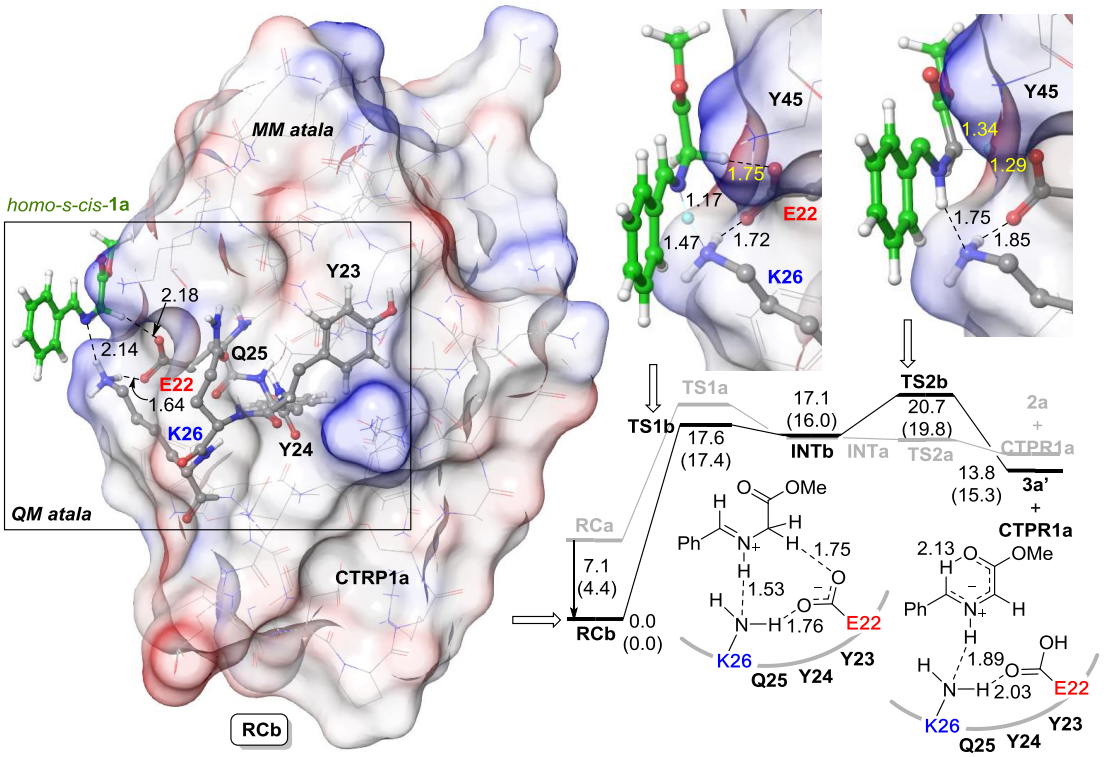
Hala ere, 2.3. Eskeman azalduta agertzen zen moduan, **1a** iminak *homo-s-cis* konformazioa ere eduki dezake N-C_α lotura bakunaren inguruko errotazioaren bidez. Hortaz, hurrengo lerroetan azalduko den moduan, E22/K26 diadak aurrez azalduriko antzeko elkarrekintza sortu dezake **3a'** azometino iluroaren katalisia emanez (2.6. Irudia).

Lehenik, *homo-s-cis* konformazioa duen **1a** imina CTPR proteinara hurreratzen da **RCb** konplexua eratzeko. Hasierako egitura honetan, K26a-ren amino taldeko hidrogeno batek eta E22-ko oxigeno batek hidrogeno lotura bat eratzen dute (1.64 Å). Era berean, K26-ko beste hidrogeno batek erreaktiboaren N-arekin (2.14 Å) eta E22-ko beste oxigenoak iminaren C_α atomoan aurkitzen den hidrogeno batekin (2.18 Å)

hidrogeno lotura ahulagoak eratzen dituzte. **RCb** konplexu honetan aurkitzen diren elkarrekintzak aurreko **RCa**-n ematen zirenekin alderatzen baditugu, *homo-s-trans* edota *homo-s-cis* konformazioen eta proteinaren artean ematen den elkarrekintza desberdina dela ikus daiteke. Alde batetik, **RCa**-n K26-ak sp^2 hibridazioa duen oxigenoarekin elkarrekiten duten bitartean, aminoazido honek iminaren N atomoarekin du interakzioa **RCb** konplexuan. Honela, proteinak aurkezten duen malgutasun handiak (2.4. Irudia) CTPR1a-n ematen den hasierako geometriaren distortsio esanguratsua ahalbidetzen du. Honen ondorioz, bi konplexuen erreaktibitatea desberdina da, **RCb**-ren bitartez **3a'** azometino iluroa eratzen den bitartean, **RCa**-ren bidez, ordea, **3a** sortzen da. Gainera, konformazio berri honen egitura aurrekoa baino 4.4 kcal/mol azpitik aurkitzen da Gibbs energiari dagokionez.

K26-aren amonio taldearen orientazio berri honek *homo-s-cis* **1a**-ren N protonazioa ahalbidetzen du, **RCb**-a baino 16kcal/mol ezegonkorragoa den **INTb**-ren eraketa endergonikoa emanez. Prozesu hau, 17.4 kcal/mol-eko aktibazio energi duen trantsizio-egoera berantiar baten, **TS1b**, bitartez ematen da (C=N--H 1.17 Å eta H----NH_2-K26 1.47 Å) Bestalde, **INTb**-n eta **RCb**-n ematen diren elkarrekintzei dagokienez, C_α posizioko hidrogenoaren eta E22-ko oxigenoaren arteko distantziaren murrizketa bat ematen da 2.18 Å-tik 1.75 Å-ra.

Ondorengo urratsean, E22-an aurkitzen den karboxilato taldeak iminio artekari honetako, **INTb**, α posizioa desprotonatu dezake **3a'** azometino iluroa emanez. Prozesu hau, 3.8 kcal/mol-eko aktibazio energia duen **TS2b** sinkronoaren bitartez ematen da; C_a---H (1.34 Å) eta H---O-E22 (1.29 Å). Eratutako **3a'** azometino iluroak elkarrekintza zuzena eratzen du K26-arekin hidrogeno lotura bidez (1.89 Å) konplexu bat eratuz, non E22-ak K26-arekin hidrogeno lotura baten bidezko elkarrekintza duen (2.03 Å).



2.6. Irudia. CTPR1a proteinako E22-K26 diadak katalizatutako *homo-s-cis* **1a** iminak **3a'** azometino iluroa emateko erreakzioaren profil energetikoa. Bien arteko konparaketa errazteko 2.5. Irudian erakutsitako profila, grisez, ere azaltzen da. Gainerako xehetasunak 2.5. Irudian adierazten dira.

2.6. Irudian bi erreakzio mekanismoen arteko alderaketa azaltzen da, bi profil energetikoak antzekoak direla baieztatuz. Beraz, honen bitartez, **3a** eta **3a'** azometino iluroen eraketa konpetitiboa azaldu daiteke.

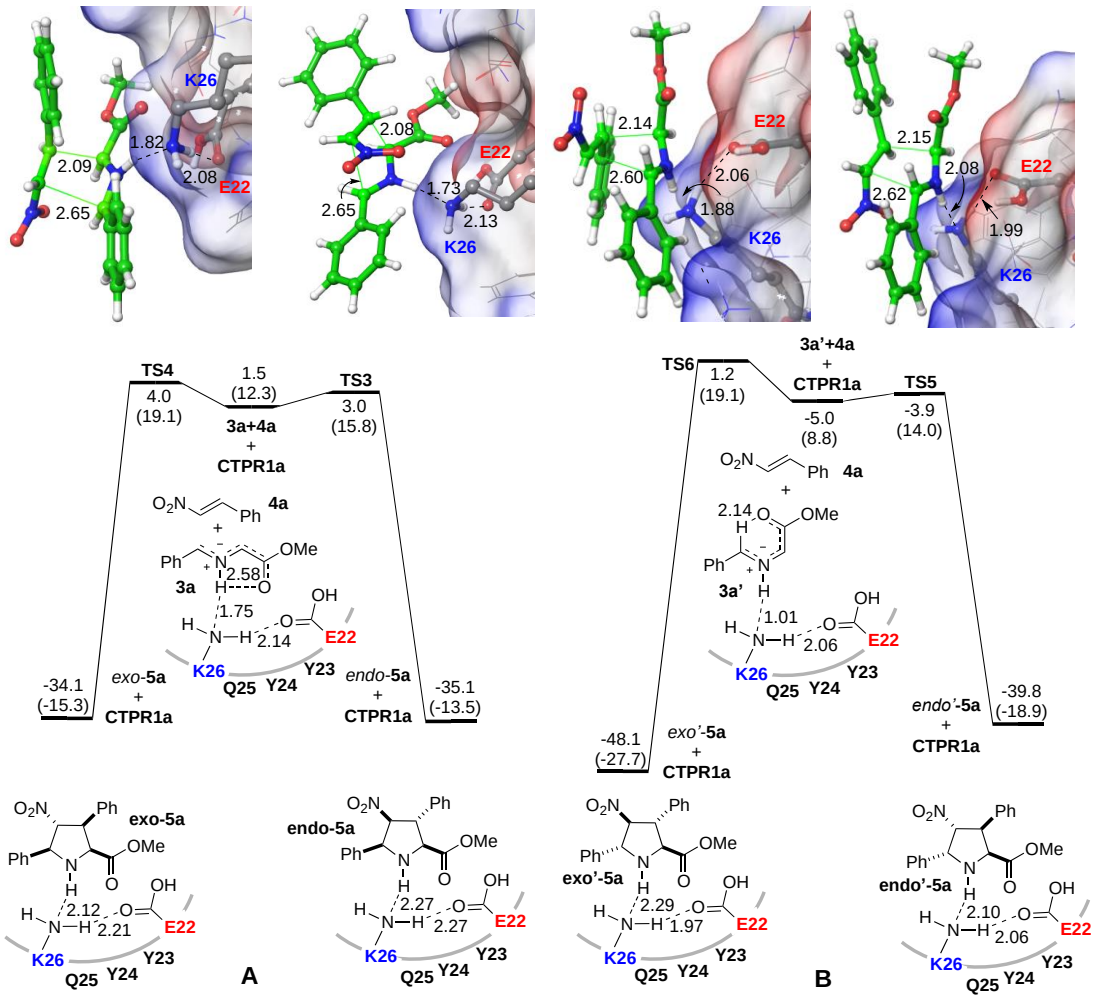
2.4.2 E22-K26 diada katalitikoak sustatutako (3+2) zikloadizio erreakzioaren azterketa konputazionala

2.3. Eskeman proposatu den moduan, **3a** eta **3a'** 1,3-dipoloak eratu ostean, bi espezie erreaktibo hauek *exo*-**5a** eta *endo*-**5a** edo *endo'*-**5a** eta *exo'*-**5a**, hurrenez hurren, eraldatu daitezke. Hala ere, zikloadizio erreakzio hauekin sartu baino lehen, **1a** iminaren *homo-s-trans* konformazioaren bitartez lortutako **2a** enola prototropia pseudoperizikliko baten bitartez **3a** azometino ilurora transformatu daitekeela aipatu beharra dago (2.1. Eskema). Gure analisi konputazionalaren arabera, **3a** eta CTPR1a-ren arteko konplexua **2a** enol egitura baino 3 kcal/mol gorago aurkitzen da Gibbs energia askeari so eginez.

Hortaz, **4a** (*E*)-nitroestirenoa, CTPRa proteinak eta **3a** zein **3a'** 1,3-dipoloek eratzen duten konplexuetara hurbiltzen da (3+2) zikloadizioa emateko (2.7. Irudia). Katalisiaren azken urrats hau **TS3-6** prozesu kontzertatu baina nahiko asinkronoen bidez gertatzen da, azometino iluroen eta nitroalkenoen beste hainbat (3+2) zikloadizioetan ematen diren [$\pi 4_s + \pi 2_s$] Woodward-Hoffmann topologiekin alderatzen badugu.[3,4]

exo-**5a** eta *endo*-**5a** produktuak ematen dituen mekanismoari bagagozkio, **4a** β-nitroestirenoak **3a**-CTPR1a konplexuarekin erreakzionatzen du **TS3** eta **TS4** egitura konpetitiboen bitartez *endo*-**5a** eta *exo*-**5a** produktuak emateko. *endo*-**5a** ematen duen **TS3**-ak, *exo*-**5a** eratzen duen **TS4**-ak baina aktibazio energia baxuagoa du, 3.5 eta 6.8 kcal/mol, hurrenez hurren. Bi trantsizio-egituretan ematen diren distantzien analisiak egitura hauen izaera kontzertatu baina nahiko asinkronoa baieztatzen du (**TS3**; 2.09/2.65 Å eta **TS4**; 2.08/2.65 Å).

Bestalde, **3a'**-CTPR1a konplexuaren eta **4a** dipolarofiloaren arteko erreakzioak, **TS5** eta **TS6** trantsizio-egoeren bitartez *exo'*-**5a** eta *endo'*-**5a** produktuak ematen ditu. Modu honetan, *endo'*-**5a** sortzen duen erreakzioa 5.2 kcal/mol datzan **TS5**-aren bitartez ematen da eta *exo'*-**5a** eratzeko erreakzioa, ordea, 10.3 kcal/mol-eko aktibazio energia duen **TS6**-aren bitartez. *endo*-**5a** eta exo **5a** konposatuak ematen dituzten prozesuen antzera, prozesu hauek kontzertatuak baina nahiko asinkronoak dira, berauetan ematen diren distantziak **TS5** eta **TS6**-etan ondorengoak izanik; 2.15/2.62 Å eta 2.14/2.50 Å, hurrenez hurren.



2.7. Irudia. E22/K26 diadak katalizatutako **3a** zein **3a'** azometino iluroen eta **4a** dipolarofiloaren erreakzioen bidezko, *endo/exo*-**5a** (**A**) eta *endo'/exo'*-**5a** (**B**) aduktuen eraketa prozesuen profil energetikoak. Profil energetikoetan azaltzen diren balioak, energia erlatiboei eta Gibbs energia askeei (parentesi artean) dagozkie eta **RCa** (2.5. Irudia) eta **RCb** (2.6. Irudia) hartzen dira erreferentzia puntutzat.

2.7. Irudian aurkeztutako lau mekanismoetan, CTPRa proteinaren eta dipolo edota zikloaduktuen arteko elkarrekintza erreakzio koordenadetan zehar mantentzen da, ligandoko NH taldearen eta K26-ren nitrogenoaren atomoaren bidez. Horrez gain, K26-an dagoen amino taldeak eta E22-ko karboxilo talde neutroak diada katalitikoa egonkortzen duen hidrogeno lotura bat eratzen dute. Hala ere, CTPR1a eta produktuen artean ematen den elkarrekintza artekari zein trantsizio-egituretan ematen dena baino ahulagoa da. Hortaz, zikloaduktuak eratu ondoren, hauek sistematik errazago askatu daitezke E22/K26 diada katalitikoaren protonazio egoera berreskuratuz eta honela,

prozesu katalitikoa berriro hasiz. Bestalde, gure analisi konputazionalean prozesu desberdinentzat lortutako aktibazio energien arabera, E22-K26 diada katalitikoak **1a + 2a** → **5a** prozesuarekin erlazionatutako lau zikloaduktuak eratzeko gaitasuna izango luke.

Ondorioz, QM/MM bidez burututako azterketa konputazionalak emaitza esperimentalak balioesten ditu, CTPR1a proteinan aurkitzen diren diada katalitikoaren zein monaden aktibitatea baieztatuz. Gainera, CTPR1a proteinan eginiko eta zuzenki zein ez-zuzenki E22-K26 diadaren egonkortasuna baldintzatzen duten mutazioek *endo'*-**5a** eta *exo'*-**5a** aduktuen sintesian efektu zuzena izango dute. Aldiz, hurrengo sekzioan azaltzen den moduan, gune azido edota basiko bakanak *endo*-**5a** eta *exo*-**5a** zikloaduktuen sintesia soilik katalizatu dezakete.

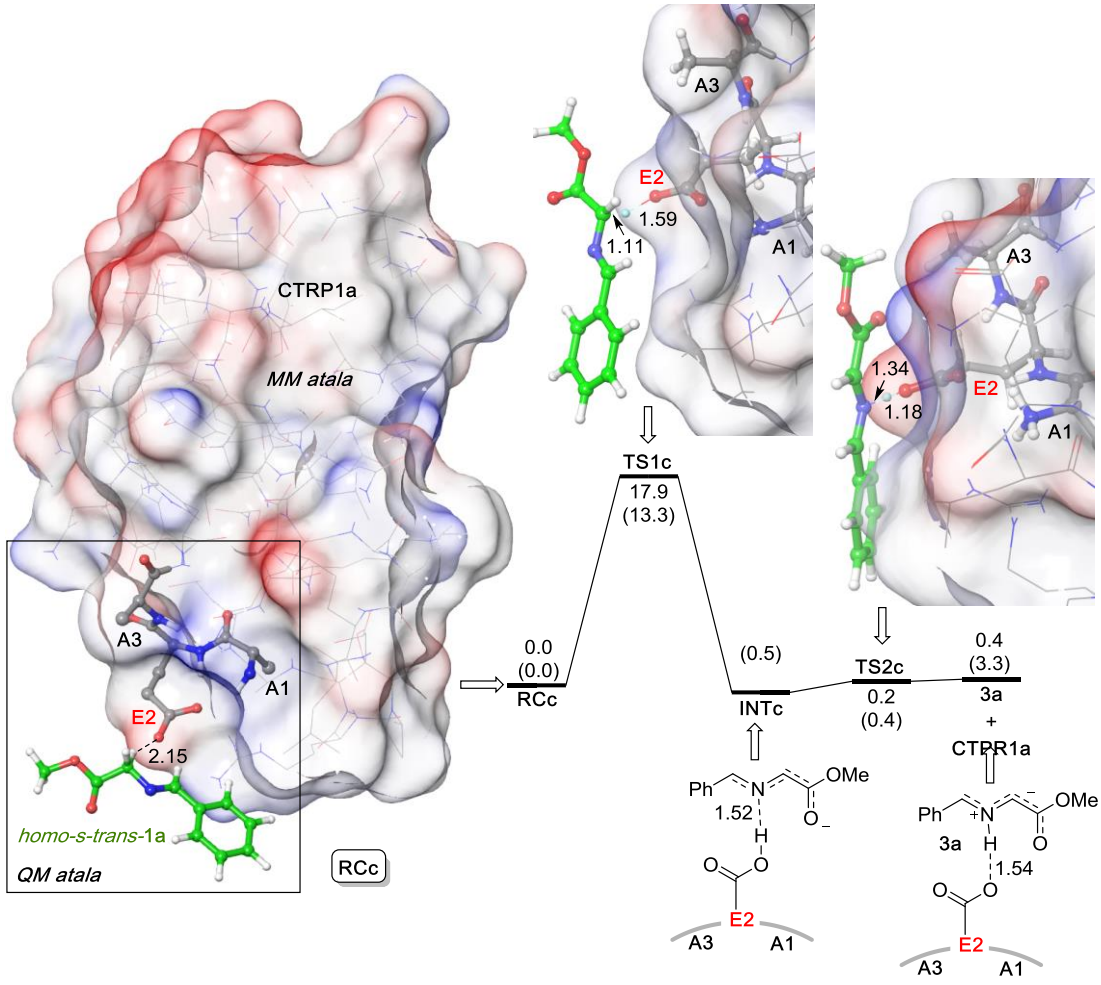
2.4.3 Monada azido edota basikoek katalizatutako 3a azometino iluroen eraketa erreakzioaren analisi mekanistikoa

CTPR1a proteinan zazpi gune azido (E2, D16, D18, E19, E29 eta D31) eta bi gune basiko (K13 and R33) aurkitu daitezke eta hauek, monada eran lan eginez Huisgenen erreakzioa katalizatzeko gai izan daitezke. Hortaz, gune azido bat, E2, eta basiko bat, K13, aukeratuak izan ziren hauen potentzial katalitikoaren ebaluaziorako. Hasiera batean hauek **3a** zein **3a'** azometino iluroen eraketa proposatu genuen. Burutako analisi konputazionalak, aldiz, mekanismo hauen bitartez **3a** artekariak soilik sortu daitezkeela erakusten du. E2 eta K13 aminoazidoen erreaktibitatea aztertu zenean, bi monada hauek ez ziren gai **1a** iminaren *homo-s-cis* konformazioarekin elkarrekintza egonkor bat sortzeko. *homo-s-cis* konformazioaren eta gune azido zein basikoen arteko elkarrekintza optimizatu zenean, orekak beti *homo-s-trans* konformaziora eramaten zuen egitura, modu honetan, soilik **3a** 1,3-dipoloa eratzeko behar zen aitzindaria lortzen zelarik.

E2 gune aktiboaren modelizaziorako A1, E2 eta A3 aminoazidoak eta erreaktiboa sartu ziren QM atalean, gainontzeko proteina MM bidez aztertzen zelarik. Lehenik, *homo-s-trans* **1a** CTPR1a proteinara hurbiltzen zen **RCc** konplexua eratuz, non E2-aren karboxilo taldeak eta C_α-H hidrogeno batek hidrogeno lotura bat eratzen duten (2.8. Irudia). Horrela, hasierako orientazio honek azido glutamikoaren bidezko substratuko

C_α-H protoietako baten abstrakzioa ahalbidetzen du **TS1c** trantsizio-egituraren bitartez, 13kcal/mol inguruko aktibazio energia duen prozesu batean. 2.5. Irudian azaltzen zen diadak katalizatutako prozesuarekin alderatuz, prozesu honek trantsizio-egitura goiztiar batean du oinarria non transferitutako protoia C_α-tik (1.11 Å) gertuago dagoen E2-aren oxigenotik baino (1.59 Å). Honela, **INTc** eratzen da eta hau, 1,3-dipoloaren nitrogenoaren artean eta protonatutako E2-aren artean ematen den elkarrekintza bitartez egonkortzen da.

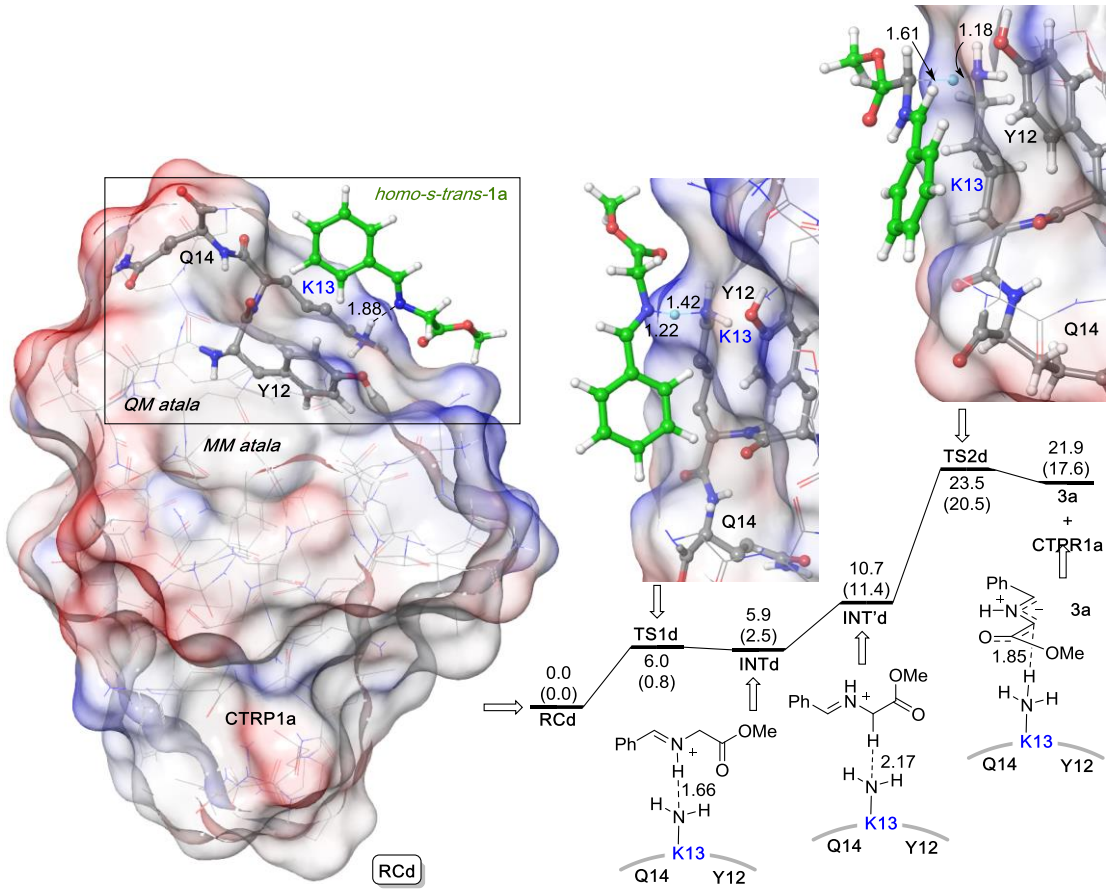
Hurrengo urratsean, sp^2 hibridazioa duen enolatoaren nitrogenoaren E2 bidezko protonazio prozesua ematen da. TS2c-aren bitartez ematen den etapa honetan ez da aktibazio energiarik behar eta **3a** azometino iluroaren sintesia du emaitzatzat. Dipolo honen eraketa, produktuko E2 aminoazidoaren eta NH taldearen arteko elkarrekintzaz egonkortzen den arren (1.54 Å), 3.3 kcal/mol-eko prozesu endergonikoan oinarritzen da. Hortaz, prozesu honen mekanismoa aurretiaz 2.5. Irudian deskribatutako prozesuaren antzekoa da, ezberdintasun nagusia E2 gune katalitikoak duen emaile-hartzaile duala izanik (2.8. Irudia).



2.8. Irudia. CTPR1a proteinan aurkitzen den E2 aminoazidoak katalizatutako *homo-s-trans* **1a** iminaren transformazioa **3a** azometino iluroa eratzeko. QM atala ligandoaz (berdez adierazita) eta A1-E2-A3 hondarrez (grisez adierazita) osatuta dago eta "*ball-and-stick*" moduan irudikatuta dago. Gainontzeko xehetasunak 2.5. Irudian adierazten dira.

Gune basikoen monada erako katalisiaren ebaluaketa burutzeko, antzeko analisi konputazionala gauzatu zen kasu honetan K13 aminoazidoa aukeratuz. Hortaz, QM atala definitzeko, ligandoa eta Y12, K13 zein Q14 aminoazidoak aukeratu ziren, gainontzeko proteina MM bidez deskribatzen zelarik. Erreakzio mekanismo honetan *homo-s-trans* konformazioa duen **1a** imina CTPR1a proteinara gerturatzen da, lisinaren amonio taldearen eta iminaren sp^2 hibridazioa duen nitrogenoaren artean elkarrekintza bat sortuz eta erreakzioaren hasiera puntua den **RCd** konplexua eratuz. Lehenik, aipatutako nitrogeno honen protonazioa ematen da, 0.8 kcal/mol-eko aktibazio energia (Gibbs energia askeari dagokionez) duen **TS1d**-aren bitartez. Osatutako **INTd** honetan, ligandoaren NH-aren eta K13-aren artean elkarrekintza bat sortzen da, hortaz,

ezinbestekoa da honen orientazioa aldatzea **INT'd** sortuz. Modu honetan, N-posizioan protonatutako iminio egiturako C_α posizioan aurkitzen den hidrogeno batekin elkarrekintza bat eratzen du, hurrengo urratsa ahalbidetzen duen konformazioa emanez.



2.9. Irudia. CTPR1a proteinan aurkitzen den E2 aminoazidoak katalizatutako *homo-s-trans* **1a** iminaren transformazioa **3a** azometino iluroa eratzeko. QM atala ligandoaz (berdez adierazita) eta Y12-K13-Q14 hondarretaz (grisez adierazita) osatuta dago eta "ball-and-stick" moduan irudikatuta dago. Gainontzeko xehetasunak 2.5. Irudian adierazten dira.

Beraz, mekanismo honen bigarren urratsean, lisina katalitikoaren amino talde neutroak, iminio katioiaren C_α posizioan aurkitzen den hidrogeno baten abstrakzioa ematen du, 17.6 kcal/mol-eko prozesu endergoniko baten bitartez. Bigarren trantsizio-egoera hau, **TS2d**, prozesu guztiaren etapa mugatzailea da, Gibbs energiari dagokionez 9.1 kcal/mol-eko aktibazio energia baitu. Ondorioz, monada basikoek burutako mekanismo honetan, K13-ak, **1a** imina protonatzen du eta ondoren, C_α posizioko hidrogenoaren abstrakzioa burutzen du.

Hortaz, CTPR proteinetan banaka aurkitzen diren gune azido zein basikoek beraien kabuz **3a** 1,3-dipoloaren eraketa katalizatu dezakete eta modu honetan, (3+2) zikloadizioaren bitartez *endo* zein *exo* zikloaduktuak emateko gai dira 2.7. Irudian agertzen den mekanismoaren antzeko prozesu baten bitartez. Analisi konputazionalaren bidez lortutako emaitza hauek, *exo*' eta *endo*' zikloaduktuen katalisia diada katalitikoak soilik eman dezakeela konfirmatzen dute eta mutazioetan gune azido edota basikoak ezabatzean ikusten den *Huisgenasa* aktibitatearen murrizketa ere azaldu dezakete (2.2. Taula).

2.5 METIL N-BENTZILIDENO GLIZINATOAREN ETA CTPR PROTEINEN ARTEKO ELKARREKINTZAREN EMN BIDEZKO ANALISIA

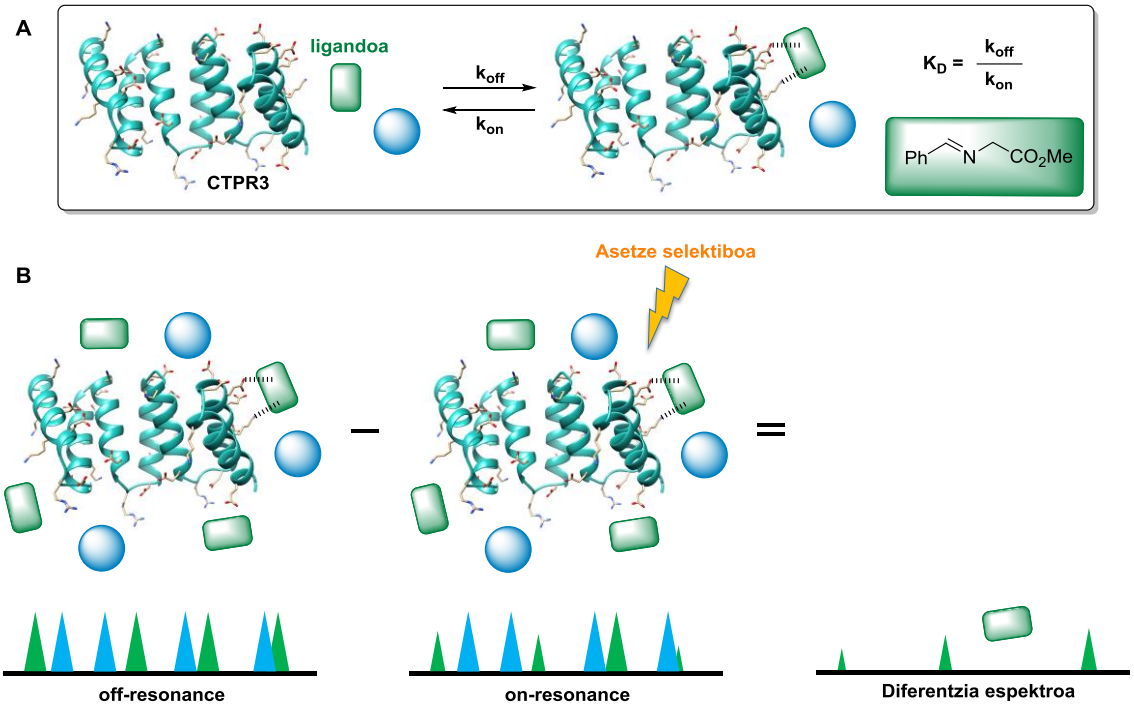
Azken urteotan, asetasun transferentzia diferentzia EMN esperimentuak (STD-NMR) eta 2D ^{1}H-^{15}N korrelazio esperimentuei esker proteina eta ligandoen arteko elkarrekintzen analisia erraztu da.[22 - 24] Hortaz, CTPR eta **1a** iminaren arteko elkarrekintzak ulertu nahian, STD-EMN eta 2D ^{1}H-^{15}N korrelazio esperimentuak burutu ziren Jesús Jiménez-Barbero (CIC bioGUNE) profesorearen ikerketa taldearekin lankidetza burutzu.

Lehenik eta behin, EMN analisian erabili beharreko balditzen optimizazioa burutu zen, tenperatura eta disolbatzailea esaterako. Gogoratu beharra dago, **1a** aitzindariaren 1,3-dipoloaren hidrolisia ekiditeko helburuarekin (3+2) zikloadizio guztiak THF-a erabiliz burutu zirela. Hala ere, EMN esperimentuen optimizazioak fosfatodun buffer akuoso baten erabilera iradoki zuen, proteinaren estruktura tertziarioa egonkortzeko eta EMN analisia errazteko.

Hasiera batean, **1a** iminaren eta hiru errepikapeneko proteinen, CTPR3a eta CTPR3, arteko elkarrekintza STD-EMN esperimentuen bitartez aztertu zen. Esperimentu

22. Williams, M.; Davites, T., *Protein-Ligand Interactions: Methods and Applications*. 2nd ed.; Humana Press.: New York, 2013.
23. Meyer, B.; Peters, T., *Angew. Chem., Int. Ed.* **2003**, *42*, 864-890.
24. Fernández-Alonso, M. C.; Berbis, M. A.; Canales, A.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J., New Applications of High-resolution NMR in Drug Discovery and Development. In *New Applications of NMR in Drug Discovery and development*, Garrido, L.; Beckmann, N., Eds. Royal Society of Chemistry: 2013; pp 7-42.

hauek, ligandoak proteinara era ahul batean loturik daudenean, lotuta eta aske dauden ligandoen artean dagoen oreka egoera dute oinarritzat (2.10A Irudia). Modu honetan, STD-NMR esperimentuek bi analisi nagusi dituzte; alde batetik proteina egoera asegabe batean burutzen den *off resonance* deituriko ^{1}H NMR espektroa eta bestetik, proteina selektiboki asetzen den esperimentua, *on-resonance* deiturikoa, zeinetan erradiazio selektibo baten bitartez proteinaren erresonantzia soilik biltzen den. Hortaz, asetze prozesua jaso duen ligandoaren seinaleak soilik ematen dituen desberdintasun espektro bat lortu daiteke (2.10B Irudia).[25] Gure kasuan, aztertutako proteinak CTPR3 edota CTPR3a izango lirateke, ligandoa ordea, **1a** imina.

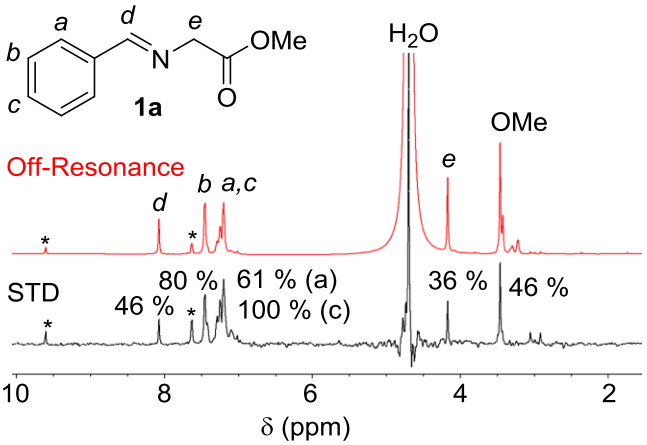


2.10. Irudia. A) CTPR3 proteinaren aske eta loturiko **1a** ligandoaren arteko oreka egoera. **B**) STD-NMR esperimentuaren deskribapena, non *on-resonance* eta *off-resonance* espektroak lortuz diferentzia espektroa ateratzen den.

STD-NMR esperimentu hauetan 1.5:50 ratio molarra erabili zen CTPR3a/CTPR3:**1a** nahastearentzat. Analisi hauek, STD erantzun argia erakutsi zuten CTPR3a/CTPR3 proteinan, bi kasuetan antzeko espektroak lortuz. 2.11. Irudian erakusten den moduan, STD seinale horiek lortu ziren bentzilidenoaren a-d protoientzat, zein glizinato egiturarentzat. Ondorioz, emaitza hauek CTPR3 proteinen eta **1a**

25. Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J., *J. Chem. Educ.* **2011**, *88*, 990-994.

ligandoaren arteko gertutasuna erakusten dute. Bestalde, esperimentu hauetan zehar, inguruneko ur kantitate txikiak **1a** iminaren bentzaldehidorako hidrolisia bultzatzen zuen. Erreakzio sekundario honen efektua tenperatura 278 K-era murriztuz jaitsi bazitekeen ere, aldehidoarentzat lortutako STD seinaleak ezin izan ziren guztiz desagerrarazi.



2.11. Irudia. 1a-ren 1.5 mM eta CTPR3a-ren 30µM dituen nahastearen STD-NMR espektroa (*on-resonance* frekuentzia δ 0.53 ppm) eta ^{1}H-NMR erreferentzia espektroa (*off-resonance* frekuentzia δ -25 ppm) 278 K-etan eta D_2O-an neurtuta (600 MHz). Izartxo batez adierazitatko seinaleak bentzaldehidoaren seinaleari dagozkio. STD erantzun erlatiboa **1a**-ren ^{1}H seinaleen portzentai moduan adierazten da.

1a iminaren eta CTPR proteinen arteko elkarrekintzak aztertzeko bigarren analisia ^{1}H-^{15}N TROSY EMN esperimentuak erabiliz burutu zen. Hasieran batean, azterketa hau CTPR3 eta CTPR3a proteinekin egitea zen ideia nagusia, baina CTPR3a proteinekin esperimentuak burutu zirenean seinale zabalak lortzen ziren hauek jasaten zuten aldaketa konformazional bat zela eta. 2.4. Irudian aipatutako moduan, B-faktoreek CTPR3a serieek portaera desberdin bat zutela erakusten zuten. Hortaz, serie honen ^{1}H-^{15}N TROSY NMR bidezko azterketa deuseztatu zen. Hala eta guztiz ere, CTPR3 proteinan TPR motiboaren sekuentzia antzekoa izatea eta proteina honetan ^{1}H eta ^{15}N seinaleak aurretik esleituak zeudela kontuan hartu behar dira.[26] Ondorioz, ^{1}H-^{15}N TROSY

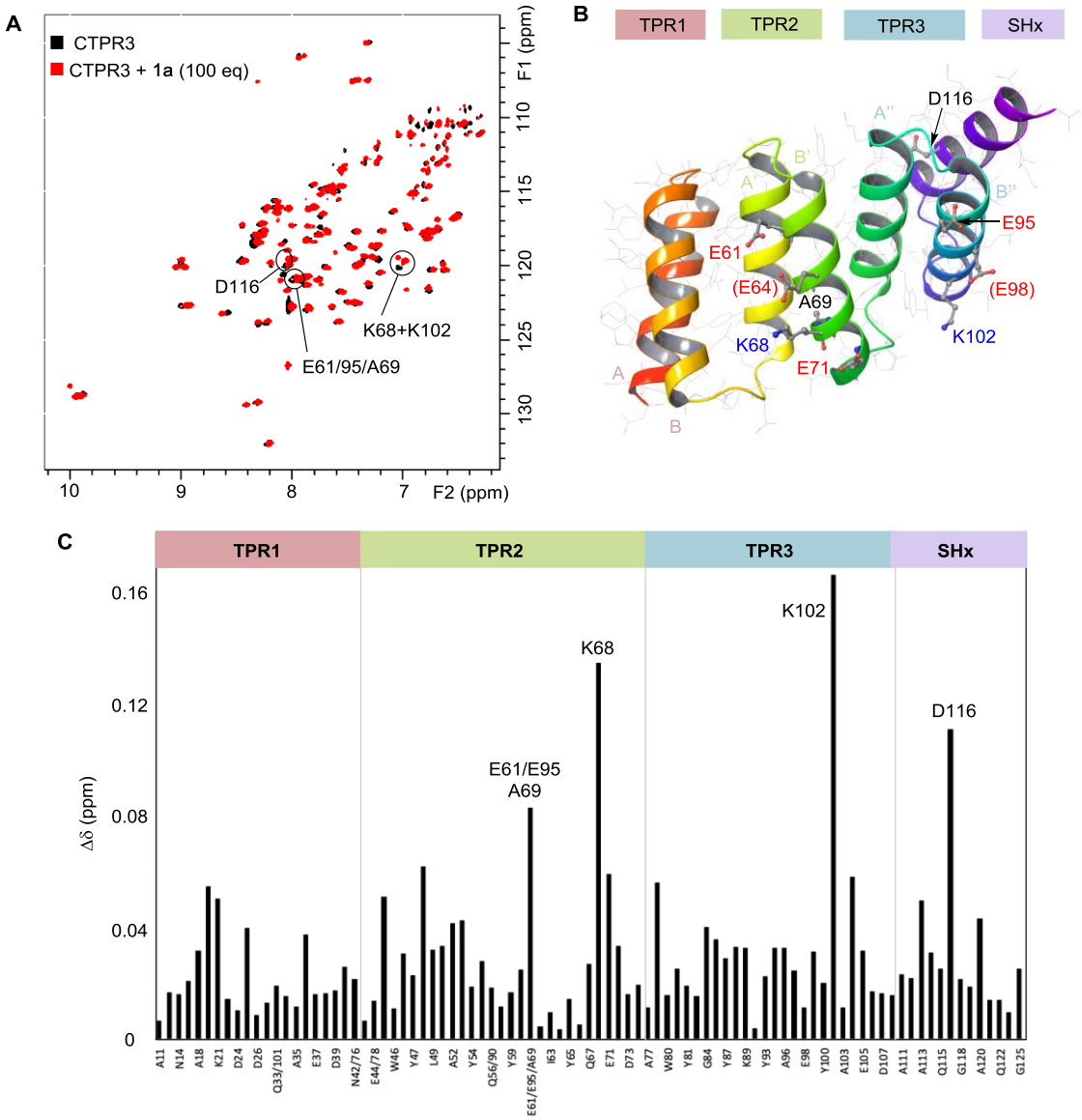
26. Main, E. R. G.; Stott, K.; Jackson, S. E.; Regan, L., *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5721-5726.

EMN esperimentuak aurrez deskribatuta eta aztertzeko erraztasun gehiago ematen zituen CTPR3 proteina erabiliz burutu ziren (2.1. Irudia).

1a ligandoak ^{15}N-z markaturiko CTPR3 proteinaren egituran eragindako desplazamendu kimikoaren perturbazioa (CSP) ^{1}H-^{15}N TROSY bidez aztertu zen (2.11. Irudia). STD-EMN esperimentuetan gertatutakoaren antzera, **1a** iminaren bentzaldehido zein glizina esterrerako deskonposizioa ere gertatzen da. Arrazoi hau kontuan izanik, glizina metil esterra eta bentzaldehidoa bakarka ere aztertu ziren kontrol esperimentu moduan (2.13. Irudia).

Burututako lehenengo esperimentuek, **1a** iminak eragindako zenbait CTPR3-ko ^{1}H-^{15}N seinaleetan hainbat ppm-ko desplazamendua erakutsi zuten, modu honetan perturbazio nagusienak jasotzen zituzten aminoazidoak zehaztu zitezkeelarik (2.12A Irudia). Aurrez eginiko STD-NMR analisiak proteinen eta ligandoaren arteko elkarrekintza konfirmatu zutenez, ^{1}H-^{15}N TROSY espektroan behatutako desplazamendu kimikoa gehienbat **1a** ligandoari leporatu zitzaion. Hala ere, desplazamendu kimiko honetan bentzaldehidoaren kontribuzioa ere ezin da baztertu.

Atal honetako emaitzen eztabaida hasi aurretik, EMN atalean adierazitako aminoazidoen nomenklaturak atal esperimentalean jakinarazitako eta klonazio prozesutik eratorritako GAMDPGNS-N sekuentzia gehitzen duela adierazi behar da. **1a** imina gehitu ondorengo seinaleen azterketak bi pikoren perturbazioa eragiten du, batez ere, positiboki kargatutako K68 eta K102-renak hain zuzen ere (aurreko nomenklaturaren arabera bigarren zein hirugarren errepikapenetako K26 posizioak). Emaitza hauek, posizio hauek mutatzerakoan lortutako etekin kimikoen eta TON/TOF balioen murrizketa azaltzen dute. Gainera, TPR2 etaTPR3 errepikapenetan kokaturiko B' zein B'' helizeetan dauden gune basiko hauek katalizatzaile moduan lan egiteko duten gaitasuna erakusten dute emaitzek. Honetaz gain, 2.12C Irudian ikusten den bezala K68 eta K102 hondarrak E64 eta E98-etatik gertu daudela kontuan izan behar da, hauek baitira diada katalitikoa osatzeko gai izango ziren aminoazidoak. Hala ere, beste gune basikoak lisina hauen mugimenduarekin alderatzen baditugu, esperimentu hauen bitartez elkarrekintza adierazgarririk behatzen ez dela esan genezake (2.12C Irudia).

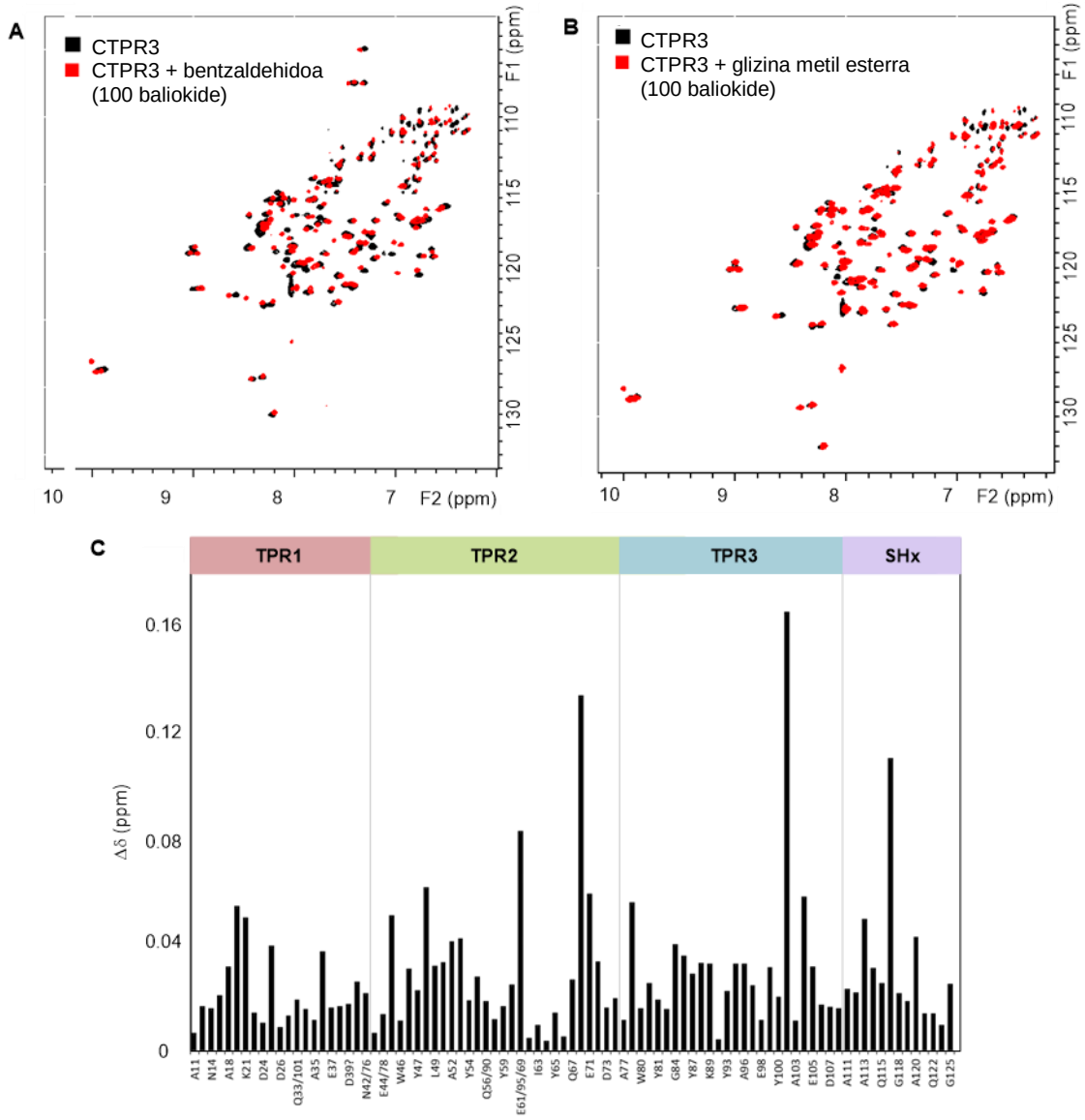


2.12. Irudia. A) apo CTPR3 proteinaren (beltzez) eta **1a** ligandoaren 100 baliokide gehitu ostean (gorriz) lortutako ^{1}H-^{15}N TROSY espektroen gainezarpen espektroa. **B**) CTPR3 proteinaren X-izpien difrakzio bidez lortutako estruktura non gehien eraldatutako posizioak azaltzen diren (pdb kodea: 1NA0, ikus 2.1A Irudia). K68 eta K102 posizioetatik gertu dauden, baina ia desplazamendurik jasaten ez duten E64 zein E98-a parentesi artean adierazten dira. Gune azido eta basikoak, gorriz eta urdinez, hurrenez hurren, nabarmenduta agertzen dira. **C**) **1a** (100 baliokide) gehitu ostean CTPR3 proteinako aminoazidoetan lorturiko bataz-besteko ^{1}H eta ^{15}N seinaleen desplazamendu kimikoaren perturbazioaren grafikoa ($\Delta\delta=[(\Delta\delta^2_H+0.14\cdot\Delta\delta^2_N)/2]½$, ppm-tan).[27]

27. Williamson, M. P., *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1-16.

Aipatutako aminoazidoez gain, badira karga negatiboa duten hainbat aminoazidok erakusten dituzten desplazamenduak, modu honetan, hauen monada izaera iradokiz. Hauen artean, TPR2 errepikapeneko B' helizean aurkitzen den E61 eta TPR3 errepikapeneko B'' helizean dagoen eta aurrekoaren baliokidea den E95 dira nagusienak. Hala eta guztiz ere, piko hauen desplazamenduak A69-aren desplazamendutik banatzea ez zen posible izan. Era berean, D116, Y19, N48, E71, A79 eta L104 aminoazidoek 2σ baino handiagoak ziren CSP balioak erakutsi zituzten (2.12C Irudia). Aitzinean aipatu bezala, glizina metil esterra eta bentzaldehidoa erabili ziren kontrol esperimentu moduan. 2.13. Irudian aurkeztutako analisi hauetan ikusten den bezala, bentzaldehidoak, **1a** iminaren antzeko espektroak ematen zituen eta glizina metil esterrak kasuan ordea ez da seinale esanguratsurik lortzen.

EMN bidez lorturiko emaitza hauen arabera, CTPR3 proteinetan aurkitzen diren B' zein B'' helizeetan aurkitzen diren hainbat gune azido eta basikoek **1a** imina azometino iluroetan transformatzeko gaitasuna izango lukete. Ziurrenik, analisi hauetan lortutako emaitzak CTPR1 proteinan aurkitzen diren aminoazidoetara ere aplikatu daitezke. Ondorioz, STD zein ^{1}H-^{15}N TROSY EMN esperimentuen bidez lorturiko emaitza hauek, aurrez azaldutako emaitza esperimentalak zein konputazionalak balioztatzen laguntzen dute, CTPR proteinen *Huisgenasa* aktibitatea erakutsiz. Horrez gain, lan hau *Huisgenasa* aktibitate diastereo- zein enantioselektiboa burutu dezaketen proteinen diseinua burutzeko hasiera puntu bezala hartzen da, honela, entzimek burutu ditzaketen erreakzio kopurua handituz.



2.13. Irudia. A) apo CTPR3 proteinaren (beltzez) eta bentzaldehidoaren 100 baliokide gehitu ostean (gorriz) lortutako ^{1}H-^{15}N TROSY espektroen gainezarpen espektroa. **B**) apo CTPR3 proteinaren (beltzez) eta glizina metil esterraren 100 baliokide gehitu ostean (gorriz) lortutako ^{1}H-^{15}N TROSY espektroen gainezarpen espektroa. **C**) Bentzaldehidoa (100 baliokide) gehitu ostean CTPR3 proteinako aminoazidoetan lorturiko bataz-besteko ^{1}H eta ^{15}N seinaleen desplazamendu kimikoaren perturbazioaren grafikoa ($\Delta\delta=[(\Delta\delta^2_H+0.14\cdot\Delta\delta^2_N)/2]½$, ppm-tan).

2.6 ONDORIOAK

Kapitulu honetan azaldu eta eztabaidatutako azterketen ondoren, hurrengo ondorioak laburtu daitezke:

1. Erabilitako kontsentsu-tretratrikopetido errepikapen (CTPR) proteinek, bat zein hiru errepikapenetako CTPR eta CTPRa multzoek, N-bentzilideno glizinatoaren eta (*E*)-β-nitroestirenoaren arteko (3+2) zikloadizio erreakzioa katalizatzeko eta beraz, *Huisgenasa* aktibitatea burutzeko gaitasuna erakutsi dute.
2. Aipatutako erreakzioek lau zikloaduktu errazemiko desberdin sortu ditzakete antzeko proportzioan. Katalisiaren diastereoselektibitatea CTPR aldaera desberdinen malgutasunarekin jokatuz aldatzea lortu da. Hortaz, CTPRa proteina malguagoen erabilerak lau zikloaduktu desberdinen sintesia ahalbidetzen du.
3. Gune azido zein basiko desberdinetan buruturiko mutazioek, hauek (3+2) zikloadizioen katalisian duten garrantzia konfirmatzen lagundu dute. Sekuentzietan burututako mutazio guztien bitartez *endo*' eta *exo'* produktuen sintesiaren inhibizioa ematen zen, *endo* eta *exo* prolinak soilik lortzen zirelarik.
4. QM/MM analisi konputazionalaren bitartez, esperimentalki lorturiko emaitzen arrazionalizazioa lortu da. Alde batetik, E22-K26 diadak iminaren *homo-s-cis* zein *homo-s-tran*s konformazioen egonkortze eta erreaktibitatearen bitartez (3+2) lau zikloaduktuen sintesia ahalbidetzen du. Bestetik, gune azido zein basiko bakunak, monadak, **3a** azometino iluroa eratzeko gai dira eta beraz, *exo*-**5a** eta *endo*-**5a**-ren katalisia eman dezakete soilik.
5. EMN esperimentuek, ligandoak proteinetan aurkitzen diren aminoazido desberdinekin dituen elkarrekintzen azterketaren bitartez, CTPR proteinen *Huisgenasa* aktibitatea balioesten dut