



Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

SCHOOL OF CHEMISTRY

Dr Cristina Trujillo & Prof Stephen Connon

THEORETICAL DEVELOPMENT OF NEW CLASS PHASE TRANSFER CATALYSTS

IÑIGO IRIBARREN AGUIRRE

SEPTEMBER 17, 2024

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
PHD IN CHEMISTRY

Declaration

I hereby declare that this Thesis is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at <http://www.tcd.ie/calendar>.

I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at <http://tcd-ie.libguides.com/plagiarism/ready-steady-write>.

Signed:  _____

Date: 20-12-2023

Abstract

This thesis stems from the imperative need to integrate computational chemistry into the process of designing asymmetric organocatalysts. Traditionally, computational chemistry has been employed as a retrospective tool to analyse results rather than in a prospective manner for predicting results and leading the rational designing process. With this goal in mind, the thesis aims to develop computational strategies that leverage the predictive potential of computational chemistry, thereby initiating a paradigm shift in catalyst design.

Organised into five chapters, the thesis tackles various aspects of this approach, addressing and resolving key challenges in catalyst design.

The first chapter analysed the most popular conformational analysis software packages, identifying their strengths and weaknesses. The efficiency of each package in different facets of the conformational analysis process is evaluated using five established criteria. A comprehensive list of pros and cons is presented to facilitate the selection of the most suitable software.

In the second chapter, the nature of ion-pair interactions is reevaluated. The study examines different models and systems typically grouped under the same category. Three intermediate categories between the pure ion-pair and the hydrogen bond were identified and analysed; namely long-range ion-pair, hydrogen-bond-assisted ion-pair and charge-assisted hydrogen bond.

Chapter three investigates the different interaction positions of a cinchona-based catalyst with different anions, providing a binding mode for these complexes. Additionally, three different cyanation reactions are studied, modifying the catalysts to improve the selectivity of the process.

Chapter four integrates all previously studied methodologies to examine the asymmetric production of protected amino acids. Collaborating with experimental colleagues, the chapter highlights the importance of rational design in improving catalyst selectivity. This is achieved by strategically introducing a second anchorage point and validating computational analysis conclusions through experimental verification.

The final chapter of this thesis introduces a new automatic workflow designed for the accurate description of the transition states involved in a reaction with flexible organic molecules. This innovative workflow seeks to study the reaction steps in an unbiased and more detailed manner, aligning closely with the experimental results compared to manually generated transition states.

This research work combined the design of computational strategies and the re-evaluation of basic chemistry concepts to improve the rational design of asymmetric catalysts. It demonstrates the predictive power of computational studies and their ability to guide the design process effectively.

Acknowledgements

A Spanish proverb says that *“it is a trait of the high-born to be grateful”*. I might not be high-born but my parents made a huge effort to educate me and they would be really disappointed if I did not thank everyone that made this thesis possible.

First and foremost, thanks to Dr. Cristina Trujillo, 'la Cristi', the best supervisor I could have asked for. Not only for all the things you taught me during these (very intense) years but also for all the advice and support and for being much more than just a mentor. Keep in mind that this is not just the beginning and that you will not get rid of me that easily, I cannot wait to see how many crazy collaboration ideas we have in the future.

Many thanks as well to Prof. Ibon Alkorta, not only for teaching me so much during my master thesis but also for sending me the email that started everything with a simple 'por si quieres ir a pasar una temporada a Dublin' and thought that Cristina and I would get along quite well. How right you were!

Thanks as well to all the teachers and lecturers who made learning enjoyable for me. Especially thanks to Dr Majiare Fernandez, who made me fall in love with chemistry in high school, to Prof. Manuel Yañez for being the best theoretical chemistry professor (and who is my scientific grandad now), to Dr Julia Contreras for the time I spent with her during my master, to Dr Merche Montero-Campillo for all the things she taught me during the master, to Prof Isabel Rozas for being a scientific reference and giving me so many nice opportunities and, of course, to Dr Goar Sánchez for all the collaborations, knowledge, origami and diverse freak-stuff that we have shared. This thesis is also yours.

I would also like to say thank you to someone who will never read this message, the person who has had the biggest influence on science over the last century and who has made scientific knowledge accessible to everyone, Alexandra Elbakyan. Without her, this and many more theses worldwide would have not been possible.

I must also thank all the people who have offered me plenty of 'extra-curricular' opportunities during my PhD. Dr John O'Donogue, for all the outreach activities, posters, banners, and different commissions; Dr Noelle Scully, for keeping me in mind whenever a new computational demonstrator opportunity appeared; and Dr Niamh Mc Goldrick, for all her help with bureaucracy at the beginning of the PhD and all the activities we have done together. These activities have meant wonderful mental and economic relief from the stressful, underpaid, PhD life during these years.

I would like to express my deepest gratitude to my friends because I would definitely not be here without them. Being the lucky person I am, this list is quite long. I would also like to take the opportunity to apologize for the following grammatical errors because, sadly, I do not speak any of the following languages; but I want to take the opportunity to thank each one of you in your mother tongue.

First of all thanks to all the current and past members of the TRG happy family for making good days brighter and bad days more bearable. The Dublin division, Nika, Eric, Diana and Rica; the Manchester division, Tim and James and the adopted member, Max. Спасибо Нике! мой товарищ, моя «бобер курва», мой любимый звездный читатель. У меня никогда не будет времени поблагодарить вас за все хорошие моменты, которые вы подарили мне в

офисе. Diana, mulțumesc pentru toate „întrebarile protesti”, continuă să vină! Tim, auch wenn du erst seit ein paar Monaten hier bist, bist du für mich eine sehr wichtige Stütze geworden. Danke, dass du genauso seltsam bist wie ich. James, thank you for keeping the aesthetics of the future figures safe, you are the beacon of figure aesthetics now. Max, mon petit frère scientifique, merci pour tous les moments passés ensemble à Dublin et Madrid et pour tous les futurs.

Thanks as well to my experimental collaborators and other members of the Prof Cannon group; Simon, Lee, Lorenzo, Pierre and Ian. And to all the other groups and friends in the office, and in other groups and offices, for all the celebrations and pints together. Especially Susie, Ellen, Aine, Laura and Ines. It has been a very intense journey together.

Ahora a un nivel más personal, gracias a "Spanish Crew" de Trinity; Natalia, Marc y Adrian, por tantos 'breaks para un coffee' y por aguantar todas mis quejas semanales, lo voy a echar mucho de menos (y a vosotros más). Gracias también a todos los demás expatriados de Dublín, en especial Laura y Sara, por hacer que esta ciudad fuera más como estar en casa estos años.

Por supuesto, infinitas gracias a los miembros de 'el Casoplón del coletas'; por referente científico y personal. A Rubén, no solo por ser mi enciclopedia particular y el primer lugar al que recurrir, sino por ser mi más mi inestimable compañero de Arams. A mi 'gemelo' Darío, porque no podría imaginar un doctorado sin ciclos de cine insólito y porque ha escuchado interminables audios de mi parte sin quejarse y ha estado siempre dispuesto a todo. Y a Eva, que ha sido quien realmente me ha enseñado a valorar la cultura popular de calidad (mediaset) y que es, además, de las mejores científicas que conozco.

Gracias a ti Mario, por todo lo que has hecho, y sigues haciendo, por mi estos años. Me alegra muchísimo ver cómo hemos madurado y crecido (juntos y separados) y me enorgullece muchísimo poder seguir teniéndote en mi vida. Como dice una canción que siempre me recuerda a ti 'momentu zailenetan zein une goxoetan hor zaude zu nire alboan' y espero que eso nunca cambie.

Gracias también a todas mis amigas, que me habéis hecho ser como soy hoy. Las que habéis crecido conmigo; Rubén, Asiáin, Pablo, Raquel, Marta, Silvia, Amaia, Gara, Laura, Eneko, Andrea, Bea, Fátima y todas las demás. Hacéis que volver a casa sea como viajar atrás en el tiempo y no llevase ya más de una década fuera. A las que llegasteis después pero os hicisteis indispensables Rosa, Marta (Artacho y Torrecilla), Elena, Naia, Isabel e Irene. Nunca pensé que una residencia fuese a darme a gente tan maravillosa y no puedo más que dar las gracias por todo lo que habéis sido estos años. Eskerrik asko zuei ere, Inaki, Ander eta Enara, ikaragarritzko lagunak zaretelako eta landa-etxeen antolatzaile onenak. Y, gracias a todas aquellas a las que, seguro, he olvidado nombrar.

Muchísimas gracias, por supuesto, a toda mi familia, por impulsarme y apoyarme tantos años y tener que ver cómo me movía de un sitio para otro, prácticamente sin pasar por casa. Muchísimas gracias por los más de 1460 mensajes de buenos días que he recibido durante estos años y más de 208 domingos recibiendo llamadas que servían para poner fin a una semana y recuperar fuerzas para la siguiente. Gracias por todo.

Finally, thanks to the past me. You made it, you should be proud!

This thesis has been a very intense journey but I would do it again.

Contents

Abstract	iii
Introduction	1
Asymmetric organocatalysis development	2
Phase Transfer Catalysis	7
Aims and Objectives	9
References	11
Computational methods	17
Origin of Quantum Chemistry	18
Hartree-Fock method	19
Density Functional Theory	20
Local Density Approximation	21
Generalised Gradient Approximation	22
Basis sets	23
Non-Covalent Interactions	25
Quantum Theory of Atoms In Molecules	25
Natural Bonding Orbitals analysis	26
Non-Covalent Interaction index and Independent Gradient Model	27
Electron Density Shift	28
Energy Decomposition Analysis	28
References	31
1 Efficiency and Suitability when Exploring the Conformational Space of Phase-Transfer Catalysts	35
1.1 Introduction	36
1.2 Computational Details	39
1.3 Evaluation criteria	43
1.4 Results and Discussion	47
1.4.1 Unconstrained Conformational Analysis	47
1.4.2 Constrained Conformational Analysis	57
1.4.3 Final Grades	59
1.5 Conclusions	61
1.5.1 Pros and Cons	61
1.5.2 General Recommendations	63
References	64

2	Revisiting Ion-Pair Interactions in Phase Transfer Catalysis: From Pure Ionic Compounds to Real Catalyst Systems	67
2.1	Introduction	68
2.2	Computational Details	70
2.3	Results and Discussion	72
2.3.1	Ion-Pair	72
2.3.2	Imine- and Iminium-based Models	75
2.3.3	Guanidine- and Guanidinium-based Models	81
2.3.4	Ammonia- and Ammonium-based Models	83
2.3.5	Phase Transfer Catalysts	87
2.4	Conclusions	90
	References	92
3	Improving phase-transfer catalysis by enhancing non-covalent interactions	95
3.1	Introduction	96
3.2	Computational Details	100
3.2.1	Binding Mode Study and Cyanation Reaction of (E)-chalcone	100
3.2.2	Cyanation Reaction of (2E)-1-Phenyl-2-buten-1-one and (2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one	101
3.3	Results and Discussion	104
3.3.1	Conformational Analysis, Binding Mode Study, and Cyanation Reaction Investigation of (E)-Chalcone	104
3.3.2	Conformational Analysis, Binding Mode Study, and Cyanation Reaction Investigation of (2E)-1-Phenyl-2-buten-1-one and (2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one	110
3.4	Conclusions	124
	References	125
4	Dynamic kinetic resolution of azlactones for the production of bis- protected amino acids	129
4.1	Introduction	130
4.2	Computational Details	134
4.3	Results and Discussion	136
4.3.1	Preliminary Study	136
4.3.2	Additional Anchor Point	142
4.3.3	Future Work	150
4.4	Conclusions	152
	References	154
5	Computational Design: Transition State Studies, Enantioselectivity Calculations, and Workflow Automatisa-tion.	157
5.1	Introduction	158
5.2	Computational Details	161
5.2.1	Workflow	161
5.2.2	Enantioselectivity Calculation	163

5.3	Results and Discussion	165
5.3.1	Conformer Generation	165
5.3.2	Conformer Filtration and Selection	166
5.3.3	Reoptimisation Cycles	168
5.3.4	Future Work	171
5.4	Conclusions	173
	References	174
	Conclusions	177
A	Appendixes	181
A.1	Supporting Information	181
A.2	Published papers	185

List of Schemes

1	A- Computational study on the effect of an imidazolium salt counterion and basicity on the favoured product. B- Selective Mannich reaction previously performed experimentally leading to the <i>syn</i> -product, top, and new, computationally design anti-Mannich reaction leading to the <i>anti</i> -product, down.	4
2	Phase-Transfer Catalysis mechanism scheme described by Starks.	7
3.1	Catalytic asymmetric cyanation studied in the synthetic literature.	97
3.2	Catalysts and cyanation reactions of the different ketones under study. Ketone 1 - (E)-chalcone, ketone 2 - (2E)-1-Phenyl-2-buten-1-one and ketone 3 - (2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one.	99
3.3	Asymmetric conjugate cyanation of (E)-Chalcone.	108
4.1	General scheme of Dynamic Kinetic Resolution reaction rates (left) and the corresponding energy scheme of the different processes (right).	130
4.2	Hydrolysis of azalactone 1 to produce a bis-protected amino acid.	131
4.3	a) Scheme of the DKR of an α -substituted azalactone. b) Scheme of the new protective group (tetrachloroisopropoxyester) and the thalamide product. c) Peptide growth mechanism and intermediate involved.	131
4.4	Reaction mechanism of the production of protected amino acids using azalactone as starting material.	133
4.5	Reaction under study and catalyst tested experimentally to validate the experimental conditions.	136
4.6	Nucleophilic attack of the phenolate to the azlactone; the first step of the reaction and responsible for the enantioselectivity of the product.	139
4.7	Suggested movement of electrons and of charge over during the nucleophilic attack of the phenolate to the azlactone and the ring opening process.	150

List of Figures

1	Evolution and historical highlights of organocatalysis.	2
2	Different asymmetric organocatalysis activation methods grouped by type of activation mechanism.	3
3	Commonly used descriptors and representations in ML models applied to catalyst design.	6
4	a) Chiral cinchona alkaloid-based scaffold. b) Spiro-quaternary ammonium salts-based scaffold. c) varied chiral backbones of CPA within anionic PTC.	8
5	Different σ -hole interactions described. From left to right, Halogen Bond, Chalcogen Bond, Pnictogen Bond and Tetrel Bond.	9
6	Workflow of the traditional (top) and computational approach (bottom) of organocatalyst design.	10
7	Effect of p and d polarisation functions to a s -orbital, left, and a p -orbital, right.	24
1.1	Chiral bifunctional PTC-catalyst scaffold and modifications under study.	36
1.2	Proposed binding mods for chiral bifunctional phase transfer catalysts.	37
1.3	Top: Definition of the three possible conformations for the thiourea and squaramide substituted catalysts: <i>anti-anti</i> , left, <i>anti-syn</i> , middle and <i>syn-syn</i> , right. Down: Population of the possible conformations of thiourea and squaramide substituted catalysts when $R = EWG$	37
1.4	Software packages under study.	38
1.5	Workflow for the different software used in this study.	39
1.6	Representative atoms and bonds of the scaffold of the catalysts under study and the different substituents used in the clusterisation process.	44
1.7	Conformational ensembles generated by all methods under study for cat_squa_anthra . Ballon (26), RDKit (110), wSterimol (95) and CREST (209).	48
1.8	Statistical data for the of the conformational space explored by all the packages under study for cat_OH_anthra , top left; clusters generated by all the programs, top right; clusters divided by packages, middle'; and 3D ensembles generated by each software package, bottom.	49
1.9	3D-plot of the different conformers forming every cluster generated by the packages under study for cat_OH_anthra	50
1.10	Gantt chart of the different clusters obtained for cat_OH_anthra and the individual software contributions to each of them.	51
1.11	<i>Space exploration</i> grades for each software calculated over all the studied catalysts.	52

1.12	Energy and RMSD comparison between the studied conformers optimised with the different software packages with respect to the DFT-optimised structures for cat_OH_ph catalyst.	52
1.13	<i>Energy accuracy</i> grades for each software calculated over the three selected catalysts.	53
1.14	<i>Structure accuracy</i> grades for each software calculated over the three selected catalysts.	53
1.15	Most stable conformations for cat_OH_py (left) and cat_thio_tBu (right) for each method. Shown in parenthesis are the RMSD values of all the structures with respect to the DFT one, using the same representative atoms used in the clusterisation process.	54
1.16	<i>Prediction capacity</i> grades for each software calculated over all the catalysts under study.	55
1.17	<i>Tunability</i> grads for each software under study.	56
1.18	Relative energy of the generated conformers for cat_squa_py with all the studied software (left). The thick lines represent those conformers that accumulate more than 99% of the Boltzmann population. Shown in bars is the population of the generated conformers with respect to the %Conformers grouped in 5% intervals (right). The area below the curve, the cumulative population of the conformers, is generated with respect to the %Conformers.	57
1.19	Number of conformers of each catalyst for each software under study and DFT. .	58
1.20	Statistical data of the constrained conformational space explored by all the packages under study for cat_thio_py , top left; clusters generated by all the programs, top right; clusters divided by packages and 3D ensembles generated by each software package, bottom.	58
1.21	Relative energy of the generated constrained conformers for cat_squa_py with all the studied software (left). The thick lines represent those conformers that accumulate more than 99% of the Boltzmann population. Shown in bars is the population of the generated conformers with respect to the %Conformers grouped in 5% intervals (right). The area below the curve, the cumulative population of the conformers, is generated with respect to the %Conformers.	59
1.22	Global grades for each studied software for all the criteria analysed.	60
2.1	Traditional considered interactions between the catalyst and the anion in PTC. .	68
2.2	Different types of ion-pairing interactions. HB-assisted ion-pair and Strict ion-pair.	69
2.3	Representation of the ionic salt models used to investigate strict ion-pairing interactions, left, and anions used, right.	72
2.4	LMO-EDA energy partition terms for the different salts under study, top. LMO-EDA energy partition terms for the $[N(CF_3)_4]^+$ system, bottom.	73
2.5	NBO E(2) values corresponding to the different ionic salts studied, top, and NBO E(2) values corresponding to the different $[N(CF_3)_4]-X$ systems studied, bottom.	74
2.6	QTAIM molecular graph corresponding to $N(CF_3)_4-Cl$	75
2.7	Different imine- and iminium-based derivatives under study.	76
2.8	LMO-EDA energy partition terms for the different Iminium-based complexes under study. Top - Iminium- CF_3 , middle - Iminium CH_3 and bottom - Iminium-tBu.	78

2.9	NBO E(2) values corresponding to the different Iminium-based complexes under study.	79
2.10	QTAIM molecular graphs for Imine_ <chem>CH3-CN</chem> (left) and Iminium_ <chem>CH3-CN</chem> (right).	79
2.11	NBO charges of the monomers and different Iminium-based complexes under study. Legend: charge difference between the complexes and the monomers is depicted.	80
2.12	Gunidine- and guanidinium-based systems under study.	82
2.13	Ammonia- and ammonium-based models under study.	83
2.14	LMO-EDA energy partition terms for the different ammonium-based complexes under study. Top - Ammonium- <chem>CF3</chem> , middle - Ammonium- <chem>CH3</chem> and bottom - Ammonium- <chem>tBu</chem>	84
2.15	NBO E(2) values corresponding to the different ammonia_ _{based} complexes under study.	85
2.16	QTAIM molecular graph for ammonia_ _{based} _ <chem>CH3-Cl</chem> and ammonium_ <chem>CH3-Cl</chem>	85
2.17	NBO charges to the different Ammonium-based complexes under study.	86
2.18	Non-covalent interactions established within the model <chem>Me3NH-CH2COOMe</chem> studied by Houk et al.	86
2.19	PTC catalysts under study, 1 and 2	87
2.20	Top - QTAIM exponential correlation of the different models under study. Bigger dots correspond to the salts and <chem>N(CF3)4-X</chem> systems. Middle - Laplacian values for all the different compounds under study. Bottom - Energy density values for all the different compounds under study.	88
2.21	Different types of intermolecular non-covalent interactions under study in the work, within the phase transfer catalysis topic.	90
3.1	Traditionally interactions considered between the catalyst and the anion in PTC.	96
3.2	Two different interactions analysed in the present study.	98
3.3	Left: Dihedral angles considered for the conformational analysis. Right: Lowest-energy optimised conformer for the catalyst under study, conf_1	104
3.4	MEP of the catalyst under study and different cat-anion complexes under study.	105
3.5	Interaction energies, ΔG_{int} , for all the different cat-anion complexes under study.	105
3.6	Exponential relationships of the intermolecular distance between the catalyst and the anion, in Å, with ρ_{BCP}	106
3.7	3D-NCI plots of the interactions found for the most stable complexes complex_X_1 . From left to right: <chem>Cl^-</chem> , <chem>Br^-</chem> , <chem>MeCO2^-</chem> and <chem>CN^-</chem> . RDG isosurfaces provide a 3D representation of interaction regions. The sign of the second eigenvalue of the density Hessian matrix ($\nabla^2\rho$) is used to differentiate repulsive ($\lambda_2 > 0$) from attractive ($\lambda_2 < 0$) interactions.	107
3.8	Left: Free-energy profiles for the cyanation of 1 in the presence of cat_OMe_ph . Right: Free-energy profiles for the cyanation of 1 in the presence of cat_OH_ph	108
3.9	Optimised geometric for both TSs in the presence of cat_OMe_ph (left) and cat_OH_ph (right) for the major enantiomer in each reaction (<i>S</i> and <i>R</i> respectively).	109

3.10	Catalyst scaffold and substituents (anthra, cf3, ph and tbu) under study and most important rotatable bonds are highlighted.	110
3.11	Conformational analysis of cat_OMe_anthra . From left to right: conformer ensemble, Boltzmann population of each conformer at DFT level and 3D structures of the two most abundant conformers.	111
3.12	a) α - β -unsaturated ketones under study: (2E)-1-Phenyl-2-buten-1-one left and (2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one right. b) <i>s-cis</i> and <i>s-trans</i> conformers' populations of each α - β -unsaturated ketones.	112
3.13	3D structures of the cat-CN (left) and the cat-ketone (right) complexes.	113
3.14	3D structures of the TSs coming from cat-CN (left) and from cat-ketone (right) complexes.	114
3.15	Conjugate cyanation reactions under study. a) (2E)-1-Phenyl-2-buten-1-one cyanation. b) (2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one cyanation.	114
3.16	Different reaction paths under study. From left to right <i>s-cis</i> - <i>R</i> , <i>s-cis</i> - <i>S</i> , <i>s-trans</i> - <i>R</i> , <i>s-trans</i> - <i>S</i>	115
3.17	<i>R:S</i> ratio of the transition states under study for the conjugated cyanation of α - β -unsaturated ketone 1	115
3.18	TS 3D structure of the most stable TS conformer for each $R_2 = OMe$ catalyst under study: <i>R s-trans</i> for cat_OMe_anthra , <i>R s-cis</i> for cat_OMe_cf3 , <i>S s-cis</i> for cat_OMe_ph , and <i>R s-cis</i> for cat_OMe_tbu	116
3.19	Superposition of the <i>S s-cis</i> TS for ketone 1 . In green cat_OH_anthra , in blue cat_OH_cf3 , in purple cat_OH_ph and in orange cat_OH_tbu	117
3.20	NBO E(2) stabilisation energies corresponding to the different HBs present in the transition state in reaction 1. Left $R_2 = OMe$, $CN^- \rightarrow Cat^+$ [$LP(N) \rightarrow BD^*(OH)$], Center: $R_2 = OH$ $CN^- \rightarrow Cat^+$ [$LP(N) \rightarrow BD^*(OH)$], Right: $R_2 = OH$ $Ketone \rightarrow Cat^+$ [$LP(O) \rightarrow BD^*(OH)$].	118
3.21	Exponential relationships of the intermolecular distance, in Å, and the corresponding $\rho(BCP)$ between the catalyst and the anion CN^- (orange), the catalyst and the ketone (blue and purple) and the CN^- anion and the CN^- and the ketone (green).	119
3.22	3D structures of the optimised TSs under study for the cyanation of ketone 2	119
3.23	<i>R:S</i> ratio of the transition states under study for the conjugated cyanation of α - β -unsaturated ketone 2	120
3.24	3D structures of the TS leading to enantiomer <i>R</i> with ketone 2 in the <i>s-cis</i> conformation.	121
3.25	Superposition of the most abundant binding mode for ketone 2 . In green cat_OH_anthra , in blue cat_OH_cf3 , in purple cat_OH_ph and in orange cat_OH_tbu	121
3.26	NBO E(2) stabilisation energies corresponding to the different HBs present in the transition state in reaction 1. Left $R_2 = OMe$, $CN^- \rightarrow Cat^+$ [$LP(N) \rightarrow BD^*(OH)$], Center: $R_2 = OH$ $CN^- \rightarrow Cat^+$ [$LP(N) \rightarrow BD^*(OH)$], Right: $R_2 = OH$ $Ketone \rightarrow Cat^+$ [$LP(O) \rightarrow BD^*(OH)$].	122

3.27	Exponential relationships of the intermolecular distance, in Å, and the corresponding $\rho(BCP)$ between the catalyst and the anion CN^- (blue), the catalyst and the ketone (red) and the CN^- anion and the CN^- and the ketone (green).	123
4.1	Conformational analysis catalyst and their active species. Left to right: cat_ph , cat_anthra , cat_cf3 , cat_tbu	137
4.2	From left to write and top to bottom: 2D scheme of the azlactone derivative under study, Boltzmann population of the different conformers obtained, structure of the three most stable conformers.	138
4.3	3D structures of cat-phenolate (left) and cat-azlactone (right).	138
4.4	Computational and experimental <i>major:minor</i> ratios for, top to bottom, cat_ph , cat_anthra , cat_cf3 and cat_tbu	140
4.5	Experimental vs computational % <i>ee</i> error calculated with 0.20 kcal/mol (light blue) and 0.68 kcal/mol (dark blue) DFT free energy error and $\pm 10\%$ confidence area (grey).	140
4.6	3D structures for TS_R (left) and TS_S (right) under the presence of cat_ph	141
4.7	a) Schematic view of the NCIs established in the TS under study under the presence of cat_ph . b) 3D molecular graph of the TS showing the web of HBs established between the azlactone and the catalyst and their corresponding BCPs as black spheres.	141
4.8	Modified catalysts under study after adding a second HB-donor: cat_conh2 (left), cat_conhph (right).	143
4.9	Molecular Electrostatic Potential (MEP) (a.u.) of cat_conh2 and cat_conhph showing the critical points corresponding to the two different HB-donors.	143
4.10	Different binding modes studied for the cat-phenolate complexes; from left to right NH , OH and 2HB	144
4.11	3D structures of the different conformers of the cat-phenolate complexes between the phenolate and cat_conh2 (up) and cat_conhph (down).	145
4.12	Boltzmann population of the different conformers of the cat-phenolate complexes between the phenolate and cat_conh2 (left) and cat_conhph (right).	145
4.13	Molecular graphs of cat_conh2-phenolate and cat_conhph-phenolate complexes in front conformation and ρ_{BCP} , H and E(2) values associated with each HB between them.	146
4.14	3D structures of the two different binding modes under study for the TS under the presence of cat_conh2 . BM1 left, BM2 right.	147
4.15	Optimised 3D structures for TS_R (right) and TS_S (left) with cat_conh2	148
4.16	Modified catalysts under study after adding a second HB-donor: cat_conh2 (left), cat_conhph (right).	149
4.17	Computational and experimental <i>major:minor</i> ratios for, top to bottom, cat_conh2 , cat_conhph , cat_conhph-no2 and cat_conhph-obr	149
4.18	New modified catalysts for future studies: cat_conhc6conhph-rr and cat_conhc6conhph-ss	151
5.1	Three different approaches for computationally-led organocatalyst design.	158

5.2	General workflow of the different design approaches. a) Rational design, b) automatic design and c) ML design.	159
5.3	TS optimisation workflow scheme.	161
5.4	Nucleophilic attack to the azlactone from the two different faces, left, and interactions and distances fixed in the different orientations of the binding mode used as input, right.	161
5.5	Schematic Potential Energy Surface representing the assumptions done for the study of these reaction. Two different reactions path for each enantiomer of the azlactone and a conformational ensemble for each TS under study.	164
5.6	Binding modes considered for the study. From left to right $R O_{back}$, $R O_{front}$, $S O_{back}$, $S O_{front}$	165
5.7	Conformational ensembles generated by CREST. From left to right $R O_{back}$, $R O_{front}$, $S O_{back}$, $S O_{front}$	166
5.8	Remaining conformers after RMSD filtration using different procedures. a) Filtered together, b) filtered separately.	167
5.9	Top: aligned 3D conformers of cat_conhph generated from $R O_{back}$ starting point at the different theory levels of each reoptimisation cycle. Bottom: Population and cumulative population of the different conformers. Structures marked with * are not optimised, only filtered.	169
5.10	Most stable conformer at each round of the reoptimisation cycle. Left to right $R O_{back}$, $R O_{front}$, $S O_{back}$ and $S O_{front}$. Each round is represented in a different colour: CREST in red , fixed b3lyp/3-21g in the gas phase in orange and fixed m06-2x/6-31g(d) in toluene in green	170
5.11	Experimental and theoretical <i>major:minor</i> ratio and <i>ee</i> values at the different levels of theory.	171

List of Tables

1.1	Number of conformers for each catalyst studied with all the different conformational analysis software.	47
1.2	Results on prediction of the most stable conformation of a catalyst by each studied software.	55
1.3	Results on <i>tunability</i> the different software packages under study.	56
2.1	Ionic salts and $[N(CF_3)_4]\text{-X}$ DFT Interaction energies, ΔE in kJ/mol. Distances in Å.	73
2.2	Electron density, $\rho(r)_{BCP}$, Laplacian, $\nabla^2\rho(r)_{BCP}$, and Energy density, H_{BCP} , at the bond critical point for the ionic systems studied.	75
2.3	Iminium charged and Imine complexes DFT Interaction energies, ΔE in kJ/mol. Distances in Å. Distances between brackets () refer to the second shortest N—O distance in those systems where two O atoms of the PO_4Me_2 anion are interacting with the other monomer. Those systems marked with * were optimised with fixed X-H distance because a H transfer was produced otherwise.	77
2.4	DFT Interaction energies in three different solvents and gas phase, for Ionic salts and Iminium charged complexes, ΔE in kJ/mol. Distances in Å.	81
2.5	Cinchona-based catalyst (1) and Guanidinium-based catalyst (2) DFT Interaction energies, ΔE in kJ/mol. Distances in Å.	87
4.1	Density values at the BCP, ρ_{BCP} (a.u.), Energy density, H (a.u.), and Second order perturbation energy, $E(2)$ (kcal/mol), of the different HB in TS_R and TS_S between the azlactone and cat_ph	142
4.2	Density values at the BCP, ρ_{BCP} (a.u.), Energy density, H (a.u.), and Second order perturbation energy, $E(2)$ (kcal/mol), of the different HB in the different conformers under study.	146

Acronyms

*BD** anti-bonding.

LP Lone Pair.

LV Lone Vacancy.

ee enantiomeric excess.

AARON Automated Reaction Optimiser for New Catalysts.

ACE Asymmetric Catalyst Evaluation.

ADF Amsterdam Density Functional.

ASO Average Steric Occupancy.

BCP Bond Critical Point.

BM Binding Mode.

BOA Born-Oppenheimer approximation.

CatVS Catalyst Virtual Screening.

CC Coupled Cluster.

CCM Continuous Chirality Measure.

CCP Cage Critical Point.

CEP Calculated Electronic Parameter.

CGF Contracted Gaussian Functions.

ChB Chalcogen Bond.

CI Configuration Interaction.

CM Coulomb Matrix.

CP Critical Points.

CPA Chiral Phosphoric Acids.

DENOPTIM DE Novo OPTimisation of In/organic Molecules.

DFT Density Functional Theory.

DKR Dynamic Kinetic Resolution.

DMSO dimethylsulfoxide.

EC Entrance Channel.

EDA Energy Decomposition Analysis.

EDB-EDA Electron Density Based Energy Decomposition Analysis.

EDG electron donating groups.

EDS Electron Density Shift.

EWG electron withdrawing groups.

FDA Food and Drug Administration.

FF Force Field.

GA Genetic Algorithm.

GGA Generalised Gradient Approximation.

GS Ground State.

GSK-EDA Generalised Kohn-Sham Energy Decomposition Analysis.

GTO Gaussian Type Orbitals.

HB Hydrogen Bond.

HDBSCAN Hierarchical Density-Based Spatial Clustering of Applications with Noise.

HEP Huynh Electronic Parameter.

HF Hartree-Fock.

IGM Independent Gradient Model.

IP Ion-Pair.

KR Kinetic Resolution.

LCAO Linear Combination of Atomic Orbitals.

LDA Local Density Approximation.

LMO-EDA Localised Molecular Orbital Energy Decomposition Analysis.

LSDA Local Spin-Density Approximation.

MEP Molecular Electrostatic Potential.

ML Machine-Learning.

MP Möller-Plesset.

NACP Nuclear Attractor Critical Point.

NAO Natural Atomic Orbitals.

NBO Natural Bonding Orbital.

NCI Non-Covalent Interaction.

NCIi Non-Covalent Interaction Index.

NEDA Natural Energy Decomposition Analysis.

NHO Natural Hybrid Orbitals.

NLS Natural Lewis Structure.

PCM Polarizable continuum model.

PES Potential Energy Surface.

PnB Pnictogen Bond.

pre-TS pre-transition state.

PTC Phase-Transfer Catalysis.

QTAIM Quantum Theory of Atoms In Molecules.

RCP Ring Critical Point.

RDG Reduced Density Gradient.

RMSD Root Mean Square Deviation.

RMSE Root Mean Square Error.

RVS Reduced Variational Space analysis.

SCF Self-Consistent Field.

SDF Structure-Data File.

SE Schrödinger Equation.

SEP Semiempirical Electronic Parameter.

SLATM spectrum of London and Axilrod-Teller-Muto.

SMD Solvation Model based on solute electron Density.

SOAP Smooth Overlap of Atomic Positons.

STO Slater Type Orbitals.

TDSE time-dependent Schrödinger Equation.

TEP Tolman Electronic Parameter.

TISE time-independent Schrödinger Equation.

TS transition state.

TSCN Trimethylsilyl Cyanide.

TSFF transition state force fields.

TtB Tetrel Bond.

UMAP Uniform Manifold Approximation and Projection.

vdW van der Waals.

XB Halogen Bond.

Introduction

Asymmetric Organocatalysis Development

Asymmetric organocatalysis has been a dominant and very challenging field in organic chemistry since it was separated from 'alchemy', and is one of the most extensively investigated, owing to its importance for the pharmaceutical industry. Significant advancements in this technology could have an invaluable beneficial environmental and economic impact. For instance, in 2015, all chiral drugs approved by the Food and Drug Administration (FDA) but one were required to be enantiomerically pure.¹

Asymmetric catalysis has traditionally been divided into two main families: transition-metal complex catalysis and biocatalysis, using enzymes as catalysts. But that dichotomy has been diluted during the last two decades since a third category of catalysis emerged: organocatalysis; defined as the use of small organic molecules, organocatalysts, to catalyse organic transformations. This branch, characterised by its robustness and environmental sustainability, has become the standard for modern-day asymmetric catalysis, aligning with contemporary industrial demands regarding environmental degradation.² The application of small organic molecules for catalysis requires mild reaction conditions and facile catalyst substrate separation processes, offering solutions to limitations associated with the respective metal-based catalysts.³ In addition, the advantages of this new type of catalysis, organocatalysis, are unquestionable when applied to the production of pharmaceutical compounds due to its inherent lack of transition metals, which can be extremely toxic (Os, Sn, Pb, Hg, Pd), causing several diseases.⁴⁻⁷

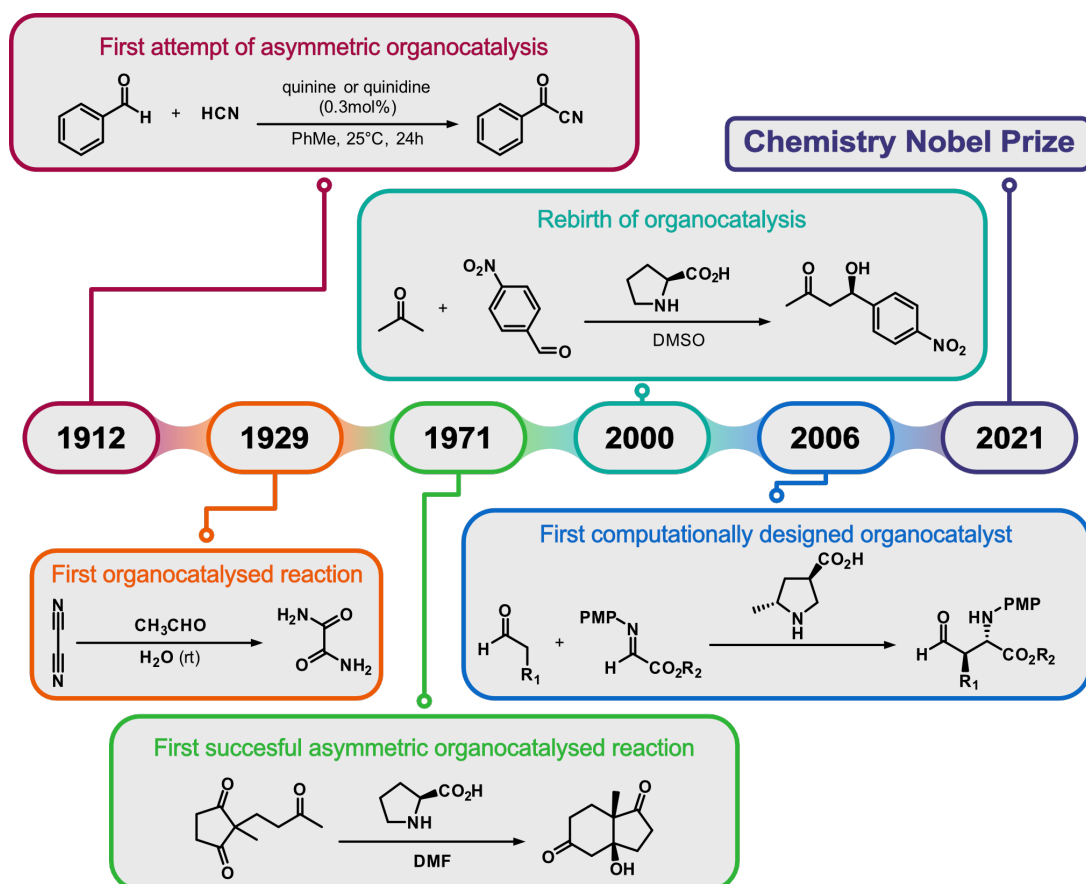


Figure 1. Evolution and historical highlights of organocatalysis.

Despite such merits, the growth of organocatalysis was stagnant in its early years, leaving a vacancy in its research and chemists sceptical of its ability to be a worthwhile candidate. The

first organocatalytic study dates back to 1860,⁸ when cyanogen was converted to oxamide in the presence of aqueous acetaldehyde. However, regardless of such early exposure to organocatalysis, its ability to encourage enantioselectivity was only detected by Bredig in 1912^{9,10} with the addition of hydrogen cyanide to benzaldehyde in the presence of a chiral cinchona alkaloid catalyst (Figure 1).

Immediately after, in 1929, Wolfgang Langebenck reported the formation of oxamides using cyanogen in the presence of aldehyde (Figure 1).¹¹ But it was in 1971, when Hajos and Perish led the way to asymmetric organocatalysis, publishing the first and highly enantioselective catalytic aldol reactions,^{12,13} the Hajos-Parrish-Eder-Sauer-Wiechert reaction; using a simple amino acid, proline, as the catalyst.

Regrettably, despite the significant impact of this compelling milestone in the area of asymmetric organocatalysis, the subsequent three decades witnessed a dormant phase. However, between 1970 and 2000, there was a wealth of interest and development of asymmetric organocatalytic concepts. Notably, Phase-Transfer Catalysis (PTC),^{14,15} Brønsted acid catalysis,¹⁶ tertiary amine catalysis¹⁷ and phosphine catalysis,¹⁸ emerged as crucial elements in contemporary organocatalytic research, although it went largely unnoticed at the time.

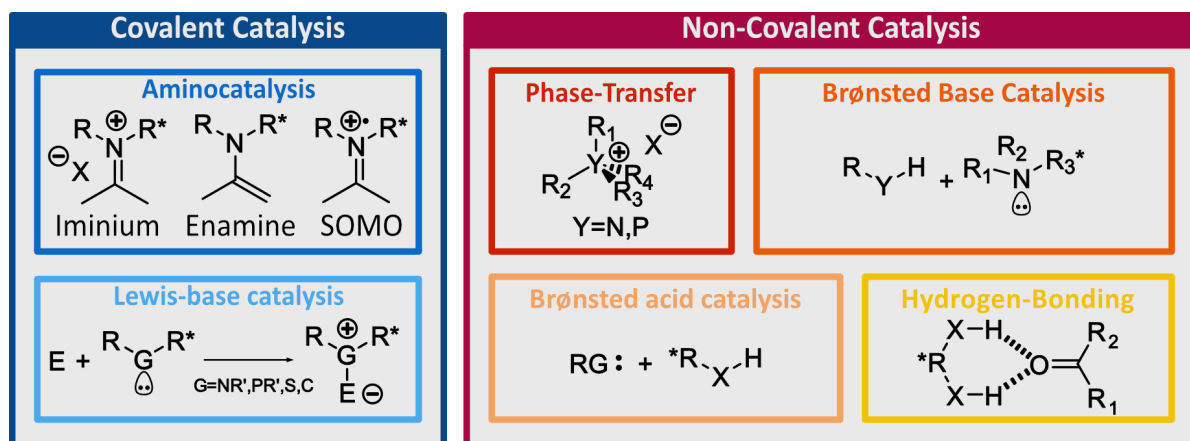


Figure 2. Different asymmetric organocatalysis activation methods grouped by type of activation mechanism.

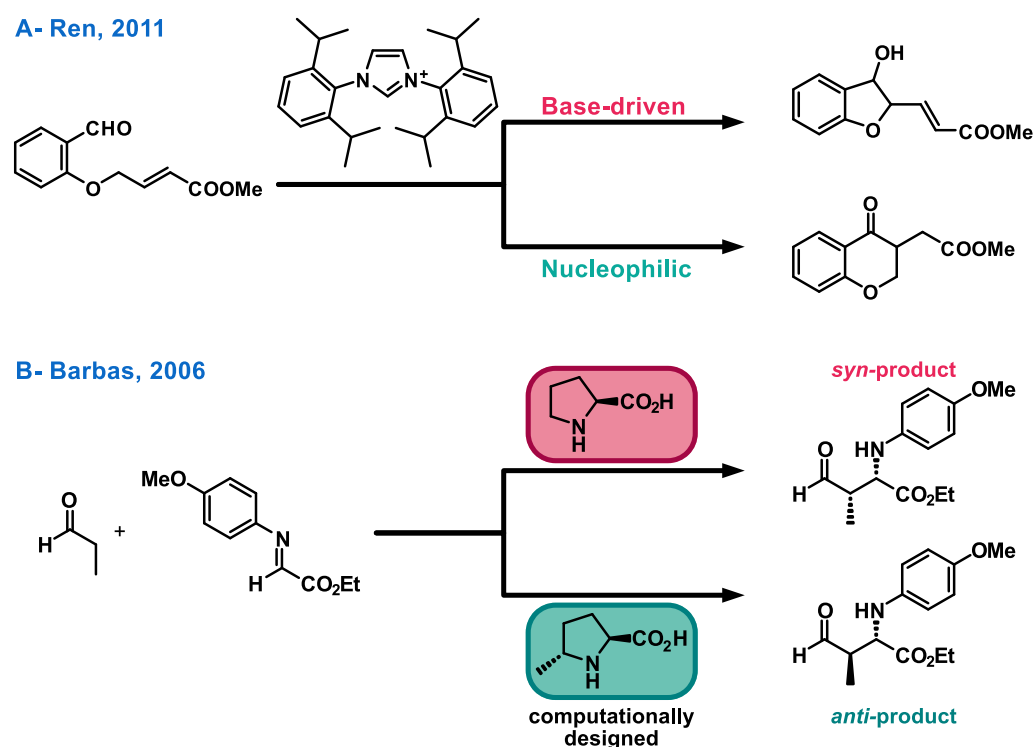
In the 2000s, a paradigm shift occurred, affirming that relatively simple organic molecules could serve as highly effective and remarkably enantioselective catalysts for a variety of fundamentally important transformations.¹⁹ This transformative period was mainly governed by the influential and persistent work of List^{20–22} and MacMillan.^{23–27} Their research focused on enamine and iminium ion-based organocatalysis,^{28,29} simultaneously emphasised the novelty and potential of the field. They demonstrated that organocatalysts are not only versatile but also capable of addressing existing gaps in organic chemistry. This rediscovery induced an immense growth in research activities in organocatalysis, particularly in the last decade, leading to the development of several families with different activation mechanisms, as illustrated in Figure 2.

However, while the field of asymmetric organocatalysis has been continuously expanding, the ability to unravel the mechanistic details has failed to keep pace; mainly because experimental screening has been one of the most common methods employed in catalyst design. It was not until 2004, that the first asymmetric organocatalytic reaction, the Hajos-Parrish reaction, was studied computationally by Houk and co-workers, elucidating the origin of the activity and selectivity of prolines, later expanded to other amino acids.³⁰

With the increase of computational power during these years, numerous reactions have been computationally studied to provide valuable mechanistic insights, elucidating the different selectivities or effects produced by the catalysts. At that stage, computational chemistry was predominantly employed as a retrospective³¹ tool for rationalising chemical phenomena; used, in the particular case of asymmetric organocatalysis, at the end of the design process, once the catalyst was optimised to obtain high selectivity values, to explain the mechanism behind this selectivity.

This retrospective approach was used by Prof Ren and co-workers to explain the role of the counteranion of an imidazolium salt in organocatalysis (Scheme 1 A), revealing that as the basicity of the counteranion increases, the formation of free carbene accelerates accordingly.³² These computational results correlate perfectly with the experimental findings and were used to explain the behaviour of the catalytic cycle.

However, this traditional approach significantly underestimates the predictive potential of computational chemistry. Along with the aforementioned increase of computational power over the last decades, a slow change of paradigm has emerged; substituting this retrospective approach with a prospective one that accounts for the predictive potential inherent in computational chemistry.



Scheme 1. A- Computational study on the effect of an imidazolium salt counterion and basicity on the favoured product. B- Selective Mannich reaction previously performed experimentally leading to the *syn*-product, top, and new, computationally design anti-Mannich reaction leading to the *anti*-product, down.

One of the pioneer examples of this new strategy within the field of asymmetric organocatalyst design was performed by Dr Barbas and co-workers in 2006 (Figure 1).³³ In this project, the authors computationally designed an amino acid to perform an asymmetric anti-Mannich reaction, obtaining astonishing selectivities. They computationally studied the mechanism of a normal Mannich reaction, leading to the *syn*-product (Scheme 1 top) and identified the key

factors that made the process highly enantioselective; finding that the process was guided by the predominance of the (*E*)-enamine, the s-trans conformation of the enamine in the TS, the proton-transfer from the carboxylic acid to imine nitrogen and the attachment of the enamine through the *si* face of the imine.

After analysing all the factors that led to the production of the *syn*-product, they computationally designed a catalyst that reverses the process and produces the *anti*-product (Scheme 1 down). This catalyst was predicted to give a 95:5 *anti:syn* diastereoselectivity and 98%*ee* and the subsequent experimental validation found a quantitative agreement with this prediction, obtaining excellent diastereo and enantioselectivities.

This new prospective approach in which the catalysts are studied computationally and the modifications performed on the basis of these results gave birth to what is known as the rational design of catalysts. Since the first successful example aforementioned, several catalysts have been designed accordingly, taking a candidate catalyst, studying its mechanism, and consequently modifying it in order to obtain the desired outcome.

This approach has been successfully employed in various studies to theoretically enhance the selectivity of different organocatalysts. For instance, Sarotti et al. demonstrated its efficacy in 2020,³⁴ and our own study, presented in this thesis,³⁵ also utilised this method. In the former study, a rational design approach was employed to create a novel organocatalyst for the asymmetric Fiedel-Craft alkylation reaction of indoles, resulting in a significant improvement in process selectivity.

However, this rational approach is not the sole method that has emerged with the development of computational chemistry; different automatic approaches have also been developed,³⁶ facilitating the process of refining a catalyst by computationally testing a large number of possible modifications and candidates, identifying the most important and efficient features, rather than rationally selecting a few of them.

This automatic *in silico* design operates by performing a screening of a library of compounds, using different computational methods such as quantum mechanics or molecular mechanics, to select the most promising catalytic scaffolds for the selected reaction and then designing an optimal catalyst by maximising their efficiency based on the predicted values of the different modifications performed.

Several software packages have been developed under this framework in the last fifteen years like Asymmetric Catalyst Evaluation (ACE),³⁷ which can describe transition regions as “linear combinations of reactants and products” in a reaction through the automation of the conformational search, geometry optimisations, and energy evaluation of transition states using empirical molecular mechanics; Automated Reaction Optimiser for New Catalysts (AARON),³⁸ that can be used to screen processes of both organocatalysts and transition metal catalysts and to generate of intermediates and transition states, being able to determine their catalytic performance; Catalyst Virtual Screening (CatVS),^{39,40} a virtual screening tool that uses Q2MM and generates transition state force fields (TSFF) to identify the most selective ligands; DE Novo OPTimisation of In/organic Molecules (DENOPTIM),⁴¹ a virtual screen and design of organic, inorganic, organometallic catalysts, and supramolecular compounds is enabled by this particular Java package that includes information on the most suitable catalysts in terms of synthetic viability.

In addition to the rational and automatic design of catalysts, a new approach has emerged over the last few years, the Machine-Learning (ML) approach. This alternative to catalyst design

is very attractive since it can potentially accelerate the design process by implementing some of the latest developed ML algorithms.

The ML approach is a vast field that has exploded over the past few years and includes very diverse techniques and models that can be applied to the catalyst design. However, one of the most common and successful is the use of classic molecular descriptors⁴² (Figure 3) to train different ML models that can be used for the evaluation of new catalysts for a given reaction, accelerating the design process.

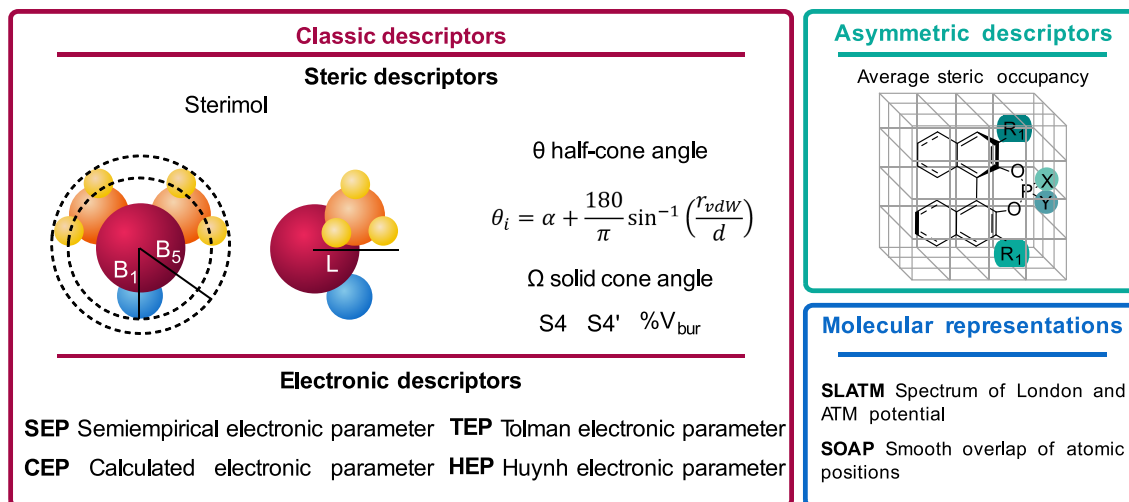


Figure 3. Commonly used descriptors and representations in ML models applied to catalyst design.

Although these descriptors have increased their popularity in recent times, some of them have been around for a few decades, like Sterimol parameters,⁴³ developed in the 70s and that have been used in many different projects. Newer versions, such as wSterimol parameters,^{44,45} have been developed more recently to describe the systems with greater precision, taking into account the different conformations in which the molecule can be found. In addition to steric descriptors, several electronic descriptors have been widely used like Tolman Electronic Parameter (TEP),^{46,47} Calculated Electronic Parameter (CEP),⁴⁸ Semiempirical Electronic Parameter (SEP)⁴⁹ or Huynh Electronic Parameter (HEP);⁵⁰ accounting for the electronic characteristic of the system in different ways (Figure 3).

However, even though classical descriptors perform very well on catalysts, the enantioselectivity is not taken into account and, therefore, new descriptors have been designed for asymmetric catalysis like Average Steric Occupancy (ASO)⁵¹ and Continuous Chirality Measure (CCM)^{52–55} (Figure 3). These asymmetric descriptors have been used in combination with the aforementioned classic descriptors in the design of new asymmetric catalysts.

In addition to the molecular descriptors, different representations have been developed to describe molecules in a way that is readable for the different ML models and that retains the needed information for the prediction of the desired property; the so-called quantum or atomistic ML. Some of the most commonly used molecular representations are Coulomb Matrix (CM),^{56,57} spectrum of London and Axilrod-Teller-Muto (SLATM)⁵⁸ or Smooth Overlap of Atomic Positons (SOAP)⁵⁹ (Figure 3).

Prof. Corminboeuf and co-workers successfully used and modified the SLATM representation for the prediction of the enantioselectivity of different organocatalysts, establishing a protocol to

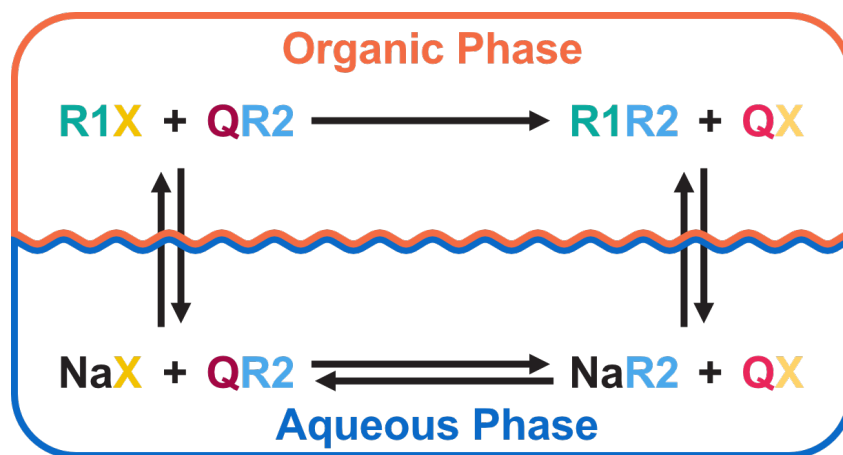
improve molecular representations within the atomistic ML field.⁶⁰ In addition to this protocol, Prof. Corminboeuf’s group has set the state of the art in the use of ML methods for the design and evaluation of new asymmetric catalysts, developing a whole suit of packages, strategies and databases with this purpose.^{61–68} Among the most prolific are NaviCatGA,⁶¹ a genetic algorithm for catalyst optimisation, the B2R2 package⁶² for the generation of different reaction representations and the OSCAR database,⁶⁸ a large repository of catalysts with diverse chemical properties and functional groups.

In recognition of the revolutionary impact that asymmetric organocatalysis has represented, Benjamin List and David MacMillan were awarded the Chemistry Nobel Prize in 2021 as precursors of this field.

Phase Transfer Catalysis

One of the most successful and prolific types of organocatalysis over the last two decades has been Phase-Transfer Catalysis (PTC), a concept introduced for the first time in 1971 by Charles M. Starks⁶⁹ and gaining prominence since the last decade of the century. The mechanism proposed by Starks for the PTC at that time is depicted in Scheme 2. As the name suggests, PTC includes several different procedures, all sharing the common characteristic of occurring in two different and immiscible phases that keep the reactants separated. These phases could traditionally be either an organic and an aqueous phase, or an organic phase and a solid phase. In both cases, the catalyst, **Q**, facilitates the transfer of some of the reactants (**R2** in the scheme) from one phase, usually the aqueous or solid phase, to the other, organic, where the isolated reactant, **R2**, is insoluble. This process greatly accelerates the reaction rate.

Since Dolling’s groundbreaking asymmetric alkylation in 1984,⁷⁰ which marked the first instance of PTC competing with organometallic catalysts, numerous organic reactions have been catalysed using this method.^{71–76}



Scheme 2. Phase-Transfer Catalysis mechanism scheme described by Starks.

In its early stages, PTC primarily relied on the cinchona alkaloid scaffold (Figure 4-a),⁷⁷ despite such scaffolds presenting plausible merits, they also exhibited limitations such as strenuous catalytic modifications. These limitations established a promising starting point for Maruoka et al.⁷⁸ in the development of spiro-quaternary ammonium salts PTC (Figure 4-b). The unique scaffold was found to be catalytically rewarding and vigorous under mild PTC conditions.⁷⁹ How-

ever, within this decade, alternative catalytic strategies are beginning to emerge. For instance, Chiral Phosphoric Acids (CPA)^{80–85} derivatives were incorporated within PTC on account of their efficiency and tremendous structural versatility (Figure 4-c).

Meanwhile, Hydrogen Bond (HB) mediated asymmetric organocatalysis continued to prosper throughout this decade and has been extensively explored since its pivotal role in organic synthetic reactions at the dawn of the millennium. HB-Donor organocatalysts based on the thiourea motif were dominant in the early 2000s,⁸⁶ but in the late 2000s, squaramide-based organocatalysts began to gain acknowledgement earning the scaffold popularity.

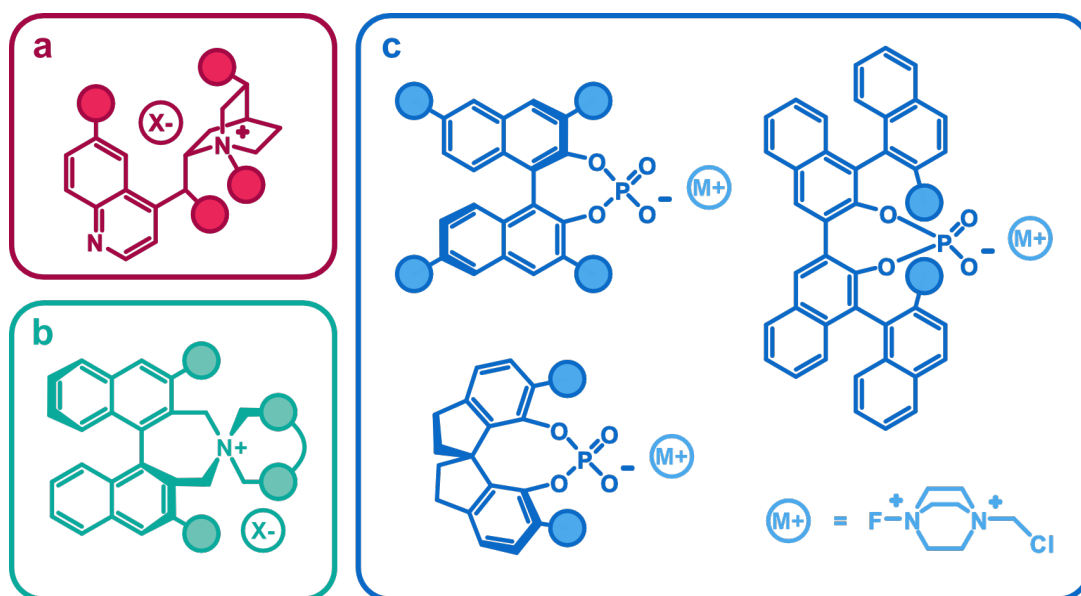


Figure 4. a) Chiral cinchona alkaloid-based scaffold. b) Spiro-quaternary ammonium salts-based scaffold. c) varied chiral backbones of CPA within anionic PTC.

The considerable success of HB-Donors as organocatalysts has gradually cultivated an interest in other non-covalent interactions as promising activation modes, such as σ -hole⁸⁷ interactions. This concept, introduced by Clark in 2005,⁸⁸ referred originally to the positive area generated in the central part of a halogen atom X , located on the opposite side concerning the covalent bond and that is surrounded by a higher electron density belt created by the three pairs of electrons occupying the halogen Lone Pair (LP) orbitals. This concept was later expanded to different positive areas generated, opposite to a covalent bond, in atoms of different chemical groups, generating different σ -hole families (Figure 5); namely Halogen Bond (XB), Chalcogen Bond (ChB), Pnictogen Bond (PnB), and Tetrel Bond (TtB).

The σ -hole catalysis field is still in its infancy, currently being assessed via experimental and theoretical means.^{89,90} Despite such interactions being conceptually new within catalysis, they surpass HB in directionality and rigidity⁹¹ and are very promising substitutes of HB-Donors as catalysts.

The combination of attractive Non-Covalent Interaction (NCI), like the aforementioned, with traditional PTC scaffolds, like cinchona alkaloids, represents a very promising opportunity to enhance their selectivity and design catalysts based on specific attractive interactions, rather than mere steric hindrance.

This aforementioned combination has been successfully used in different reactions like the rational design of a styrene-based chiral organocatalyst for the enantioselective cyclization of 2-

benzothiazolimines performed by Liu et al.⁹² This work was built upon previous theoretical work and rationally modified the catalysts by introducing a thiourea-tertiary amine with multiple H bonding sites on an oxindole-based styrene scaffold, conferring the newly designed catalyst with enantioselectivities up to 98%*ee*.

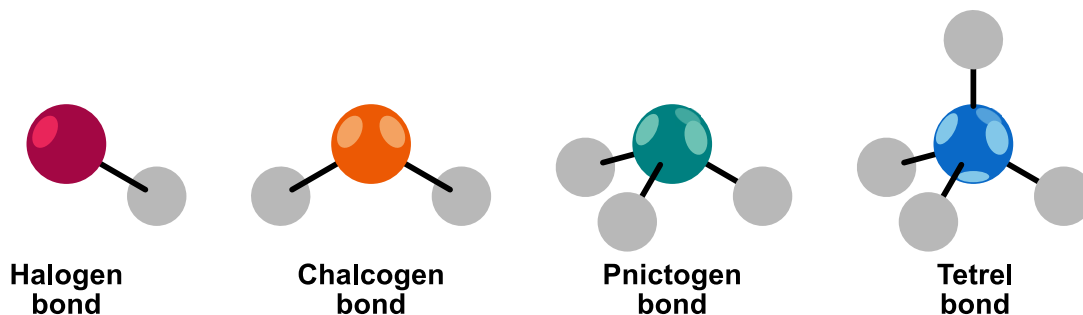


Figure 5. Different σ -hole interactions described. From left to right, Halogen Bond, Chalcogen Bond, Pnictogen Bond and Tetrel Bond.

Several cinchona-based catalysts have been studied in different computational and experimental collaborations between Dr. Trujillo and Prof. Connon, modifying this well-known scaffold with several HB-donors like squaramide^{93–95} and thiourea,^{94,96} obtaining very good selectivities and highlighting the power of combining experimental and computational studies when designing asymmetric organocatalysts based on attractive NCIs.

Aims and Objectives

The synergy between experimental and computational chemistry research has facilitated the development of several qualitative and quantitative models to study enantioselective catalytic processes and design new catalysts, predicting their activity.⁹⁷ However, as mentioned previously, computational chemistry has predominantly functioned as a retrospective tool in most of organocatalytic studies over the last few decades (Figure 6 top), serving primarily to elucidate the mechanism of the reaction.

This traditional approach in which the experiments are run first and the computational studies at the end ignores the predictive potential of computational chemistry and, most importantly, only allows the designing of new catalysts by trial and error, since the mechanism of action of the catalyst is unknown until the very last step.

The objective of the present thesis and some other projects over the last few years is to change this paradigm, letting computational chemistry lead the catalyst optimisation process (Figure 6 bottom); based on mechanistic studies rather than brute force. Once the most key features for high selectivity were described, characterised, and understood, the prospective, rational, and *in silico* design of new catalysts will take place, followed by the experimental testing of the most promising one, creating an iterative process from which optimum catalysts will emerge.

In order to achieve this ambitious goal, this thesis was divided into several sections that will address the different stages of a catalyst design process. The first challenge to be solved with flexible organic molecules, such as the cinchona-based catalysts studied here, is the enormous number of conformations that they can produce and how to find the most relevant and abundant ones. In the first chapter of this thesis, a rigorous benchmark of the most commonly used

conformational analysis software packages was performed to determine the most efficient way of exploring the conformational space of these flexible catalysts.



Figure 6. Workflow of the traditional (top) and computational approach (bottom) of organocatalyst design.

Once the isolated catalysts were studied and the most relevant conformations were found, the interaction between the catalyst and the anionic reactant was analysed since, as can be seen in Scheme 2, the complex between the catalyst and the reactant is the first species formed in this process. Traditionally it has been considered that these complexes were bonded by an ion-pair interaction; however, the nature of this interaction was reevaluated in the second chapter of this thesis, revealing the true nature of the different interactions that can be established and proving guidelines for the correct nomenclature of the different types that can be formed.

After studying the different interactions between the catalysts and the anion involved in a PTC reaction, a case study was selected in order to validate the designing approach that is adopted and understand the mechanism of action of these catalysts. In the third chapter, the enantioselective cyanations reaction of three different α,β -unsaturated ketones were studied in the presence of different cinchona-based catalysts, presenting different EWG and EDG, to analyse the electronic and the steric effects of different groups in the selectivity of the catalysts.

In the fourth chapter, a more challenging reaction, the production of enantiomerically pure protected amino acids, was studied, in collaboration with our experimental colleagues. The validated methodologies tested and analysed during the previous chapters were applied to this reaction, starting an iterative cycle of computational studies, experimental validation and rational design of a new round of catalysts that significantly improved the catalyst performance.

Finally, as a response to challenges encountered at various stages of this design approach, a more automatised computational workflow for studying the TSs involved in an enantioselective reaction was designed, tested, and optimised in the last chapter of this thesis, thus concluding the comprehensive computational study of organocatalysed asymmetric reactions.

References

- (1) A. Calcaterra and I. D’Acquarica, *Journal of Pharmaceutical and Biomedical Analysis*, 2018, **147**, 323–340.
- (2) D. W. MacMillan, *Nature*, 2008, **455**, 304–308.
- (3) M. Fèvre, J. Vignolle, Y. Gnanou and D. Taton, *Polymer Science: A Comprehensive Reference*, 2012.
- (4) K. S. Egorova and V. P. Ananikov, *Organometallics*, 2017, **36**, 4071–4090.
- (5) P. B. Tchounwou, C. G. Yedjou, A. K. Patlolla and D. J. Sutton, in *Experientia Supplementum*, Springer Basel, 2012, pp. 133–164.
- (6) J.-J. Kim, Y.-S. Kim and V. Kumar, *Journal of Trace Elements in Medicine and Biology*, 2019, **54**, 226–231.
- (7) F. Nessa, S. Khan and K. A. Shawish, *Indian journal of pharmaceutical sciences*, 2016, **78**, 111.
- (8) J. F. von Liebig, *Justus Liebigs Annalen der Chemie*, Verlag Chemie, 1927, vol. 453.
- (9) P. S. Fiske, Ph.D. Thesis, ETH Zurich, 1911.
- (10) H. Pellissier, *Tetrahedron*, 2007, **63**, 9267–9331.
- (11) W. Langenbeck, *Justus Liebigs Annalen der Chemie*, 1929, **469**, 16–25.
- (12) Z. G. Hajos and D. R. Parrish.
- (13) Z. G. Hajos and D. R. Parrish, *The Journal of organic chemistry*, 1974, **39**, 1615–1621.
- (14) U. H. Dolling, P. Davis and E. J. Grabowski, *Journal of the American Chemical Society*, 1984, **106**, 446–447.
- (15) T. Ooi, M. Kameda and K. Maruoka, *Journal of the American Chemical Society*, 1999, **121**, 6519–6520.
- (16) M. S. Sigman and E. N. Jacobsen, *Journal of the American Chemical Society*, 1998, **120**, 4901–4902.
- (17) H. Wynberg and E. G. Staring, *Journal of the American Chemical Society*, 1982, **104**, 166–168.
- (18) J. M. Longmire and X. Zhang, *Tetrahedron letters*, 1997, **38**, 1725–1728.
- (19) K. A. Ahrendt, C. J. Borths and D. W. MacMillan, *Journal of the American Chemical Society*, 2000, **122**, 4243–4244.
- (20) J. W. Yang, M. T. Hechavarria Fonseca, N. Vignola and B. List, *Angewandte Chemie International Edition*, 2005, **44**, 108–110.
- (21) J. Zhou and B. List, *Journal of the American Chemical Society*, 2007, **129**, 7498–7499.
- (22) S. Hoffmann, A. M. Seayad and B. List, *Angewandte Chemie*, 2005, **117**, 7590–7593.
- (23) S. Lee and D. W. MacMillan, *Tetrahedron*, 2006, **62**, 11413–11424.
- (24) R. K. Kunz and D. W. MacMillan, *Journal of the American Chemical Society*, 2005, **127**, 3240–3241.

- (25) J. B. Tuttle, S. G. Ouellet and D. W. MacMillan, *Journal of the American Chemical Society*, 2006, **128**, 12662–12663.
- (26) D. A. Nicewicz and D. W. MacMillan, *Science*, 2008, **322**, 77–80.
- (27) S. G. Ouellet, A. M. Walji and D. W. Macmillan, *Accounts of Chemical Research*, 2007, **40**, 1327–1339.
- (28) B. List, R. A. Lerner and C. F. Barbas, *Journal of the American Chemical Society*, 2000, **122**, 2395–2396.
- (29) T. Bui and C. F. Barbas III, *Tetrahedron Letters*, 2000, **41**, 6951–6954.
- (30) P. H.-Y. Cheong, K. Houk, J. S. Warriar and S. Hanessian, *Advanced Synthesis & Catalysis*, 2004, **346**, 1111–1115.
- (31) Z. Tu, T. Stuyver and C. W. Coley, *Chemical Science*, 2023.
- (32) S. Wei, X.-G. Wei, X. Su, J. You and Y. Ren, *Chemistry–A European Journal*, 2011, **17**, 5965–5971.
- (33) S. Mitsumori, H. Zhang, P. Ha-Yeon Cheong, K. Houk, F. Tanaka and C. F. Barbas, *Journal of the American Chemical Society*, 2006, **128**, 1040–1041.
- (34) G. G. Gerosa, M. O. Marcarino, R. A. Spanevello, A. G. Suárez and A. M. Sarotti, *The Journal of Organic Chemistry*, 2020, **85**, 9969–9978.
- (35) I. Iribarren and C. Trujillo, *Physical Chemistry Chemical Physics*, 2020, **22**, 21015–21021.
- (36) M. Foscato and V. R. Jensen, *ACS catalysis*, 2020, **10**, 2354–2377.
- (37) C. Corbeil, S. Thielges, J. Schwartzentruber and N. Moitessier, *Angewandte Chemie-German Edition*, 2008, **120**, 2675–2678.
- (38) Y. Guan, V. M. Ingman, B. J. Rooks and S. E. Wheeler, *Journal of chemical theory and computation*, 2018, **14**, 5249–5261.
- (39) A. R. Rosales, J. Wahlers, E. Limé, R. E. Meadows, K. W. Leslie, R. Savin, F. Bell, E. Hansen, P. Helquist, R. H. Munday et al., *Nature Catalysis*, 2019, **2**, 41–45.
- (40) A. R. Rosales, T. R. Quinn, J. Wahlers, A. Tomberg, X. Zhang, P. Helquist, O. Wiest and P.-O. Norrby, *Chemical Communications*, 2018, **54**, 8294–8311.
- (41) M. Foscato, V. Venkatraman and V. R. Jensen, *Journal of chemical information and modeling*, 2019, **59**, 4077–4082.
- (42) D. J. Durand and N. Fey, *Chemical reviews*, 2019, **119**, 6561–6594.
- (43) A. Verloop, W. Hoogenstraaten and J. Tipker, *Drug design*, 1976, **7**, 165–207.
- (44) T. Piou, F. Romanov-Michailidis, M. Romanova-Michaelides, K. E. Jackson, N. Semakul, T. D. Taggart, B. S. Newell, C. D. Rithner, R. S. Paton and T. Rovis, *Journal of the American Chemical Society*, 2017, **139**, 1296–1310.
- (45) A. V. Brethome, S. P. Fletcher and R. S. Paton, *ACS Catalysis*, 2019, **9**, 2313–2323.
- (46) C. A. Tolman, *Chemical reviews*, 1977, **77**, 313–348.
- (47) O. Diebolt, G. C. Fortman, H. Clavier, A. M. Slawin, E. C. Escudero-Adan, J. Benet-Buchholz and S. P. Nolan, *Organometallics*, 2011, **30**, 1668–1676.

- (48) L. Perrin, E. Clot, O. Eisenstein, J. Loch and R. H. Crabtree, *Inorganic chemistry*, 2001, **40**, 5806–5811.
- (49) K. D. Cooney, T. R. Cundari, N. W. Hoffman, K. A. Pittard, M. D. Temple and Y. Zhao, *Journal of the American Chemical Society*, 2003, **125**, 4318–4324.
- (50) H. V. Huynh, *Chemical reviews*, 2018, **118**, 9457–9492.
- (51) A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow and S. E. Denmark, *Science*, 2019, **363**, eaau5631.
- (52) H. Zabrodsky, S. Peleg and D. Avnir, *Journal of the American Chemical Society*, 1992, **114**, 7843–7851.
- (53) H. Zabrodsky, S. Peleg and D. Avnir, *Journal of the American Chemical Society*, 1993, **115**, 8278–8289.
- (54) H. Zabrodsky and D. Avnir, *Journal of the American Chemical Society*, 1995, **117**, 462–473.
- (55) A. F. Zahrt and S. E. Denmark, *Tetrahedron*, 2019, **75**, 1841–1851.
- (56) B. Huang and O. A. von Lilienfeld, *Nature chemistry*, 2020, **12**, 945–951.
- (57) M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. Von Lilienfeld, *Physical review letters*, 2012, **108**, 058301.
- (58) O. A. von Lilienfeld, K.-R. Müller and A. Tkatchenko, *Nature Reviews Chemistry*, 2020, **4**, 347–358.
- (59) S. De, A. P. Bartók, G. Csányi and M. Ceriotti, *Physical Chemistry Chemical Physics*, 2016, **18**, 13754–13769.
- (60) S. Gallarati, R. Fabregat, R. Laplaza, S. Bhattacharjee, M. D. Wodrich and C. Corminboeuf, *Chemical Science*, 2021, **12**, 6879–6889.
- (61) R. Laplaza, S. Gallarati and C. Corminboeuf, *Chemistry-Methods*, 2022, **2**, e202100107.
- (62) P. van Gerwen, A. Fabrizio, M. D. Wodrich and C. Corminboeuf, *Machine Learning: Science and Technology*, 2022, **3**, 045005.
- (63) R. Laplaza, S. Das, M. D. Wodrich and C. Corminboeuf, *Nature Protocols*, 2022, **17**, 2550–2569.
- (64) S. Gallarati, R. Laplaza and C. Corminboeuf, *Organic Chemistry Frontiers*, 2022, **9**, 4041–4051.
- (65) R. Laplaza, J.-G. Sobez, M. D. Wodrich, M. Reiher and C. Corminboeuf, *Chemical Science*, 2022, **13**, 6858–6864.
- (66) M. D. Wodrich, M. Chang, S. Gallarati, Ł. Woźniak, N. Cramer and C. Corminboeuf, *Chemistry–A European Journal*, 2022, **28**, e202200399.
- (67) S. Das, R. Laplaza, J. T. Blaskovits and C. Corminboeuf, *Angewandte Chemie*, 2022, **134**, e202202727.
- (68) S. Gallarati, P. van Gerwen, R. Laplaza, S. Vela, A. Fabrizio and C. Corminboeuf, *Chemical Science*, 2022, **13**, 13782–13794.
- (69) C. M. Starks, *Journal of the American Chemical Society*, 1971, **93**, 195–199.

- (70) U. H. Dolling, P. Davis and E. J. J. Grabowski, *Journal of the American Chemical Society*, 1984, **106**, 446–447.
- (71) C. M. Starks and R. M. Owens, *Journal of the American Chemical Society*, 1973, **95**, 3613–3617.
- (72) S. Juliá, J. M. del Mazo, L. Avila and J. Elguero, *Organic Preparations and Procedures International*, 1984, **16**, 299–307.
- (73) E. V. Dehmlo, *Angewandte Chemie International Edition in English*, 1977, **16**, 493–505.
- (74) L. J. Mathias and R. A. Vaidya, *Journal of the American Chemical Society*, 1986, **108**, 1093–1094.
- (75) A. McKillop, J.-C. Fiaud and R. Hug, *Tetrahedron*, 1974, **30**, 1379–1382.
- (76) H. Freedman, *Pure and Applied Chemistry*, 1986, **58**, 857–868.
- (77) T. Ooi and K. Maruoka, *Angewandte Chemie International Edition*, 2007, **46**, 4222–4266.
- (78) T. Ooi, M. Kameda and K. Maruoka, *Journal of the American Chemical Society*, 2003, **125**, 5139–5151.
- (79) M. Kitamura, S. Shirakawa, Y. Arimura, X. Wang and K. Maruoka, *Chemistry–An Asian Journal*, 2008, **3**, 1702–1714.
- (80) F. Li, T. Korenaga, T. Nakanishi, J. Kikuchi and M. Terada, *Journal of the American Chemical Society*, 2018, **140**, 2629–2642.
- (81) Á. M. Pálvölgyi, F. Scharinger, M. Schnürch and K. Bica-Schröder, *European Journal of Organic Chemistry*, 2021, **2021**, 5367–5381.
- (82) J. Zhang, P. Yu, S.-Y. Li, H. Sun, S.-H. Xiang, J. Wang, K. N. Houk and B. Tan, *Science*, 2018, **361**, eaas8707.
- (83) J. Zhang, Y.-Y. Wang, H. Sun, S.-Y. Li, S.-H. Xiang and B. Tan, *Science China Chemistry*, 2020, **63**, 47–54.
- (84) I. Čorić and B. List, *Nature*, 2012, **483**, 315–319.
- (85) A. G. Woldegiorgis and X. Lin, *Beilstein Journal of Organic Chemistry*, 2021, **17**, 2729–2764.
- (86) Y.-L. Sun, Y. Wei and M. Shi, *ChemCatChem*, 2017, **9**, 718–727.
- (87) P. Politzer, J. S. Murray and T. Clark, *Physical Chemistry Chemical Physics*, 2013, **15**, 11178–11189.
- (88) T. Clark, M. Hennemann, J. S. Murray and P. Politzer, *Journal of molecular modeling*, 2007, **13**, 291–296.
- (89) L. Lu, Y. Lu, Z. Zhu and H. Liu, *Journal of Molecular Modeling*, 2020, **26**, 1–12.
- (90) S. Benz, A. I. Poblador-Bahamonde, N. Low-Ders and S. Matile, *Angewandte Chemie*, 2018, **130**, 5506–5510.
- (91) H. Yang and M. W. Wong, *Molecules*, 2020, **25**, 1045.
- (92) S.-J. Liu, Z.-H. Chen, J.-Y. Chen, S.-F. Ni, Y.-C. Zhang and F. Shi, *Angewandte Chemie International Edition*, 2022, **61**, e202112226.

- (93) C. Trujillo, I. Rozas, A. Botte and S. J. Connon, *Chemical Communications*, 2017, **53**, 8874–8877.
- (94) F. Manoni, U. Farid, C. Trujillo and S. J. Connon, *Organic & Biomolecular Chemistry*, 2017, **15**, 1463–1474.
- (95) C. Trujillo, S. A. Cronin and S. J. Connon, *European Journal of Organic Chemistry*, 2022, **2022**, e202100818.
- (96) A. G. Collar, C. Trujillo, B. Lockett-Walters, B. Twamley and S. J. Connon, *Chemistry–A European Journal*, 2019, **25**, 7275–7279.
- (97) Q. Peng, F. Duarte and R. S. Paton, *Chemical Society Reviews*, 2016, **45**, 6093–6107.

Computational methods

Since theoretical chemistry was born at the beginning of the 20th century, it has grown exponentially and it has been applied in combination with experiments in all the different disciplines in chemistry. However, its applicability has always been limited by the state of computational technology. The combination of technological progress over the last decades and new quantum methodologies has allowed computational chemists to study systems that were impossible a few years ago.

Origin of Quantum Chemistry

One of the first, and most important, steps in the development of quantum mechanics was made by Louis-Victor de Broglie in 1923,¹ who generalised the idea that light waves present a corpuscular nature and declared that all elementary particles with momentum, p , are associated with a particular wavelength, λ ; expressing it as:

$$\lambda = \frac{c}{\nu} = \frac{2\pi c}{\omega} = \frac{2\pi\hbar}{p} \quad (1)$$

Three years later, Erwin Schrödinger determined that, if all the elementary particles express a wavelike behaviour, they should be governed by a wave function; writing one of the fundamental equations of quantum mechanics, the Schrödinger Equation (SE):²

$$i\hbar \frac{\delta\Psi}{\delta t} = -\frac{\hbar^2}{2m} \left(\frac{\delta^2\Psi}{\delta x^2} + \frac{\delta^2\Psi}{\delta y^2} + \frac{\delta^2\Psi}{\delta z^2} \right) + U(x, y, z)\Psi = \hat{H}\Psi \quad (2)$$

If the SE equation can be simplified, neglecting the relativistic effects and expressing it in atomic units ($\hbar = m_e = e = c = 1$), to describe any quantum system, the non-relativistic time-dependent Schrödinger Equation (TDSE) emerges (Eq. 3). This equation can be written in terms of the electron coordinates, (r_i) , nuclei coordinates, (R_i) and the space-spin coordinates $x_i = (r_i, \eta_i)$, expressing the Hamiltonian as $\hat{H}(r_i, R_i, t)$ and the wavefunction as $\Psi(r_i, R_i, t)$.

$$i\hbar \frac{\delta}{\delta t} \Psi(x_i, R_i, t) = \hat{H}(r_i, R_i, t) \Psi(x_i, R_i, t) \quad (3)$$

However, chemical systems are usually studied as stationary points, getting rid of the time variable and describing them using the most famous and common form of the SE, the time-independent Schrödinger Equation (TISE) (equation 4). Here, one eigenvalue, E , is associated with each of the wavefunction, Ψ , eigenstates of the Hamiltonian, $\hat{H}$.

$$\hat{H}\Psi = E\Psi \quad (4)$$

The Hamiltonian, $\hat{H}$ represents the total energy of a system and can be expressed as the sum of the different energy contributions: the kinetic energy operator for electrons, $\hat{T}_e$, and nuclei, $\hat{T}_n$; the electron-electron, $\hat{V}_{ee}$, and nuclei-nuclei, $\hat{V}_{nn}$, potential operator; and the electron-nuclei Coulomb attraction, namely external potential, $\hat{V}_{ext}$ (Equation 5).

$$\begin{aligned} \hat{H} &= -\sum_{i=1}^N \frac{\hbar^2}{2} \nabla_i^2 - \sum_{A=1}^M \frac{\hbar^2}{2m_A} \nabla_A^2 + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{i=1}^M \sum_{j>i}^M \frac{Z_A Z_B}{R_{ij}} - \sum_{i=1}^N \sum_{A=i}^M \frac{Z_A}{r_{iA}} = \\ &= \hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{nn} + \hat{V}_{ext} \end{aligned} \quad (5)$$

However, all the aforementioned versions of the SE present the same problem; none of them can be exactly solved for systems formed by more than two given particles. In chemical words, the SE can only be exactly solved for the Hydrogen atom or mono-electronic cations.

This is known as the many-body problem and, to overcome it, Max Born and Julius Robert Oppenheimer proposed, in 1927, their famous approximation; the Born-Oppenheimer approximation (BOA). This approximation is founded on the fact the lightest atomic nuclei in the universe, 1H , is about 1800 times more massive than the electron, allowing the separation of the electronic and nuclear components of the Hamiltonian (Eq. 6) and the wavefunction (Eq. 7), because the electronic cloud will instantaneously adapt to any nuclear configuration.

$$\hat{H} = \hat{H}_{elec} + \hat{H}_{nuc} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ext} + (\hat{V}_{nn}) \quad (6)$$

$$\Psi(r, R) = \Phi(r, R)\chi(R) \quad (7)$$

This approximation removes the kinetic contribution of the nuclei, $\hat{T}_{nn}$, now static, and forces the repulsive potential operator between the nuclei, $\hat{V}_{nn}$, to be constant. In the wavefunction, Ψ , the electronic wavefunction, Φ , depends now on the electron coordinates, r , and the static nuclear coordinates, R , while the nuclear wavefunction, χ , only depends on the nuclear positions, R .

Hartree-Fock method

The separation of the electronic and nuclear components of the SE is not enough to solve the many-body problem, different assumptions, approximations and idealizations have to be done to obtain an approximate solution. In 1927 Douglas Rayner Hartree introduced the concept of Self-Consistent Field (SCF);^{3,4} expressing, in a system with two electrons, the movement of the first electron, e_1 , based on a potential field formed by the nucleus of the system and the second electron, e_2 , obtaining an effective Hamiltonian, $\hat{H}_1^{eff}$, for electron 1 (Eq. 8).

$$\hat{H}_1^{eff}\Psi^1(r) = E_1\Psi_1(r) \quad (8)$$

However, the exact solution is not obtained and the variational principle (Eq. 9) must be successively applied, minimising the error of the trial until obtaining the desired accuracy.

$$\Psi_{trial}^*\hat{H}\Psi_{trial} = E_{trial} \geq E_{min}\langle\Psi|\hat{H}|\Psi\rangle \quad (9)$$

When larger, N-electron, systems are described, the wavefunction must be approximated by the product of N particle functions, known as the Hartree product (Eq. 10).

$$\Psi(r) = \prod_{i=1}^N \varphi_i r_i \quad (10)$$

However, due to the fermionic nature of electrons, this approximation does not satisfy the Pauli exclusion principle and antisymmetric functions must be used to amend this problem.

Soon after Hartree published his approximation, in 1930, Vladimir A. Fock, originally Fok (Фок), and John C. Slater described a replacement for the Hartree product (Eq. 10), the

Slater determinant^{5,6} (Eq. 11), expressing the wavefunction as the determinant of one-electron wavefunctions matrix and satisfying the antisymmetric principle.

$$\phi_0 = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(r_1) & \varphi_2(r_1) & \dots & \varphi_N(r_1) \\ \varphi_1(r_2) & \varphi_2(r_2) & \dots & \varphi_N(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(r_N) & \varphi_2(r_N) & \dots & \varphi_N(r_N) \end{vmatrix} \quad (11)$$

This wavefunction, along with the Fock operator (Eq. 12), can be used to rewrite the SE as shown in Eq. 13 and can be solved by using the variational principle (9) 'self-consistently'. This way of solving the SE is known as Hartree-Fock (HF) method.

$$\hat{F}(r) = \hat{h}_i(r) + \sum_i \hat{J}_i(r) - \hat{K}_i(r) \quad (12)$$

$$\hat{F}(r)\phi_0(r) = E_0\phi_0(r) \quad (13)$$

The Fock operator, $\hat{F}(r)$ is expressed in terms of the kinetic energy and the electron-nucleus attraction, $\hat{h}_i(r)$, the Coulomb operator, $\hat{J}_i(r)$ and the exchange operator, $\hat{K}_i(r)$.

However, in terms of accuracy, the HF method does not provide the exact energy of a system, since a single Slater determinant does not take into any consideration the electron correlation. To solve this impediment, different methods have been developed over the years, known as post-HF methods, including the Möller-Plesset (MP) methods,⁷ based on perturbations of several orders (MP^n), the Configuration Interaction (CI) method,⁸⁻¹⁰ based on including more than one electronic configuration, or the Coupled Cluster (CC) method,¹¹⁻¹³ based on the exponential cluster operator.

Density Functional Theory

An alternative picture to the wave functions methods was proposed by Pierre Hohenberg and Walter Kohn in 1964,¹⁴ the 'exact density formalism', that led to the Density Functional Theory (DFT).

This formalism declares that the Ground State (GS) energy of a system can be obtained from its electronic density, and relates the wavefunction of the GS, $\Psi_{(GS)}$, with the electronic density, ρ , in an N -electron system. Obtaining the following equation (Eq. 14) for describing the GS energy of a system.

$$E_0 = \int \dots \int \left[\int [\Psi^*(r, r_2, \dots, r_N) \hat{H} \Psi(r, r_2, \dots, r_N) dr_2 \dots dr_N] \right] \quad (14)$$

This new Hamiltonian can be split into different parts following the Born-Oppenheimer approximation (Eq. 15). By doing so, the nuclear kinetic energy can be neglected and the Hamiltonian described as the sum of the electronic kinetic energy, $\hat{T}(r)$, the electron-electron and nucleus-nucleus Coulombic repulsion, $\hat{V}(r)$ and $\hat{V}(R)$ respectively, and the electron-nucleus Coulombic attraction, $\nu(r)$, also known as external potential.

$$\hat{H}_{el}(r, R) = \hat{T}(r) + \hat{V}(r) + \hat{V}(R) + \nu(r) \quad (15)$$

When this new Hamiltonian is introduced in the GS energy equation, the first three terms can be combined in the so-called 'functional', $F[\rho(r)]$, separated from the external potential (ν), obtaining the following energy expression (Eq. 16).

$$E_0 = \int \dots \int F_x[\rho(r)] + \int_{-\infty}^{\infty} \rho(r)\nu(r)dr \quad (16)$$

While the problem of the wavefunction methodology was the impossibility of solving the equation for multiple electron systems, the problem of this new methodology is obtaining the exact electronic density distribution. This electronic distribution would mean that the energy of the system is minimised.

To find a simplified method to obtain the functional corresponding to the adequate electronic distribution, Walter Kohn, in collaboration with Lu Sham, defined the electron density as a sum of orbital densities (Eq. 17).¹⁵

$$F_x[\rho(r)] = \hat{T}[\rho(r)] + \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 + E_{XC}[\rho(r)] \quad (17)$$

This functional is expressed as the sum of three terms: the electronic kinetic energy, $\hat{T}[\rho(r)]$, that has not changed, the electron-electron interaction in the electronic density, $\frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|}$, and the exchange-correlation energy, $E_{XC}[\rho(r)]$, reflecting the interactions and exchanges between same-spin electrons among the system. The system now is described by a set of orbitals that contains non-interacting electrons, known as Kohn-Sham orbitals, that must satisfy the Schrodinger equation for the system (Eq. 18).

$$\hat{H}^{KS}\Phi_i^{KS}(r_i) = \epsilon_i\Phi_i^{KS}(r_i) \quad (18)$$

Local Density Approximation

Amongst all the terms discussed by Kohn and Sham, the exchange-correlation potential, ν_{XC} , is the only one that remains undisclosed; besides, its analytical solution is still unknown and, consequently, several approximations have been developed.

The Local Density Approximation (LDA) is one of the most common approximations to solve this problem, based only on the electron density; splitting the space in a grid and calculating the electron density at each point.¹⁶⁻¹⁸ This method (Eq. 19) calculates the two different contributions separately, as the sum of the exchange, $E_X^{LDA}[\rho]$, based on the uniform electron gas derived by Dirac¹⁹ (Eq. 20); and the correlation term, which analytical function (Eq. 21) remained unknown until 2016,²⁰ and that was previously calculated with different parameterisation schemes.^{17,18}

$$E_{XC}^{LDA}[\rho] = E_X^{LDA}[\rho] + E_C^{LDA}[\rho] \quad (19)$$

$$E_X^{LDA}[\rho] = -\frac{3}{2} \left(\frac{3}{\pi} \right)^{\frac{1}{3}} \int \rho(r)^{\frac{4}{3}} dr \quad (20)$$

$$E_C^{LDA}[\rho] = \ln \left(1 + \frac{b}{r_s} + \frac{b}{r_s^2} \right) \quad (21)$$

The terms a and b in equation 21 are constants evaluated at the MP2 perturbation theory level and $r_s = \left(\frac{3}{4\pi\rho} \right)^{\frac{1}{3}}$, known as the Wigner-Seitz radius.²¹

An alternative version of this approximation in which α and β electron densities are considered separately, as open-shell systems, was proposed;²² known Local Spin-Density Approximation (LSDA) (Eq. 22).

$$E_{XC}^{LSDA}[\rho^\alpha] = E_X^{LSDA}[\rho^\alpha] + E_C^{LSDA}[\rho^\alpha, \rho^\beta] = E_X^{LSDA}[\rho^\alpha] + \int \rho^\alpha(r) R_C[\rho^\alpha, \rho^\beta] dr \quad (22)$$

Generalised Gradient Approximation

Different alternatives to calculate the exchange-correlation were also developed, including the reduced gradient of the density,²³⁻²⁵ developing the Generalised Gradient Approximation (GGA) methods (Eq. 23); and the kinetic energy, $\tau(r)$, and the Laplacian of the density, $\nabla^2\rho$, developing meta-GGA methods (Eq. 24).

$$E_{XC}^{GGA}[\rho] = \int f(\rho, \nabla\rho) dr, \quad x = \left| \frac{\nabla\rho}{\rho^{4/3}} \right| \quad (23)$$

$$E_{XC}^{meta-GGA}[\rho] = \int f(\rho, \nabla\rho, \tau, \nabla^2\rho) dr, \quad \tau(r) = \frac{1}{2} \sum_{i=1}^{N_{occ}} |\nabla\varphi_i|^2 \quad (24)$$

These methods have been combined at a constant ratio, based on the concept of the adiabatic connection,^{22,26,27} with the exact exchange to calculate the exchange-correlation term, creating a new category known as hybrid-GGA.

Functionals

Since Kohn-Sham theory was released, multiple functionals have been developed under different approximations to increase the level of description of different systems.

The most widely used functional since it was released back in 1993 has been B3LYP, a hybrid-functional developed by Becke.²⁸ The reason why B3LYP is so widely used is that it provides a very good relation between computational cost and accuracy when compared with previous functionals or other more demanding methods like post-HF methods.

Despite being so widely used, B3LYP presents some problems describing Non-Covalent Interaction (NCI). Hence, new functionals were developed, designed to account for those interactions properly. Some examples of the most commonly used amongst those new functionals are wB97XD²⁹ or M06-2x,³⁰ a hybrid meta-GGA functional that combines 32 empirically optimised parameters into the exchange-correlation term and it is provided with double the amount of non-local exchange (54% of HF exchange).

Another option to address the poor description of NCIs is the addition of empirical dispersion corrections, like DFT-D3³¹ (Eq. 25), a damped, atom-atom empirical potential developed by Grimme.

$$E_{disp}^{DFT-D} = - \sum_{A>B} \sum_{n=6,8,\dots} s_n \frac{C_n^{AB}}{R_{AB}^n} f_n(R_{AB}), \quad f_n(R_{AB}) = \left[1 + 6 \left(\frac{R_{AB}}{s_{r,n} R_{AB}^0} \right)^{\alpha_n} \right]^{-1} \quad (25)$$

In this equation, C_n^{AB} is the average isotropic dispersion coefficient for atom pair AB, R_{AB} is their internuclear distance, s_n are the scaling factors and $f_n(R_{AB})$ is a damping function used to avoid near singularities for small internuclear distances.

This empirical potential is usually combined with the B3LYP functional to significantly improve its accuracy when studying systems that present different NCIs.

Basis Sets

In order to solve the SE (4) by using the variational method (9), an appropriate trial wavefunction describing the system has to be selected, as explained above, and a new approximation is required to obtain this initial state. The Linear Combination of Atomic Orbitals (LCAO) (Eq. 26), introduced by Lennard-Jones,³² generates an appropriate wavefunction, φ , by optimising the coefficients $c_{\mu i}$ in a basis of functions Ψ^μ .

$$\varphi = \sum_{\mu=1} c_{\mu i} \Psi_\mu(r; R_i) \quad (26)$$

However, the exact form of these functions, Ψ_μ , is unknown; requiring their mathematical representation as a set of functions, known as a basis set.

Different types of functions have been developed to describe the orbitals over the years. The Slater Type Orbitals (STO) (Eq. 27), defined in spherical coordinates, mimic the exact eigenfunctions of the H atom. However, despite their different advantages, there are no analytical methods to compute two-electron integrals with STO basis set, requiring a huge computational effort to solve numerically that makes them usable only for those methods in which these electrons are not necessary, like semi-empirical methods.

$$\eta^{STO} = Er^{n-1} e^{-\zeta r} Y_{lm}(\Theta, \phi) \quad (27)$$

The Gaussian Type Orbitals (GTO)^{33,34} (Eq. 28) were developed a few years later, solving the integration problem since their exponential dependence, $e^{-\alpha r^2}$, can be solved analytically.

$$\eta^{GTO} = x^l y^m z^n e^{-\alpha r^2} \quad (28)$$

However, these types of functions, used directly, poorly describe the atomic orbitals near the nucleus, leading to greater inaccuracies. To solve this problem, a linear combination of GTO, called primitives, must be used (Eq. 29). This linear combination of primitive Gaussian, namely Contracted Gaussian Functions (CGF), describes the orbitals more accurately as they grow in size and has been widely used to generate different basis sets since Pople introduced them.³⁵⁻³⁷

$$\eta_\tau^{GTO} = \sum_a^A d_{a\tau} \eta_a^{GTO} \quad (29)$$

Further studies on the different basis sets led to the development of split-valence basis set,

in which core and valence electrons are described by a different number of basis functions. Core electrons, being constrained within the atom and not involved in any interatomic interactions, are described by a single basis function; however, valence electrons need more functions to be described and even the addition of more diffuse and polarisation functions to describe their behaviour accurately.

The diffuse functions are extra Gaussian functions used to increase the accuracy in describing orbitals located furthest from the nucleus. These functions become particularly important when studying anionic species or systems presenting different long-range or weak NCIs.

On the other hand, polarisation functions are higher-order orbital functions added to facilitate the desymmetrisation of the atomic orbitals. These auxiliary functions break the symmetry of the orbitals, enabling an accurate representation of the different bonds present between different atoms. This is particularly important in polarised covalent bonds and NCIs. The impact of adding a p polarisation function to an s -orbital and a d polarisation function to a p -orbital is depicted in Figure 7.

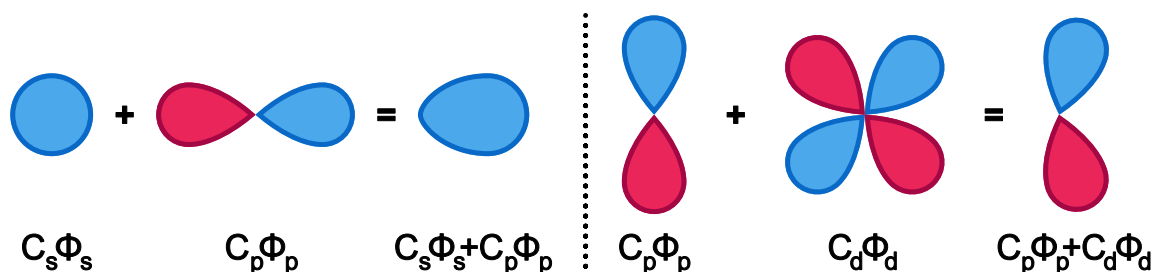


Figure 7. Effect of p and d polarisation functions to a s -orbital, left, and a p -orbital, right.

Different types of basis sets fall into this category, e.g. split-valence Pople basis sets,^{38,39} correlation-consistent basis sets developed by Dunning and coworkers⁴⁰ and Karlsruhe type basis sets.⁴¹

Each of these basis sets has a different nomenclature that represents the number and type of functions that contains. Pople basis sets are expressed using the $X-YZG$ nomenclature where X is the number of primitive functions for each core-electrons and Y and Z represent the number of functions for the valence atoms, indicating that the basis set is a split-valence (double ζ) basis set. Greater basis sets like triple- ζ , and quadruple- ζ require more numbers after the hyphen. The addition of diffuse functions is indicated with a $+$ sign before the 'g'; one $+$ indicates diffuse functions on heavy atoms and $++$ indicates that these functions are also added to hydrogen atoms. Finally, the presence of an $*$ at the end of the basis set code indicates the addition of polarisation functions to the basis set, following the same logic as diffuse functions, $*$ representing only the addition to heavy atoms and $**$ including H. Therefore, the way of representing a triple- ζ basis set that includes polarisation and diffuse functions for heavy atoms and hydrogen would be $6-31++G**$.

The Dunning type basis sets are usually expressed using the $(aug-)\text{cc-}pVNZ$ nomenclature. Here, cc core stands for Correlation-consistent, V indicates that it is split valence and N represents the ζ number, D for double- ζ , T for triple- ζ and so on. The addition of polarisation functions is indicated by p and the presence of diffuse functions is represented by adding aug- at the beginning of the name of the base set. The equivalent of the triple- ζ basis set with polarisation and diffuse functions with this nomenclature is written as $\text{aug-cc-}pVTZ$.

Finally, the Karlsruhe type basis sets use the *def2-NV(P)(D)* nomenclature where N is the ζ number for the split valence functions (*TZ* for triple- ζ , *QZ* for quadruple- ζ , etc.) and the presence of polarisation and diffusion functions is indicated by P and D respectively. Polarisation functions can be indicated as (P) when only the heavy atoms are included, P when there is one set of polarisation functions for all atoms and PP when there are two sets of polarisation functions for all the atoms.

Non-Covalent Interactions

In addition to energy calculations and geometry optimisations, another very important area of computational organic chemistry is the analysis of NCIs. There are several methodologies designed for this task, using different approaches and focusing on different aspects of the interaction.

Quantum Theory of Atoms In Molecules

The Quantum Theory of Atoms In Molecules (QTAIM)^{42–44} is a theory used to analyse the topology of the electron density distribution function, $\rho(x, y, z)$, and extract different properties of the atoms forming the system. The different properties that can be extracted from this analysis are the connectivity of the molecule, expressed as *line paths*, previously known as *bond paths*, that correspond with chemical bonds and different NCIs,⁴⁵ different types of Critical Points (CP), defined as those points of the space in which the density gradient, $\nabla\rho$, disappeared and which type is determined by its curvature.

This analysis is done by calculating the eigenvalues ($\lambda_1, \lambda_2, \lambda_3$) of the Hessian matrix, defined as the matrix of partial second derivatives of the electron density at every given point (Eq. 30).

$$\begin{pmatrix} \frac{\delta^2\rho}{\delta x^2} & \frac{\delta^2\rho}{\delta x\delta y} & \frac{\delta^2\rho}{\delta x\delta z} \\ \frac{\delta^2\rho}{\delta y\delta x} & \frac{\delta^2\rho}{\delta y^2} & \frac{\delta^2\rho}{\delta y\delta z} \\ \frac{\delta^2\rho}{\delta z\delta x} & \frac{\delta^2\rho}{\delta z\delta y} & \frac{\delta^2\rho}{\delta z^2} \end{pmatrix} \Rightarrow \begin{pmatrix} \frac{\delta^2\rho}{\delta x^2} & 0 & 0 \\ 0 & \frac{\delta^2\rho}{\delta y^2} & 0 \\ 0 & 0 & \frac{\delta^2\rho}{\delta z^2} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \quad (30)$$

The different CP can be classified using the (ω, σ) nomenclature, where ω is the number of non-zero eigenvalues and σ is the sum of the signs of the eigenvalues. Four different types of eigenvalues arise from this classification:

- **Nuclear Attractor Critical Point (NACP)** has all negative curvatures and represents a local maxima of the electron density. Correspond to the beginning and the end of a bond path (3,-3).
- **Bond Critical Point (BCP)** shows two positive curvatures and one negative curvature, corresponding to a saddle point of the electron density. It is the minimum density point along a bond path (3, -1).
- **Ring Critical Point (RCP)** is characterised by two positive curvatures and one negative curvature, corresponding to a different type of saddle point of the electron density (3, +1).
- **Cage Critical Point (CCP)** corresponds to local minima of the electron density and is characterised by all positive curvatures (3,+3).

The QTAIM method is used, mainly, to analyse and characterise different intra- and inter-molecular interactions based on the information extracted from the BCPs. The values used for characterising these interactions are usually the value of the electron density at the BCP, ρ_{BCP} , related to the strength of the interaction; the Laplacian of the density at the BCP, $\nabla^2\rho_{BCP}$, defined as the sum of the three eigenvalues, and it is usually negative ($\nabla^2\rho_{BCP} < 0$) for polar covalent bond and positive for ($\nabla^2\rho_{BCP} > 0$) for ionic bonds, hydrogen bonds and weak NCIs. Further information about the nature of the bond can be extracted from the electronic energy density at the BCP, $K(\rho_{BCP})$, defined in Eq. 31 as the sum of the local kinetic and potential energy densities at the BCP, $G(\rho_{BCP})$ and $V(\rho_{BCP})$ respectively.

$$K(\rho_{BCP}) = G(\rho_{BCP}) + V(\rho_{BCP}) \quad (31)$$

Based on these values, the different BCPs can be classified as:

- **Covalent Bond** when $V(\rho_{BCP}) \ll 0$, $G(\rho_{BCP}) \ll |V(\rho_{BCP})|$, $K(\rho_{BCP}) \ll 0$ and $\nabla^2\rho_{BCP} < 0$
- **Polar Covalent Bond** when $V(\rho_{BCP}) \ll 0$, $G(\rho_{BCP}) \ll |V(\rho_{BCP})|$, $K(\rho_{BCP}) \ll 0$ and $\nabla^2\rho_{BCP} > 0$
- **Non-Covalent Interaction** when $V(\rho_{BCP}) < 0$, $G(\rho_{BCP}) \simeq |V(\rho_{BCP})|$, $K(\rho_{BCP}) \simeq 0$ and $\nabla^2\rho_{BCP} \simeq 0$

Natural Bonding Orbitals analysis

A Natural Bonding Orbital (NBO)^{46–48} is defined as the orthonormal set of localised *maximum occupancy* orbitals that give the closest Lewis-like description of an N -electron molecular system. These NBO orbitals are a localised Natural Lewis Structure (NLS) representation of the wavefunction.

Mathematically, these orbitals are composed of a series of optimised linear combinations of Natural Atomic Orbitals (NAO), known as Natural Hybrid Orbitals (NHO) (Eq. 32).

$$h_A = \sum_k a_k \Theta_k^{(A)} \quad (32)$$

The orbitals that can be formed under these conditions can be divided into two main groups: 1-centre orbitals, defined by a single normalised NHO, $n_A = h_A$, or 2-centre orbitals, $\Omega_{AB} = a_A h_A + a_B h_B$, composed by a linear combination of two NHO. Core orbitals (CR) and lone pair (LP) orbitals fall into the first category while σ and π bonding orbitals (BD) would fall into the latter.

To satisfy the orthogonality of the set, each bonding NBO must have an anti-bonding (out-of-phase) orbital (BD*), $\Omega_{AB}^* = a_B h_A - a_A h_B$.

A proper analysis of these orbitals can provide valuable information about the strength of intermolecular interactions, making it possible to measure the stabilisation caused by the different inter- and intra-molecular NCIs. These interactions emerge in the NBO analysis as the second-order perturbation energies, $E(2)$; the stabilisation energy due to the charge transfer $i \longleftrightarrow j$, $E(2)$, between a pair of donor NBO i and acceptor NBO j (Eq. 33).

$$E(2) = \Delta E_{ij}(2) = q_i \frac{F(i, j)^2}{(\epsilon_i - \epsilon_j)} \quad (33)$$

Here, q_i represents the occupancy of the donor NBO, $F(i, j)$ are the off-diagonal NBO matrix elements and ϵ_i and ϵ_j stand for the orbital energies (diagonal elements of the NBO Fock matrix).

By measuring the second-order perturbation energy associated with an interaction, we can not only measure its strength but also analyse its nature in terms of orbitals and classify it; e.g. HB are defined by the $E(2)$ associated to the charge transfer from a lone pair orbital (LP) within the HB-acceptor to an anti-bonding sigma orbital (BD*), also expressed as σ^* , within the HB-donor (Eq. 34).

$$LP(HB - acceptor) \longrightarrow \sigma_{OH}^*(HB - donor) \quad (34)$$

Non-Covalent Interaction index and Independent Gradient Model

Despite being extremely useful for studying and analysing different NCIs, when these interactions are very weak or dispersed like, for example, π - π stacking, these methods can fail to analyse them properly. Amongst the different alternatives to analyse and visualise these interactions, two different density-based methodologies arose: Non-Covalent Interaction Index (NCIi)^{49,50} and Independent Gradient Model (IGM).^{51,52}

Both methodologies are based on the analysis of the electron density, ρ , similarly to QTAIM.

The NCIi is based on the analysis of the Reduced Density Gradient (RDG), $s(\rho)$ (Eq. 35).

$$s(\rho) = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho|}{\rho^{4/3}} \quad (35)$$

Interactions are characterised by low values of this RDG, $s(\rho)$, and their nature can be determined based on the sign of the second eigenvalue of the diagonal Hessian matrix (Eq. 30), λ_2 .⁵³ Negative values of λ_2 ($\lambda_2 > 0$) indicate the presence of a bonding interaction, e.g. HB, while positive λ_2 values ($\lambda_2 > 0$) indicate the presence of repulsive and non-bonded interactions, such as steric repulsion. Weak van der Waals (vdW) interactions are defined by negligible overlaps between the two electronic densities, resulting in small values of ρ , RDG and $\lambda_2 \approx 0$.

Moreover, later developments of the methodology⁵⁰ allowed the quantification of these interactions by integrating the electron density in the NCI region (Eq. 36), obtaining optimal values for $n = 2-2.5$.

$$I_n = \int_{\Omega_{NCI}} \rho^n(r) dr \quad (36)$$

A similar approach was taken by IGM in which the RDG was substituted by a new descriptor δg , calculated as the difference between the module of the density gradient, $|\nabla\rho|$, and the Independent density gradient, $|\nabla\rho^{IGM}|$, defined in Eq. 37. In addition to identifying and characterising the interactions, analogously to how they were characterised with NCIi, the descriptor δg is also correlated to the strength of the interaction, growing in tandem with it.

$$\delta g = |\nabla\rho^{IGM}| - |\nabla\rho| \quad (37)$$

This independent gradient, $|\nabla\rho^{IGM}|$, is obtained by calculating the IGM derivatives in the

the space components, as it is shown in Eq. 38 for x .

$$\left(\frac{\delta\rho}{\delta x}\right)^{IGM} = \sum_{i=1}^N \left|\frac{\delta\rho_i}{\delta x}\right| = \sum_{i=1}^N \frac{|x_i - x|}{r_i} \sum_{j=1}^{n_s} a_{i,j} b_{i,j} e^{-b_{i,j} r_i} \quad (38)$$

Electron Density Shift

Another density-based methodology for the study of NCIs is the Electron Density Shift (EDS) maps.⁵⁴ The EDS density, ρ^{EDS} , is defined as the difference between the density of the complex under study, ρ^{sys} , and the sum of the densities of the individual monomers forming it, ρ^{mon} , as shown in Eq. 39.

$$\rho^{EDS} = \rho^{sys} - \sum_{i=1}^N \rho_i^{mon} \quad (39)$$

Despite that this technique is visually very powerful and flexible, it lacked any numerical evaluation. However, the electron density can be directly extracted from the wavefunction of a molecule, which is very convenient and clean in terms of evaluation and numerical manipulation.

To solve this lack of quantification, we developed a new method for the quantification of the EDS.⁵⁵ The quantification of the EDS is based on separating the positive, $\rho^{sys} > \sum \rho^{mon}$, and negative, $\rho^{sys} < \sum \rho^{mon}$, density areas because, if they are considered together, the total EDS value would be 0, since the global electron density remains the same, not creating or destroying any in the process. This separation is defined, over the points of a grid, in Eq. 40.

$$\sum_{i=1}^N \rho^{EDS}(i) d\tau = \sum_{i=1}^N \rho_+^{EDS}(i) d\tau + \sum_{i=1}^N \rho_-^{EDS}(i) d\tau \quad (40)$$

The total value of the positive and negative EDS corresponds to the total number of electrons displaced upon interaction (Eq. 41). These EDS values correlate with the strength of the interaction, providing another tool to study them numerically.

$$\sum_{i=1}^N \rho_+^{EDS}(i) d\tau = n e_+^{EDS} \quad (41)$$

Energy Decomposition Analysis

A completely different approach to studying molecular systems from an energetic perspective rather than a density one are the different Energy Decomposition Analysis methods, very commonly used in the analysis of intermolecular NCIs and their nature. All these methods divide the total energy of the system into different components such as electrostatic, exchange, repulsion, dispersion or polarisation energies. Due to the semi-arbitrary selection of the components and how to perform the partition, many different types of EDA have been developed over the years for *ab initio* methods, DFT methods and even some density-based methods, all of them with different pros and cons, and suitable for different types of systems; namely Energy Decomposition Analysis (EDA),⁵⁶ Natural Energy Decomposition Analysis (NEDA),⁵⁷ Localised Molecular Orbital Energy Decomposition Analysis (LMO-EDA),^{58,59} Generalised Kohn-Sham Energy Decomposition Analysis (GSK-EDA),⁶⁰ Reduced Variational Space analysis (RVS),⁶¹ Electron Density Based Energy Decomposition Analysis (EDB-EDA),^{62,63} amongst others.

Taking one of the most commonly used EDA⁵⁶ as an example, implemented in the Amsterdam Density Functional (ADF) software package, it can be seen how the total HF interaction energy of a certain system, N , formed by N monomers is calculated as shown in Eq. 42.

$$\Delta E^{HF} = \langle \Phi_X | H_X | \Phi_X \rangle - \sum_{i=1}^N \langle \Phi_i | H_i | \Phi_i \rangle \quad (42)$$

This interaction energy can be decomposed into electrostatic, exchange, repulsion and polarisation energies (Eq. 43).

$$\Delta E^{HF} = \Delta E^{ele} + \Delta E^{ex} + \Delta E^{rep} + \Delta E^{pol} \quad (43)$$

The different energetic components of the EDA are defined by different combinations of the following approximate energy expressions:

$$\begin{aligned} E^{HF} = & \sum_i^{\alpha,\beta} \sum_j^{\alpha,\beta} h_{ij} (S^{-1})_{ij} + \frac{1}{2} \sum_i^{\alpha,\beta} \sum_j^{\alpha,\beta} \sum_k^{\alpha,\beta} \sum_l^{\alpha,\beta} \langle ij | kl \rangle \times (S^{-1})_{ij} (S^{-1})_{kl} \\ & - \frac{1}{2} \sum_i^{\alpha} \sum_j^{\alpha} \sum_k^{\alpha} \sum_l^{\alpha} \langle ik | jl \rangle \times (S^{-1})_{ij} (S^{-1})_{kl} - \frac{1}{2} \sum_i^{\beta} \sum_j^{\beta} \sum_k^{\beta} \sum_l^{\beta} \langle ik | jl \rangle \\ & \times (S^{-1})_{ij} (S^{-1})_{kl} + E^{nuc} \end{aligned} \quad (44)$$

$$E_X^{(1)} = \sum_{i \in X}^{\alpha,\beta} h_i + \frac{1}{2} \sum_{i \in X}^{\alpha,\beta} \sum_{j \in X}^{\alpha,\beta} \langle ii | jj \rangle - \sum_A \left(\frac{1}{2} \sum_{i \in A}^{\alpha} \sum_{j \in A}^{\alpha} \langle ij | ij \rangle + \frac{1}{2} \sum_{i \in A}^{\beta} \sum_{j \in A}^{\beta} \langle ij | ij \rangle \right) + E_X^{nuc} \quad (45)$$

$$E_X^{(2)} = \sum_{i \in X}^{\alpha,\beta} h_i + \frac{1}{2} \sum_{i \in X}^{\alpha,\beta} \sum_{j \in X}^{\alpha,\beta} \langle ii | jj \rangle - \frac{1}{2} \sum_{i \in X}^{\alpha} \sum_{j \in X}^{\alpha} \langle ij | ij \rangle - \frac{1}{2} \sum_{i \in X}^{\beta} \sum_{j \in X}^{\beta} \langle ij | ij \rangle + E_X^{nuc} \quad (46)$$

$$\begin{aligned} E_X^{(3)} = & \sum_{i \in X}^{\alpha,\beta} \sum_{j \in X}^{\alpha,\beta} h_{ij} (S^{-1})_{ij} + \frac{1}{2} \sum_{i \in X}^{\alpha,\beta} \sum_{j \in X}^{\alpha,\beta} \sum_{k \in X}^{\alpha,\beta} \sum_{l \in X}^{\alpha,\beta} \langle ij | kl \rangle \times (S^{-1})_{ij} (S^{-1})_{kl} \\ & - \frac{1}{2} \sum_{i \in X}^{\alpha} \sum_{j \in X}^{\alpha} \sum_{k \in X}^{\alpha} \sum_{l \in X}^{\alpha} \langle ik | jl \rangle \times (S^{-1})_{ij} (S^{-1})_{kl} - \frac{1}{2} \sum_{i \in X}^{\beta} \sum_{j \in X}^{\beta} \sum_{k \in X}^{\beta} \sum_{l \in X}^{\beta} \langle ik | jl \rangle \\ & \times (S^{-1})_{ij} (S^{-1})_{kl} + E_X^{nuc} \end{aligned} \quad (47)$$

These approximate expressions are combined for calculating the different items in Eq. 43 like:

$$\Delta E^{ele} = E_X^{(1)} - \sum_A E_A^{HF} \quad (48)$$

$$\Delta E^{ex} = E_X^{(2)} - E_X^{(1)} \quad (49)$$

$$\Delta E^{rep} = E_X^{(3)} - E_X^{(2)} \quad (50)$$

$$\Delta E^{ele} = E_X^{(HF)} - E_X^{(3)} \quad (51)$$

References

- (1) L. De Broglie, *Nature*, 1923, **112**, 540–540.
- (2) E. Schrödinger, *Annalen der physik*, 1926, **385**, 437–490.
- (3) D. R. Hartree, Mathematical Proceedings of the Cambridge Philosophical Society, 1928, vol. 24, pp. 89–110.
- (4) D. R. Hartree, Mathematical Proceedings of the Cambridge Philosophical Society, 1928, vol. 24, pp. 111–132.
- (5) V. Fock, *Zeitschrift für Physik*, 1930, **61**, 126–148.
- (6) J. C. Slater, *Physical Review*, 1930, **35**, 210.
- (7) C. Møller and M. S. Plesset, *Physical review*, 1934, **46**, 618.
- (8) E. Condon and G. Shortley, *Physical Review*, 1931, **37**, 1025.
- (9) E. Condon, *Physical Review*, 1930, **36**, 1121.
- (10) J. A. Pople, M. Head-Gordon and K. Raghavachari, *The Journal of chemical physics*, 1987, **87**, 5968–5975.
- (11) K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chemical Physics Letters*, 1989, **157**, 479–483.
- (12) J. Čížek, *The Journal of Chemical Physics*, 1966, **45**, 4256–4266.
- (13) H. J. Monkhorst, *International Journal of Quantum Chemistry*, 1977, **12**, 421–432.
- (14) P. Hohenberg and W. Kohn, *Physical review*, 1964, **136**, B864.
- (15) W. Kohn and L. J. Sham, *Physical Review*, 1965, **140**, A1133–A1138.
- (16) J. Perdew, E. McMullen and A. Zunger, *Physical Review A*, 1981, **23**, 2785.
- (17) J. P. Perdew and Y. Wang, *Physical review B*, 1992, **45**, 13244.
- (18) S. H. Vosko, L. Wilk and M. Nusair, *Canadian Journal of physics*, 1980, **58**, 1200–1211.
- (19) P. A. Dirac, Mathematical proceedings of the Cambridge philosophical society, 1930, vol. 26, pp. 376–385.
- (20) T. Chachiyo, *The Journal of Chemical Physics*, 2016, **145**.
- (21) L. A. Girifalco, *Statistical mechanics of solids*, OUP USA, 2003, vol. 58.
- (22) O. Gunnarsson and B. I. Lundqvist, *Physical Review B*, 1976, **13**, 4274.
- (23) J. P. Perdew, *Physical Review B*, 1986, **33**, 8822.
- (24) A. D. Becke, *Physical review A*, 1988, **38**, 3098.
- (25) C. Lee, W. Yang and R. G. Parr, *Physical review B*, 1988, **37**, 785.
- (26) S. Kurth, J. P. Perdew and P. Blaha, *International journal of quantum chemistry*, 1999, **75**, 889–909.
- (27) J. Harris and R. Jones, *Journal of Physics F: Metal Physics*, 1974, **4**, 1170.
- (28) A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 1372–1377.
- (29) J.-D. Chai and M. Head-Gordon, *The Journal of chemical physics*, 2008, **128**, 084106.

- (30) Y. Zhao and D. G. Truhlar, *Theoretical Chemistry Accounts*, 2007, **120**, 215–241.
- (31) S. Grimme, J. Antony, T. Schwabe and C. Mück-Lichtenfeld, *Organic & Biomolecular Chemistry*, 2007, **5**, 741–758.
- (32) J. E. Lennard-Jones, *Proceedings of the Physical Society*, 1931, **43**, 461.
- (33) R. McWeeney, *Solutions of the Schrödinger Wave Equation*, University of Oxford, 1948.
- (34) R. McWeeney, *Nature*, 1950, **166**, 21–22.
- (35) J. Almlöf and P. R. Taylor, *The Journal of chemical physics*, 1987, **86**, 4070–4077.
- (36) P.-O. Widmark, P.-Å. Malmqvist and B. O. Roos, *Theoretica chimica acta*, 1990, **77**, 291–306.
- (37) K. A. Peterson and T. H. Dunning Jr, *The Journal of chemical physics*, 2002, **117**, 10548–10560.
- (38) M. J. Frisch, J. A. Pople and J. S. Binkley, *The Journal of Chemical Physics*, 1984, **80**, 3265–3269.
- (39) P. C. Hariharan and J. A. Pople, *Theoretica Chimica Acta*, 1973, **28**, 213–222.
- (40) T. H. Dunning Jr, *The Journal of chemical physics*, 1989, **90**, 1007–1023.
- (41) A. Hellweg and D. Rappoport, *Physical Chemistry Chemical Physics*, 2015, **17**, 1010–1017.
- (42) C. Matta and R. Gillespie, *Journal of Chemical Education*, 2002, **79**, 1141.
- (43) R. F. Bader, *Accounts of chemical research*, 1985, **18**, 9–15.
- (44) P. L. A. Popelier, F. Aicken and S. O’Brien, *Atoms in molecules*, Prentice Hall Manchester, 2000, vol. 188.
- (45) E. Runge and E. K. Gross, *Physical review letters*, 1984, **52**, 997.
- (46) F. Weinhold and C. R. Landis, *Valency and bonding: a natural bond orbital donor-acceptor perspective*, Cambridge University Press, 2005.
- (47) B. D. Dunington and J. Schmidt, *Journal of Chemical Theory and Computation*, 2012, **8**, 1902–1911.
- (48) A. E. Reed, L. A. Curtiss and F. Weinhold, *Chemical Reviews*, 1988, **88**, 899–926.
- (49) J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan and W. Yang, *Journal of chemical theory and computation*, 2011, **7**, 625–632.
- (50) R. A. Boto, F. Peccati, R. Laplaza, C. Quan, A. Carbone, J.-P. Piquemal, Y. Maday and C.-G. Julia, *Journal of Chemical Theory and Computation*, 2020, **16**, 4150–4158.
- (51) C. Lefebvre, G. Rubez, H. Khartabil, J.-C. Boisson, C.-G. Julia and E. Hénon, *Physical Chemistry Chemical Physics*, 2017, **19**, 17928–17936.
- (52) C. Lefebvre, J. Klein, H. Khartabil, J.-C. Boisson and E. Hénon, *Journal of Computational Chemistry*, 2023.
- (53) G. B. Arfken, H. J. Weber and F. E. Harris, *Mathematical methods for physicists: a comprehensive guide*, Academic press, 2011.
- (54) D. Stalke, *Chemistry—A European Journal*, 2011, **17**, 9264–9278.

- (55) I. Iribarren, G. Sánchez-Sanz, I. Alkorta, J. Elguero and C. Trujillo, *The Journal of Physical Chemistry A*, 2021, **125**, 4741–4749.
- (56) P. Su and H. Li, *The Journal of Chemical Physics*, 2009, **131**, 014102.
- (57) E. D. Glendening, *The Journal of Physical Chemistry A*, 2005, **109**, 11936–11940.
- (58) P. Su and H. Li, *The Journal of Chemical Physics*, 2009, **131**, 014102, DOI: 10.1063/1.3159673.
- (59) P. Su, Z. Tang and W. Wu, *WIREs Computational Molecular Science*, 2020, **10**, e1460.
- (60) P. Su, Z. Tang and W. Wu, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2020, **10**, e1460.
- (61) W. Chen and M. S. Gordon, *The Journal of Physical Chemistry*, 1996, **100**, 14316–14328.
- (62) M. Mandado and J. M. Hermida-Ramón, *Journal of Chemical Theory and Computation*, 2011, **7**, 633–641.
- (63) N. Ramos-Berdullas, I. Pérez-Juste, C. Van Alsenoy and M. Mandado, *Physical Chemistry Chemical Physics*, 2015, **17**, 575–587.

1 | Efficiency and Suitability when Exploring the Conformational Space of Phase-Transfer Catalysts

1.1 Introduction

Conformational analysis plays a fundamental role in several research fields of organic chemistry. In 1970, Barton proposed the first definition of conformation: *"the conformations of a molecule (of defined constitution and configuration) are those arrangements in space of the atoms of the molecule which are not superimposable upon each other"*.^{1,2}

Organocatalysts are flexible organic compounds with rich structural dynamics. Similarly to enzymes, their function is directly linked to their structure and their activity may be regulated via their conformational dynamics. Consequently, it is crucial to investigate their full conformational space when studying such flexible molecules. The generation of conformations for flexible organic molecules is a challenging problem of continuing interest in the field of cheminformatics and computational drug discovery.

Traditionally, the exploration of this conformational space was performed by hand, manually generating the different conformers by rotating different bonds, following the chemical intuition to avoid unlikely conformations and save computational time. The time demand for this process escalates extremely fast with the number of flexible bonds and the number of rotations of each bond ($n_{\text{conformers}} = n_{\text{bonds}}^{n_{\text{rotations}}}$), making it unfeasible for large organic molecules. As computational power started to increase, different software packages were developed to automate conformational analysis, speeding up the process and making it as unbiased as possible. Different approaches have been taken to address both the conformer generation problem (e.g. random number generation, metadynamics, systematic rotations) and the conformer optimisation problem (e.g. Force Field (FF)s, semiempirical methods, DFT calculations). This chapter will provide an overview of widely employed software packages for sampling conformational space, with a particular emphasis on methods tailored for organic compounds of mutual interest within the field of organocatalysis, particularly in the context of PTC.

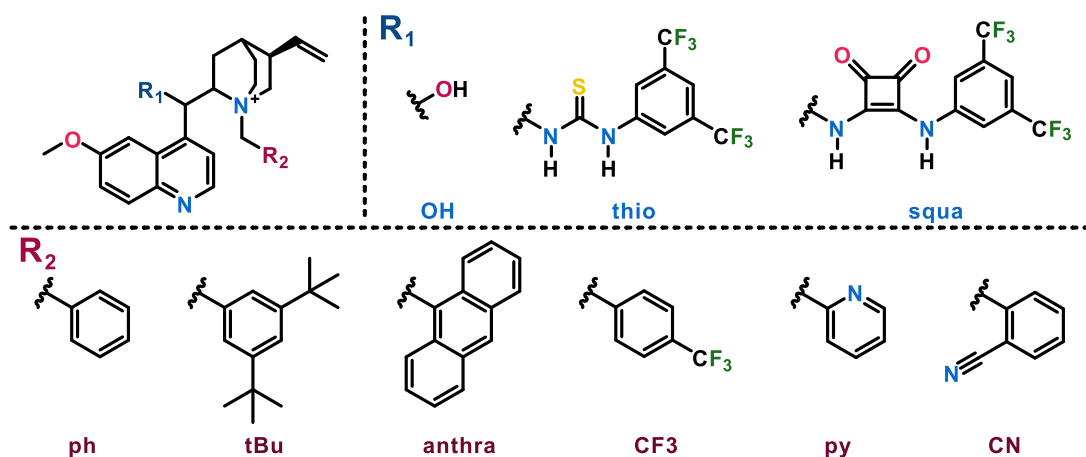


Figure 1.1. Chiral bifunctional PTC-catalyst scaffold and modifications under study.

As it was aforementioned, asymmetric PTC-catalysts are defined as small and chiral organic molecules used to induce enantioselectivity in a chemical reaction. These catalysts present well-defined chiral motifs to which a myriad of organic groups can be attached. Among the different PTC-catalyst classes developed, chiral binaphthyl and Cinchona alkaloids have been proven as the most successful examples,³⁻⁹ including the quaternary ammonium salts, and have been successfully applied to highly enantioselective conversions.

For this particular benchmark, three different HB-donors have been chosen to cover the most commonly used cinchona alkaloid-based derivatives (Figure 1.1).

Those particular phase-transfer catalysts chosen for this benchmark are well-known examples within the traditional hydrogen bonding-based catalysis field, and the binding mode via a bifurcated hydrogen bond has been confirmed in the literature (Figure 1.2).¹⁰ In particular, thiourea and squaramide units have been considered very efficient bifunctional catalysts for several important enantioselective organic transformations.^{11–17}

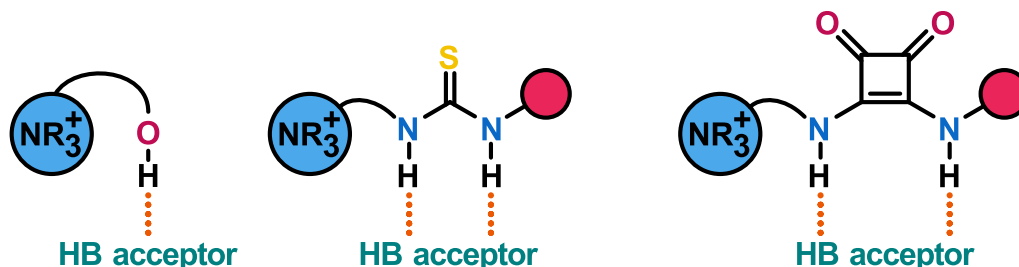


Figure 1.2. Proposed binding mods for chiral bifunctional phase transfer catalysts.

However, both squaramide (**squa**) and thiourea (**thio**) scaffolds can adopt *anti-anti*, *anti-syn* and *syn-syn* conformations. It is evident that the conformations adopted by thiourea and squaramide functional groups influence both the catalysis and binding (Figure 1.3 top).

The conformational behaviour of squaramides and thioureas has been studied in the past by Luchini et al.¹⁸ and by our group¹⁹ under different conditions. Both studies concluded that the most preferred conformer is heavily influenced by the polarity of the solvent (Figure 1.3 down and S1.1). Polar solvents (e.g. H_2O , $\epsilon = 78.4$) favour both *anti-anti* and *syn-syn* conformations, while more apolar solvents (e.g. THF , $\epsilon = 7.4$) present almost a non-existing *syn-syn* conformation and *anti-anti*, is the most populated one.

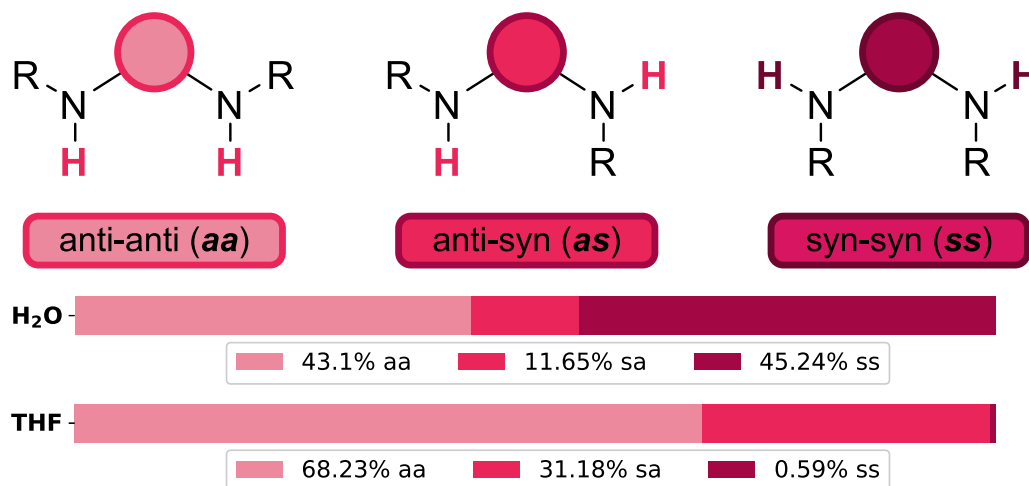


Figure 1.3. Top: Definition of the three possible conformations for the thiourea and squaramide substituted catalysts: *anti-anti*, left, *anti-syn*, middle and *syn-syn*, right. Down: Population of the possible conformations of thiourea and squaramide substituted catalysts when $R = EWG$.

Independently of these results, the presence of the anionic reactant to which the catalysts will be bonded in any PTC reaction will force the catalyst to be in an *anti-anti* conformation, as the accepted mechanism of Schreiner's catalyst states.²⁰ Therefore, it is very important to maintain that *anti-anti* conformation during the conformational analysis process since it might

happen that the most stable conformer able to establish this bifurcated HB is not structurally close to the most stable when the *anti-syn* and *syn-syn* conformations are allowed.

Given the inherent conformational problem featured within the different thiourea and squaramide substituted catalysts and the impact on the possible binding modes, two different explorations (**unconstrained** and **constrained**) of the conformational space within PTC-catalysts by means of computational method benchmarking is presented in this study.



Figure 1.4. Software packages under study.

A huge number of conformer generators based on different algorithmic approaches are available nowadays. However, for this particular study, only the most used conformational analysis software packages (Figure 1.4) were chosen to characterise the different cinchona alkaloid-based PTC catalysts: **Balloon**,^{21,22} **RDKit**,²³ **wSterimol**^{24,25} and **CREST**.^{26–28}

1.2 Computational Details

In order to use the same structure for every catalyst as an input for the different software under study, standardising the initial structure, all the catalyst structures were generated with the corresponding SMILES code (Table S1.1) and were pre-optimised using the MMFF94²⁹ FF, implemented using RDKit²³ library for Python3.³⁰ The script used for generating these 3D structures can be found in the supporting information (Scripts S1 - *catalyst_generation*).

The general workflow of each conformational analysis software package under study can be seen in Figure 1.5 and the details of each section are specified below.

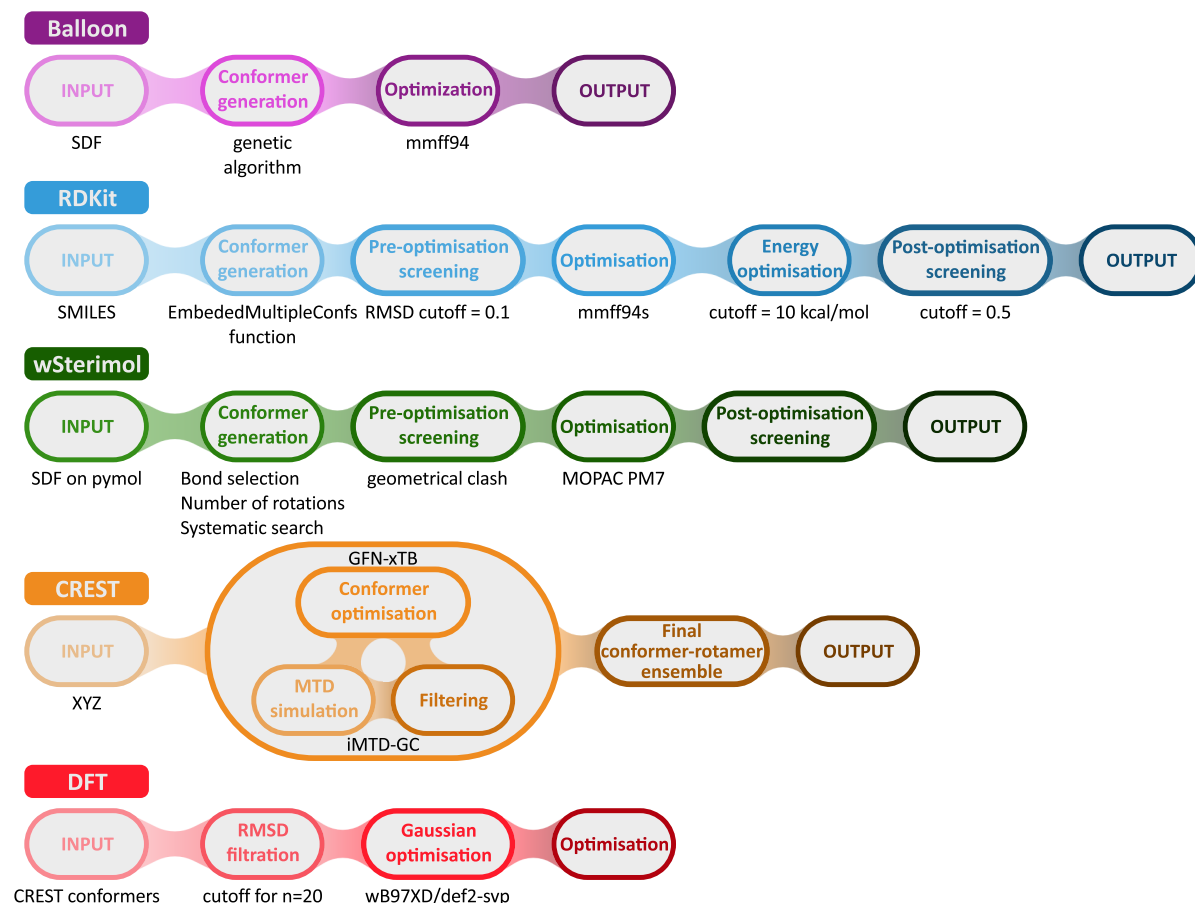


Figure 1.5. Workflow for the different software used in this study.

Balloon

The pre-optimised structure with an Structure-Data File (SDF) format was used as input. The different conformers were generated with the default parameters of the Genetic Algorithm (GA) and optimised using the MMFF94 FF. The default cutoffs were used for the different filtration processes. All the calculations were carried out with the 1.8.0, Apr 10 2021, package version of the software.

RDKit

The previously used generation script was integrated within the conformational analysis script, and the SMILES of each catalyst were used as input. The conformers were generated us-

ing the *EmbedMultipleConfs* function. This function uses distance geometry to obtain multiple sets of coordinates for a molecule by generating random numbers. The number of conformers was set to the cube of the number of rotatable bonds ($n_conformers = n_{rotatable}^3$), the pre-optimisation Root Mean Square Deviation (RMSD) filtration cutoff was set to 0.1 Å (*pruneRmsdThresh* = 0.1) and the options *useExpTorsionAnglePrefs* and *useBasicKnowledge* were activated to generate more realistic conformers by imposing experimental torsion angle preferences and basic chemistry knowledge (e.g., flat rings). A variation of MMFF94 FF available in RDKit, MMFF94s, was used for the optimisation of the different conformers. This modified FF demonstrates superior performance when planar hybridised nitrogen atoms are present in the system, such as those found in the quinoline ring, squaramide, thiourea, and similar compounds. Post-optimisation energy and RMSD was applied to the optimised conformers to avoid redundancies in the structures and were set to $E_{cutoff} = 10 \text{ kcal/mol}$ for unconstrained systems and $E_{cutoff} = 20 \text{ kcal/mol}$ for constrained systems and $RMSD_{cutoff} = 0.5$. All calculations were performed using the 2021.03.1 version of the RDKit package. The script used for the conformational analysis can be found in the supporting information (Scripts S1 - *clusterisation_UMAP*).

wSterimol

PyMOL³¹ software was used as a graphical interface to use wSterimol and the pre-optimised structure with an SDF format was used as input. wSterimol conducts a systematic conformational search across all the selected rotatable bonds. The rotatable bonds were manually selected and the number of rotations per bond was set to 3 (*ANGLE_COUNT* = 3). Those single bonds attached to a group with C_3 rotational symmetry were excluded from the selection (e.g. CH_3 , CF_3 , tBu) to avoid redundant calculations. The generated conformers were optimised at the PM7^{32,33} semiempirical level using the MOPAC³⁴ package. The parameters used for the input generation and the post-optimisation filtration were:

- *RMSD_CLUSTER_OPT* = 0.5 Å
- *ATOMIC_MODEL* = *bondi*
- *RJCT* = 0.5 (maximum overlapping radius of an atom (between 0 and 1))
- *TEMPERATURE* = 298 K
- *ENERGYWINDOW_CUTOFF* = 6.0 kcal/mol
- *PRINT_CUTOFF* = 6.0 kcal/mol

All the calculations were performed with the 1.0.3 released version²⁴ of wSterimol and version 21.138L of MOPAC2016.

CREST

CREST package is a utility that works together with the xTB³⁵ program. The conformer search is performed using the iMTD-GC algorithm, a workflow that uses a combination of RMSD based metadynamics simulations and semiempirical calculations using GFN n -xTB³⁶ (GFN2-xTB) to explore the conformational space of a given molecule and optimise those conformers. The energy limit for the metadynamics simulation was set to 6.0 kcal/mol, the RMSD cutoff

to 0.125 Å and the energy threshold between conformer pairs to 0.05 kcal/mol. Both RMSD and energy threshold parameters were very conservative for large systems such as the catalysts studied here but also ensured that none of the conformers was lost in the process. Since these calculations are based on metadynamics simulations, it is guaranteed that all the conformers generated will be physically accessible from the starting point as long as the barrier to access them is lower than 6.0 kcal/mol. All the calculations were performed using version 6.2 of the CREST package.

To evaluate the results of the different packages under study, we require reference structures and energies. Consequently, higher-level DFT calculations are essential. Given that DFT calculations are considerably more computationally expensive than any of the calculations above, it is impractical to compare the more than 12k generated conformers at this level of theory. Thus, we will only examine a subset of these conformers at this level.

Upon initial examination, it becomes evident that the software capable of the most exhaustive and complete exploration of the conformational space is CREST and, therefore, in order to avoid any biased selection, an RMSD based filtration of the conformers generated by this program is performed. This RMSD calculations were performed using *calculate_rmsd*³⁷ package. The Hungarian algorithm³⁸ was used as the atomic reordering method and the Kabsch algorithm³⁹ was used as the system rotation method.

The RMSD value for all the possible pairs of conformers of the same catalyst is calculated, obtaining an $N \times N$ matrix, where N is the number of conformers of the studied catalyst. Once the matrix is computed, a threshold is set to select only those conformers whose RMSD value is above it, maximising the structural difference between the remaining conformers. The value of this cutoff is selected so the number of remaining conformers after the filtration is reasonable for performing DFT calculations (between 15 and 22). The script used for this filtration can be found in the supporting information (Scripts S1 - *rmsd_filtration*).

DFT

The structures used as a reference for evaluation were optimised at the wb97xd⁴⁰ /def2-svp⁴¹ computational level in gas-phase and room temperature. Harmonic vibrational frequencies were computed at the same level used for the geometry optimisations to confirm the nature of the stationary points as local minima. Calculations were performed using the Gaussian 16⁴² software.

The atomic coordinates for all the generated conformers are provided, in XYZ format, and can be found in the supporting information. Table S1.2 for the unconstrained conformational analysis and Table S1.3 for constrained conformational analysis.

Clustering

The dimension reduction process was performed using the Uniform Manifold Approximation and Projection (UMAP)⁴³ library. Both the number of neighbours (*n_neighbours*) and the minimum distance (*min_distance*) were adjusted manually to obtain the optimal result for each specific set of conformers. The number of components was set to 2 (*n_components* = 2) in all the cases and the random state was set to 42 (*random_state* = 42).

The identification of the individual clusters in the 2D dataset was performed using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)⁴⁴ algorithm

and the values for the minimum number of samples (*min_samples*) and the minimum cluster size (*min_cluster_size*) were tweaked manually in order to minimise the number of points not assigned to any cluster. The corresponding values for each unconstrained and constrained system can be found in Table S1.4 and S1.5 respectively.

1.3 Evaluation criteria

In order to assess the different programs under study and evaluate their performance in the most objective and unbiased manner possible, five different evaluation criteria were established. Each of these criteria provides useful information about different aspects of each software. Depending on the requirements of the user, certain aspects may carry varying degrees of importance, and assessing them individually aids in selecting the most suitable option. These developed criteria encompass the prediction capacity, the tunability, the space exploration, the structure accuracy and the energy accuracy and are explained in detail below.

Prediction capacity

The first criterion, *prediction capacity* (Eq. 1), is defined as the ratio between the number of systems in which the most stable conformer at the software level (e.g. MMFF94s for RDKit) and the most stable conformer at the DFT reference level match ($n_{correct}$) and the number studied systems ($N = 18$).

$$Prediction\ capacity = \frac{n_{correct}}{N} \quad (1)$$

This criterion is especially important since the aim of most conformational analyses is to obtain the most stable conformer of a molecule. This is essential to ensure that the system or reaction under study is performed with the appropriate conformer.

Tunability

The second criterion accounts for the flexibility of each software and the number of different parameters that can be modified. The *tunability* (Eq. 2) is defined as the ratio between the number of parameters that can be modified in the package under study ($n_{features}$) and the total number of considered features. The evaluated features, in no particular order, include: *optimisation method*, *conformer generation method*, *similarity cutoffs*, *charge of the system*, *addition of constraints*, *addition of solvent* and *temperature*. These were chosen to cover both technical features (e.g. *similarity cutoffs*, *conformer generation* or *optimisation methods*) and features that allow the user to describe the system more accurately (e.g. *charge* and *constraints*) or match different experimental conditions (e.g. *solvent* and *temperature*).

$$Tunability = \frac{n_{features}}{N} \quad (2)$$

This criterion would be of special interest for those studies in which the user needs much control over the conditions of the analysis, either because the system is very complex and only certain parts of it should be allowed to vary, because a very specific generation or optimisation method must be used, or because the study must some reproduce experimental results.

Space exploration

The next criterion to evaluate is the *space exploration* (Eq. 3). This criterion evaluates the proportion of the conformational space explored by each package in the analysis and is defined as the average ratio between the number of individual conformers ($n_{clusters}$) generated by the software

under study and the number of individual clusters that make up the complete conformational space of the system (n_{total}) among all systems studied ($N = 18$).

$$Space\ exploration = \frac{\sum_{i=1}^N \frac{n_{clusters}^i}{n_{total}^i}}{N} \quad (3)$$

Since the conformational space of a system, depending on its size and flexibility, can be immense and the number of accessible conformers impossible to define, it was assumed that the complete conformational space of the system under study is the combination of all the conformers generated by the four software under study. In other words, it is assumed that there is no area of this conformational space left unexplored by at least one of these software packages.

The clusterisation process to identify the individual conformers that form the conformational space of the system was performed using two very well-known algorithms in ML: UMAP for the dimension reduction process in combination with HDBSCAN to identify and separate individual clusters present in the 2D dataset.

The data used for the clusterisation process were the $N_{conformers} \times N_{conformers}$ RMSD matrix formed by the RMSD values of each possible pair of conformers among all conformers generated by the four software for each system.

RMSD (Eq. 4) is one of the most frequently employed quantitative methods to measure the similarity between chemical structures. It calculates the atomic distances between the two structures under comparison, atom by atom, d_i^{ab} , and computes their Root Mean Square Error (RMSE). Therefore, similar structures will present values closer to zero because all the distances between equivalent atoms, x , in molecule A (mol_a) and molecule B (mol_b) would be very small.

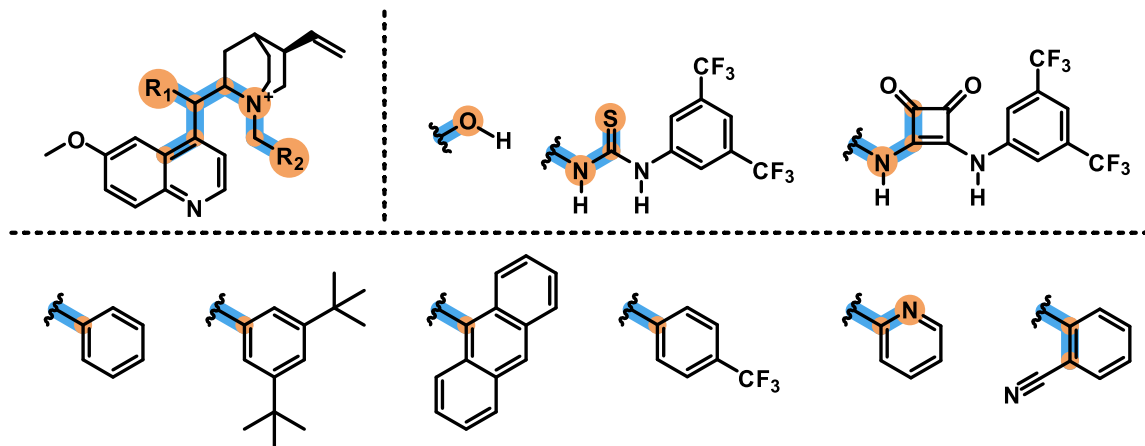


Figure 1.6. Representative atoms and bonds of the scaffold of the catalysts under study and the different substituents used in the clusterisation process.

$$RMSD(mol_a, mol_b) = \sqrt{\frac{\sum_{i=1}^{n_{atoms}} (d_i^{ab})^2}{n_{atoms}}} \quad (4)$$

In order to avoid local symmetry problems (e.g., benzene 180° rotation) and longer computational times, only the most representative atomic coordinates of the catalyst and its different substituents, highlighted in orange in Figure 1.6, were considered for the RMSD calculations. Those selected atoms preserve the direction of the substituents of the catalysts and ensure that the different conformers are distinguishable. In addition to the RMSD matrix and aiming to

refine the information, the value dihedral angles of the flexible single bonds among the selected atoms, blue in Figure 1.6, were tabulated.

Structure and Energy accuracy

The next two criteria evaluate the accuracy of the generated structures and the associated energies between each individual software and the reference level DFT. Due to the enormous amount of demanding DFT calculations needed for this criteria, only 10 conformers for each family (**OH**, **squa** and **thio**) of one particular catalyst (**ph**) were calculated.

The *structure accuracy* (Eq. 5) is defined as the average similarity between all the conformer generated by the evaluated software (cat_i^{method}) and the same conformers reoptimised at the reference DFT level (cat_i^{DFT}).

$$Structure\ accuracy \propto \frac{\sum_{i=1}^{n_{conformers}} RMSD\left(XYZ_{cat_i^{DFT}}, XYZ_{cat_i^{method}}\right)}{n_{conformers}} \quad (5)$$

As it was aforementioned, similarity and RMSD are inversely proportional; meaning that, in order to homogenise the criteria and obtain numerical results that range from 0 to 10, being 0 the worst and 10 the best performance, the results obtained from these calculations need to be transformed. The upper limit (*structure_accuracy* = 10) was set to perfectly matching structures ($RMSD = 0$) and the lower limit (*structure_accuracy* = 0 Å) was set to the highest RMSD value ($RMSD = 4.7$ Å) obtained amongst all the pairs of conformers compared (12063 pairs).

Finally, the *energy accuracy* (Eq. 6) was computed as the RMSE between the calculated energies for each conformer computed with the package under study and the reference energies at the DFT level).

$$Energy\ accuracy \propto RMSE\left(E_{cat_i^{DFT}}, E_{cat_i^{method}}\right) \quad (6)$$

Similarly to the previous criterion, it was necessary to convert the *energy accuracy* into a more easily interpretable value for grading purposes. Therefore, the same process was followed and the best grade (10) was set to perfectly matching energies ($dE = 0$ kcal/mol) and the worst grade (0) was set to the biggest energy difference ($dE = 17.34$ kcal/mol) amongst all the evaluated energies (120).

The combination of these five criteria allowed a broad evaluation of the packages under study. Furthermore, it offered diverse perspectives, providing useful information that would help in selecting the most appropriate option based on the specific needs of each study.

1.4 Results and Discussion

1.4.1 Unconstrained Conformational Analysis

The initial step of our investigation involved performing an unconstrained conformational analysis of the catalysts under examination, which amounted to a total of 18 unique catalysts (see Figure 1.1). This comprehensive analysis was executed using four distinct conformational analysis packages, along with the application of DFT calculations on individual conformers. These conformers were selected following a rigorous filtration and clusterisation process based on the RMSD measurements, which were applied to the ensembles of conformers generated through the CREST methodology. Details of all these computations can be found in the computational methods section.

After generating all the different conformational ensembles, the first aspect to analyse was the number of conformers that form them (Table 1.1). This measurement is not meaningful for evaluating the performance of the software since it does not take into account the quality of the ensemble or the similarity between the possible conformer pairs; therefore it has not been included in the evaluation criteria and *space exploration* criterion was used instead. Nevertheless, assuming that the calculation time for each conformer is similar among all the semiempirical and FF methods, it can be used as an estimated measure of the speed of each software.

This approximate value is required due to the impossibility of measuring the execution time of each software accurately. To precisely benchmark the execution time of a software package, all calculations must be performed on the exact same processor. Unfortunately, that was not possible, since some of the packages could not be installed in the computational centre and had to be installed locally.

Table 1.1. Number of conformers for each catalyst studied with all the different conformational analysis software.

Complex	Balloon	RDKit	wSterimol	CREST	DFT
cat_OH_ph	47	20	25	164	21
cat_OH_tbu	97	54	29	495	22
cat_OH_anthra	57	30	24	156	21
cat_OH_cf3	57	31	48	246	17
cat_OH_py	58	24	45	144	16
cat_OH_CN	63	43	37	185	17
cat_thio_ph	82	89	272	336	17
cat_thio_tbu	77	114	250	500	19
cat_thio_anthra	61	105	136	430	17
cat_thio_cf3	54	101	218	280	16
cat_thio_py	62	84	334	131	17
cat_thio_CN	40	92	162	432	17
cat_squa_ph	66	121	210	105	15
cat_squa_tbu	38	157	194	258	18
cat_squa_anthra	26	110	95	209	17
cat_squa_cf3	59	163	189	171	15
cat_squa_py	48	28	192	264	18
cat_squa_CN	78	143	477	253	19

As a general trend, it can be seen that two packages (**CREST** and **wSterimol**) produce

much larger conformer ensembles, while the other two (**Balloon** and **RDKit**) generated smaller ensembles. These results are not unexpected, since **CREST** uses metadynamics to generate the conformers, facilitating the appearance of intermediate structures between the conformers and the cutoffs selected were very conservative. Alternately, **wSterimol** uses a systematic search for conformers, rotating all the indicated bonds a certain number of times, generating a large number of conformers. These packages also use semiempirical methods (GFN2-xTB and PM7, respectively) to optimise the generated structures. On the other hand, both **Balloon** and **RDKit** use random numbers to generate the different conformers and, instead of using semiempirical methods, they use FFs (MMFF94 and MMFF94s respectively), giving as a result smaller conformational ensembles.

A closer inspection of the different groups that were substituted in the scaffold (R_1 and R_2) revealed that there is not a general trend in the number of conformers generated by each software when changing the HB donor ($R_1 = OH$, *thio* and *squa*). However, when focusing on the other substitution point (R_2), $R_2 = tBu$ exhibits the largest conformer ensemble sizes in most of the catalysts under study. These larger ensembles are in agreement with the fact that *tBu* is the most flexible substituent used. However, it can be expected that many of those conformers are just rotamers and not independent conformers since catalysts with *tBu* substituents present C_3 local symmetries, impossible to identify by the filtration algorithms. For instance, **wSterimol** does not present those big differences between other ensembles and *tBu* because the rotatable bonds have to be selected and those bonds that present this kind of symmetry were purposely avoided.

Space exploration

If the conformer ensembles generated by the different packages under study for a particular case (e.g., **cat_squa_anthra**) are plotted (Figure 1.7), it can be seen that both **CREST** and **wSterimol** produce much more dense conformational ensembles, showing a more exhaustive, and probably more redundant, conformer search. But, as has been aforementioned, the total number of conformers forming a particular ensemble does not provide any information about how well the conformational space was explored. Therefore, the more detailed criterion *space exploration* (Eq. 3) is used.

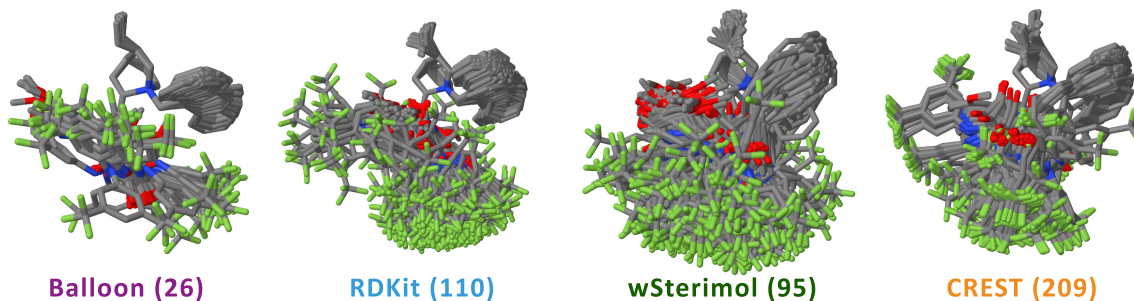


Figure 1.7. Conformational ensembles generated by all methods under study for **cat_squa_anthra**. Balloon (26), RDKit (110), wSterimol (95) and CREST (209).

To evaluate the *space exploration*, a cluster analysis was carried out using UMAP and HDBSCAN to automate the cluster identification process, as explained in the computational methods section. Ideally, a conformational analysis program should be able to explore the whole conformational space and generate as many individual clusters (different conformers) as

possible. To obtain a global picture, the number of conformers, the information obtained from the clusters, and all the different statistical data are displayed in Figure 1.8.

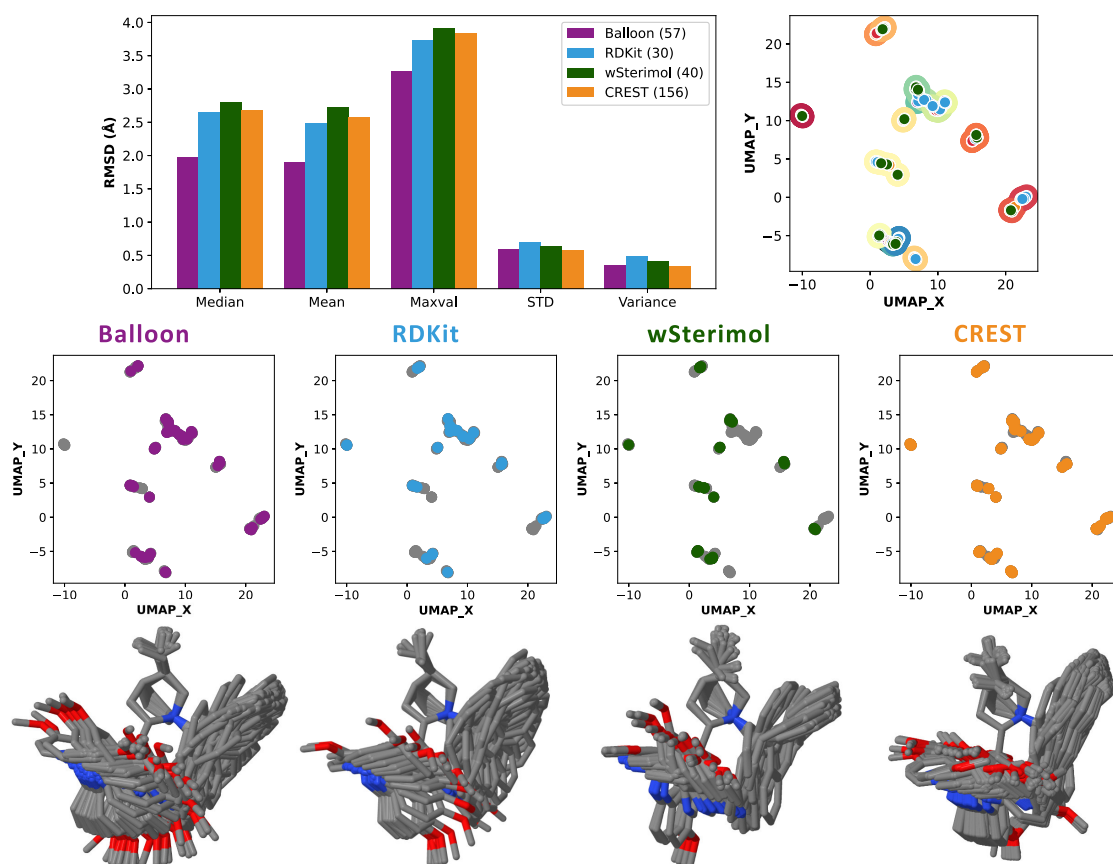


Figure 1.8. Statistical data for the of the conformational space explored by all the packages under study for **cat_OH_anthra**, top left; clusters generated by all the programs, top right; clusters divided by packages, middle; and 3D ensembles generated by each software package, bottom.

From Figure 1.8, it is clear that a very similar conformational space was explored by most of the programs. A similar distribution of points appears when each program's results are scattered individually (Figure 1.8 middle). Analysing each individual plot in detail and comparing them with the statistical values, **Balloon**, with 57 generated conformers, explores only the conformational space corresponding to the right part of the reduced dimension UMAP plots. It is discernible from the **Balloon** 3D-plot that the majority of the points correspond to two large groups of conformers, where the quinoline ring is either facing up or down. This uneven distribution and lack of exploration explain the low statistical values that **Balloon** presents among all the different statistical measurements.

On the other hand, **RDKit** and **wSterimol** with about half conformers (30 and 24, respectively) explore the same space as **Balloon** and also explore the extreme left region that did not appear previously. This smaller number of conformers and wider exploration of the conformational space are translated into higher statistical values for all the measurements, as highlighted in the graph. **RDKit** shows a more uneven distribution of the generated conformers, presenting a large concentration in the middle area, explaining the smaller statistics compared to **wSterimol**. It is reasonable to conclude that **wSterimol** is the one exploring best the conformational space and obtaining the biggest statistical values because of the systematic search implying the

rotation of all the flexible bonds.

Finally, **CREST** performs the most exhaustive exploration of the conformational space and generates the largest number of conformers (156). Despite generating so many conformers, they are distributed evenly over all the conformational space and, as a result, the statistical values obtained from such a varied set of conformers are still relatively high.

To illustrate the output from the clustering analysis in detail, **cat_OH_anthra** was chosen as a significant example, but the rest of the catalysts are featured in the Supporting Information (Figures S1.2–S1.5 and Tables S1.6 and S1.7).

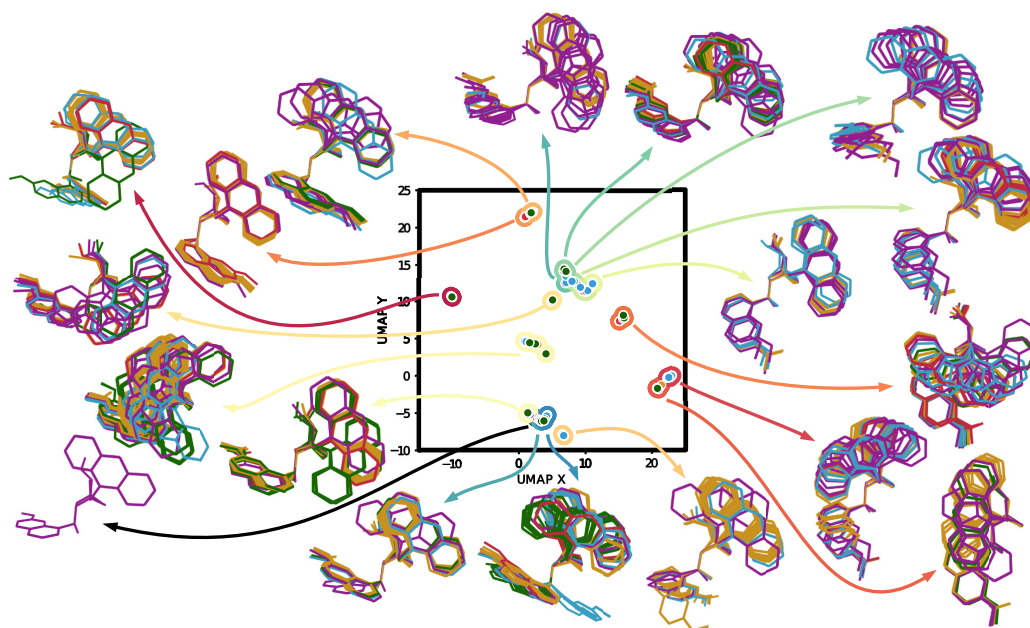


Figure 1.9. 3D-plot of the different conformers forming every cluster generated by the packages under study for **cat_OH_anthra**.

To further analyse this clusterisation process, the 3D structures of the conformers forming each cluster were extracted and superimposed (Figure 1.9), revealing the different individual conformers obtained. As a general trend, the conformers are very well collected within each cluster. Since cluster identification is a complex process, performed by HDBSCAN, some clusters would not be the same as we, as humans, would recognise and manually select. But, being a ML process, it is consistent and unbiased independently of the dataset that it is fed with.

It can happen that what a human eye would consider as a single cluster (e.g. the two light orange clusters on top of the graph or the light greenish blue clusters in the middle) appears as two independent clusters. These kinds of situations in which two clusters look the same but are mathematically different are why it is so important to use unbiased processes, such as ML to evaluate the efficiency of each software.

Another problem that can arise from this process is the appearance of outlier conformers that are not assigned to any cluster (e.g. the black dot in Figure 1.9). The parameters for the clusterisation, as explained in the computational details section, were selected to minimise these problematic conformers and keep their numbers as low as possible.

Despite these problems that can arise with this intricate analysis, HDBSCAN is a trustworthy and very widely used method to separate all conformers in individual clusters.

This cluster information is vital for the evaluation of each software since, as it was afore-

mentioned, it provides the information needed to calculate the *space exploration* criterion.

The clusters represented in Figure 1.9 can be represented as a Gantt chart to visualise how the generated conformers are distributed within the different clusters, how many conformers from each cluster, and the proportion of conformers from each cluster generated by each program (Figure 1.10 and S1.6).

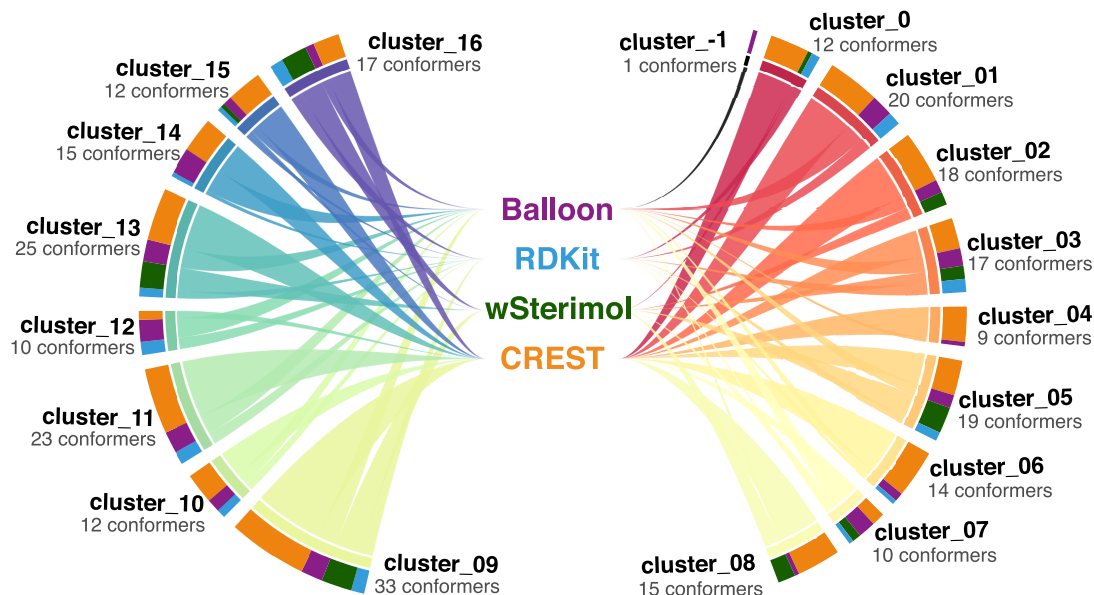


Figure 1.10. Gantt chart of the different clusters obtained for `cat_OH_anthra` and the individual software contributions to each of them.

From the analysis of the diagram generated for `cat_OH_anthra`, 17 clusters and only 1 outlier point were obtained. The biggest of those clusters is formed by 33 conformers (cluster_09) and the smallest is formed by 9 conformers (cluster_04). While cluster_09 presents conformers coming from all the studied packages, only **Balloon** and **CREST** contribute to cluster_04. As shown in the computational details, the number of clusters in which each software participates concerning the total number of clusters determines its performance when exploring the conformational space (Eq. 3). By analyzing this particular case, we can observe that **Balloon** participates in 16 of the 17 total clusters found for this system, **RDKit** explored 14, **wSterimol** 10 of them, and **CREST** 17.

Therefore, if the ratio between the explored and total conformers for this particular catalyst is calculated, the *space exploration* grade for each package would be 9.4 for **Balloon**, 8.2 for **RDKit**, 5.9 for **wSterimol** and 10 for **CREST**, which explored all individual conformers.

It is also important to take into account the number of conformers that each software needs to explore that certain percentage of the total conformational space. **Balloon** needed 57 conformers to generate those 16 individual conformers, meaning that about 72% of the generated conformers were redundant and not needed. The two most efficient packages were **RDKit** and **wSterimol** for which only 53% and 58% of the conformers were redundant. On the other hand, the most inefficient package for this particular catalyst was **CREST** for which 89% of the generated conformers were redundant.

However, the number of redundant conformers can be decreased easily by tweaking the different cutoffs that each software has for RMSD filtration without decreasing their space exploration. Due to this input dependence of this 'efficiency' measure, the total number of

conformers generated was not included in the definition of the *space exploration* criterion.

After measuring the *space exploration* of each software for each individual catalyst and averaging their values, the final grades obtained for each software are summarised in Figure 1.11, being **wSterimol** (7.01) the one that explored the conformational space the best, closely followed by **CREST** (6.97), **RDKit** (6.19) and, finally, **Balloon** (4.28).



Figure 1.11. *Space exploration* grades for each software calculated over all the studied catalysts.

Structure accuracy and Energy accuracy

The next aspect of the software to be analysed is its accuracy in optimising the generated conformers in terms of structure and energy. Due to the huge number of conformers generated ($n_{conformers} > 12k$), it is not feasible to optimise all of them at the DFT level and compare the difference between the optimised systems of each software and the re-optimised systems at the DFT level. Therefore, three catalysts were selected (**cat_OH_ph**, **cat_squa_ph**, and **cat_thio_ph**) and 10 different conformers were selected for each system by means of RMSD filtration, making sure that the conformers were as different as possible, and re-optimised at the DFT reference level.

Figure 1.12 and S1.7 display the relative energies predicted by each software for the conformer under study (light dots), the corresponding reference at the DFT-level (dark dots) and the structure difference between predicted and DFT-reference represented as the RMSD of their coordinates (bars).

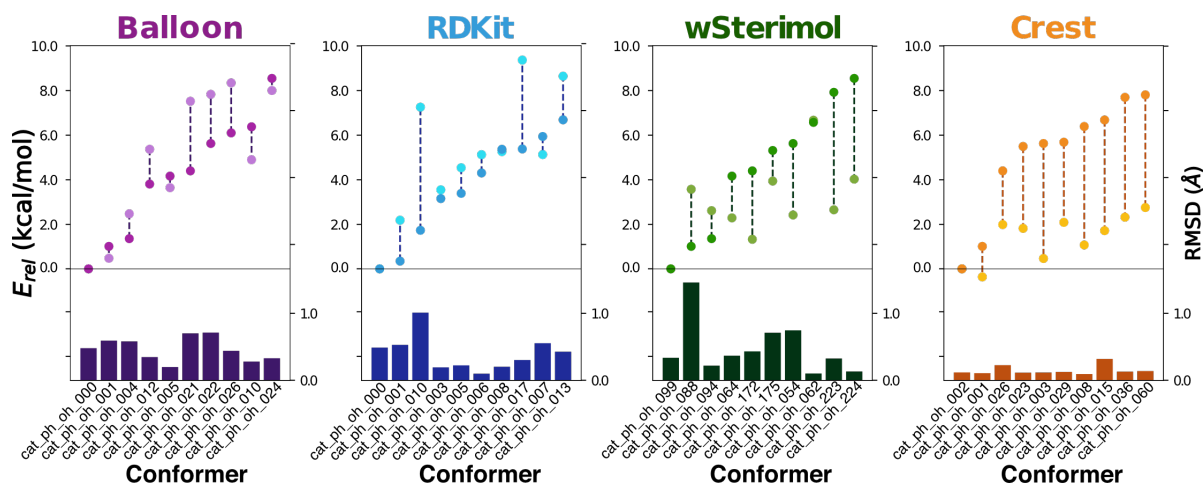


Figure 1.12. Energy and RMSD comparison between the studied conformers optimised with the different software packages with respect to the DFT-optimised structures for **cat_OH_ph** catalyst.

In terms of energy accuracy, **Balloon** and **RDKit** succeed in their predictions on average but they present significant differences in some particular cases. These big variations manifest that MMFF94 and MMFF94s FFs present some inconsistencies with some particular interactions

or substructures exhibited in those conformers. **CREST** present less accurate individual results for this particular case but is much more consistent with the trend presented, with the predicted energy always smaller than the reference and more or less consistent errors. Finally, **wSterimol** shows mixed results for this particular catalyst.

If the results for the three catalysts are analysed and the *energy accuracy* calculated (Figure 1.13), **CREST** presents the best results (7.3), followed by **wSterimol** (6.5), **RDKit** (5.4) and, finally, **Balloon** (4.5).



Figure 1.13. *Energy accuracy* grades for each software calculated over the three selected catalysts.

These results manifest that the modified MMFF94s FF from **RDKit** performs better for these systems than the default MMFF94 version used in **Balloon**. It is also clear that semiempirical methods are more accurate in predicting the energy of these catalysts than the studied FFs. It is worth mentioning that **CREST**, in general, overestimates or favours conformations where π - π interactions are established, over the rest of the conformers, despite getting the best results in terms of energy accuracy.

In terms of structural accuracy, **CREST** performs exceptionally nicely with the OH substituent, showing extremely small and consistent RMSD values. **CREST** also displays the most consistent results among the different catalysts, predicting all the structures with similar accuracy and presenting the smallest differences between the best and the worst prediction.

When analysing the three catalysts numerically (Figure 1.14) in all conformers studied, the results were overall adequate, with **CREST** being the most accurate when optimising the structure of each conformer (8.6) followed closely by **RDKit** (8.2), **Balloon** (7.7), and **wSterimol** (7.7). Generally, the **CREST** optimization method is in very good agreement with DFT-optimized systems.



Figure 1.14. *Structure accuracy* grades for each software calculated over the three selected catalysts.

Prediction capacity

Looking at the most stable conformation obtained from each method (Figure S1.8), it can be concluded that, as a general trend, the disagreement among the different software and DFT optimisations is, unfortunately, clear and critical, as it can be seen in **cat_thio_tBu** (Figure 1.15,

right). However, for $R_1 = OH$, since the scaffold is simpler, a better approximation was obtained, and the most stable conformers present a more similar conformation to the one obtained utilising DFT (**cat_OH_py**, Figure 1.15, left). From Figure 1.15 it is worth mentioning that **CREST** (**CREST** in orange, left), in general, overestimates or favours conformations where π - π interactions are established, over the rest of the conformers.

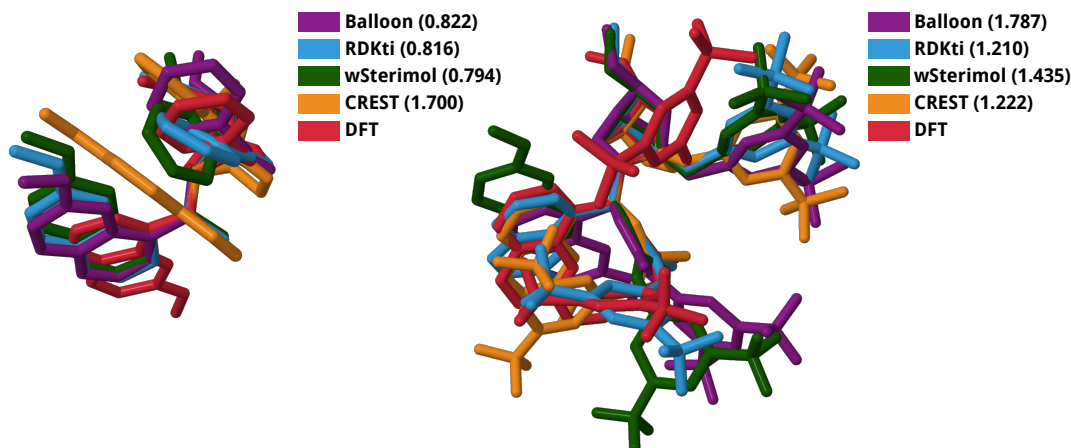


Figure 1.15. Most stable conformations for **cat_OH_py** (left) and **cat_thio_tBu** (right) for each method. Shown in parenthesis are the RMSD values of all the structures with respect to the DFT one, using the same representative atoms used in the clusterisation process.

To evaluate this disagreement between the predicted and the reference most stable conformers, the *prediction capacity* (Eq. 1) criterion was analysed, accounting for how well each software package predicts the most stable conformation of each of the 18 catalysts studied. The *prediction capacity* was calculated by comparing the most stable conformation of each individual catalyst, generated with each method, with the most stable reference conformation at the DFT level.

By examining Table 1.2, it is clear that some particular structures are much easier to predict than others. Catalysts with $R_1 = OH$ are, in general, easier to predict the most stable conformer. This outcome is expected, given that the *OH* group is much smaller than the *squa* and *thio* groups and can be involved in many fewer intermolecular interactions than the larger substituents. On the other hand, catalysts with *squa* and *thio* groups are much more flexible and establish a myriad of attractive NCIs that make these predictions more intricate. In fact, the conformational analysis performed on **cat_OH_R2** catalysts was able to successfully predict the most stable conformer on 15 occasions out of the 24 total calculations performed with that scaffold, while only 2 and 7 most stable conformers for **cat_squa_R2** and 7 **cat_thi_R2** were appropriately anticipated by the different software packages.

Focussing on R_2 , the results are also very uneven between all the different combinations, with *Ph*, *Py*, and *tBu* being the substituents presenting less accurate predictions among all the catalysts studied. This can be explained again by the fact that those catalysts establish some attractive NCIs (e.g. $\pi - \pi$) that are not well described by the FFs or semiempirical methods that the different software packages use. Meanwhile, in the substituted catalysts *anthra*, *CF3* and *CN*, the sterical issues are much more important than those NCIs in the most stable conformation.

After evaluating these results (Figure 1.16), it was obtained that **RDKit**, using the MMFF94s FF, is the software that predicts the biggest number of most stable structures correctly (8 cor-

rect predictions, 4.4 grade); followed by **CREST** (6 correct predictions, 3.3 grade), **Balloon** (5 correct predictions, 2.8 grade), and **wSterimol** (5 correct predictions, 2.8 grade).

Table 1.2. Results on prediction of the most stable conformation of a catalyst by each studied software.

Catalyst	Balloon	RDKit	wSterimol	CREST
cat_OH_anthra	✓	✓	✓	✓
cat_squa_anthra	✗	✓	✗	✗
cat_thio_anthra	✓	✓	✗	✗
cat_OH_CF3	✓	✓	✓	✓
cat_squa_CF3	✗	✗	✗	✗
cat_thio_CF3	✗	✓	✗	✓
cat_OH_CN	✓	✓	✓	✓
cat_squa_CN	✗	✓	✗	✗
cat_thio_CN	✗	✗	✓	✗
cat_OH_Ph	✓	✓	✗	✓
cat_squa_Ph	✗	✗	✗	✗
cat_thio_Ph	✗	✗	✓	✗
cat_OH_Py	✗	✗	✗	✗
cat_squa_Py	✗	✗	✗	✗
cat_thio_Py	✗	✗	✗	✓
cat_OH_tBu	✗	✗	✗	✗
cat_squa_tBu	✗	✗	✗	✗
cat_thio_tBu	✗	✗	✗	✗

The first conclusion that can be extracted from these results is that even though **CREST** and **wSterimol** obtained the best results for the *energy accuracy*, that accuracy does not correlate with the prediction of the most stable conformers of each catalyst. This discrepancy arises from the inaccurate description of the NCIs present in some of those most stable conformers, especially from the overestimation of π - π interactions between the different aromatic rings of the catalysts. The other conclusion that can be extracted is that the modified version of MMFF94 FF, MMFF94s, works much better for these kinds of systems in which planar nitrogen atoms can be found.



Figure 1.16. *Prediction capacity* grades for each software calculated over all the catalysts under study.

Tunability

The conformational analysis of a molecule can be dramatically affected by external conditions, the method used for generating the conformers, and the optimization method. Conformational analysis software should be as flexible as possible to study systems under different conditions. Also, some conformational analysis must be done under specific conditions to match the experimental results as closely as possible.

In order to evaluate the *tunability* (Eq. 2) of each software program, a list of requisites was established (see Evaluation Criteria section) to cover all the key aspects that might be important to make the software as flexible as possible, and each program was evaluated to account for how many of these features can be modified.

Table 1.3. Results on *tunability* the different software packages under study.

Software	Opt.	Gen	Cutoffs	Charge	Constrains	Solvent	Temp
Balloon	✓	✓	✓	✓	✗	✗	✗
RDKit	✓	✗	✓	✓	✓	✗	✗
wSterimol	✓	✗	✓	✓	✓	✗	✓
CREST	✓	✓	✓	✓	✓	✓	✓

The result of this analysis (Table 1.3) showed that some of the aspects studied are more commonly modified than others. Those aspects that allow modifying internal parameters of the conformational analysis process (e.g. optimisation method, filtration cutoffs, constraints) can be modified by most of the packages, providing a lot of flexibility to the workflow and allowing the user to tweak the parameters to modify the exhaustiveness of the analysis. The parameters that are used to describe the system more accurately (e.g. charge) can be modified independently of the package, ensuring the right description of the molecules under study. Finally, those parameters used to recreate experimental conditions (e.g. Solvent and Temperature) can be modified only in a few cases (both in the **CREST** package and only Temperature in **wSterimol**), therefore, if recreating experimental conditions is key for the system under study, this has to be taken into account.

The results of this analysis are collected in Figure 1.17 and show that **CREST** is the most flexible package (10), able to modify all parameters, followed by **wSterimol** (7.1) and finally, **RDKit** and **wSterimol** (5.7), being the least flexible packages.



Figure 1.17. *Tunability* grads for each software under study.

Boltzmann population

To finish the analysis of the different software under study, the statistical weight of the different conformations in terms of Boltzmann distribution was determined. This analysis explores how many of the generated conformers are statistically relevant and their Boltzmann distribution. The relative energy of the different conformers of **cat_squa_Py** and their Boltzmann population are shown in Figure 1.18.

This particular case, **cat_squa_py**, reflects the two extremely different scenarios that could occur when performing a conformational study. As it is shown in Figure 1.18, conformers generated by **Balloon** and **wSterimol** grow extremely fast at the beginning and, as a consequence, a small percentage of the generated conformers, less than 10%, represent more than 99% of the Boltzmann population (under 3 kcal/mol). However, **RDKit** and **CREST** exhibit a much

slower growth and need almost 50% of the total conformers to reach the same cumulative population. In addition, it is clear from the graphics *vide supra* that both **Balloon** and **wSterimol** reach higher energy values, and therefore, could be very appropriate if the aim of the conformational analysis under study is exploring conformations that appear only on that range of energies.

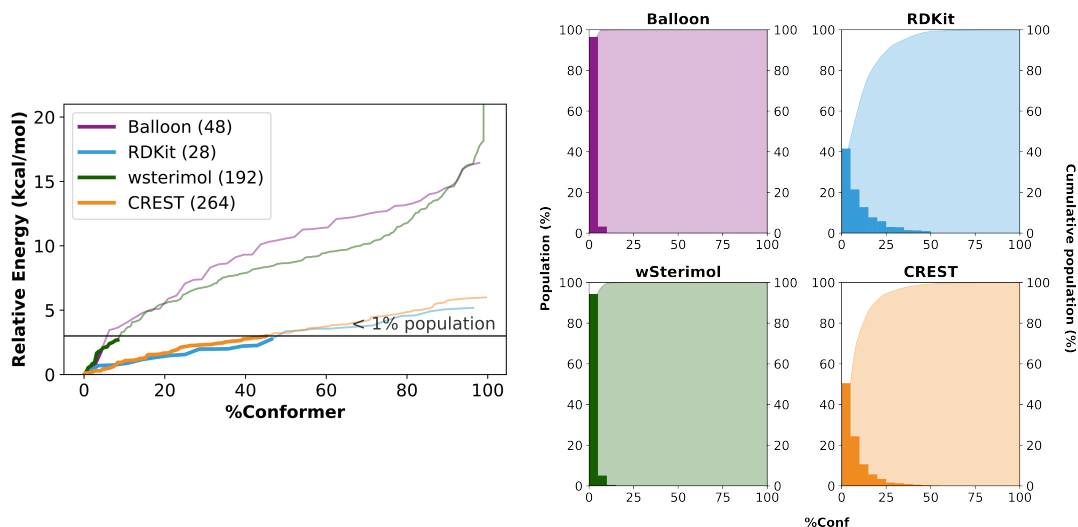


Figure 1.18. Relative energy of the generated conformers for **cat_squa_py** with all the studied software (left). The thick lines represent those conformers that accumulate more than 99% of the Boltzmann population. Shown in bars is the population of the generated conformers with respect to the %Conformers grouped in 5% intervals (right). The area below the curve, the cumulative population of the conformers, is generated with respect to the %Conformers.

This analysis was performed for all the catalysts under study (Figures S1.9 and S1.10) and lies between these two extreme cases. In general, most of the conformers explored are not statistically relevant in terms of the Boltzmann population, being more pronounced in the case of **Balloon** and **wSterimol**.

1.4.2 Constrained Conformational Analysis

It is important to highlight that if a nonconstrained conformational analysis is made with all the different software packages chosen for this benchmark, the anti-syn conformation is overall predominant for all thiourea and squaramide-based catalysts. Therefore, the outcome reached by regular conformational analysis is not the desired one, as only the *anti-anti* conformation is involved in the organocatalytic process.

A similar, less detailed, analysis of the efficiency of the software under study was performed for the constrained systems. For this analysis, only **CREST** and **RDKit** were taken into account since those are the only software in which the position of the catalyst can be properly constrained. Afterwards, a similar DFT study of the main conformations from **CREST** was performed for comparison (Figure 1.19).

To illustrate the output from the analysis, **cat_thio_py** was chosen as a significant example (Figure 1.20), and the rest are depicted in the Supporting Information (Figures S1.11–S1.14 and Tables S1.8 and S1.9). In order to obtain a global picture, the number of conformers, the outcome of the cluster information, and all the different statistical data are displayed in Figure 1.20.

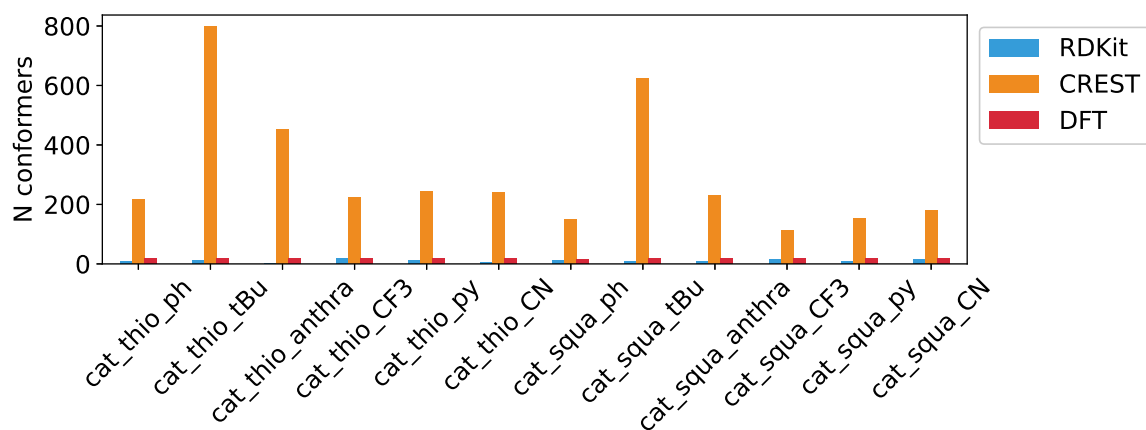


Figure 1.19. Number of conformers of each catalyst for each software under study and DFT.

Space exploration

From the clustering analysis, it is clear that **RDKit** and **CREST** explore a similar conformational space. At first glance, there is only one area (formed by several clusters and outliers) that **RDKit** has not explored compared to **CREST**. This similar exploration was performed using 10 times fewer conformers in the case of **RDKit** (24 **RDKit** and 244 **CREST**), making it much more efficient in terms of the exploration/number of conformers ratio. Regarding the statistical values, it can be seen that there is a big difference between both methods' max RMSD value, meaning that **CREST** has generated at least a certain pair of conformers much more different than **RDKit**. This big difference comes from the clusters that **RDKit** has not explored.

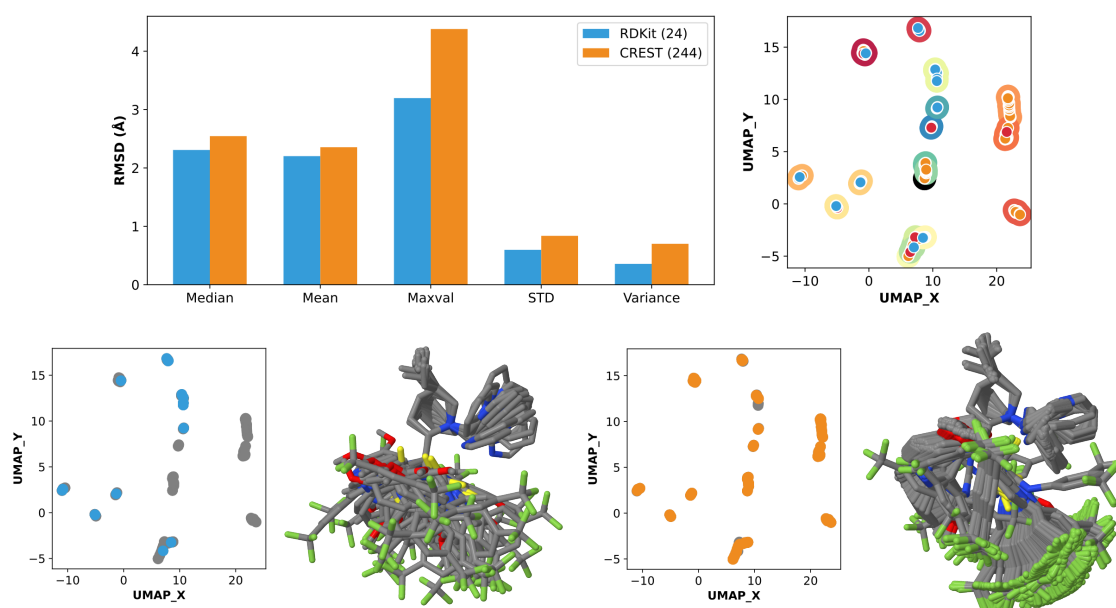


Figure 1.20. Statistical data of the constrained conformational space explored by all the packages under study for **cat_thio_py**, top left; clusters generated by all the programs, top right; clusters divided by packages and 3D ensembles generated by each software package, bottom.

Prediction capacity

If a comparison of the most stable structure from **RDKit**, **CREST**, and DFT is made (Figure S1.15), a better agreement is obtained than the one found with the unconstrained conformational analysis. It is clear from the figures that **CREST** again overestimates the interactions π - π , but in this particular constrained analysis, the main difference between the most stable structures arises from some rotations that can be accessible after DFT optimisations.

Boltzmann population

If a similar analysis of the statistical weights of the different conformations in terms of the Boltzmann distribution was performed for the constrained conformers, a comparable picture was obtained. The relative energy of the different conformers of **cat_squa_py** and their Boltzmann population are shown in Figure 1.21 and Figures S1.16 and S1.17 for the rest of the systems under study. **CREST** presents a more gradual growth, needing more than 50% of the total conformers to reach the same cumulative population that **RDKit** obtains with only 10% of the conformers searched. Overall, most of the conformers explored are not statistically relevant, being more pronounced in the case of **RDKit**.

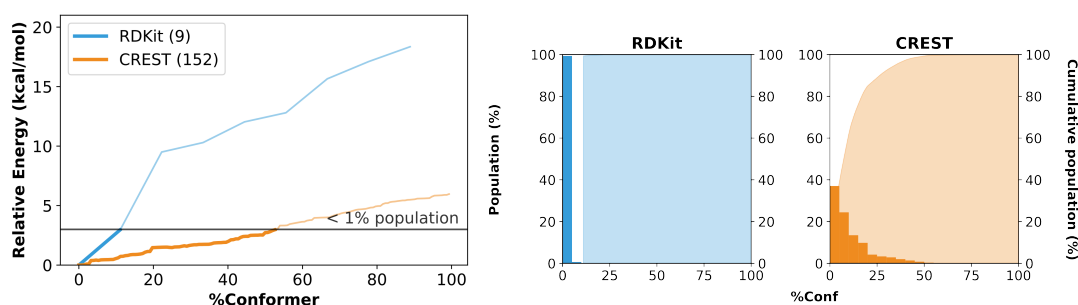


Figure 1.21. Relative energy of the generated constrained conformers for **cat_squa_py** with all the studied software (left). The thick lines represent those conformers that accumulate more than 99% of the Boltzmann population. Shown in bars is the population of the generated conformers with respect to the %Conformers grouped in 5% intervals (right). The area below the curve, the cumulative population of the conformers, is generated with respect to the %Conformers.

1.4.3 Final Grades

A summary of the grades regarding the different criteria used for each software is depicted in Figure 1.22. Overall, **CREST** is the one that obtained the best grades, standing out in its *tunability* (10), *structure accuracy* (8.6) and *energy accuracy* (7.3). However, this particular software does not perform as well when predicting the most stable conformer of a catalyst. In addition, **CREST** usually generates a huge number of conformers in order to explore the conformational space, defining it as a very exhaustive but not very efficient software. This second problem is very easy to fix by changing the different cutoffs that the package provides, decreasing dramatically the number of remaining structures.

On the other hand, **wSterimol** and **RDKit** present similar results, slightly lower for **RDKit**, especially concerning its *tunability* and *space exploration*. The advantage of **RDKit** is that it obtains very similar results in exploration with many fewer conformers than the other

two software packages, thus making its exploration process more efficient. A delicate point of **RDKit** is that being a library and not a software package, it is not as user-friendly as other software, especially for people who are not used to programming. Meanwhile, **wSterimol** is very systematic and accurate in *space exploration*, making it less prone to leaving areas without exploration. The main limiting factor for **wSterimol** would be that it is not designed specifically for conformational analysis and requires several software programs to execute the full workflow (PyMOL, MOPAC, and Gaussian). It also requires us to manually input the bonds that are to be explored. This feature presents a clear advantage, giving the user control over the exploring space but, on the other hand, complicates the automatisation and implementation of the process for different molecules.

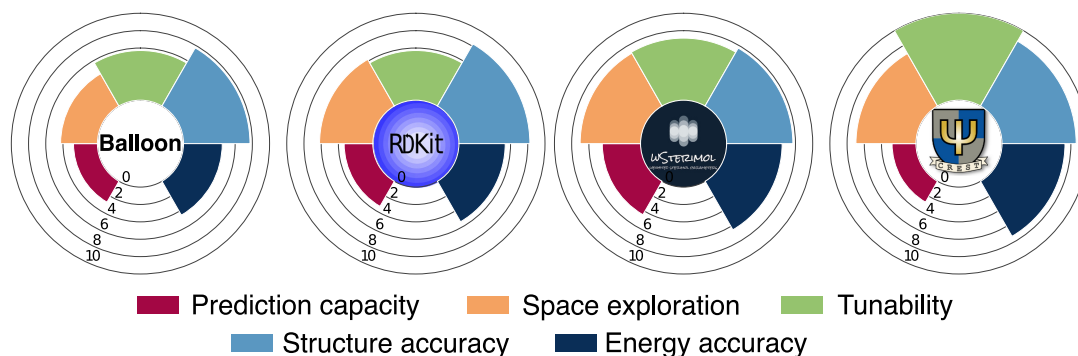


Figure 1.22. Global grades for each studied software for all the criteria analysed.

Finally, **Balloon** presents the lowest overall values (again, with the default configuration used) but it presents a very good agreement when focussing on *structural accuracy*. The strongest point of this software is its simplicity and the small number of resources that are needed to perform the analysis. This makes **Balloon** a very good option for generating conformers of a molecule that are going to be used as a starting point for re-optimisation at a higher level of theory.

1.5 Conclusions

In this study, a complete exploration of the conformational space of PTC-catalysts by means of computational method benchmarking is presented. For this particular study, only the most applied conformational analysis approaches have been chosen to characterize the different cinchona alkaloid-based PTC-catalysts.

This guiding benchmarking study aimed to rigorously compare the performance of different conformational methods, determining the strengths of each method and providing recommendations regarding suitable choices of methods for conformational analysis.

When an unconstrained conformational analysis was performed, the most stable conformers correspond to the *anti-syn* conformation for the thiourea and squaramide-based catalysts, providing a non-realistic scenario for the catalytic process to occur. Furthermore, if the most stable conformer from every package was compared with the one optimised by means of DFT methods, large disagreements are obtained. Among all the different approaches applied for this benchmarking study, **RDKit** provides the best balance between efficiency and computing time.

When the constrained scenario was considered, only *anti-anti* conformers were studied. **RDKit** provided again an equitable outcome regarding capability and performance.

1.5.1 Pros and Cons

Since the main goal of our study was to provide general knowledge of the different features of a variety of the most common free software for conformational analysis, a general list of pros and cons for each software package to facilitate the user the selection of the most adequate software is presented as well as general recommendations.

Balloon

Pros

- Remarkably user-friendly (for terminal users). As simple as running one command in the terminal.
- Multiple possible modifications (cutoffs, atomic charges, conformer generation method, symmetry, etc.).
- Easy to automate and include in workflows and scripts.

Cons

- Inferior in terms of the outcome obtained.
- Fast energy growth: a small number of statistically relevant conformers.

RDKit

Pros

- Very versatile and has numerous options in addition to purely conformational analysis.
- Multiple possible modifications (cutoffs, atomic charges, conformer generation method, symmetry, etc...).
- Easy to automate and include in workflows and scripts.

Cons

- It is a library, not software per se, so it has to be scripted from scratch.
- Steeper learning curve than the rest.

wSterimol

Pros

- Systematic conformer search ensures that with well-defined dihedrals, no conformer will be missing.
- Graphical interface. Good for beginners and users not used to working with a terminal.
- Energy optimization can be done directly in DFT (with Gaussian).

Cons

- Difficult to automate. It is needed to input the atoms forming the rotatable dihedrals manually.
- Fast energy growth: a few numbers of statistically relevant conformers.

CREST

Pros

- Easy to automate, works within XTB, and config. files can be reused.
- Many possible modifications and generation and optimisation options.
- Energy optimization can be done directly in DFT The use of metadynamics ensures that the conformers are physically related and are accessible within a selected energy barrier.

Cons

- Difficult to tune the parameters to obtain a reasonable number of conformers.
- Intermediate learning curve for specific studies. relevant conformers.

1.5.2 General Recommendations

Balloon: for a quick, and easy, qualitative study or to obtain a general idea of the possible conformers of a molecule.

RDKit: for including it in highly automated workflows (molecule generation, conformers generation, analysis of results, etc.) in Python and for reusable codes.

wSterimol: for intense systematic studies or studies in which the rotation of only some specific dihedral angles needs to be studied.

CREST: for exhaustive studies in which the accessibility of the conformers and their "physical" meaning are important.

Based on the good results obtained and their capacity for being generalised and automated, both **CREST** and **RDKit** would be the preferred recommendation for general users to include in their daily workflow and research projects.

References

- (1) D. H. R. Barton, *The Principles of Conformational Analysis: Nobel Lecture*, Norstedt, 1970.
- (2) A. Mazzanti and D. Casarini, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2012, **2**, 613–641.
- (3) S. Shirakawa and K. Maruoka, *Angewandte Chemie International Edition*, 2013, **52**, 4312–4348.
- (4) T. Ooi and K. Maruoka, *Angewandte Chemie International Edition*, 2007, **46**, 4222–4266.
- (5) J. Novacek and M. Waser, *European Journal of Organic Chemistry*, 2013, **2013**, 637–648.
- (6) S. Kaneko, Y. Kumatabara and S. Shirakawa, *Organic & Biomolecular Chemistry*, 2016, **14**, 5367–5376.
- (7) S. Liu, Y. Kumatabara and S. Shirakawa, *Green Chemistry*, 2016, **18**, 331–341.
- (8) A. Gollandaj, A. Ahmad and D. Ramjugernath, *Advanced Synthesis & Catalysis*, 2017, **359**, 3676–3706.
- (9) J. Schörgenhumer, M. Tiffner and M. Waser, *Beilstein Journal of Organic Chemistry*, 2017, **13**, 1753–1769.
- (10) J. Alemán, A. Parra, H. Jiang and K. A. Jørgensen, *Chemistry–A European Journal*, 2011, **17**, 6890–6899.
- (11) J. P. Malerich, K. Hagihara and V. H. Rawal, *Journal of the American Chemical Society*, 2008, **130**, 14416–14417.
- (12) J. W. Lee, T. H. Ryu, J. S. Oh, H. Y. Bae, H. B. Jang and C. E. Song, *Chemical communications*, 2009, 7224–7226.
- (13) Y. Qian, G. Ma, A. Lv, H.-L. Zhu, J. Zhao and V. H. Rawal, *Chemical communications*, 2010, **46**, 3004–3006.
- (14) H. Konishi, T. Y. Lam, J. P. Malerich and V. H. Rawal, *Organic letters*, 2010, **12**, 2028–2031.
- (15) D.-Q. Xu, Y.-F. Wang, W. Zhang, S.-P. Luo, A.-G. Zhong, A.-B. Xia and Z.-Y. Xu, *Chemistry–A European Journal*, 2010, **16**, 4177–4180.
- (16) A. R. Brown, W.-H. Kuo and E. N. Jacobsen, *Journal of the American Chemical Society*, 2010, **132**, 9286–9288.
- (17) L. Dai, S.-X. Wang and F.-E. Chen, *Advanced Synthesis & Catalysis*, 2010, **352**, 2137–2141.
- (18) G. Luchini, D. M. Ascough, J. V. Alegre-Requena, V. Gouverneur and R. S. Paton, *Tetrahedron*, 2019, **75**, 697–702.
- (19) M. R. Garcia, I. Iribarren, I. Rozas and C. Trujillo, *Chemistry–A European Journal*, 2023, e202203577.
- (20) A. Wittkopp and P. R. Schreiner, *Chemistry–A European Journal*, 2003, **9**, 407–414.
- (21) M. J. Vainio and M. S. Johnson, *Journal of chemical information and modeling*, 2007, **47**, 2462–2474.

- (22) J. S. Puranen, M. J. Vainio and M. S. Johnson, *Journal of computational chemistry*, 2010, **31**, 1722–1732.
- (23) G. Landrum et al., 2016.
- (24) T. Piou, F. Romanov-Michailidis, M. Romanova-Michaelides, K. E. Jackson, N. Semakul, T. D. Taggart, B. S. Newell, C. D. Rithner, R. S. Paton and T. Rovis, *Journal of the American Chemical Society*, 2017, **139**, 1296–1310.
- (25) A. V. Brethome, S. P. Fletcher and R. S. Paton, *ACS Catalysis*, 2019, **9**, 2313–2323.
- (26) P. Pracht, F. Bohle and S. Grimme, *Physical Chemistry Chemical Physics*, 2020, **22**, 7169–7192.
- (27) S. Grimme, *Journal of chemical theory and computation*, 2019, **15**, 2847–2862.
- (28) P. Pracht and S. Grimme, *Chemical science*, 2021, **12**, 6551–6568.
- (29) T. A. Halgren, *Journal of Computational Chemistry*, 1996, **17**, 490–519.
- (30) G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA, 2009.
- (31) L. Schrödinger and W. DeLano, *Google Scholar There is no corresponding record for this reference*.
- (32) P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *The Journal of physical chemistry*, 1994, **98**, 11623–11627.
- (33) J. J. Stewart, *Journal of molecular modeling*, 2013, **19**, 1–32.
- (34) J. J. Stewart, *MOPAC2009. Stewart Computational Chemistry*, 2009.
- (35) C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2021, **11**, e1493.
- (36) S. Spicher and S. Grimme, *Angewandte Chemie International Edition*, 2020, **59**, 15665–15673.
- (37) J. C. Kromann, *Calculate Root-mean-square deviation (RMSD) of Two Molecules Using Rotation*, version 1.4, 2020.
- (38) H. W. Kuhn, *Naval research logistics quarterly*, 1955, **2**, 83–97.
- (39) W. Kabsch, *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 1976, **32**, 922–923.
- (40) J.-D. Chai and M. Head-Gordon, *Physical Chemistry Chemical Physics*, 2008, **10**, 6615–6620.
- (41) A. Schäfer, H. Horn and R. Ahlrichs, *The Journal of Chemical Physics*, 1992, **97**, 2571–2577.
- (42) M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji et al., *Wallingford, CT*, 2016.
- (43) L. McInnes, J. Healy and J. Melville, *arXiv preprint arXiv:1802.03426*, 2018.
- (44) L. McInnes, J. Healy and S. Astels, *Journal of Open Source Software*, 2017, **2**, 205.

2 | Revisiting Ion-Pair Interactions in Phase Transfer Catalysis: From Pure Ionic Compounds to Real Catalyst Systems

2.1 Introduction

Asymmetric ion-pairing catalysis, a prominent form of non-covalent organocatalysis, takes advantage of specifically charged intermediates or reagents. By forming an Ion-Pair (IP) between the catalyst and the substrate, reactivity and selectivity can be significantly enhanced.^{1,2} The emergence of chiral ion-pairs has played a pivotal role in the development of asymmetric catalysis, enabling precise control and improving reactivity across a diverse range of chemical reactions. However, the pursuit of a general approach to achieving exceptional enantioselectivity in asymmetric ion-pairing catalysis faces notable challenges stemming from the inherent directionless nature associated with ion-pairing interactions, in contrast to the discernible behaviour exhibited by HB and covalent interactions.^{1,3–10} Enantioselective ion-pairing catalysis encompasses a diverse array of small organic molecules, amongst which quaternary ammonium salts have emerged as the most extensively studied class of chiral catalysts within the Phase-Transfer Catalysis (PTC) framework.

These catalysts feature a nitrogen atom positioned at the centre of a tetrahedron formed by carbon atoms, with three sterically hindered faces, resulting in a single accessible face for anion interaction, thereby facilitating the establishment of ion-pairs (Figure 2.1a). The successful use of electrostatic interactions between the ammonium cation and the counterion holds significant promise in the induction of chiral information, thus enabling asymmetric induction (Figure 2.1b).⁴

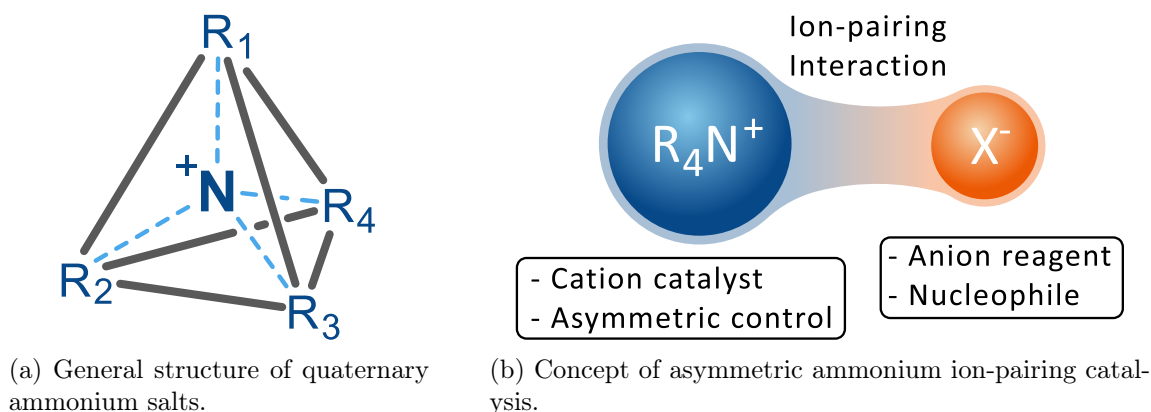


Figure 2.1. Traditional considered interactions between the catalyst and the anion in PTC.

The introduction of a second interaction point, such as a Hydrogen Bond (HB) donor, has been reported to influence the ion-pairing process of PTC and competes with the establishment of a HB interaction. As a result, two distinct modes of activation for these bifunctional catalysts have been proposed, wherein the catalyst can act either through the formation of an IP or a HB.

In this context, as will be seen in the next chapter, our group conducted a comprehensive theoretical investigation aimed at elucidating the preferred binding mode and its impact on enantioselectivity. Contrary to the conventional notion of a "strict ion-pair" in bifunctional catalysis, our study revealed that a definitive binding mode through a HB is energetically more favourable. This highly stabilising and directional interaction was found to profoundly influence the control of enantioselectivity in reactions. Intriguingly, our computational analysis did not provide conclusive evidence supporting the sole existence of a discernible strict IP interaction,

underscoring the significance of hydrogen bonding in these systems.¹¹ Consequently, these findings prompt the need for an updated definition of ion-pairing within the field of phase-transfer catalysis.

The definition of Ion-Pair has evolved over time, driven by advances in our understanding of electrostatic interactions and molecular behaviour. Historically, IP started with Bjerrum’s pioneering work in 1926,¹² where he proposed a definition based on the electrostatic work required to separate two ions (Eq. 1). Subsequently, Anslyn and Dougherty introduced a physical-organic definition in 2006,¹³ emphasising the spatial proximity and the energetic strength of the electrostatic attraction between ions. Their definition considers the comparison between the electrostatic energy and the thermal energy available for ion separation.

$$W_{+-}(r) = \frac{-z_i z_j e^2}{\epsilon r} \quad (1)$$

In addition, the influence of Coulomb’s law on ion-pairing interactions emphasises the role of the distance and dielectric constant in determining the strength of these interactions. Coulomb’s law (Eq. 2) describes the attractive potential energy between two ions (q_1 and q_2). The magnitude of the electrostatic interaction is inversely related to the distance between the ions (r) and the dielectric constant of the medium (ϵ_0). Thus, strong ion-pairing interactions are preferred at short distances and in nonpolar solvents.^{1,14}

$$F = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \quad (2)$$

Given the influence of this interaction in the structural stabilisation of ionic complexes, especially at short distances, the concept of ion-pairing is ubiquitously used in literature as the key factor governing catalytic activity and selectivity in ionic complexes, with little emphasis on other major potential sources of stabilisation. This is especially significant in experimental works where theoretical insights are unavailable.

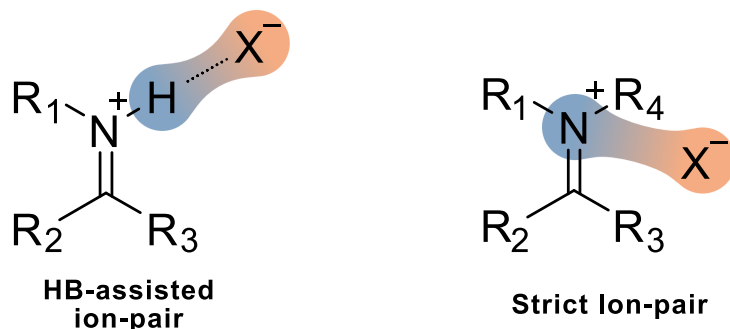


Figure 2.2. Different types of ion-pairing interactions. HB-assisted ion-pair and Strict ion-pair.

Thus, unravelling the true nature of relevant ion-pairs found in literature is key to understanding *where* and *how* ion-pairing interactions are key players in the formation of these complexes or whether they are mere spectators in a network of equally stronger types of interactions. This will provide a new framework for using the concept of ion-pairing in the emerging field of asymmetric organocatalysis. To achieve this, the nature of relevant ion-pair interactions was investigated and thoroughly characterised in this chapter, challenging different systems in which, in theory, only strict ion-pair or HB-assisted ion-pairing interactions can be established within PTC catalyst (Figure 2.2) and comparing them with their neutral counterparts.

2.2 Computational Details

The structures of the salts and the different models of the catalysts under study were optimised at the wB97x-D/def2-tzvp^{15–17} computational level and the real catalysts were optimised at wB97x-D/def2-svp. Harmonic frequencies were computed at the optimisation level to confirm that the relaxed structures correspond to local minima. In some structures, marked with * in the text, the distance $N-H$ was fixed to avoid H transfer and to keep all complexes consistent for analysis. Single-point calculations were performed to improve the energy description at the computational level of wB97x-D/def2-qzvp^{15–17} for salts and models and wB97x-D/def2-tzvp computational level for the real catalysts under study. Structure optimisations and energy calculations were performed using the Gaussian16 software.¹⁸ All the calculations were performed in the gas phase and at room temperature (298K). The optimised structures can be found in Table S2.1.

The interaction energies (ΔE) were calculated as the difference between the free energy of the optimised complexes ($E_{complex}$) and the energy of each optimised monomer ($E_{monomer}$) as shown in Eq. 3. The interaction energies can be found in Table S2.2.

$$\Delta E = E_{complex} - \sum E_{monomer} \quad (3)$$

An energy decomposition analysis was performed based on the LMO-EDA¹⁹ scheme at the M06-2X/KTZVP computational level. The interaction energy was obtained as the sum of different energetic terms, as shown in Eq. 4

$$E_{int} = E_{elec} + E_{exc} + E_{rep} + E_{pol} + E_{disp} \quad (4)$$

where E_{elec} is the electrostatic term that describes the classical Coulombic interaction of the occupied orbitals of one monomer with those of the other. The term E_{exc} is attractive and denotes the contribution of the exact exchange energy between the orbitals of one monomer and the orbitals of the others, E_{rep} is repulsive and is associated with the Pauli exclusion principle, and the terms E_{pol} and E_{disp} correspond to the terms of polarisation and dispersion, respectively. These calculations were carried out using GAMESS software (version 2018-R2).²⁰ Numerical LMO-EDA values are collected in Table S2.3.

Due to some technical problems with the solvation model when running EDA calculations, the energy decomposition analysis for the systems in solution was calculated using the EDB-EDA^{21,22} at the wB97x-D/def2-tzvp level of theory. Solvent effects were included in the optimisation by means of a continuum method, the Solvation Model based on Density (SMD) approach²³ in Gaussian16. The use of an implicit solvent model allows us to consider the electrostatic and non-electrostatic cavitation, solvation, and structural effects of the solvent while, at the same time, allowing for an undisturbed and in-detail analysis of any inter- and intramolecular interactions governing structural stabilisation of our catalysts. The interaction energies are obtained as the sum of the energy terms expressed in Eq. 5

$$E_{int}^{EDA-NCI} = E_{elec} + E_{exc} + E_{rep} + E_{pol} \quad (5)$$

where E_{elec} , E_{exc} , and E_{rep} represent the same magnitudes as in LMO-EDA and the term E_{pol} correspond to the term of polarisation and it is calculated using Eq. 6

$$E_{pol} = E_{def} + E_{exc_def} \quad (6)$$

where E_{def} is the electron density deformation energy, E_{exc_def} is the deformation energy associated with the exchange-correlation density.

The NBO²⁴ methodology was used to evaluate atomic charges in the different systems and to evaluate the strength of individual interactions by measuring the charge transfer between occupied and unoccupied orbitals within different fragments. The calculations were performed at the same level of theory as the geometry optimisations: wB97x-D/def2tzvp for the salts and small models and wB97x-D/def2svp for the real catalysts. Intermolecular NBO charge transfer values can be found in Table S2.4 and NBO charges in Table S2.6. The NBO calculations yield the E(2) value, which will be thoroughly discussed in this manuscript. As explained in the Computational Methods chapter, E(2) is defined as second-order perturbative energy, determined by the analysis of all possible interactions between Lewis-type NBO donors and non-Lewis-type NBO acceptors, with their energetic importance estimated through second-order perturbation theory. *"For each donor NBO (i) and acceptor NBO (j), the donor-acceptor stabilisation energy E(2) associated with i → j delocalisation is calculated as E(2) = ΔE_{ij}(2) = q_i^{F(i,j)²}/(ε_j - ε_i), where q_i represents the donor orbital occupancy (2 for closed shell, 1 for open shell), ε_i and ε_j denote the diagonal elements (orbital energies), and F(i, j) stands for the off-diagonal NBO Fock matrix element."*²⁵

The QTAIM methodology^{26,27} was used to analyse the electron density of the systems within the AIMAll software, and the electron density at the BCPs between the systems and the different anions upon complexation was extracted. The calculations were performed at the same level of theory as the geometry optimisations: wB97x-D/def2tzvp for the salts and small models and wB97x-D/def2svp for the real catalysts. The densities at the bond critical points, ρ_{BCP} , bond distances, d , Laplacian value at the BCP, $\nabla^2\rho_{BCP}$ and energy densities of th BCP, H_{BCP} , can be found in Table S2.5 and 3D molecular graphs of all the systems in Figure S2.4.

2.3 Results and Discussion

To establish a comprehensive understanding of ion-pair interactions, this study investigates various scenarios, starting with a discussion of ionic compounds to define the concept of an ion-pair. Subsequently, a wide range of ion-pair catalyst models is examined, as well as their neutral forms for comparison, followed by the characterisation of real PTC catalysts. To ensure homogeneity and comparability, all calculations were performed in the gas phase, and the effect of the solvent was studied for one particular case.

By examining these systems, we aimed to unravel the intricate interplay between HB and IP interactions by discerning the subtle differences that influence the nature and strength of hydrogen-bonding and its role in facilitating ion-pair formation, shedding light on their impact on the overall catalytic activity and selectivity.

2.3.1 Ion-Pair

In this part of the study, our primary objective is to explore notable instances of ionic salts characterised by polar bonding²⁸ between atoms (e.g. between Na^+ and Cl^- , CN^- —through both terminations— and $\text{PO}_4(\text{CH}_3)_2^-$ through O atoms). Furthermore, an additional cation, namely $[\text{N}(\text{CF}_3)_4]^+$, was introduced to encompass a structural motif comparable to the predominant PTC scaffold (Figure 2.3). The aim of studying these systems was to obtain an accurate definition of a strict ion-pair and, therefore, to shed light on the nature of the interactions of the following systems studied by comparison.

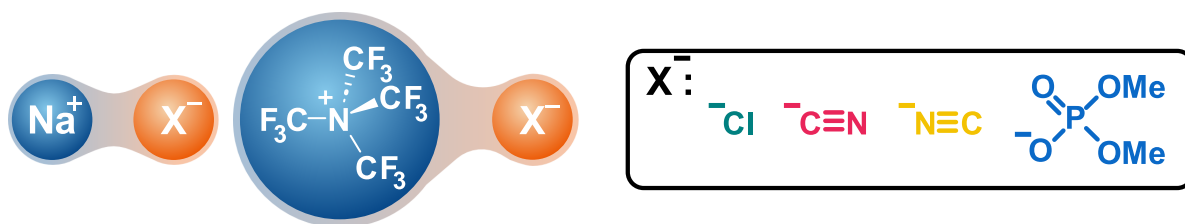


Figure 2.3. Representation of the ionic salt models used to investigate strict ion-pairing interactions, left, and anions used, right.

The interaction energies of the studied complexes are presented in Table 2.1. The obtained values fall within the expected range for pure ionic interactions, providing evidence for the electrostatic attraction between the cation and the negatively charged anion. Additionally, the calculated distances between the ions are in excellent agreement with the reports found in the literature, further validating the accuracy of the computational approach.²⁹ In the case of complexes including the $[\text{N}(\text{CF}_3)_4]^+$ cation, the interaction energies are comparatively lower and present cation-anion distances around 4 Å, primarily attributable to the steric hindrance arising from its larger molecular structure, thereby highlighting the establishment of a longer-range ion-pair interaction.

Characterisation of the ionic interactions

To characterise and define what can be considered an ion-pair from a theoretical perspective, an exhaustive analysis of the interactions established within the different systems under study was carried out. First, to investigate the physical origins of the interactions involved, an analysis

Table 2.1. Ionic salts and $[\text{N}(\text{CF}_3)_4]\text{-X}$ DFT Interaction energies, ΔE in kJ/mol. Distances in Å.

Ionic Complex	ΔE	$d(\text{Na}^+\text{X}^-)$
Na-Cl	-546.4	2.39
Na-CN	-521.0	2.27
Na-NC	-539.1	2.14
Na- PO_4Me_2	-586.1	2.26
Ionic Complex	ΔE	$d(\text{N}^+\text{X}^-)$
$\text{N}(\text{CF}_3)_4\text{-Cl}$	-316.5	4.20
$\text{N}(\text{CF}_3)_4\text{-CN}$	-315.3	4.25
$\text{N}(\text{CF}_3)_4\text{-NC}$	-320.7	4.11
$\text{N}(\text{CF}_3)_4\text{-PO}_4\text{Me}_2$	-313.3	3.92

of the LMO-EDA was performed, which allows the determination of the contribution of the different energy components to the global stability of the systems (Figure 2.4). This approach has been extensively used in the literature to describe and analyse different NCIs, such as those reported in the present project.³⁰ These contributions will provide insight into the nature and magnitude of interactions within the systems under study.

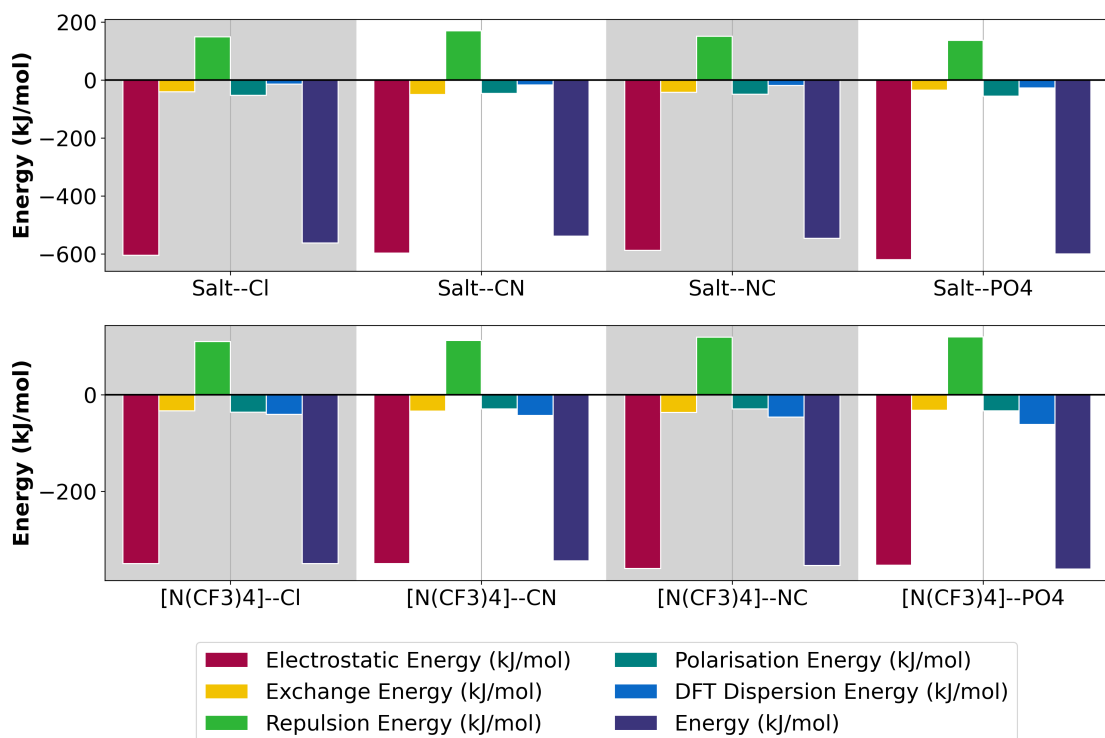


Figure 2.4. LMO-EDA energy partition terms for the different salts under study, top. LMO-EDA energy partition terms for the $[\text{N}(\text{CF}_3)_4]^+$ system, bottom.

From the analysis of the various contributions to the total energy, it is evident that the electrostatic term corresponding to a quasiclassical Coulombic interaction between monomers constitutes the major component of the overall energy in these systems. This observation further reinforces the fundamental nature of the electrostatic interactions in governing the stability and behaviour of ionic compounds. A similar trend was observed across both types of systems, encompassing both ionic and $\text{N}(\text{CF}_3)_4\text{-X}$ complexes. However, it is noteworthy that the total

energy and the electrostatic term associated with the $N(CF_3)_4-X$ systems are approximately half the corresponding values observed for the proper ionic compounds in the ion-pair configuration, emphasising that although both cases involve electrostatic interactions, the latter exhibits remarkably weaker interactions in comparison, being the distances between anion and cation increased accordingly. More importantly, the dispersion term becomes more negative in the case of $N(CF_3)_4-X$ salts, being almost negligible in the case of ionic compounds, but the repulsion term is lower since the distance between the anion and the cation increases sensibly.

Next, an NBO analysis was performed, along with a comprehensive study of the charges, both isolated and upon complexation (Figure 2.5, Figure S2.1.1 and Tables S2.4.1, S2.4.2 and S2.6.2). Since ionic bonds do not involve electron sharing, low values for $E(2)$ were correspondingly found.

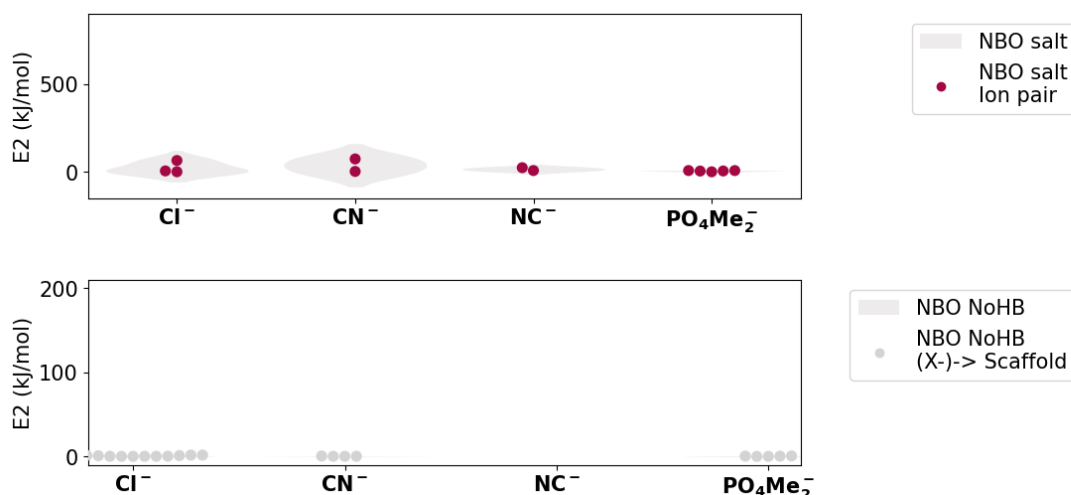


Figure 2.5. NBO $E(2)$ values corresponding to the different ionic salts studied, top, and NBO $E(2)$ values corresponding to the different $[N(CF_3)_4]-X$ systems studied, bottom.

In the scenario of a polar ionic pair, all possible interactions between the Lone Pair (LP) of the anion and the Lone Vacancy (LV) (unoccupied borane-like valence orbital) of the cation and its contributed energy were considered (Figure 2.5 top). We observed extremely low charge transfer processes for almost all the studied $N(CF_3)_4-X$ salts originating from the LP of the anion ($LP(X^-)$ to the C-F anti-bonding sigma orbital ($BD^*(C-F)$) of the scaffold (Figure 2.5 bottom).

Upon analysing the charge of the monomer and its variation upon complexation (Figure S2.1.1), we found that the N atom's charge remains unchanged upon interaction with the anions, while C3 becomes slightly more negatively charged, in agreement with the charge transfer process described by the NBO analysis. Through analysis of the monomer, it becomes evident that the positive charge is not confined to the N atom but delocalised within the C atoms of the scaffold.

Finally, we turned our attention to investigating the electronic behaviour of the intermolecular interactions studied. In this regard, QTAIM has been widely shown to be an excellent tool for analysing various properties of the electron density and the nature of chemical bonding within a molecule, including bond strengths, and understanding the presence of charge transfer or electronic delocalisation within the molecule.³¹

First, the electron density at the BCP, $\rho(r)_{BCP}$, of the relevant bond was studied, as well as their corresponding Laplacian values, $\nabla^2\rho(r)_{BCP}$ and the energy density at the BCP, H_{BCP} (Table 2.2). In the topological analysis, $\nabla^2\rho(r)_{BCP}$, plays a very important role in the character-

isation of chemical bonding, providing the physical basis for the well-known electron-pair model of Lewis.^{26,32} As explained in the computational methods chapter, a positive $\nabla^2\rho(r)_{BCP}$ indicates the dominance of excess kinetic energy, corresponding to local depletions, and is associated with NCIs (lower positive values) and ionic bonds (higher positive values).

The values obtained for the electron density and $\nabla^2\rho(r)_{BCP}$ in Table 2.2 reinforce our previous analysis of the ionic nature of the salts studied and will be used as a reference for analysing the nature of the remaining systems.

Table 2.2. Electron density, $\rho(r)_{BCP}$, Laplacian, $\nabla^2\rho(r)_{BCP}$, and Energy density, H_{BCP} , at the bond critical point for the ionic systems studied.

Anion	Cation	$\rho(r)_{BCP}(a.u.)$	$\nabla^2\rho(r)_{BCP}$	$H_{BCP}(a.u.)$
Cl^-	Na^+	0.0356	0.1899	0.0049
NC^-	Na^+	0.0343	0.2424	0.0045
CN^-	Na^+	0.0383	0.1764	0.0071
$PO_4Me_2^-$	Na^+	0.0264	0.1769	0.0071
$PO_4Me_2^-$	Na^+	0.0257	0.1718	0.0070

The QTAIM analysis was performed for the $N(CF_3)_4-X$ systems under study, and the corresponding values are provided in the Supplementary Information (Table S2.5.2) for clarity. Of these, a representative example was chosen to highlight the presence and values of BCPs, as shown in Figure 2.6. When the density values are analysed at the various BCPs, it becomes apparent that the interactions within the ammonium salts are comparatively weaker. Furthermore, the Laplacian values indicate the presence of non-covalent interactions; however, they do not reach the magnitude observed in the ionic salts ($\nabla^2\rho(r)$ ranges between 0.1 and 0.2 a.u.).

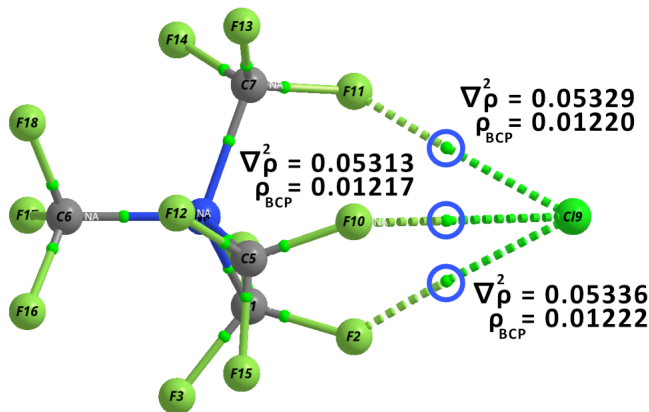


Figure 2.6. QTAIM molecular graph corresponding to $N(CF_3)_4-Cl$.

2.3.2 Imine- and Iminium-based Models

As noted above, a previous study conducted by our group considered an alternative interaction, a HB interaction, contrasting it with what was known or proposed in the literature as a strict ion-pair.¹¹ However, the latter interaction was found to be energetically less favourable, as demonstrated by its lower energy interaction. As a result, for the current investigation, we focus solely on the interaction between the anions involved and the HB donor, if feasible.

To continue our investigation, we explored systems in which the tetra-bonded N atom within the scaffold corresponds to a strong HB-donor and, therefore, facilitates the formation of a

traditional hydrogen-bonding interaction. This encompassed the study of both neutral and positively charged catalyst-based systems, aiming to discern and characterize the variations between the neutral cases, and the interactions encountered in charged systems. Furthermore, various substituents, including CF_3 , CH_3 , and $t\text{Bu}$, were considered to account for multiple electronic and steric scenarios (Figure 2.7).

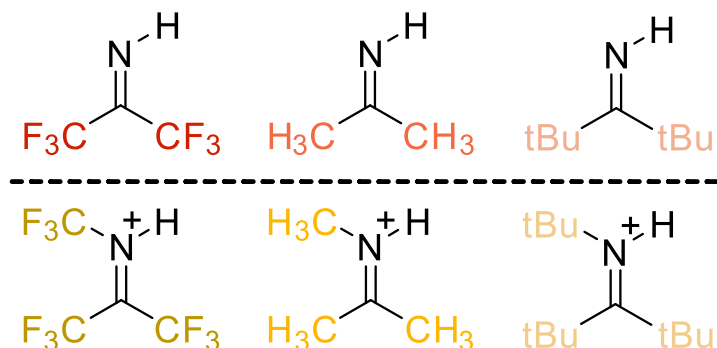


Figure 2.7. Different imine- and iminium-based derivatives under study.

The calculated interaction energies and anion-cation distances, for positively charged and neutral species, are summarised in Table 2.3. As expected, the interaction energies obtained are all negative, indicating favourable intermolecular interactions. Notably, when comparing the neutral species (imine derivatives) to their positively charged counterparts (iminium derivatives), a significant increase in interaction energies is observed. In terms of the different substituents studied, it is evident that CF_3 exhibits the most negative values, indicating the strongest interactions within the formed complexes. Conversely, $t\text{Bu}$ demonstrates the weakest interaction energies. In general, the intermolecular interactions studied are more favourable when Cl^- is involved as the counteranion.

Changing our focus to the distance between the N^+ (or N for neutral compounds) and X^- ions, it is notable that the distances in all neutral complexes surpass those observed in their charged counterparts. Additionally, by comparing the distances of the cation-anion in the iminium complexes with the previously studied $[\text{N}(\text{CF}_3)_4]\text{-X}$ systems, we find that all distances are consistently shorter than the latter.

An analysis of the different anions reveals that systems containing Cl^- exhibit the longest intermolecular distances for both neutral and cationic complexes. Considering the influence of the various substituents within the main scaffold, the distances follow the trend of $t\text{Bu} > \text{CH}_3 > \text{CF}_3$. Expectedly, the presence of a bulky electron-donating substituent, such as $t\text{Bu}$, leads to the highest distance between the ions. On the contrary, when the CF_3 substituent is involved, the shortest distance is observed, as a consequence of its electron-withdrawing nature, which strengthens the existing HB.

Characterisation of the Non-Covalent Interactions

Building upon the preceding section, this section focuses on the characterisation of interactions using LMO-EDA, NBO, and QTAIM analyses. Regarding the results obtained from the LMO-EDA analysis, it is evident that in the charged system scenario, the electrostatic term makes the most significant contribution to the total energy (Figure 2.8), which aligns with the expected dominance of Coulombic interactions in charged systems. However, even though the distances

Table 2.3. Iminium charged and Imine complexes DFT Interaction energies, ΔE in kJ/mol. Distances in Å. Distances between brackets () refer to the second shortest N—O distance in those systems where two O atoms of the PO_4Me_2 anion are interacting with the other monomer. Those systems marked with * were optimised with fixed X-H distance because a H transfer was produced otherwise.

Complex	ΔE	$d(\text{N}^+ \text{X}^-)$
Iminium_ $\text{CF}_3\text{-Cl}^*$	-535.3	2.68
Iminium_ $\text{CF}_3\text{-CN}^*$	-527.5	2.52
Iminium_ $\text{CF}_3\text{-NC}^*$	-527.3	2.41
Iminium_ $\text{CF}_3\text{-PO}_4\text{Me}_2^*$	-518.6	2.34
Iminium_ $\text{CH}_3\text{-Cl}$	-456.3	2.86
Iminium_ $\text{CH}_3\text{-CN}$	-445.5	2.69
Iminium_ $\text{CH}_3\text{-NC}$	-449.0	2.57
Iminium_ $\text{CH}_3\text{-PO}_4\text{Me}_2$	-456.9	2.55
Iminium_ $^t\text{Bu-Cl}$	-397.7	3.04
Iminium_ $^t\text{Bu-CN}$	-387.1	2.79
Iminium_ $^t\text{Bu-NC}$	-393.8	2.68
Iminium_ $^t\text{Bu-PO}_4\text{Me}_2$	-399.2	2.62
Imine_ $\text{CF}_3\text{-Cl}$	-97.4	3.07
Imine_ $\text{CF}_3\text{-CN}$	-90.0	2.90
Imine_ $\text{CF}_3\text{-NC}$	-91.9	2.77
Imine_ $\text{CF}_3\text{-PO}_4\text{Me}_2$	-84.7	2.74 (3.81)
Imine_ $\text{CH}_3\text{-Cl}$	-61.2	3.42
Imine_ $\text{CH}_3\text{-CN}$	-49.3	3.27
Imine_ $\text{CH}_3\text{-NC}$	-52.9	3.11
Imine_ $\text{CH}_3\text{-PO}_4\text{Me}_2$	-52.3	3.01 (4.41)
Imine_ $^t\text{Bu-Cl}$	-65.1	3.40
Imine_ $^t\text{Bu-CN}$	-50.7	3.26
Imine_ $^t\text{Bu-NC}$	-55.5	3.09
Imine_ $^t\text{Bu-PO}_4\text{Me}_2$	-56.5	3.06 (4.99)

are not much larger, in most of the systems studied, the electrostatic term is lower compared to that found for ionic salts, similar to that found within the quaternary ammonium salts, $\text{N}(\text{CF}_3)_4\text{-X}$. However, it becomes apparent that iminium/imine systems, attractive terms, polarisation, exchange and dispersion have increased significantly as a result of the establishment of HBs upon complexation (almost -200 kcal/mol). It is noteworthy that in the preceding section, these terms were found to be extremely low in $\text{N}(\text{CF}_3)_4\text{-X}$ complexes and almost negligible in the ion-pair systems examined.

Additionally, it should be highlighted that the electron-electron repulsion term has also increased notably, surpassing its contribution in the previous case. On the contrary, in the case of the neutral counterpart, the repulsion term emerges as the primary contributor to the total energy (Figure S2.2.3 and S.2.2.4). This significant Pauli repulsion effect leads to an increased distance between the interacting species, resulting in a weaker interaction.

The NBO analysis, as depicted in Figure 2.9 and Figures S2.3.3 and S2.3.4, reveals that the highest values correspond to the presence of charge transfer interactions specifically associated with HB formation for both cases, for charged and neutral systems, respectively. These interactions involve the donation of electron density from the Lone Pair ($LP(\text{X}^-)$) of the anion to any sigma anti-bonding orbital involving the theoretically positively charged nitrogen atom ($BD^*(\text{N-H})$). The obtained values of $E(2)$, which represent the stabilisation energy due to charge transfer between the aforementioned orbitals, follow a trend similar to that of the interaction energy concerning different substituents, e.g. $\text{CF}_3 > \text{CH}_3 > ^t\text{Bu}$. Furthermore,

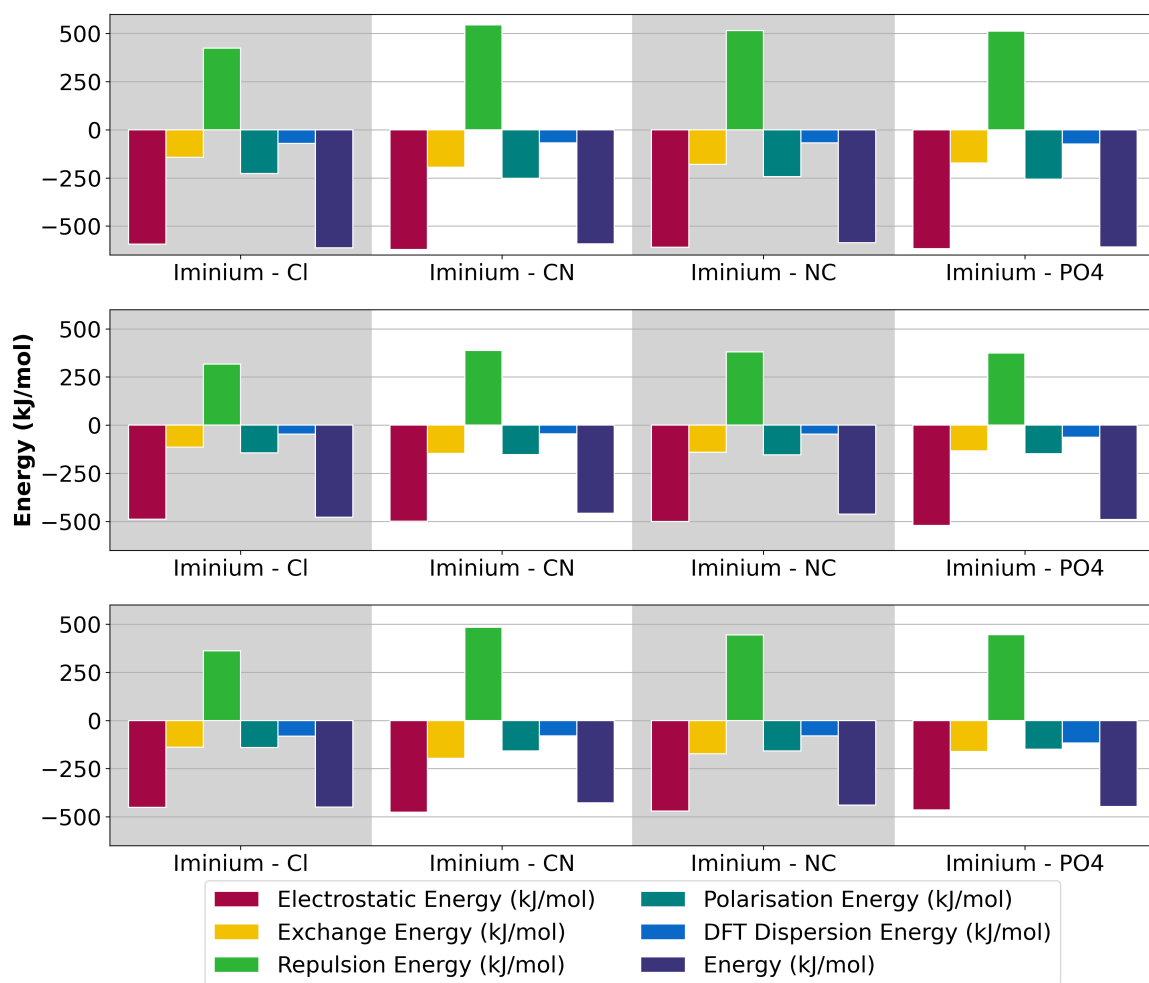


Figure 2.8. LMO-EDA energy partition terms for the different Iminium-based complexes under study. Top - Iminium- CF_3 , middle - Iminium CH_3 and bottom - Iminium- tBu .

considering the charged systems, we find that the $E(2)$ values are higher than their neutral counterparts because of the positive charge distribution within the molecule. Nevertheless, the nature of the charge transfer interaction remains unchanged.

The obtained results from the QTAIM analysis provide compelling support for the findings obtained from the previous analyses (Figures S2.4.3 and S2.4.4 and Tables S2.5.3 and S2.5.4). Examination of the electron density, $\rho(r)_{BCP}$, at various BCPs within the studied complexes, reaffirms the nature of the NCIs established upon complexation. In addition, the positive $\nabla^2\rho(r)_{BCP}$ values derived from the QTAIM analysis further validate the non-covalent nature of the interactions observed in all the cases under study. Interestingly, the values obtained for $\nabla^2\rho(r)$ are consistently similar for both neutral and charged compounds; however, these values are consistently an order of magnitude lower than those observed in ionic compounds (Figure 2.10). This observation suggests that the presence of HBs is influenced by the surrounding charges, and consequently the systems present a clear HB-assisted IP rather than a strict IP interaction.

It has been established that the charge carried within the PTC catalysts structure is vital, and therefore it becomes crucial to conduct an in-depth investigation of the charges associated with the monomer in its isolated state and upon complexation.

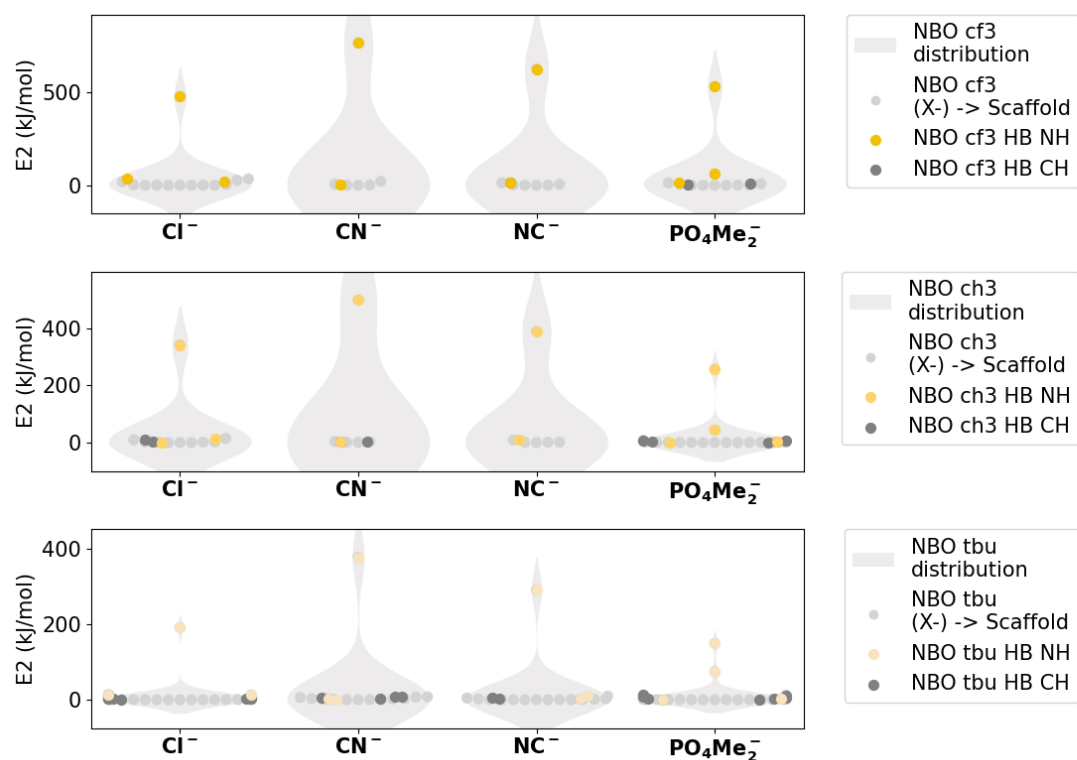


Figure 2.9. NBO E(2) values corresponding to the different Iminium-based complexes under study.

Firstly, a comparative analysis of the charges between the charged monomers (Figure 2.11) and their neutral counterparts (Figure S2.1.2) was conducted. In the isolated monomers, the **C1** atom, which is directly attached to nitrogen, was revealed to exhibit a significant increase in positive charge in all the cases studied, from neutral to charged monomers. In contrast, the nitrogen remains relatively unchanged in the CF₃ case and slightly more negative in the presence of CH₃ and ^tBu substituents, in agreement with the electron-donating nature of those substituents. Upon complexation with the anion, visible modification is observed in the **C1** atom's charge for most cases, indicating a charge transfer from the anion to **C1**. Consequently, the **C1** atom becomes less positively charged in all cases, with a more pronounced effect observed in the scenario involving the CF₃ substituents.

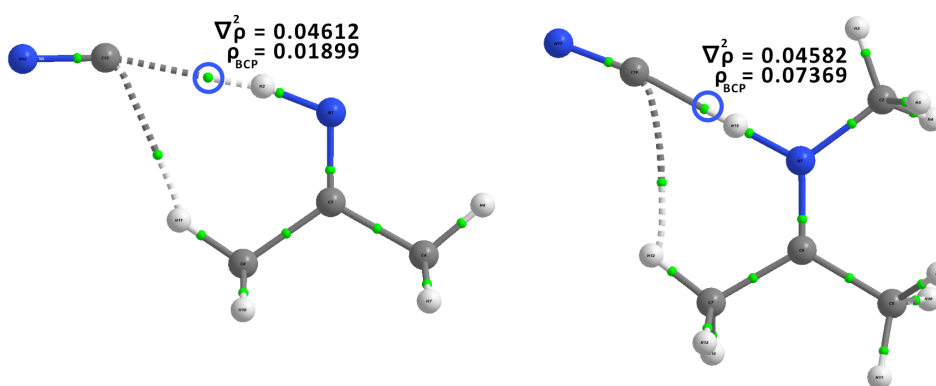


Figure 2.10. QTAIM molecular graphs for Imine_CH₃-CN (left) and Iminium_CH₃-CN (right).

A similar behaviour is observed for the neutral case (Figure S2.1.3), in which **C1** presents the highest charge difference between the monomer and the complex. The aforementioned variation in the charge distribution within the **C1** atom can also be seen in the NBO calculations. Specifically, the charge transfer from the Lone Pair ($LP(X^-)$) of the anion to the anti-bonding orbital of the $N-C$ bond ($BD^*(N-C)$) highlights this distinction (Table S2.4.4). Nevertheless, it is important to emphasise that the magnitude of this charge transfer is significantly low when compared to the corresponding value associated with the HB, (Figure 2.9, light grey, Table S2.4.4). Specifically, the highest value obtained for E(2) $BD^*(N-C)$ is 27.0 kJ/mol, while E(2) involving $BD(N-H)$ ranges from 292.1 to 768.1 kJ/mol.

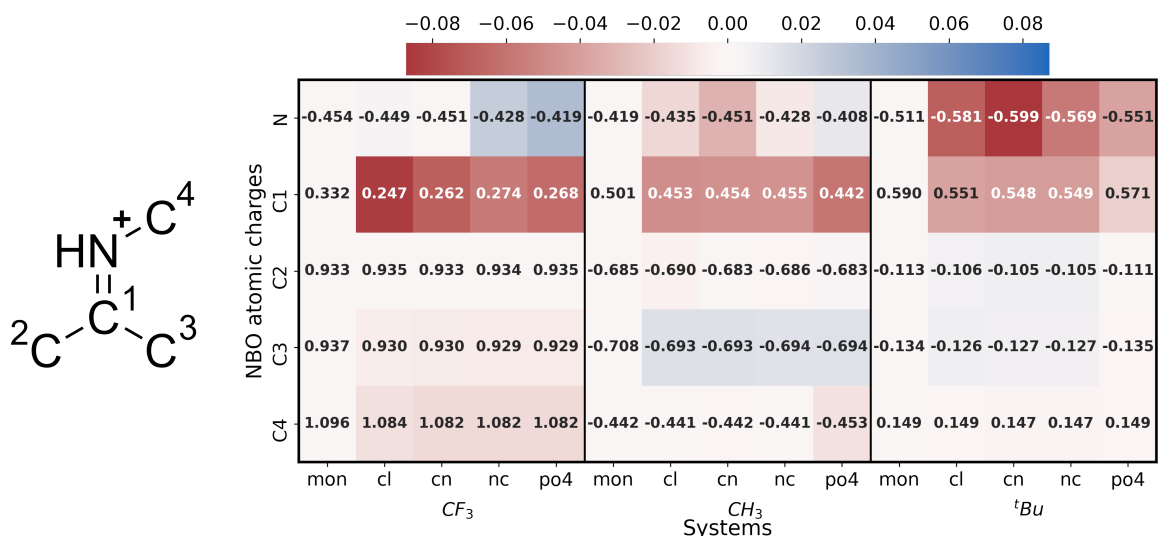


Figure 2.11. NBO charges of the monomers and different Iminium-based complexes under study. Legend: charge difference between the complexes and the monomers is depicted.

Finally, it is important to note that the positive charge in these systems is not localised in a single atom but is shared among multiple atoms. This shared positive charge results in an increased positive character (+0.1 e higher) of the hydrogen atom (NH) when the NH transitions from neutral to cationic structure. Consequently, the establishment of HBs with the different anions studied is arguably strengthened and favoured because of the increased positive charge exhibited by the hydrogen (H) atom involved.

The outcome regarding the nature of the different interactions upon complexation studied within the iminium-based complexes is far from surprising since it has been previously theoretically described in the literature as “charge-assisted hydrogen bonds”.^{33,34}

Solvent Effects

The present investigation was primarily conducted in the gas phase to focus on the characterisation of the nature of the diverse interactions inherent within ionic compounds. Nonetheless, to achieve a comprehensive evaluation, a supplementary investigation involving solvents was undertaken, as illustrated in Table 2.4.

As anticipated, the calculated interaction energies exhibited a substantial reduction across all the solvents examined, with particularly pronounced effects observed in the case of dimethylsulfoxide (DMSO) and water, across all Iminium- and Imin-based complexes studied. When the

changes in intermolecular distances between both charged atoms, Na^+ and X^- atoms, caused by solvent were examined, it was observed that the separation between cations and anions augmented uniformly in all cases, aligning with theoretical expectations. Furthermore, as the distance increased, the interaction energy presented upon complexation weakened.

Table 2.4. DFT Interaction energies in three different solvents and gas phase, for Ionic salts and Iminium charged complexes, ΔE in kJ/mol. Distances in Å.

Cation	Anion	Solvent	ΔE	$d(Na^+X^-)$
Na	Cl	H ₂ O	-61.4	2.52
Na	Cl	DMSO	-64.9	2.52
Na	Cl	toluene	-250.6	2.44
Na	Cl	gas	-546.4	2.39
Iminium_ _{CF₃}	Cl	H ₂ O	-69.6	2.87
		DMSO	-72.2	2.86
		toluene	-268.0	2.75
		gas	-535.3	2.68
Iminium_ _{CH₃}	Cl	H ₂ O	-40.2	3.08
		DMSO	-42.3	3.08
		toluene	-209.3	2.94
		gas	-456.3	2.86
Iminium_ _{^tBu}	Cl	H ₂ O	-19.5	3.72
		DMSO	-19.1	3.69
		toluene	-167.7	3.22
		gas	-397.7	3.04

A partition energy analysis was conducted for the complexes studied under solvation conditions (Figure S2.5). The EDB-EDA calculation reveals that both repulsion energy and polarisation are the most affected terms, while the corresponding electrostatic term experiences only a slight decrease. This decrease in electrostatics arises from the longer distance between the ions, as evident from the EDB-EDA calculated for the gas-phase optimised geometry in solvent (left side of Figure S2.5).

More importantly, when implicit solvent molecules are included, the repulsion term decreases; however, the polarisation term becomes positive and repulsive, particularly in the cases of DMSO and water. This alignment with the obtained low interaction energy is notable. Consequently, the sharp decrease in interaction energy can primarily be attributed to the repulsive polarisation force.³⁵

2.3.3 Guanidine- and Guanidinium-based Models

Guanidinium-based scaffolds have long been recognised for their ability to function as general acid catalysts, and therefore we decided to conduct an extensive computational investigation to gain a deeper understanding of the interactions that occur upon complexation. To ensure clarity and avoid clutter, we present the key findings from our analysis of both neutral, guanide-based, and charged, guanidinium-based, complexes (Figure 2.12) in the Supplementary Information (Figures S2.1-2.4 and Tables S2.2-S2.6), utilising the same methodologies as before, e.g. LMO-EDA, QTAIM and NBO.

The calculated interaction energies for both the neutral and charged complexes were found to be negative, as anticipated. Furthermore, the interaction energies for the neutral complexes were consistently lower than those for the charged complexes. In particular, an unexpected case that warrants attention concerns the guanidinium ^tBu complexes, which exhibited interaction

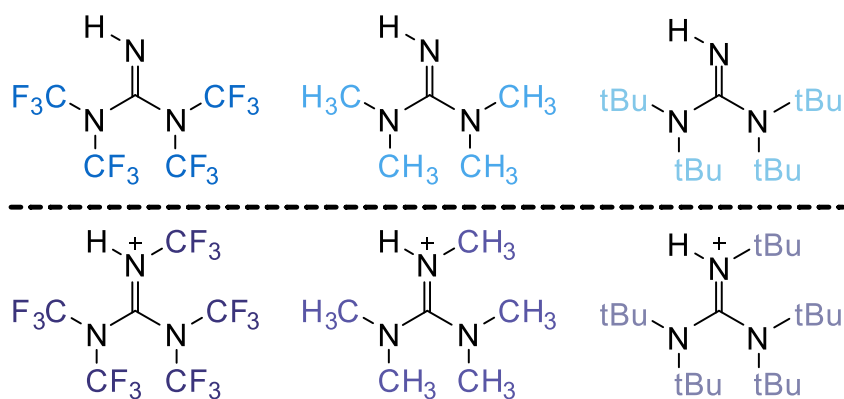


Figure 2.12. Gunidine- and guanidinium-based systems under study.

energies that were one order of magnitude lower than the rest of the charged complexes under investigation. When the monomer and the complexes were carefully examined, it became evident that accommodating the anion within such a bulky species requires substantial deformation energy. However, looking at the LMO-EDA total energies the results are consistent amongst the three substituents and present the expected stability trend. The distance between positively charged N^+ and negatively charged X^- ranged from 2.65 to 3.43 Å, higher than in the cases previously analysed.

Furthermore, our analysis of the interactions established upon complexation revealed that the established HBs are notably stronger in the charged species. Most cases exhibited Laplacian values, $\nabla^2\rho(r)_{\text{BCP}}$, consistent with a NCI nature. However, some particular cases (five amongst all the different complexes under study) showed a larger positive Laplacian value at the BCP that corresponds to the $\text{X}^- \cdots \text{HN}$ HB, as indicated in Table S2.5.5. In these particular cases, the H atom was in close proximity to the anion, almost not bonded to the guanidinium, resulting in a highly positive charge and the emergence of a slight ionic character between the H and the involved anion.

By analysing the NBO calculations performed in all systems, it can be corroborated that the established HBs correspond to the main charge transfers process from the Lone Pair of the anion to the anti-bonding orbital of the N-H bond ($LP(\text{X}^-) \rightarrow \text{BD}^*(\text{N-H})$) and the weak ones can be found from the LP of the anion to the N of the scaffolds ($LP(\text{X}^-) \rightarrow \text{BD}^*(\text{N-C})$) (Tables S2.4.5 and S2.4.6 and Figures S2.3.5 and S2.3.6).

Finally, the distribution of the different atomic charges was analysed. As we transitioned from neutral monomers to charged complexes, for the CF_3 case, the C-centred atom becomes more positive due to the electron-withdrawing nature of the substituents. Conversely, in the CH_3 and tBu cases, the positive charge was delocalised amongst the carbon and nitrogen atoms within the scaffold. Remarkably, in all cases, the HB of the NH moiety experiences an increase in positive charge, which facilitates and reinforces the HB interaction. Upon complexation, a charge distribution analysis revealed that for neutral compounds, the only significant change was observed in the carbon atom for CF_3 systems, which becomes less positive. However, for the charged counterparts, a more heterogeneous picture emerges, with the carbon (**C**) and **N1** atoms being the most affected. In the case of CF_3 and CH_3 systems, the central carbon atom generally becomes slightly more positive, while **N1** becomes more negative. Conversely, for the tBu substituent systems, both the carbon (**C**) and nitrogen (**N1**) atoms present an increase in negative charge and, therefore, become less positive. The variations in charge exhibited by

the diverse atoms during complexation align remarkably well with the NBO model, as depicted in Figures S2.1.4 and S2.1.5. Our findings indicate a discernible but not substantial charge transfer from the examined anions to the **N1** and **C** atoms ($LP(X^-) \rightarrow BD^*(NC)$). It is important to note that the magnitude of this charge transfer is consistently lower compared to the corresponding transfer observed in the HB interaction ($LP(X^-) \rightarrow BD^*(NC)$).

2.3.4 Ammonia- and Ammonium-based Models

In the subsequent phase of our study, we examined more realistic systems in which the theoretical establishment of HBs is hindered due to the presence of four substituents bonded to the nitrogen atom and the absence of any strong HB-donor (Figure 2.13).

These quaternary ammonium salts and their neutral counterparts serve as representative examples for exploring alternative modes of interaction in the absence of direct strong HB-donors.

The interaction energy values for both the cationic and neutral scenarios are compiled in Tables S2.2.7 and S2.2.8. Consistent with expected behaviour, charged species exhibit significantly higher values ΔE compared to their neutral counterparts, highlighting that the introduction of positive charges enhances the strength of interactions within these systems. Furthermore, charged systems consistently exhibit interaction energies lower than those of ionic compounds, and the separation distance between the positively charged ion N^+ and the negatively charged ion X^- varied between 3.35 and 4.50 Å, exhibiting the greatest distances observed amongst the various complexes investigated in this study.

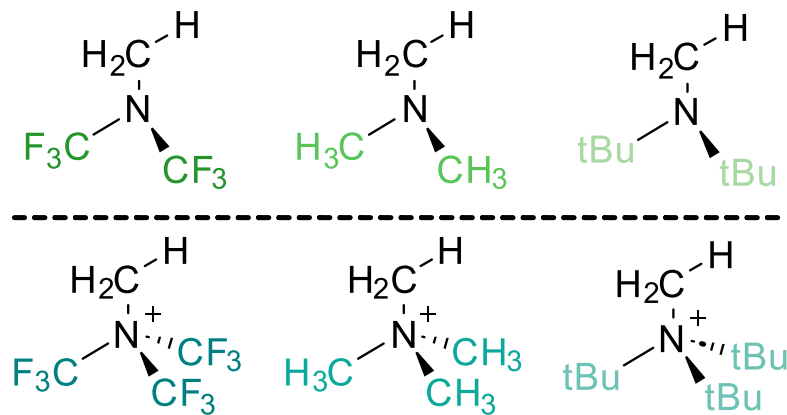


Figure 2.13. Ammonia- and ammonium-based models under study.

Characterisation of the Non-Covalent Interactions

When an LMO-EDA analysis is performed, in Figure 2.14 and Figures S2.2.7 and S2.2.8, the predominance of the electrostatic term for the charged complexes is clear.

A comparison of these values with those obtained in the previous section (iminium complexes) reveals a lower interaction energy than that observed for the strict IP systems discussed in the first section. However, compared to the $N(CF_3)_4-X$ systems, the values are found to be higher.

Furthermore, we find that the terms associated with polarisation, exchange, and dispersion, although lower than those observed for the iminium/imine systems, are higher than those observed for the ionic systems, indicating their non-negligible contribution. Additionally, the repulsion term is lower than that of the iminium-charged systems but higher than the $N(CF_3)_4-X$

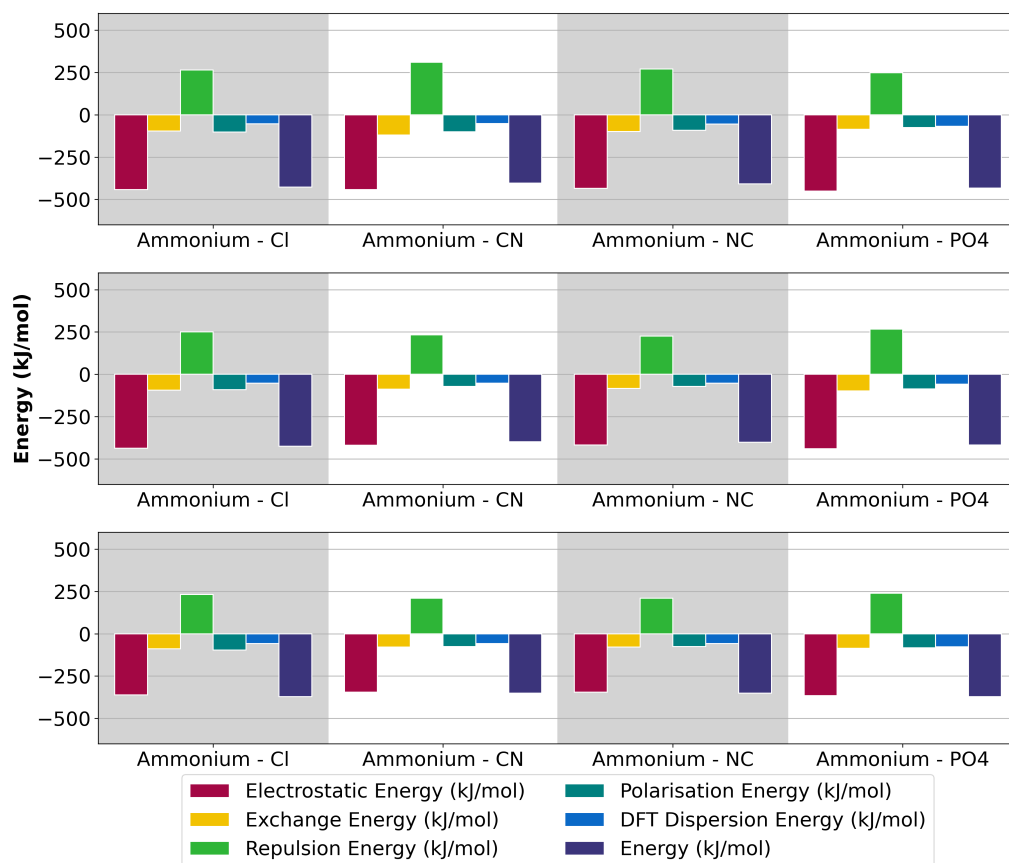


Figure 2.14. LMO-EDA energy partition terms for the different ammonium-based complexes under study. Top - Ammonium- CF_3 , middle - Ammonium- CH_3 and bottom - Ammonium-tBu.

systems. Thus, it appears that the ammonium salts studied herein exhibit a behaviour that lies between the HB-assisted IP and the longer-range IP.

The NBO analysis reported the presence of two primary types of charge transfers in the system under investigation. The first type is commonly associated with HBs and involves the donation from the LP of the anion to the anti-bonding orbital of the H-C bond ($LP(X^-) \rightarrow BD^*(C-H)$). The second type involves the donation from the LP of the anion to the centrally positioned nitrogen atom within the tetrahedral structure ($LP(X^-) \rightarrow BD^*(NC)$), which is assumed to carry a positive charge in ammonium-based models. The NBO results clearly indicate the establishment of weak to moderate HB interactions upon complexation, as well as a weak charge transfer from the LP of the anion to the orbital involving the central nitrogen atom (Figure 2.15 and S2.3.8).

The QTAIM analysis presented in Table S2.5.7 and S2.5.8 provides additional support for the aforementioned findings. Overall, the charged complex exhibits higher density and Laplacian values, $\nabla^2\rho(r)_{BCP}$, compared to the neutral ones, albeit falling within the expected range for such interactions. Figure 2.16 represents the molecular graphs specifically associated with ammonia_based_ $\text{CH}_3\text{-Cl}$ and ammonium-based_ $\text{CH}_3\text{-Cl}$. These graphs visually depict the distinct hydrogen bond interactions formed between the Cl^- ion and the hydrogen atoms of the CH_3 groups. The figures illustrate the electron density, $\rho(r)_{BCP}$, at various BCPs and also display the corresponding Laplacian values. The densities identified at the BCPs are comparatively lower than those observed within the iminium complexes because of the weaker HB-donor.

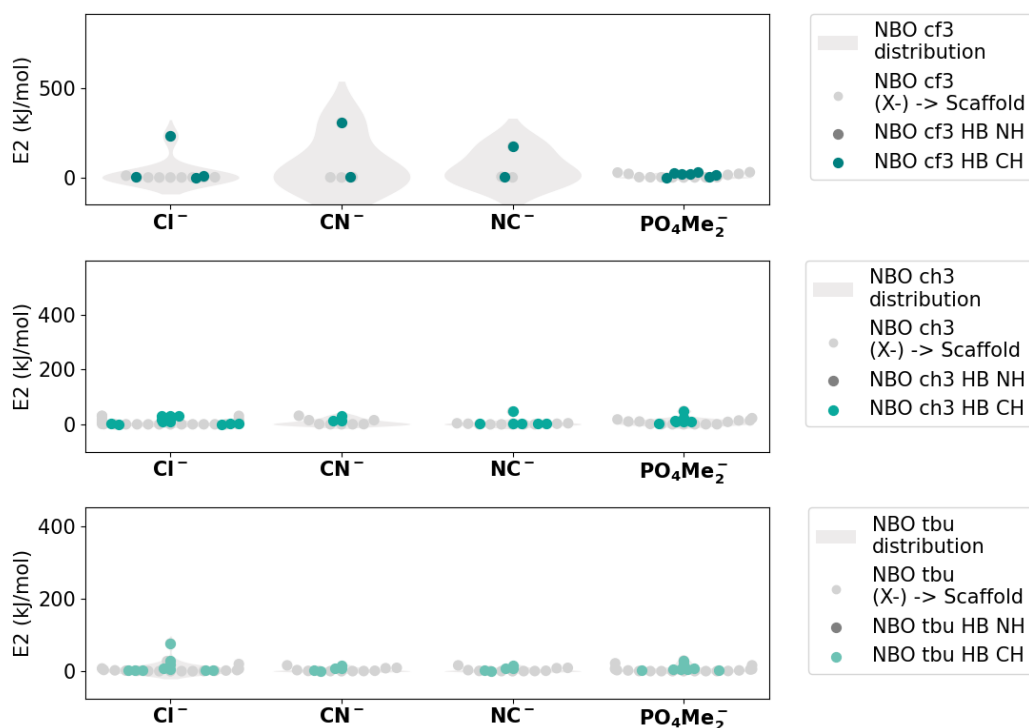


Figure 2.15. NBO E(2) values corresponding to the different ammonia_based complexes under study.

Additionally, the Laplacian values exhibit a similar range, yet consistently remain one order of magnitude lower when compared to the tight ion-pair formations detected in the salts.

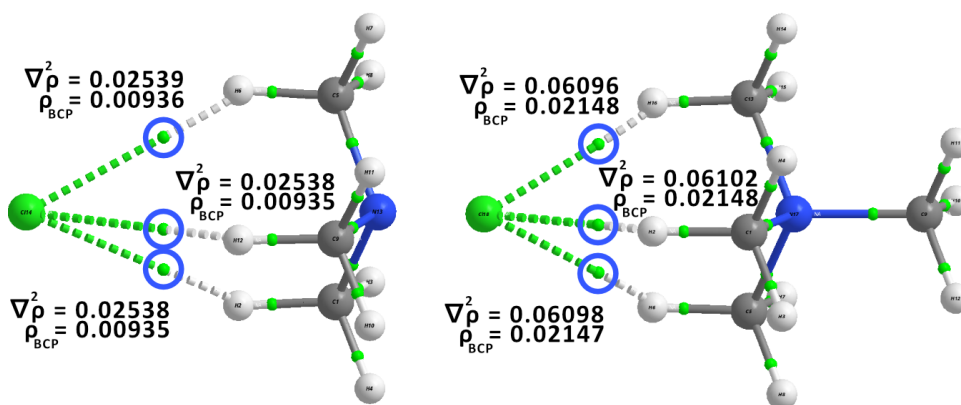


Figure 2.16. QTAIM molecular graph for ammonia_based_CH3-Cl and ammonium_CH3-Cl.

To finish this section, the charge difference within the monomers and upon complexation for both, neutral and charged complexes will be discussed (Figure 2.17 and S2.1.6 and S2.1.7).

The charge distribution of the monomers varies as they transition from neutral to charged states. Both the **N** and **H** atoms in the methyl group exhibit an increase in positive charge.

Upon complexation, a similar trend is observed for both neutral and charged compounds. However, this trend is more pronounced in complexes based on ammonium. The charge on the central **N** atom remains constant, whereas the charge on the **C1** atom experiences a significant decrease, becoming more negative in all cases, due to the interaction with the anion. In the case of the $\text{R}=\text{CH}_3$ substituent, **C1**, **C2**, and **C3** are interacting with the anion and so their charges

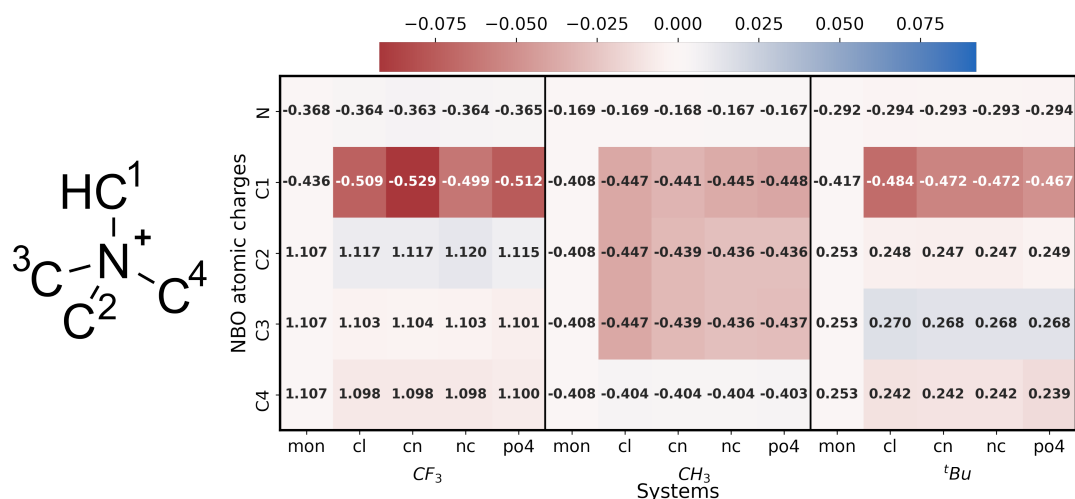


Figure 2.17. NBO charges to the different Ammonium-based complexes under study.

become slightly more negative.

The findings of our study are in perfect alignment with the literature. Specifically, in a seminal work published in 2002, Cannizzaro and Houk conducted a comprehensive evaluation of the interactions present in the Me_3NH-CH_2COOMe complex.³⁶ This complex, which served as an IP model, was examined using the MP2 level of theory.

The study identified the presence of an ion-pair arrangement stabilised by HBs (Figure 2.18) but also placed the positive charge shared amongst the different C atoms. The work by Cannizzaro and Houk provided significant insights into the nature and stability of IP complexes involving HB interactions, but it falls short of describing the nature of the interactions of the broad range of possible ion-pair interactions. As discussed, our current investigation builds upon the work of the authors by focusing on expanding the scope and applicability of the concept of ion-pair interactions in a broader context of PTC.

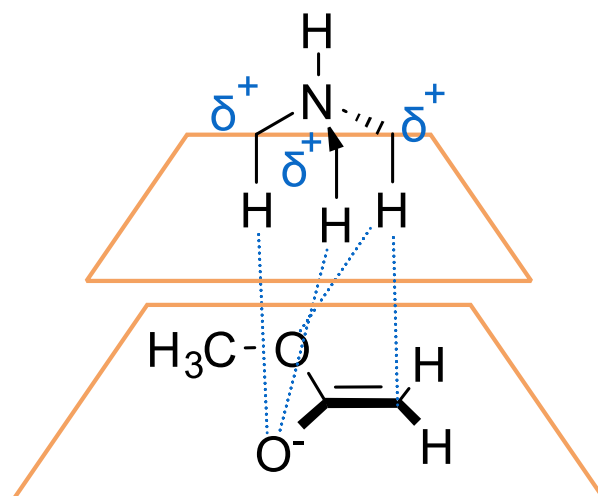


Figure 2.18. Non-covalent interactions established within the model Me_3NH-CH_2COOMe studied by Houk et al.

2.3.5 Phase Transfer Catalysts

As a final investigation, two real phase transfer catalysts were studied, **1** and **2**, Figure 2.19.

In this investigation, our focus was directed solely towards analysing the possible ion-pair interactions established in both systems. The primary objective of this scrutiny was to ensure an equitable comparison with the previously elucidated models.

The interaction energies obtained for both **1** and **2** are smaller than the models used in the previous sections based on ammonium and guanidinium (Table 2.5). Furthermore, the distances between the anion and the theoretically positively charged atom, N^+ , are smaller than those observed in the corresponding similar models examined earlier, both models based on ammonium and guanidinium. These distances range between 3.39 and 3.60 Å for **1** and, 2.54 and 2.97 Å for **2**, respectively.

We performed the same analysis of the interactions upon complexation, e.g. LMO-EDA, NBO, and QTAIM (Figures S2.2.9-10 and S2.3.9-10 and Tables S2.5.9-10 and S2.6.9-10).

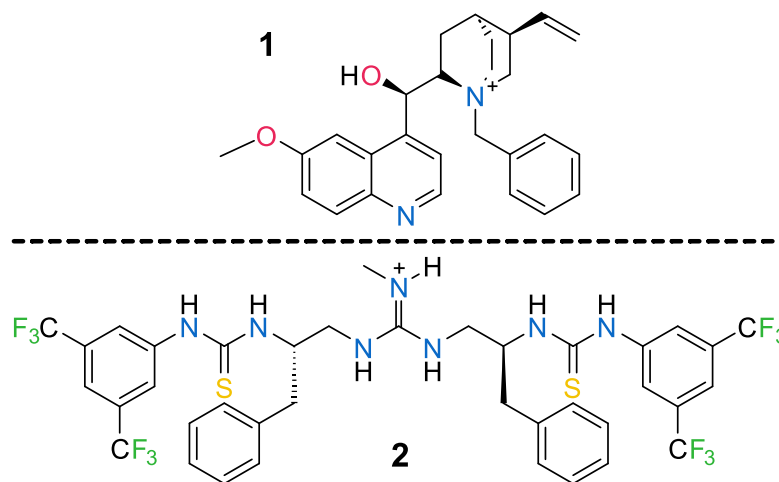


Figure 2.19. PTC catalysts under study, **1** and **2**.

All LMO-EDA values related to the nature of the interaction are found to be close to those found for the models studied previously, models based on guanidinium and ammonium, presenting a behaviour between the HB-assisted IP and the longer-range IP, as expected (Figure S2.2.9-10).

Table 2.5. Cinchona-based catalyst (**1**) and Guanidinium-based catalyst (**2**) DFT Interaction energies, ΔE in kJ/mol. Distances in Å.

Complex	ΔE	N^+X^-
1 -Cl	-304.2	3.55
1 -CN	-308.9	3.60
1 -NC	-312.7	3.39
1 -PO ₄ Me ₂	-225.5	3.47 (3.90)
2 -Cl	-326.2	2.97
2 -CN	-347.7	2.69
2 -NC	-331.0	2.64
2 -PO ₄ Me ₂	-188.4	2.54

In both cases, complexation leads to the formation of a web of weak HBs, which is closely aligned with the observed NBO interactions. For catalyst **1**, a network of hydrogen bonds forms between the anion and the C-H groups of the quinuclidine ring. The E(2) values in this

arrangement are quite comparable to the range observed in the ammonium-based ($R=CH_3$) system, approximately 10 kJ/mol. On the other hand, **2** exhibits two distinct HB-donors, resulting in two types of HBs, much like the small guanidinium-based model. The $E(2)$ values for this system are approximately -50 kJ/mol for NH as an HB-donor, while those that involve the CH group are weaker, around -10 kJ/mol.

Finally, to obtain a complete picture of the electron density, $\rho(r)_{BCP}$, values at the BCP involving the anion interaction for all the complexes studied, a graph was generated by plotting the density at the BCPs against the distance between the atoms involved in the interaction. For this particular correlation, all complexes were taken into account, including previous sections and current (Figure 2.20-top). Only the BCPs corresponding to the same pair of atoms involved were depicted for comparison. The results clearly demonstrate a nearly perfect exponential correlation, indicating a consistent pattern of NCIs across all complexes. Regarding the different anions under study, $PO_4Me_2^-$ systems present the highest density values, corresponding, therefore, to the strongest HB interactions.

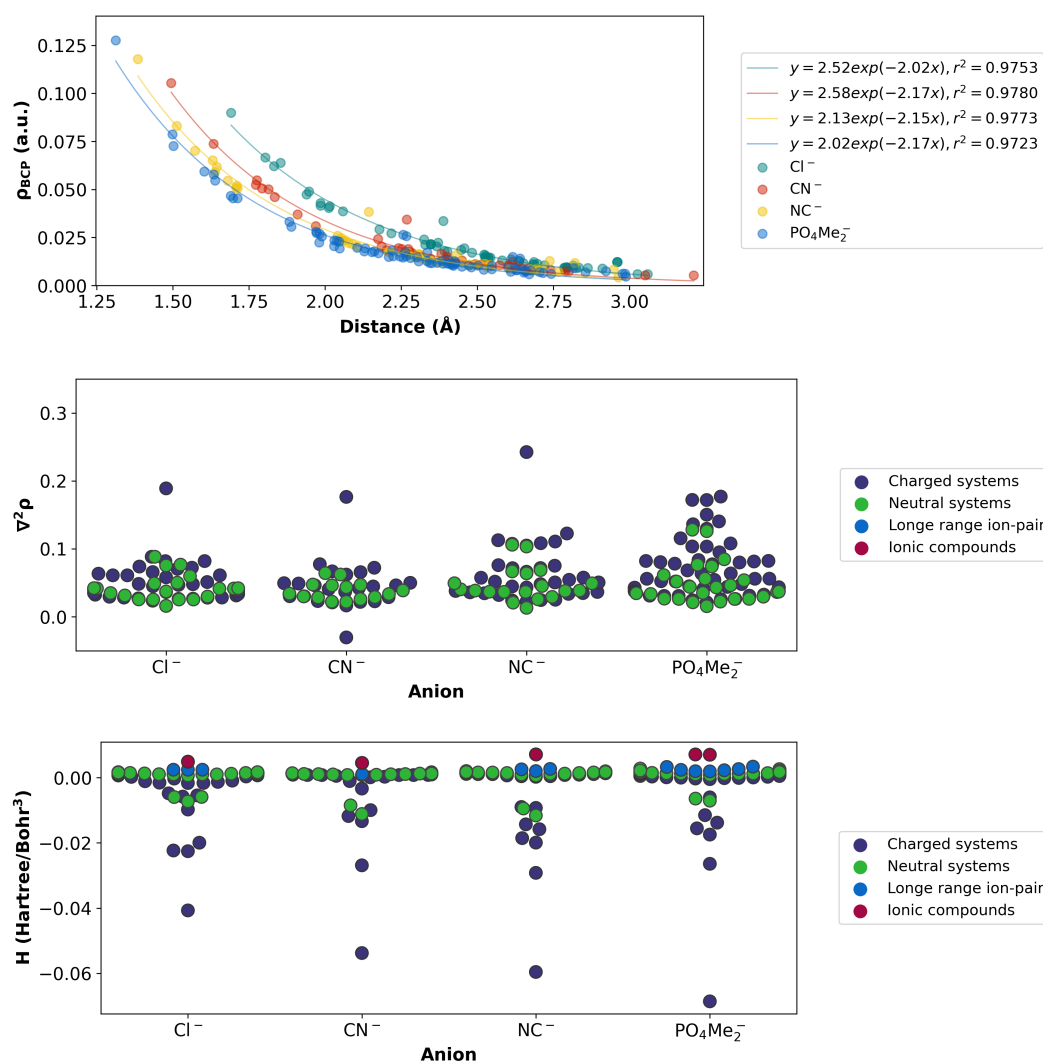


Figure 2.20. Top - QTAIM exponential correlation of the different models under study. Bigger dots correspond to the salts and $N(CF_3)_4-X$ systems. Middle - Laplacian values for all the different compounds under study. Bottom - Energy density values for all the different compounds under study.

More interestingly, the Laplacian values, $\nabla^2\rho(r)_{BCP}$, in all systems were investigated (refer to Figure 2.20-middle). Notably, the values obtained for all the systems analysed were found to be lower compared to those corresponding to the ionic complexes.

Finally, and in order to obtain a complete picture of the nature of the interactions upon complexation, the value for the energy density, H_{BCP} ,³⁷ was analysed for all the complexes under study (Figure 2.20-bottom). It is clear that the outcome is in excellent agreement with the Laplacian values, showing an H positive value for the polar compounds, while it is becoming slightly negative for the hydrogen bonds, and a few more negative values corresponding to a very strong HB, slightly covalent in nature, for the cases in which the phosphate is involved.

The general results from these analyses are in perfect agreement with the findings obtained from both previously studied models, guanidinium-based and ammonium-based. They demonstrate an intermediate behaviour between HB-assisted IP and longer-range IP interactions, underscoring the consistency and validity of the conclusions drawn from this analysis.

2.4 Conclusions

In this work, a comprehensive investigation of ion-pair interactions in the context of PTC was conducted. This study first focused on the examination of polar ionic compounds to define the concept of an IP in this context. This was then followed by the analysis of a diverse range of IP catalyst models to gain insight into the factors governing their interactions. Finally, it focused on the characterisation of real PTC catalysts, providing crucial insights into the behaviour of these systems in real applications.

This examination of these systems aimed to unravel the intricate interplay between hydrogen bonding and IP interactions, discerning the subtle differences that influence the nature and strength of hydrogen bonding and its role in facilitating IP formation.

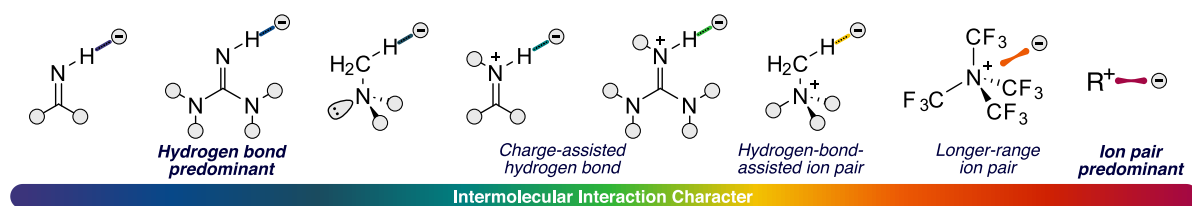


Figure 2.21. Different types of intermolecular non-covalent interactions under study in the work, within the phase transfer catalysis topic.

In particular, the thorough investigation of a wide range of PTC catalyst-based systems revealed distinct interaction characteristics: a strict ion-pair nature exhibited by salts, longer IP interactions, HB-assisted IP, and finally, charged-assisted HBs (Figure 2.21). It was demonstrated that the presence of HB-donors, either weak or strong, facilitates the formation of ion-pairs through one or an intricate network of HBs. It also provides directionality and structural stability to the IP interaction that, by definition, lacks such features. On the other hand, in the absence of these donors, longer-range IP spanning approximately 4 Å are established, whose electronic nature and distance-dependent behaviour we have characterised and discussed in depth.

In conclusion, this thorough investigation by means of state-of-the-art density-based theoretical methods revealed that the true interactions governing ion-pair systems undoubtedly depend on the existence of moieties capable of forming HBs. Given the presence of such structural motifs, structural stability and catalytic selectivity in PTC will be governed by one or several of these highly directional and charge-independent NCIs; in their absence, we have demonstrated that classical IP interactions are formed.

Therefore, a preliminary discussion of the baseline structure and potential intermolecular interactions of the PTC catalysts under study is imperative to unravel the nature of these systems: understanding the true interactions within the catalyst and how these go beyond the classic interpretation of ion-pairs is key to rationalise experimental observations and unlock the full potential that these systems hold.

To assist synthetic chemists in understanding ion-pair interactions, it is suggested to examine not only the system structure, where the presence of different interactions, such as HB is evident, but also consider the relative importance between coulombic interaction and other energy contributions. This involves naming the interaction under study based on their relative importance, as indicated by the scale proposed in Figure 2.21. This straightforward guideline

provides a practical means to discern whether newly designed synthetic systems predominantly involve coulombic interactions or if other NCIs play a significant role.

Finally, performing computational calculations to unequivocally determine the nature of these interactions is encouraged. This computational approach improves the precision in identifying and understanding the interaction between coulombic forces and other contributing factors in ion-pair systems.

References

- (1) K. Brak and E. N. Jacobsen, *Angewandte Chemie International Edition*, 2013, **52**, 534–561.
- (2) C. C. Sorensen, C. T. Kozuszek, M. A. Borden and F. A. Leibfarth, *ACS Catalysis*, 2023, **13**, 3272–3284.
- (3) M. Mahlau and B. List, *Angewandte Chemie International Edition*, 2013, **52**, 518–533.
- (4) D. Qian and J. Sun, *Chemistry – A European Journal*, 2019, **25**, 3740–3751.
- (5) J. E. Gillespie, A. Fanourakis and R. J. Phipps, *Journal of the American Chemical Society*, 2022, **144**, PMID: 36178308, 18195–18211.
- (6) T. E. Schirmer and B. König, *Journal of the American Chemical Society*, 2022, **144**, PMID: 36240496, 19207–19218.
- (7) L. Zong and C.-H. Tan, *Accounts of Chemical Research*, 2017, **50**, PMID: 28379012, 842–856.
- (8) D. Parmar, E. Sugiono, S. Raja and M. Rueping, *Chemical Reviews*, 2017, **117**, PMID: 28737901, 10608–10620.
- (9) R. J. Phipps, G. L. Hamilton and F. D. Toste, *Nature Chemistry*, 2012, **4**, 603–614.
- (10) J.-F. Brière, S. Oudeyer, V. Dalla and V. Levacher, *Chem. Soc. Rev.*, 2012, **41**, 1696–1707.
- (11) I. Iribarren and C. Trujillo, *Phys. Chem. Chem. Phys.*, 2020, **22**, 21015–21021.
- (12) Y. Marcus and G. Heftler, *Chemical Reviews*, 2006, **106**, PMID: 17091929, 4585–4621.
- (13) D. A. D. E. V. Anslyn, *Modern Physical Organic Chemistry*, CA: University Science Books, Sausalito, 2006.
- (14) K. Dong, S. Zhang and J. Wang, *Chem. Commun.*, 2016, **52**, 6744–6764.
- (15) J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.
- (16) F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.
- (17) F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
- (18) M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji et al., *Wallingford, CT*, 2016.
- (19) P. Su and H. Li, *The Journal of Chemical Physics*, 2009, **131**, 014102, DOI: 10.1063/1.3159673.
- (20) M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. A. Montgomery Jr, *Journal of Computational Chemistry*, 1993, **14**, 1347–1363.
- (21) M. Mandado and J. M. Hermida-Ramón, *Journal of Chemical Theory and Computation*, 2011, **7**, 633–641.
- (22) N. Ramos-Berdullas, I. Pérez-Juste, C. Van Alsenoy and M. Mandado, *Physical Chemistry Chemical Physics*, 2015, **17**, 575–587.
- (23) A. V. Marenich, C. J. Cramer and D. G. Truhlar, *The Journal of Physical Chemistry B*, 2009, **113**, 6378–6396.

- (24) E. D. Glendening, C. R. Landis and F. Weinhold, *Journal of Computational Chemistry*, 2019, **40**, 2234–2241.
- (25) F. Weinhold, *Discovering chemistry with natural bond orbitals*, John Wiley & Sons, 2012.
- (26) X. Fradera, M. A. Austen and R. F. W. Bader, *The Journal of Physical Chemistry A*, 1999, **103**, 304–314.
- (27) P. S. V. KUMAR, V. RAGHAVENDRA and V. SUBRAMANIAN, *Journal of Chemical Sciences*, 2016, **128**, 1527–1536.
- (28) L. Zhao, S. Pan and G. Frenking, *The Journal of Chemical Physics*, 2022, **157**, 034105.
- (29) P. D. McCaffrey, R. J. Mawhorter, A. R. Turner, P. T. Brain and D. W. H. Rankin, *The Journal of Physical Chemistry A*, 2007, **111**, PMID: 17579381, 6103–6114.
- (30) P. Su, Z. Tang and W. Wu, *WIREs Computational Molecular Science*, 2020, **10**, e1460.
- (31) P. L. A. Popelier, *The QTAIM Perspective of Chemical Bonding*, John Wiley & Sons, Ltd, 2014, ch. 8, pp. 271–308.
- (32) R. F. W. Bader, S. Johnson, T.-H. Tang and P. L. A. Popelier, *The Journal of Physical Chemistry*, 1996, **100**, 15398–15415.
- (33) C. Heering, B. Nateghi and C. Janiak, *Crystals*, 2016, **6**, DOI: 10.3390/cryst6030022.
- (34) J. Ni and J. J. Pignatello, *Environ. Sci.: Processes Impacts*, 2018, **20**, 1225–1233.
- (35) H.-G. M. Hait D, *ChemRxiv*, 2023.
- (36) C. E. Cannizzaro and K. N. Houk, *Journal of the American Chemical Society*, 2002, **124**, PMID: 12059242, 7163–7169.
- (37) D. Cremer and E. Kraka, *Angewandte Chemie International Edition in English*, 1984, **23**, 627–628.

3 | Improving phase-transfer catalysis by enhancing non-covalent interactions

3.1 Introduction

As shown in the general introduction, the quest for enantiomerically pure organic products has long been a focal topic in chemistry during the last few decades, particularly in medicinal chemistry, where chiral molecules play a pivotal role in drug synthesis. In addition, the recovery and reuse of catalysts are perspectives. Traditional approaches to the enantioselective production of chiral compounds involve the use of organometallic compounds or enzymes. However, since the pioneering work of Hajos and Perish in the 1970s,¹ showcasing the first highly enantioselective reaction catalysed by an organic molecule, the field has witnessed significant advancements. Due to the absence of transition metals, organocatalytic methods are attractive for the preparation of pharmaceutical compounds in which levels of certain metal-ion contamination are tightly controlled.

One of the most efficient and commonly used types of organocatalysis is Phase-Transfer Catalysis (PTC), detailed explained in the Introduction section of this thesis. Since Starks first introduced the term in 1971,² it has been widely applied in organic reactions in different immiscible phases. Asymmetric PTC, based on the use of different structurally well-defined chiral organic molecules as catalysts, has also attracted the attention of chemists and has been recognised as a versatile strategy for the preparation of chiral functional molecules. PTC presents several advantages over other types of asymmetric catalysis, such as organometallic or enzymatic catalysis, including milder and greener reaction conditions, resistance to moisture and oxygen, and absence of heavy metals. As a result, many novel organic transformations have been achieved with high enantioselectivity.³⁻⁷

Several catalyst scaffolds have been developed based on the Cinchona alkaloid and different chiral binaphthyls, including quaternary-onium and phosphoric acids, which have emerged among the most successful examples in PTC.

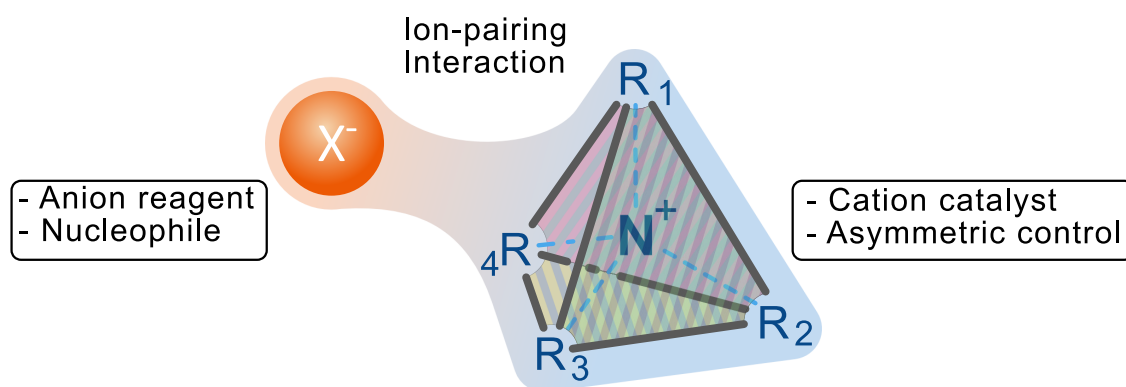


Figure 3.1. Traditionally interactions considered between the catalyst and the anion in PTC.

The traditional mechanism of action of PTC was described by Corey et al.⁸ in 1997, stating a stereochemical rationale in which the nitrogen of a Cinchona alkaloid quaternary salt is the centre of a tetrahedron and three of the four faces are sterically hindered by the alkaloid quinuclidine structure. Consequently, the PTC catalyst should be structured to exclude the anion approach to three of the faces of this tetrahedron, while the fourth face should be sufficiently open to assist in the ion-pairing interaction (Figure 3.1).

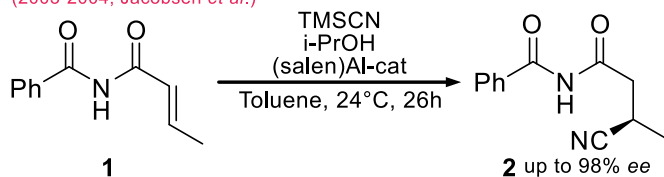
As described in the extensive synthetic literature,⁶ it was hypothesised that the ionic reactant in a PTC **cat_ion** complex is usually situated proximally to the charged nitrogen or

phosphorous atom, establishing a strict ion pairing interaction (Figure 3.1),^{9,10} and therefore in the proposed general mechanism of asymmetric PTC (Figure 3.1) the nucleophilic addition of an achiral anion to prochiral electrophiles occurs, and a new stereocentre is created.

One notable example of these nucleophilic additions is the cyanation of unsaturated prochiral compounds, such as α,β -unsaturated ketones.^{11–13} These transformations have been extensively studied in recent years, as their reaction products are considered highly desirable building blocks for pharmaceutical compounds.¹¹ Jacobsen^{14,15} reported the first highly enantioselective of one of these unsaturated compounds using chiral Al-Salen and bimetallic cooperative catalyst systems for the enantioselective conjugate additions of Trimethylsilyl Cyanide (TSCN) to α,β -unsaturated imides (Scheme 3.1, A).

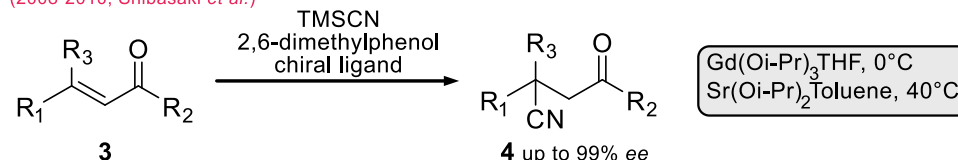
A- Conjugate Addition of TMS-CN of α,β -unsaturated imides with Salen-Aluminum Catalyst

(2003–2004, Jacobsen *et al.*)



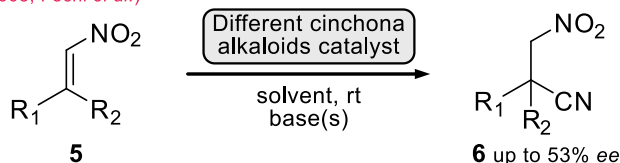
B- Asymmetric Conjugate Cyanation of α,β -unsaturated N-acylpyrrole and enones catalysed by gadolinium complex and SrII complexes with D-Glucose derived ligand

(2008–2010, Shibasaki *et al.*)



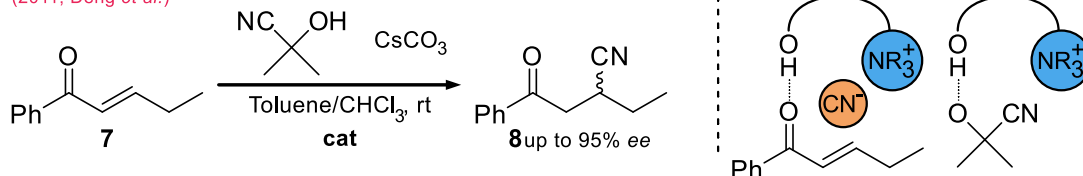
C- Conjugate addition of cyanide β,β' to disubstituted nitroolefins

(2008, Fochi *et al.*)



D- Proposed modes of activation by the bifunctional chiral catalysts

(2011, Deng *et al.*)



Scheme 3.1. Catalytic asymmetric cyanation studied in the synthetic literature.

Shibasaki^{16,17} and co-workers reported two bifunctional chiral catalysts derived from gadolinium, strontium, and different carbohydrate-based chiral ligands for a highly enantioselective 1,4-addition of cyanide to α,β -unsaturated N-acyl pyrroles and enones (Scheme 3.1, B), using TSCN as a cyanide source. In 2008, Fochi¹⁸ studied the first organocatalysed PTC conjugate addition of a cyanide ion to β,β' -disubstituted nitroolefins that leads to the efficient formation of an all-carbon quaternary stereocentre (Scheme 3.1, C) was studied by Fochi.¹⁸ Recently, Deng *et al.* presented the development of highly enantioselective catalytic 1,4-additions of cyanide with readily accessible and easy-to-handle cyanation reagents (Scheme 3.1, D).¹⁹ The strict ion-pair binding mode described in this study has been used in more recent projects such as the

one performed by Champagne and co-workers in 2022,²⁰ where they analysed different binding modes of conjugate cyanation of β -trifluoromethylated chalcones.

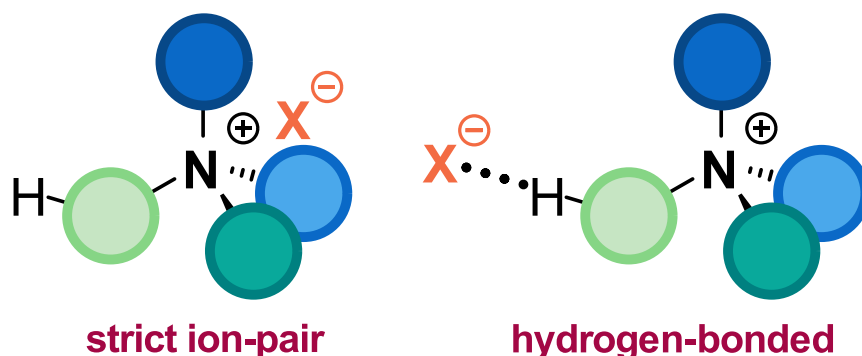


Figure 3.2. Two different interactions analysed in the present study.

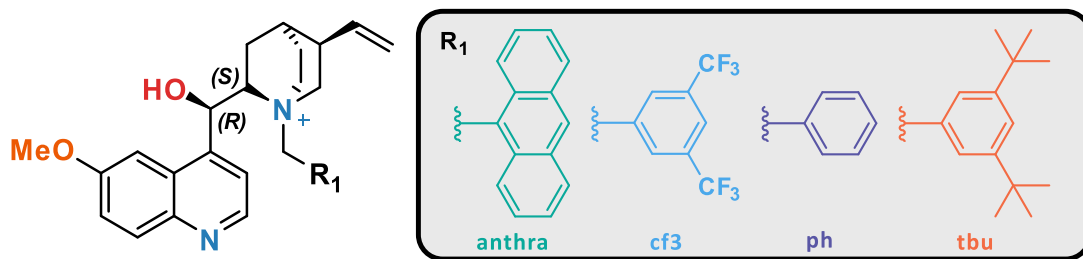
Although the field of asymmetric organocatalysis is currently growing exponentially, an understanding of the mechanistic details involved in most of these reactions has often lagged far behind the pace of catalyst development, which in turn slows the rational design of catalysts. Therefore, continuous efforts should be made toward the design and development of new catalyst classes, as well as understanding the existing relationships between the structure of the catalyst and its ability to transfer stereochemical information.^{10,21–26}

In this context, our objective was to provide a definitive binding mode for bifunctional PTC and uncover the origin of the selectivity achieved by these catalysts. To address this, we present a theoretical study that explores the various interactions that can be established between well-known alkaloid quinine-derived catalysts and three different cyanations of α,β -unsaturated ketones.

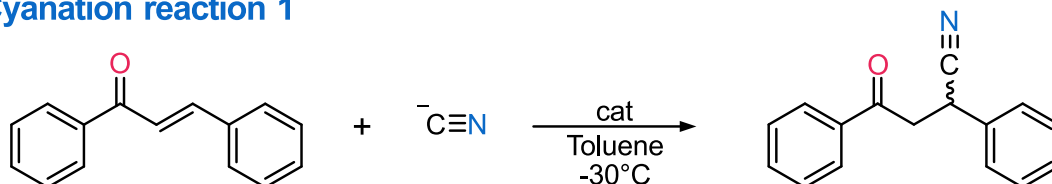
This chapter is divided into two different sections; the first section being a preliminary study to shed light on the binding mode question, studying two different activation modes for bifunctional PTC (Figure 3.2) and then using the new proposed binding mode to study one of these cyanation reactions, the asymmetric cyanation of E-chalcone (Scheme 3.2). The first binding mode studied is that in which a strict IP interaction is established (Figure 3.2 left), analogous to the one studied by the aforementioned authors; alternatively, the establishment of an intermolecular HB between the PTC counteranion and potential HB-donors was studied (Figure 3.2 right). These different binding modes were studied under the presence of different anions of interest in organocatalysis: Cl^- , Br^- , $MeCO_2^-$ and CN^- .

The comparative study and the results obtained from it that will be discussed in this chapter led to the development of the previous chapter in order to solve the need for a clear classification of all the interactions integrated under the ion-pair umbrella term.

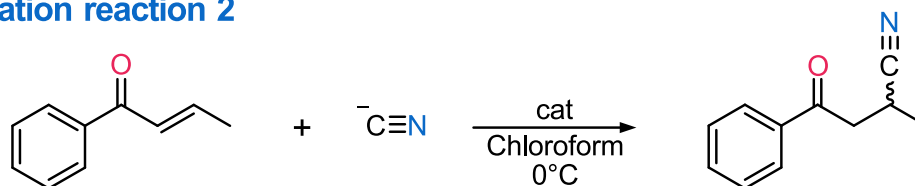
Due to the importance of the resulting products of these asymmetric cyanation reactions, the second half of this project investigated the conjugate asymmetric cyanation of other two distinct α,β -unsaturated ketones under these PTC conditions (Scheme 3.2). The investigation aimed to understand the impact of different substituents, varying in size and electron-donating or -withdrawing character, on the performance and selectivity of the phase transfer catalyst. Furthermore, the study explores the effects of introducing a second anchor point within the catalyst, providing insights into its influence on the reaction mechanism and the overall enantioselectivity of the process.



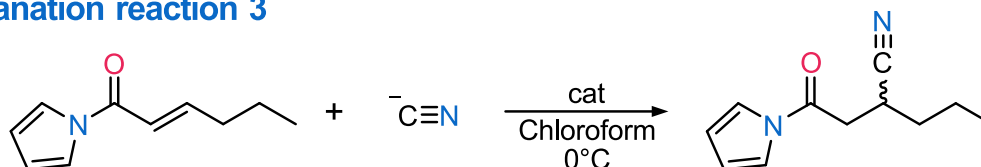
Cyanation reaction 1



Cyanation reaction 2



Cyanation reaction 3



Scheme 3.2. Catalysts and cyanation reactions of the different ketones under study. Ketone 1 - (E)-chalcone, ketone 2 - (2E)-1-Phenyl-2-buten-1-one and ketone 3 - (2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one.

Finally, as the different NCIs established between the catalyst and the reactants are the driving force of the selectivity, this chapter conducts an extensive analysis using various methodologies. The goal was to unravel the origin of the enantioselectivity of these catalysts, providing valuable insights for the rational design of future catalysts. This comprehensive computational study aligns the project with the contemporary approach in organocatalysis, where computational studies serve as a prospective tool for designing new catalysts rather than solely explaining experimental results retrospectively.

3.2 Computational Details

As was mentioned in the Introduction, this chapter is a combination of two different side projects, one being the continuation of the former, and, therefore, the computational details are separated into two different parts.

3.2.1 Binding Mode Study and Cyanation Reaction of (E)-chalcone

The structures of the complexes were optimised at the M062X/aug-cc-pVDZ²⁷⁻²⁹ computational level. Harmonic vibrational frequencies were computed at the same level used for geometry optimisation in order to confirm that the stationary points are local minima. Due to convergence issues, a 6-31g(d)^{30,31} basis set was used to perform the energy profile, instead of the previously used aug-cc-pVDZ basis set, since this specific Pople basis set was widely used in the framework of theoretical asymmetric catalysis.²⁶ A benchmark between both base sets was performed for both the conformational study and the catalyst-anion interaction study (Table S3.1), obtaining, in general, the same trend.

Single point energies for the lowest energy small basis set calculations were computed using the M062X/aug-cc-pVTZ computational level for the conformational analysis and the catalyst-anion complexes and M062X/6-311+g(d,p) for the reaction steps.

The effects of the solvent (toluene) were included in optimisation using a continuum method, the Solvation Model based on solute electron Density (SMD) approach³² and the refined SMD18³³ version for Br atoms.

The reported free energies ($\Delta G_{def2tzvp}$) in this document were computed by adding the thermal correction ($TC_{def2svp}$) from the small basis set calculations to the electronic energy obtained from the higher level single point calculations ($E_{def2tzvp}$), as shown in Eq. 1.

$$G_{def2tzvp} = E_{def2tzvp} + TC_{def2svp} \quad (1)$$

The interaction free energies ($\Delta G_{complex}$) were calculated as the difference between the free energy of the optimized complex and the energy of each monomer in its optimized geometry, as expressed in Eq. 2.

$$\Delta G_{complex} = G_{complex} - \sum G_{monomer} \quad (2)$$

All calculations were carried out using the Gaussian16 software package³⁴ and the optimised structures for all studied catalysts, complexes and reaction steps can be found in Table S3.2.

The Molecular Electrostatic Potential (MEP) of the isolated monomers was calculated on the electron density isosurface of 0.001 a.u. This isosurface was shown to resemble the van der Waals surface.³⁵ These calculations were carried out using Gaussian16 software and the numerical results were analysed using the Multiwfn program³⁶ and plotted using Jmol.³⁷

Based on the Curtin-Hammett principle³⁸ and according to the calculated Gibbs free interaction energies of the transition states, the theoretical *ee* value was determined based on Eq. 3 where $[R]$ and $[S]$ represent the concentrations of the different enantiomers.

$$ee = \frac{[R] - [S]}{[R] + [S]} \quad (3)$$

The *ee* value was determined as the ratio of the rate constants at $T = 243$ K as described in Eq. 4),³⁹ where $\delta\Delta G^\ddagger$ is the difference of Gibbs free activation energy between enantiomers, R is the gas constant, and T is the temperature.⁴⁰

$$ee = \frac{\exp\left(\frac{-\delta\Delta G^\ddagger}{RT}\right) - 1}{\exp\left(\frac{-\delta\Delta G^\ddagger}{RT}\right) + 1} \quad (4)$$

The NBO methodology⁴¹ version 7 was used to assess the strength of various NCI present in the systems, measuring the charge transfer between the occupied *LP* from the HB-acceptor and the empty *BD** orbitals of the HB-donor. The extracted E (2) energies can be found in Table S3.5.1.

The QTAIM^{42,43} methodology was used to analyse the electron density of the systems, carried out within the AIMAll software.⁴⁴ The electron density at the BCP between the catalysts and the different anions was examined. Results can be found in Table S3.6.1.

3.2.2 Cyanation Reaction of (2E)-1-Phenyl-2-buten-1-one and (2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one

The conformational analysis of all the catalysts was performed employing the RDKit Python library,⁴⁵ a reliable tool for catalysts of this nature.⁴⁶ The analysis was carried out using the *EmbedMultipleConfs* function. The number of conformers was specified as nr^3 , where *nr* represents the number of rotatable bonds. A pre-optimisation RMSD filtration threshold of 0.1 Å was applied and the *useExpTorsionAnglePrefs* and *useBasicKnowledge* options were activated. The generated conformers were then optimized using the MMFF94s force field.⁴⁷ During the optimisation process, an energy threshold of 10 kcal/mol was applied. Subsequently, the post-optimization RMSD filtration threshold was iteratively adjusted so the final number of remaining conformers was 10. All computations involving RDKit were carried out using version 2021.03.1 of the package.

All systems and complexes under study were optimised at the ω W97X-D⁴⁸/aug-cc-pVDZ^{28,29} computational level. Harmonic vibrational frequencies were calculated at the same level to validate the nature of stationary points as local minima.

Single point energies for the small basis set optimised geometries were computed using the ω W97X-D/def2-TZV(P) computational level for the conformational analysis and the complex interaction energies. To account for the effect of the solvent, a continuum approach based on the SMD³² was applied, using chloroform as the solvent.

All calculations were carried out using the Gaussian16 software package³⁴ and the optimized structures for all studied systems and complexes can be found in Table S3.3.

On the basis of the Curtin–Hammett principle,³⁸ the enantiomeric excess (*ee*) of a reaction is given by Eq. 5.

$$ee = \frac{k_{favoured} - k_{disfavoured}}{k_{favoured} + k_{disfavoured}} \quad (5)$$

However, there are different ways^{38,49,50} of calculating these reaction rates (*k*), depending on the characteristics of the reaction, the conformer interconversion rates, and the number of individual reaction paths considered. Since this is a very complicated hot topic and no general agreement on how to choose one of them when there are several transition states corresponding

to each product has been reached within the community, the two most common methods for calculating the *ee* have been used and compared in this paper.

For the first method, it is assumed that the reaction rate is much slower than the interconversion of the different conformers of the transition state and, therefore, they can be treated as an ensemble that follows a Boltzmann distribution. This method uses the average free energy ($\Delta\overline{G}^\ddagger$, Eq. 6)⁴⁹ of the transition states leading to each enantiomer to calculate the enantiomeric excess.

$$\Delta\overline{G}_i^\ddagger = \sum_j^{N_{TS}^i} p_j G_j^\ddagger \quad (6)$$

where p_i (eq. 7) is the population of the different TS leading to each enantiomer.

$$p_j = \frac{e^{(-\Delta G_j^\ddagger/RT)}}{\sum_j^{N_{TS}^i} e^{(-\Delta G_j^\ddagger/RT)}} \quad (7)$$

Finally, the reaction rate is calculated using these $\Delta\overline{G}^\ddagger$ values as shown in Eq. 8.

$$k_i = e^{(-\Delta\overline{G}_i^\ddagger/RT)} \quad (8)$$

In the second method,⁵⁰ it is assumed that all the TS happen in parallel, in different reaction paths that are not interconnected and the efficient reaction rate (k_{eff}) is calculated as the sum of the individual reaction rates for each enantiomer (Eq. 9).

$$k_i^{eff} = e^{(-\Delta G_{i,eff}^\ddagger/RT)} = \sum_j^{N_{TS}^i} e^{(-\Delta G_{i,j}^\ddagger/RT)} \quad (9)$$

The *ee* results calculated with both methods for all the catalysts under study can be found in Table S3.4. For the discussion in the main text, the first situation was assumed and the *ee* was calculated using the average free energy ($\Delta\overline{G}_i^\ddagger$, Eq. 6).

The NBO7 methodology⁴¹ was used to assess the strength of various NCIs present in the systems, measuring the charge transfer between the occupied Lone Pair (*LP*) from the HB-acceptor and the empty anti-bonding orbitals (*BD**) of the HB-donor.

All possible interactions between the different HB-acceptors (*LP* of N atom in *CN*[−] (*LP*(*N*)) and *LP* in O atom in the ketone (*LP*(*O*))) and the HB-donors of the catalyst (anti-bonding σ orbital of O-H bonds ($\sigma^*(OH)$)) have been analysed and their charge transfer energy, *E*(2), evaluated. All these values can be found in Table S3.5.2,3 and the sum of all the individual charge transfer *E*(2) in each HB (Eq. 10) is reported in the main text.

$$E(2)_{total} = \sum_{i=0}^N E(2)_{LP^i(X) \rightarrow \sigma^*(YZ)} \quad (10)$$

The QTAIM methodology was used to analyse the electron density of the systems, conducted within the AIMAll software.⁴⁴ The electron density at the BCP between the catalysts and the different reactants, as well as along the reaction coordinate in the different transition states, was examined. Results can be found in Table S3.6.2,3.

Similarly to what was done in the previous reaction, the NCI were analysed using the NBO

and the QTAIM methodologies at the theoretical optimisation level.

3.3 Results and Discussion

The results of this project are divided into two different subsections. The first part will explore the different binding modes between Cinchona-based catalysts and different anions and investigate the cyanation reaction of an α,β -unsaturated ketone, (E)-Chalcone.

The second will explore two additional cyanation reactions, studying the effect of different substituents in the catalysts over the selectivity and mechanism of the reaction, and the generalisability of the conclusions extracted from the previous study.

3.3.1 Conformational Analysis, Binding Mode Study, and Cyanation Reaction Investigation of (E)-Chalcone

This first section is divided into several components, focusing on different aspects of the systems and how they affect the mechanism of the reaction.

First, a conformational analysis of the catalysts is performed, exploring the different structures obtained by the rotation of the most relevant free bonds of its structure. Then, a comprehensive analysis of the complex formed between the catalyst and different anions was performed, exploring different binding mode positions and comparing their relative stabilities.

Finally, the cyanation reaction of (E)-Chalcone was used as a case study to understand the underlying mechanism of action of this family of catalysts, and a modification is proposed to enhance the selectivity of the process.

Conformational Analysis of the Catalyst

We began by exploring the low-energy chemical space of the catalyst by rotating three different parts of the catalyst, as plotted in Figure 3.3. Relative energies are summarised in Table S3.1.1.

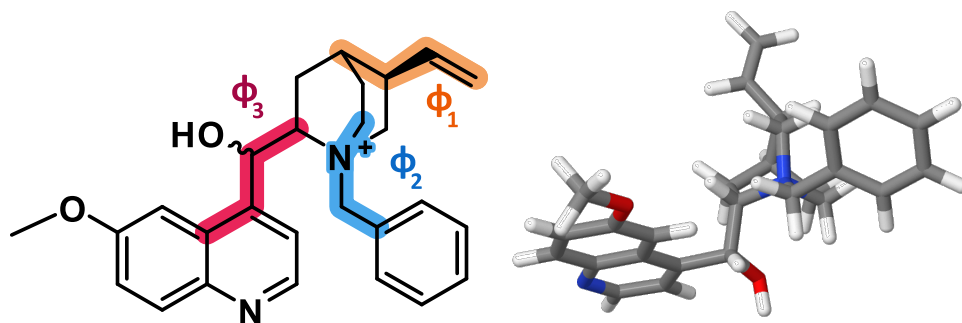


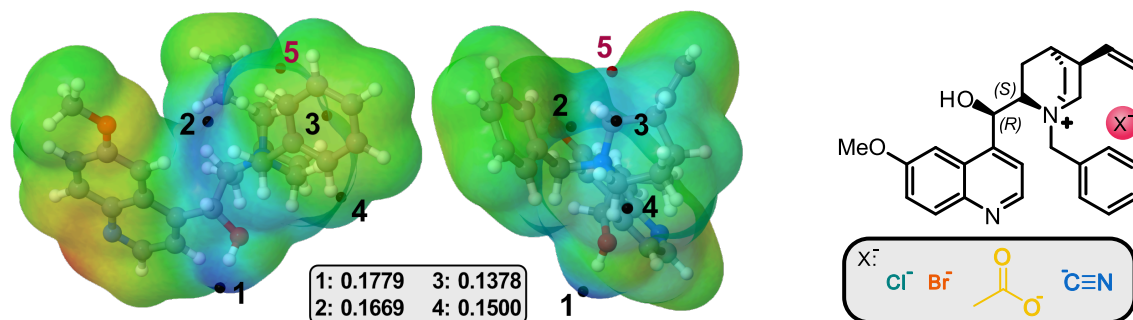
Figure 3.3. Left: Dihedral angles considered for the conformational analysis. Right: Lowest-energy optimised conformer for the catalyst under study, **conf_1**.

From the conformational analysis, 12 different conformers (Figure 3.3 and S3.1) were obtained. The relative free energies range from 0.1 to 8.6 kcal/mol. Following the different conformations, it is clear that compared to catalyst **conf_1** (no rotations made, Figure 3.3 right) when only φ_2 or φ_3 is rotated (**conf_5** or **conf_6**, respectively) the energy difference is ca. 3.3 kcal/mol, and therefore the value of ca. 7.2 kcal/mol is obtained when two dihedral angles are changed (**conf_11** and **conf_12**). Rotations of the dihedral angle φ_1 do not cause significant changes in energy outcome. **Conf_1** is the most stable one among all the 12 conformations studied, in complete agreement with studies reported in the literature by means of computational methods⁵¹ and X-ray techniques.⁷ The Boltzmann population of all conformers

was calculated (Table S3.7.1), which revealed that **conf_1** and **conf_2** constitute the majority of conformers, 47% and 46%, respectively. The difference between both conformers is the rotation of the vinyl group, φ_1 , and it is not relevant as it is not involved in the binding side.

Catalyst-Anion Interactions

With the optimal conformations in hand, we began by calculating the MEP surface of the isolated catalyst to analyse the areas susceptible to anion interactions. The values obtained for the maxima of the MEP on the 0.001 a.u. electron density isosurface are plotted in 3.4a.



(a) Molecular Electrostatic Potential of the catalyst under study. Potential maxima values (V_{max}) are represented as black dots, in a.u., and extra binding position, found in complexes, as a red dot.

(b) Structure of the different **cat-anion** complexes analysed in the present study.

Figure 3.4. MEP of the catalyst under study and different **cat-anion** complexes under study.

Four MEP maxima values (V_{max} , black dots) were found within the catalyst, one corresponding to the OH group (**pos_1**) and three localised around the quinuclidine ring (**pos_2**, **pos_3** and **pos_4**), where the positive charge is theoretically located (3.4a). It should be noted that the minimum in **pos_2** corresponds to the open face of the tetrahedron of the quinuclidine structure (Figure 3.3, left), the position in which, theoretically, the anionic reactant is located to establish an ion-pair with the catalyst. All V_{max} values are close, but the most positive corresponds to the OH group, meaning that it, therefore, corresponds to the most electrophilic area of the molecular isosurface. When the previous results are considered, it is clear that the hydrogen bond becomes a significant competitive interaction for the anion, and the interaction strength of the anionic reactant in both positions must be studied.

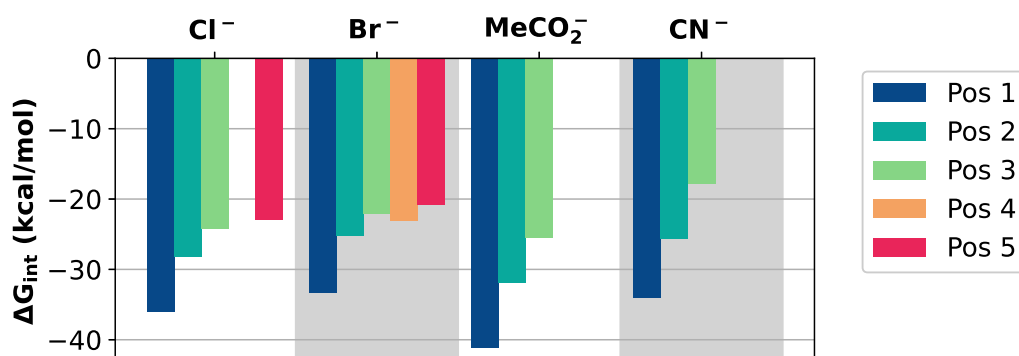


Figure 3.5. Interaction energies, ΔG_{int} , for all the different **cat-anion** complexes under study.

In order to provide insight into the different interactions that can be established with the

catalyst, a thorough computational study upon complexation was performed (3.4b). The nomenclature chosen for the formed complexes is **complex_X_N**, where **X** refers to the different anions involved ($X = Cl^-$, Br^- , $MeCO_2^-$ and CN^-) and **N** corresponds to the different positions obtained from the maxima values found in the MEP of the catalyst ($N = 1-4$, shown in 3.4b). For both halogen anions ($X = Cl^-$, Br^-), an additional complex $N = 5$ was found in a position that does not correspond to a MEP maxima position (Figure S). The free energy values of the interaction (ΔG_{int}) are presented in Figure 3.5 and range from -17.8 to -41.2 kcal/mol.

Regarding the different anions, the acetate anion complexes present the strongest interactions while the bromide ion complexes show the weakest interaction energy amongst all of them. It is clear that within the same anion–complex interaction, the position presenting an establishing intermolecular HB (**pos_1**) exhibits the strongest interaction. If a comparison is made between the HB-bound complex (**pos_1**) and the one exhibiting a strict IP interaction, in which both charged atoms are closest (**pos_2**), the former is, in general, more than 7 kcal/mol more stable compared to the latter, even though the $X^- \cdots N^+$ distance is shorter (Table S3.8).

Characterisation of the Intermolecular Interactions

Subsequently, the QTAIM analysis was carried out in order to analyse and compare the strength of the NCI established. The main electron density properties at the BCPs of the different HBs found in these complexes within the BCP analysis are summarised in Table S3.6.1. Molecular graphs for all the complexes are summarised in Figures S3.2.1 and S3.2.2. The values of the electron density at the BCP, ρ_{BCP} , were found to range from 0.079 to 0.004 a.u. The large ρ_{BCP} values indicate strong hydrogen bonds, while the positive values of $\nabla^2\rho_{BCP}$ are characteristic of closed-shell systems.

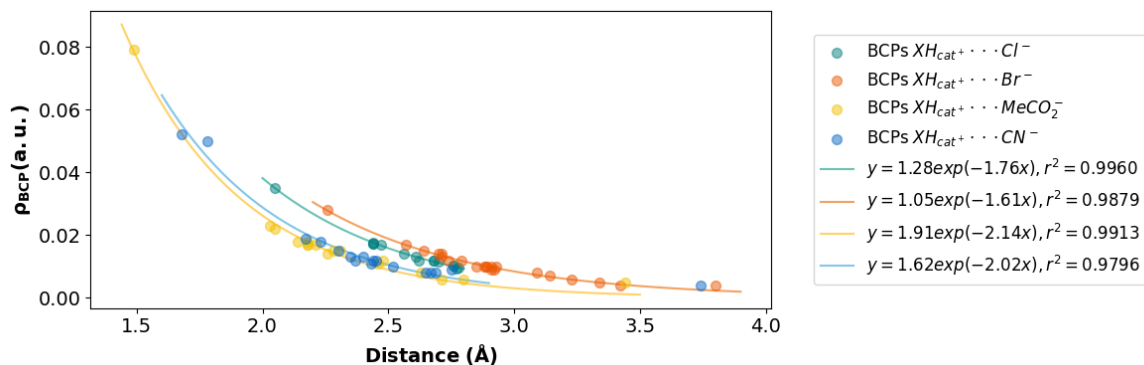


Figure 3.6. Exponential relationships of the intermolecular distance between the catalyst and the anion, in Å, with ρ_{BCP} .

From the molecular graphs and the analysis of the density at the BCPs, it is clear that the different anions can establish either a strong intermolecular HB with the hydroxy group of the catalyst ($OH \cdots X^-$) or mild HB assisted IP with weaker hydrogen bond donors such as CH groups ($[N^+]CH \cdots X^-$) from the quinuclidine ring of the catalyst. The values of both ρ_{BCP} and $\nabla^2\rho_{BCP}$ increase with the intermolecular hydrogen bond established between the anion and the OH group of the catalyst as shown in Table S3.5.1. In addition, the strength of the most relevant intermolecular HB ($OH \cdots X^-$) is also dependent on the nature of the anion. It becomes weaker in the order of $MeCO_2^- > CN^- > Cl^- > Br^-$, evidenced by the interaction energy,

the binding distance, and the electron density. That pattern is in complete agreement with the electronegative trend; the greater electronegativity of the HB-acceptor will lead to an increase in HB strength.

Exponential relationships between the intermolecular distance with the anion and ρ_{BCP} in each BCP were found (Figure 3.6), showing the decay of the electron density with distance.

To provide a visual description of the electron density overlap established in the intramolecular interactions, 3D-NCI plots for the most representative examples (**complex_X_1**) were obtained and are presented (Figure 3.7). Besides, the 2D-NCI plots are depicted in Figure S3.3.

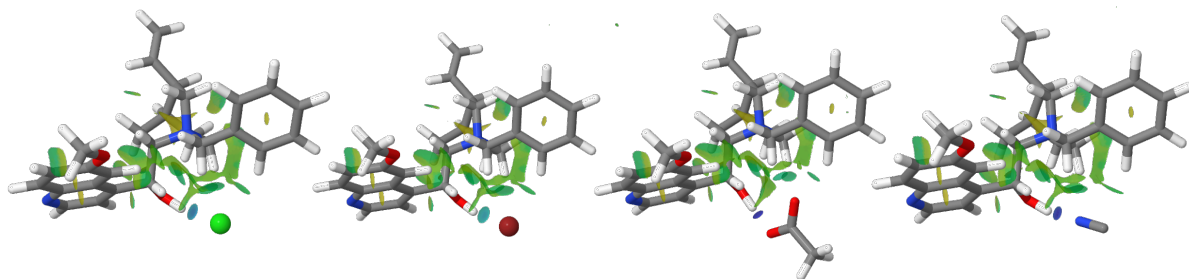


Figure 3.7. 3D-NCI plots of the interactions found for the most stable complexes **complex_X_1**. From left to right: Cl^- , Br^- , $MeCO_2^-$ and CN^- . RDG isosurfaces provide a 3D representation of interaction regions. The sign of the second eigenvalue of the density Hessian matrix ($\nabla^2\rho$) is used to differentiate repulsive ($\lambda_2 > 0$) from attractive ($\lambda_2 < 0$) interactions.

In all cases, a green area corresponding to the values of $\lambda_2 \approx 0$ (weakly attractive) appears between the anions and the catalyst indicating the interactions taking place. In the case of $MeCO_2^-$ and CN^- , two blue areas $\lambda_2 < 0$ (strongly attractive) are shown to coincide with the position of the established hydrogen bond interaction. Those are consistent with the ΔG_{int} values shown for both anion complexes. Furthermore, the Cl^- complex presents small blue areas, which are weaker compared to the other two anions but stronger than those in the bromide case. The 2D-NCI plots obtained show an intense peak corresponding to both anions $MeCO_2^-$ and CN^- , with more negative values of λ_2 .

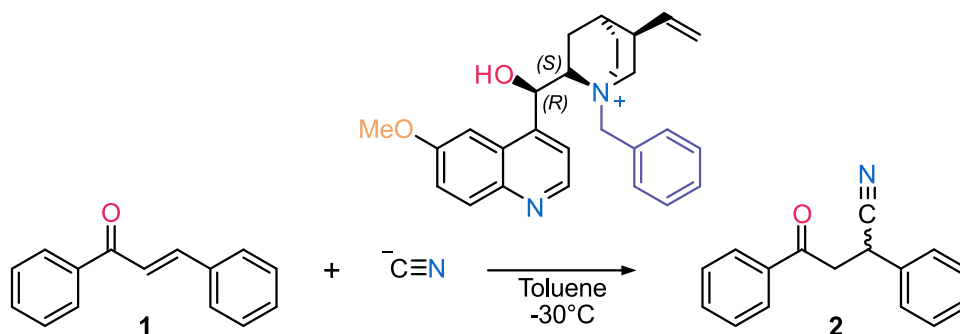
Finally, NBO analysis was used to identify and characterise the intermolecular charge transfer between the occupied LP molecular orbitals of the anion ($LP(X^-)$) and the empty BD^* orbitals within the catalyst ($\sigma^*(OH)$) upon complexation. In Table S3.5.1 the second-order perturbation energies $E(2)$ extracted from the found complexes are reported. In all anion-HB established in the series, the orbital interaction decreases following the interaction energy trend ($MeCO_2^- > CN^- > Cl^- > Br^-$). The smallest contributions of the series come from the ion pair interaction corresponding to those complexes in which both ions present the larger interatomic distance (**complex_X_3** and **complex_X_5**).

Cyanation Reaction Investigation of (E)-Chalcone

Having obtained a clearer understanding of the different interactions that can be established between the catalyst and the different anions, an application was studied to improve the PTC performance of **cat_OMe_ph**.

We began by studying the pathway of the enantioselective conjugate cyanation of α,β -unsaturated ketone, (E)-Chalcone (**1**), in the presence of the catalyst studied previously, **cat_OMe_ph**, and then a theoretical modification of the catalyst was made, **cat_OH_ph**.

(Scheme 3.3). In order to improve the enantioselectivity of this reaction, an additional anchorage point (C6'-OH) was added. Although the reaction was not performed experimentally, the reactants, as well as the conditions, were shown to be widely used in previous synthetic studies (Scheme 3.1) but not under the promotion of this particular catalyst, **cat_OMe_ph**.



Scheme 3.3. Asymmetric conjugate cyanation of (E)-Chalcone.

Only the first step, the addition of CN^- , was studied since it is responsible for the stereochemistry of the global process. Therefore the global product of the free-energy profile is the cat-bound enolate (**2**) shown in Scheme 3.3.

The free-energy profile for the formation of both enantiomers of **cat_OMe_ph**-bound enolate product (**2**) (Figure 3.8, left) indicates that the barrier corresponding to the C-C bond formation is very similar for both pathways, around 19 kcal/mol ($\delta\Delta G = 0.1 \text{ kcal/mol}$). It is clear from the free-energy profile that no steric issues are presented by the two transition state (TS)s (*R* and *S*) and therefore both are energetically accessible by the cyanide.

Regarding the cat-enolate complexes (**2**), they present similar energies to the pre-transition state (pre-TS) assembly, with the corresponding enantiomer *S* being slightly more stable than the *R* one.

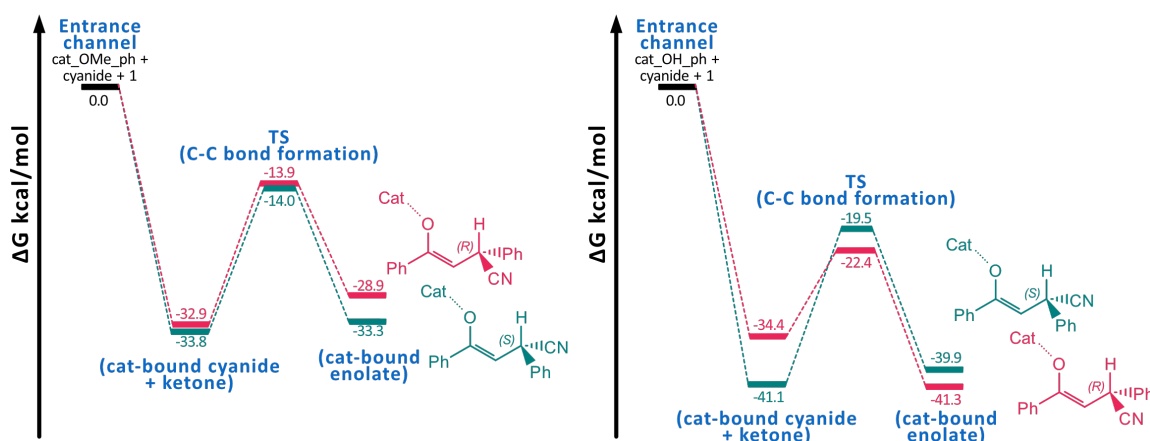


Figure 3.8. Left: Free-energy profiles for the cyanation of **1** in the presence of **cat_OMe_ph**. Right: Free-energy profiles for the cyanation of **1** in the presence of **cat_OH_ph**.

When the OMe group is replaced by a HB-donor group, OH (**cat_OH_ph**), a second anchoring point is added, forcing the intermolecular hydrogen bond between the two moieties, the new hydroxy group of **cat_OH_ph** and the oxygen of (E)-Chalcone, respectively. The free energy profile study was performed (Figure 3.8, right), showing that the energetic barrier between the two enantiomers had increased up to 2.8 kcal/mol. Regarding the four energy

pathways studied, in three of them, there is a possibility of reversibility, as the products are slightly less stable than the reactants. However, it should be noted that cat-bound enolate is not the final product of the reaction, since a protonation has to take place and will very likely stabilise the final product.

To investigate the origin of the observed asymmetric induction, we examined the transition states associated with the key stereocentre forming the addition step for both major enantiomers (Figure 3.9). The binding mode is analogous for both TSs. Inspection of the calculated TS leading to the major antipode provides an understanding of why the simple modification, 6'-hydroxylation, is superior in this process and also sheds light on the catalyst-substrate interactions. **Cat_OH_ph** forms a discernible non-covalent network identified by QTAIM calculations (Figure S3.2.2) in the TS. This, together with the binding of the OH proton to one of the carbonyl groups, aligns the electrophile.

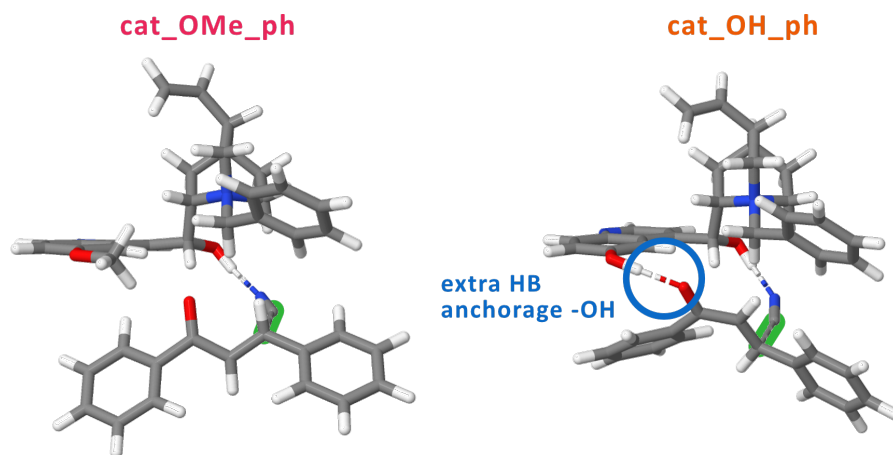


Figure 3.9. Optimised geometric for both TSs in the presence of **cat_OMe_ph** (left) and **cat_OH_ph** (right) for the major enantiomer in each reaction (*S* and *R* respectively).

3.3.2 Conformational Analysis, Binding Mode Study, and Cyanation Reaction Investigation of (2E)-1-Phenyl-2-buten-1-one and (2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one

The results of this second section are divided into several key components, focusing on different crucial aspects of the reaction process that must be studied.

First, a comprehensive analysis of the conformational analysis in the different modified catalysts under study and the two different α,β -unsaturated ketones is presented. This analysis sheds light on the relative stability, abundance, and accessibility of the different conformers.

Next, the diverse complexes between the catalysts and the different reactants are explored. Two different binding modes, previously seen in the literature, are compared, and a definite binding mode is identified.

Finally, two distinct asymmetric conjugate cyanation reactions of α,β -unsaturated ketones are explored, elucidating the underlying mechanisms and identifying key factors influencing the selectivity of the process and favoured enantiomeric product. Four different modifications in R_1 position are studied (**anthra**, **cf3**, **ph** and **tbu**) as well as two different modifications in R_2 position (**OMe** and **OH**). The different NCIs present in the systems are analysed, and their role in the process is analysed.

Conformational Analysis of the Catalysts

The initial phase of this project involved conducting a conformational analysis of the different catalysts under investigation (see Figure 3.10) to determine their most stable conformation. An automatic conformational search, pre-optimisation, and RMSD filtration using an RDKit implementation⁴⁶ (see computational details) was performed following the procedure described in the first chapter of this thesis. The resulting conformers were re-optimised by means of DFT method to obtain accurate energies for comparison, the energy values for each conformer are listed in Table S3.7.2.

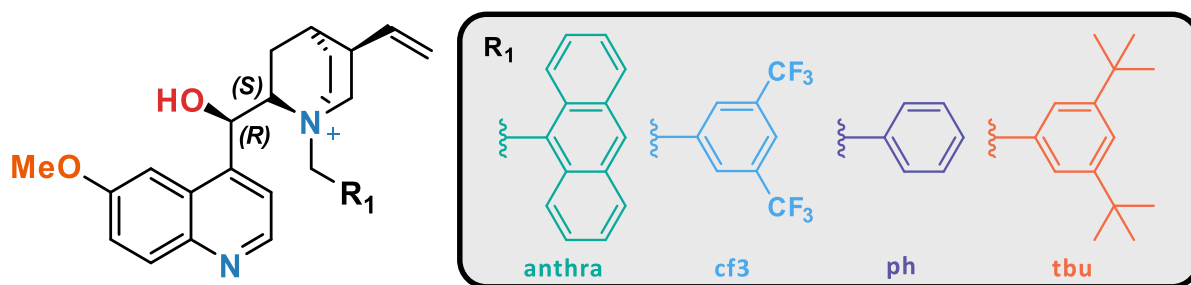


Figure 3.10. Catalyst scaffold and substituents (**anthra**, **cf3**, **ph** and **tbu**) under study and most important rotatable bonds are highlighted.

The analysis revealed that similar conformational ensembles are obtained for all the generated conformers, with the largest changes occurring due to the rotation of φ_1 , φ_2 , and φ_3 (as shown in Figure S3.1.2). However, we found that not all of these dihedral angles exhibit the same degree of rotational freedom. Specifically, φ_1 can be found in nearly any angle between 0° and 360° , while φ_2 generally collapses into two distinct positions, causing the $R_2 = \text{OMe}$ group to point in opposite directions. Additionally, φ_3 remains mostly static across the various conformers due to the steric hindrance created by the quinuclidine substructure. These findings are illustrated

in Figure 3.11 left and Figure S3.1.2.

Ideally, in order to have an efficient catalyst, its active conformer and its most stable conformer should match, facilitating the interaction with the reactant and enhancing its performance. However, that is not always the case with organic catalysts that present such rotational freedom and in which the solvent and the temperature have such a strong effect on the conformational stability. Hence, the active conformation of an organic catalyst must not be insignificant, even if it is not the most abundant one.

Looking at **cat_OMe_anthra** (Figure 3.11 centre, right), it is observed that the most abundant conformer (*conf_173*, *Boltzmann_population* > 66%) is, as well, the active form of this catalyst.

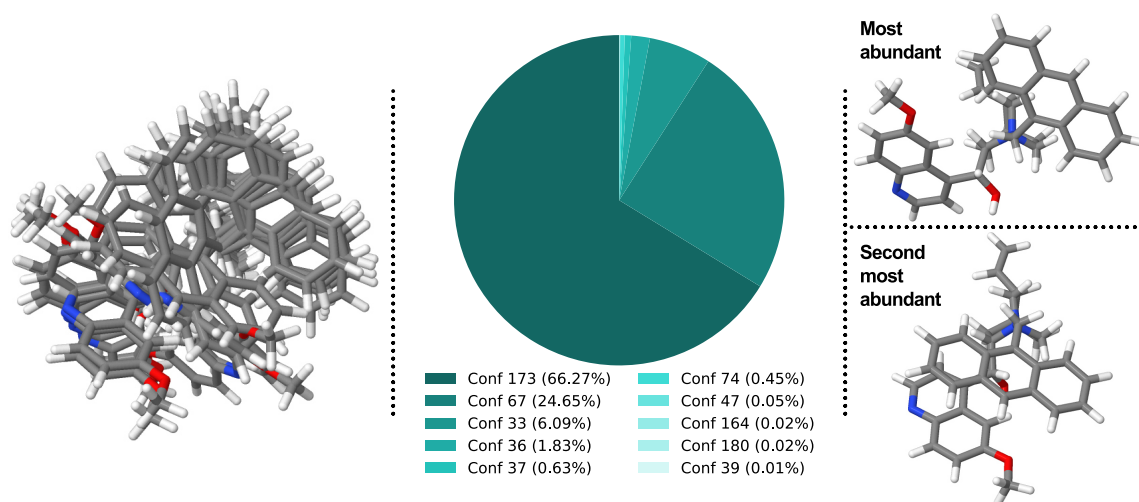


Figure 3.11. Conformational analysis of **cat_OMe_anthra**. From left to right: conformer ensemble, Boltzmann population of each conformer at DFT level and 3D structures of the two most abundant conformers.

From all the studied catalysts, this ideal situation where the active conformer and the most abundant conformer coincide happens in five of the systems under study (**cat_OMe_anthra**, **cat_OMe_ph**, **cat_OH_anthra**, **cat_OH_ph** and **cat_OH_tbu**). Moreover, the most abundant conformer in three of those ideal cases can also be considered as the sole conformer present in solution (**cat_OH_anthra** > 99%, **cat_OH_ph** > 95% and **cat_OH_tbu** > 99%). In the case of those catalysts in which the most abundant conformer and the active one are different, the active one is always either the second or the third most abundant and it presents a non-negligible population, being adequate to be considered as catalysts (**cat_anthra_cf3** > 19%, **cat_anthra_tBu** > 14% and **cat_OH_cf3** > 9%).

Taking these results into consideration, it can be assured that all the catalysts under study can be considered valid catalysts from the conformational perspective since their active conformation is the most abundant or not negligible in all cases.

However, catalysts are not the sole components that must be examined from a conformational perspective. The α - β -unsaturated ketones are also capable of generating different conformers and must be studied in order to determine which of those conformers must be taken into account for a thorough and rigorous study of the reaction.

Conformational Analysis of α - β -unsaturated ketones

Generally, α,β -unsaturated ketones are commercially available only in their *trans* isomer and, therefore, this is the only isomer that was studied. Additionally, in terms of the relative position of the two double bonds, *s-trans* conformers are usually more stable than their *s-cis* counterparts. However, due to the presence of (big) aromatic rings within the structures (Figure 3.12a) this assumption is not valid for these systems and, therefore, the relative stability of the two different conformations must be studied by means of DFT calculations.

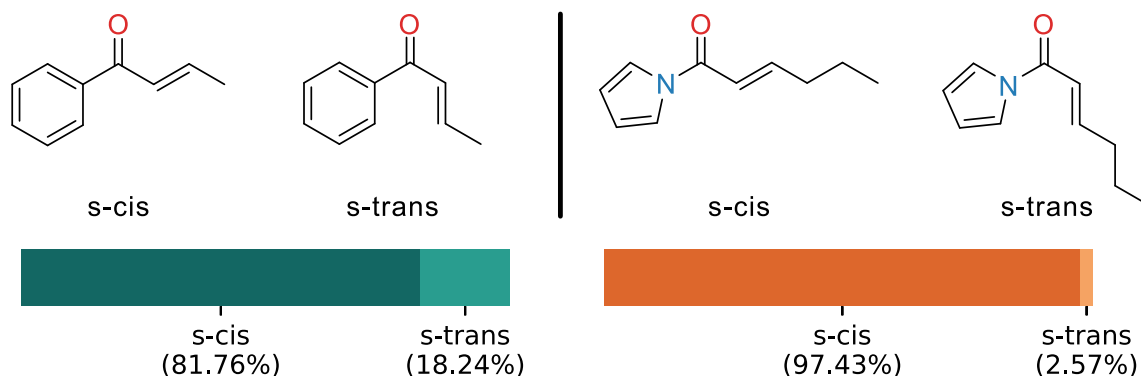


Figure 3.12. a) α,β -unsaturated ketones under study: **(2E)-1-Phenyl-2-buten-1-one** left and **(2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one** right. b) *s-cis* and *s-trans* conformers' populations of each α,β -unsaturated ketones.

Considering the conjugated nature of these ketones, only *s-cis* and *s-trans* conformers need to be considered, as any intermediate structure will collapse in one of those conformers almost immediately. Consequently, there is no need to perform a complete conformational analysis of these molecules.

The distribution of the population in both ketones is similar (Figure 3.12b), being the *s-cis* conformer the most stable with an 81.76% abundance in the case of **(2E)-1-Phenyl-2-buten-1-one** (ketone 1) and 97.43% in the case of **(2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one** (ketone 2). The Boltzmann population of the different conformers of these ketones is crucial because they determine the accessibility of the various possible binding modes that will be studied and, therefore, must be included in the calculation of the enantiomeric excess (%*ee*) of the process.

Looking at the Boltzmann populations of the studied α,β -unsaturated ketones it can be observed that the *s-trans* conformer is much less abundant than the *s-cis* conformer. These results, opposite to what usually happens with smaller substituents, where the *s-trans* conformer is the most abundant, are caused by the steric clash between the aromatic substituents and the unsaturated bond in the *s-trans* conformation.

Transforming an α,β -unsaturated ketone from the *s-cis* conformer to the *s-trans* implies breaking the delocalised *pi* system and, therefore, the rotational barriers or the process needs to be calculated in order to determine whether the interconversion or the two conformers is possible. The calculated barriers for these rotations are 4.26 kcal/mol for **(2E)-1-Phenyl-2-buten-1-one** (1) and 3.81 kcal/mol in the case of **(2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one** (2). Hence, the interconversion of the two conformers of the ketones is possible. Therefore, since the two different conformers of the ketone are accessible it is reasonable to assume that the transition states of the reaction will follow a Boltzmann distribution.

Binding Mode Study

After verifying the validity of all catalysts for this specific study and assessing the stability of various conformers of the reactants, the subsequent phase involves investigating diverse interactions and potential binding modes between the catalysts and the reactants.

In a previous investigation carried out by our group,⁵² it was established that, among the multiple interaction sites within the catalyst, the most stable and, thus, the only significant binding mode between the catalyst and the anionic reactant CN^- was through the formation of a hydrogen bond between the OH group within the catalyst and the anion. Hence, this interaction site was the sole focus of this study.

Nevertheless, an alternative scenario could involve the ketone, rather than the CN^- anion, being bonded to the OH position. To ensure the prevention of this situation, it becomes imperative to analyse the binding energy of both the **cat-ketone** and **cat-CN** complexes (as illustrated in Figure 3.13).

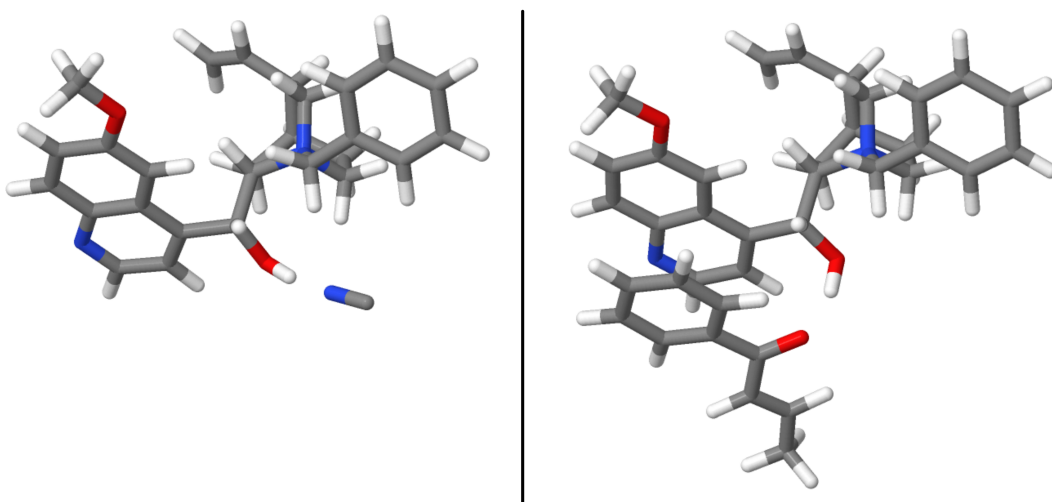


Figure 3.13. 3D structures of the **cat-CN** (left) and the **cat-ketone** (right) complexes.

The binding energies of complexes (20.3 kcal/mol for complex **cat-CN** and 3.6 kcal/mol for complex **cat-ketone**) show that **cat-CN** is one order of magnitude more stable than **cat-ketone**, even though the latter combines a HB with a π - π interaction. Consequently, as previously mentioned, the only possible binding modes that were studied were those in which the CN^- anion is forming a HB with the catalyst.

Finally, even though the complex **cat-ketone** is much less stable than the complex **cat-cn**, a transition state coming from each complex was explored (Figure 3.14).

The first binding mode, analogue to the binding mode studied in our recent study aforementioned,⁵² comes from the complex **cat-CN**. This structure of this binding mode presents almost no deviation when compared to the isolated **cat-CN** complex without the ketone, making it very geometrically accessible and with a small energy penalty.

The second contemplated binding mode is based on a previously proposed configuration in the literature,²⁰ wherein the catalyst forms a HB with the O in the ketone. In this binding mode, the CN^- is located as close as possible to the positive N within the quinoline structure. This transition state presents a remarkable change in conformation when compared to the structure of the **cat-ketone** complex, in which the ketone was establishing a π - π interaction that is no longer present in the TS. This huge conformational change in the structure, losing a stabilising

interaction, has a big energy penalty, affecting the stability of this TS.

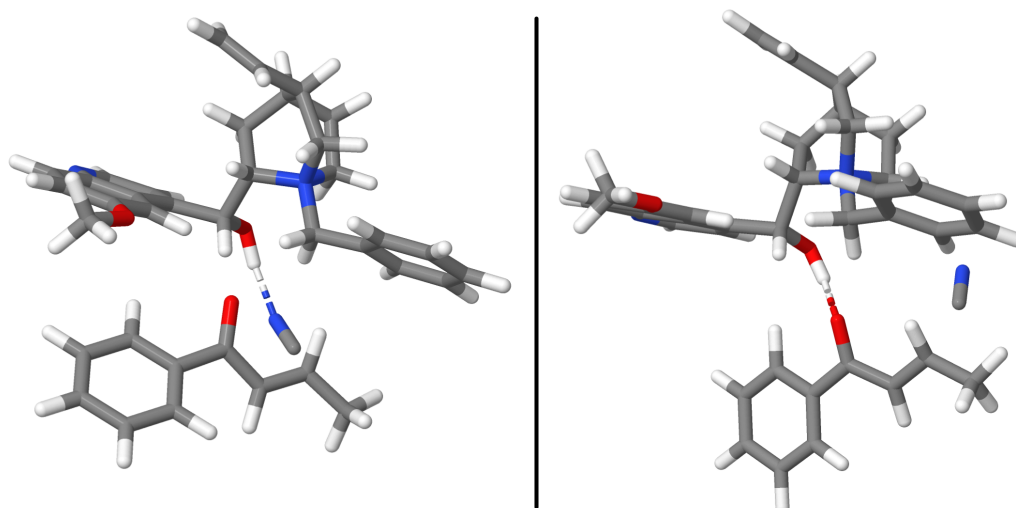


Figure 3.14. 3D structures of the TSs coming from **cat-CN** (left) and from **cat-ketone** (right) complexes.

When the two binding modes are compared and their Boltzmann populations calculated, it is obtained that the first TS, where the CN^- is establishing the HB with the catalyst, represents 99.75% of the population whilst the second one. In contrast, the second TS, involving the HB establishment with the ketone, only represents 0.25% of the population. It is important to note that not only does this second TS originate from a complex that is statistically insignificant, but even when assuming its formation, the resulting population is negligible. Consequently, this binding mode was excluded from consideration in this study.

Conjugate Cyanation

Once the conformers of the catalyst and the reactants were studied and a definite binding mode was established between the catalyst and the reactants, the cyanation reaction of the two different α - β -unsaturated ketones was studied.

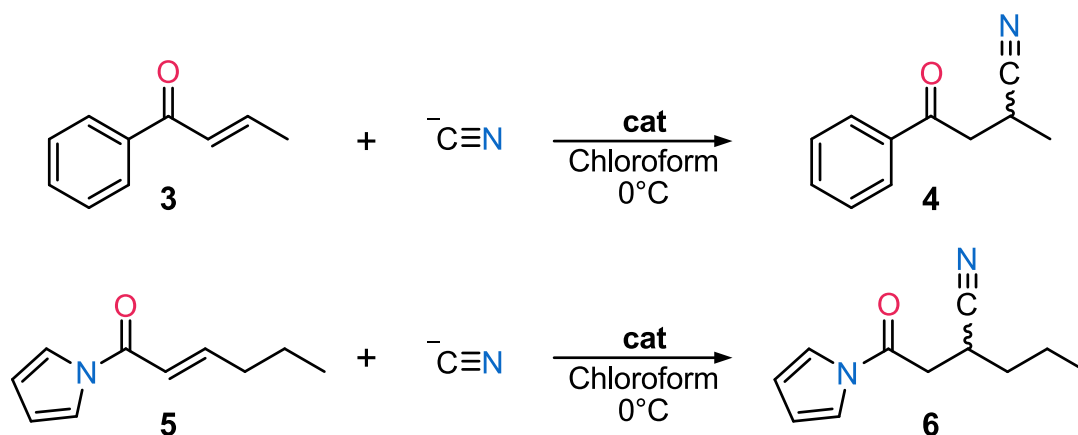


Figure 3.15. Conjugate cyanation reactions under study. a) **(2E)-1-Phenyl-2-buten-1-one** cyanation. b) **(2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one** cyanation.

Considering that the initial step of the reaction (cyanide attack) is responsible for the chirality of the product, it was the sole focus of our investigation. Four different possibilities, two for

each conformer of the α - β -unsaturated ketone, were studied to calculate the %ee of the reaction: *s-cis* conformer leading to product *R* (*R s-cis*) and product *S* (*S s-cis*) and the analogous paths with the *s-trans* conformer (*R s-trans* and *S s-trans*) (Figure 3.16).

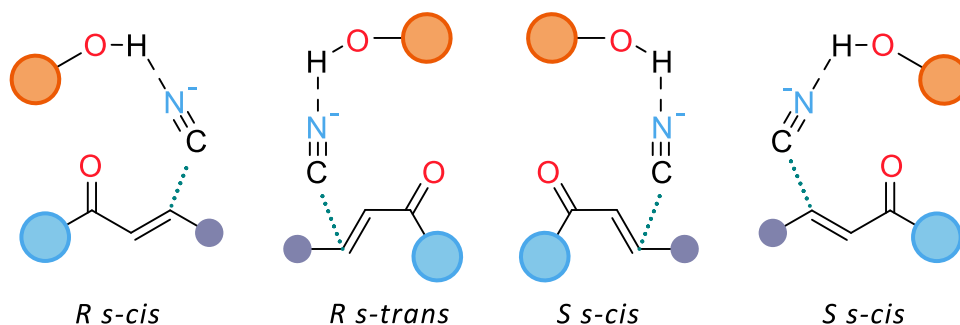


Figure 3.16. Different reaction paths under study. From left to right *s-cis* - *R*, *s-cis* - *S*, *s-trans* - *R*, *s-trans* - *S*.

(2E)-1-Phenyl-2-buten-1-one cyanation

We began by studying the conjugated cyanation of (2E)-1-Phenyl-2-buten-1-one, **1**, (Figure 3.15-a) using different modifications of the very well known cinchona based phase-transfer catalysts aforementioned (**cat_OMe_ph**, **cat_OMe_anthra**, **cat_OMe_cf3** and **cat_OMe_tbu**) and the theoretical modification previously proposed⁵² for each of them (**cat_OH_ph**, **cat_OH_anthra**, **cat_OH_cf3** and **cat_OH_tbu**).

The free-energy profile of the four different binding modes shown in Figure 3.16 were studied for the initial step of the reaction (PES values in Table S3.9.2) and the selectivity of each catalyst calculated based on the average free energy ($\Delta\overline{G}_i^\ddagger$) of the different transition states obtained (see Computational Details section).

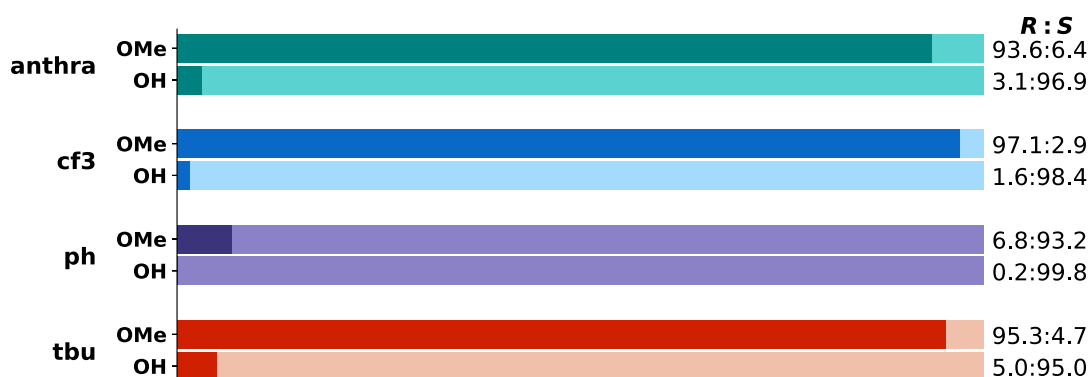


Figure 3.17. *R:S* ratio of the transition states under study for the conjugated cyanation of α - β -unsaturated ketone **1**.

The calculated *R:S* ratio for the four studied transition states for each catalyst is shown in Figure 3.17.

In terms of $R_2 = \text{OMe}$, these catalysts consistently exhibit high enantioselectivity, as evidenced by the obtained %ee values (**cat_OMe_anthra**=87%ee *R*, **cat_OMe_cf3**=94%ee *R*, **cat_OMe_ph**=86%ee *S*, **cat_OMe_tbu**=91%ee *R*). However, it is crucial to note that, due to the exponential nature of the reaction rates (Eq. 8), minor errors in energy calculations

can result in significant changes in %*ee* within this range, making the quantitative results in the low-medium *ee* unreliable but highlighting their lack of perfect selectivity, as it will be explained more extensively in the next chapter.

Additionally, it is evident that **cat_OMe_ph** displays distinct behaviour compared to the other catalysts, favouring the opposite enantiomer of the product (*S* instead of *R*).

This distinct behaviour arises from the differing stability of the most stable TS conformations depending on the catalyst.

Upon comparing the 3D structures of the most stable TS in each case, Figure 3.18, two different groups can be identified based on the NCIs established between the catalyst and the reactants. The first one, formed by **cat_OMe_anthra** and **cat_OMe_ph**, establishes a π - π interaction between the aromatic ring R_1 and the phenyl ring within the ketone. However, each of these TSs promotes a different enantiomer, presenting the ketone in a similar position but with opposite enantiomers (*s-trans* when $R_1 = anthra$ and *s-cis* when $R_1 = ph$). If the 3D structures are observed, this change is produced by the size of the R_1 aromatic ring. Since the phenyl ring is much smaller than the anthracene ring, the CN^- has enough space to be positioned below it, however, this position with the anthracene ring is not as favoured because of its bigger size and, therefore, the other conformation of the ketone is more stable.

On the other hand, the other two catalysts, **cat_OMe_cf3** and **cat_OMe_tbu**, fail to establish these π - π interactions because of the bulky R_1 substituents. In the first case, $R_1 = cf3$, the ketone does not interact with the catalyst in any remarkable way, avoiding all the repulsive interactions; while the second case, $R_1 = tbu$, is able to establish an attractive π - π between the ketone and the quinoline ring of the catalyst, further stabilising this conformation.

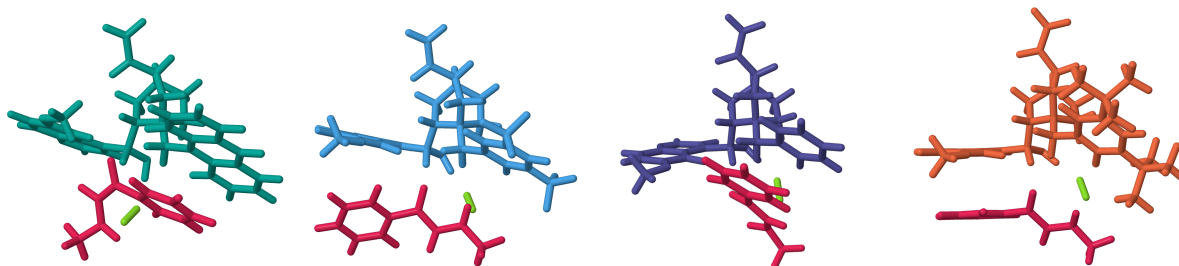


Figure 3.18. TS 3D structure of the most stable TS conformer for each $R_2 = OMe$ catalyst under study: *R s-trans* for **cat_OMe_anthra**, *R s-cis* for **cat_OMe_cf3**, *S s-cis* for **cat_OMe_ph**, and *R s-cis* for **cat_OMe_tbu**.

Despite presenting different stabilities, the studied TSs present a consistent characteristic observed among all of them; the position of the oxygen of the ketone, forming a web of charge-assisted hydrogen bonds with the quinuclidine ring.

This HB web determines the primary position of the ketone in all the TSs and the attractive secondary NCIs, like the π - π with between R_1 substituents and the ketone in **cat_OMe_anthra** and **cat_OMe_ph** or the repulsive sterical clashes (e.g. **cat_OMe_cf3**) determines the relative stability of the different TS conformers. It is also clear that despite the %*ee* values being quite similar amongst all the studied substituents in position R_1 , those catalysts that can establish the π - π interactions with the ketone in that position produce lower selectivities.

Despite the theoretical selectivities of the first set of catalysts with $R_2 = OMe$ are not exceedingly low, the NCIs interactions established between the ketone and the catalyst are, sometimes, not sufficient to hold its position amongst all the TS conformers and favour the

production of only one enantiomer of the product. Considering that the primary issue identified is the freedom of the ketone and its capacity to establish different NCIs with different parts of the catalyst, the most evident and rational design choice involves attempting to constrain this movement. This would entail securing the position of the ketone through interactions that are stronger than those already existing. The most promising candidate for achieving this is the *OMe* group, which can be replaced by an *OH* group capable of potentially forming an additional hydrogen bond, as shown in the aforementioned study.⁵² This additional interaction would anchor the ketone in place, theoretically boosting the selectivity.

Consequently, the catalysts were modified, and all the previously studied stationary points were reoptimised using these newly modified catalysts. The *R:S* ratio of the different transition states obtained for the modified catalysts ($R_2 = OH$) can be found in Figure 3.17, illustrating a dramatic change in the situation after modification compared to the original catalyst.

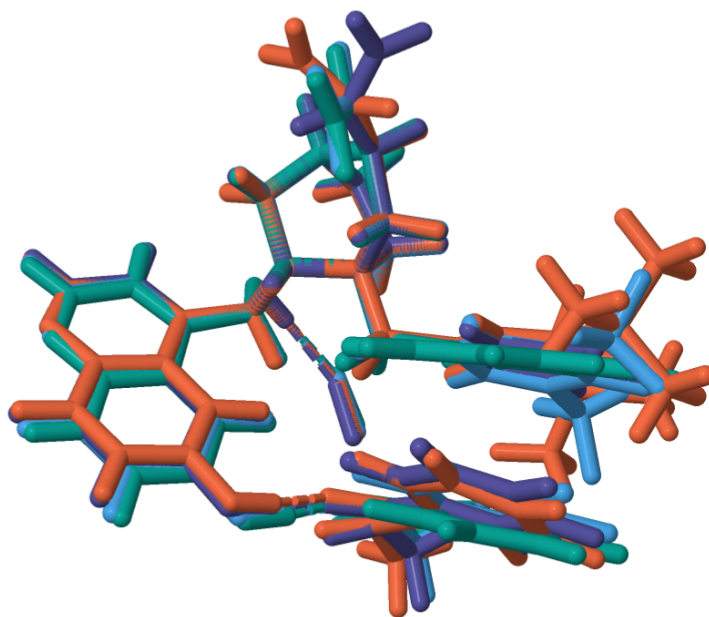


Figure 3.19. Superposition of the *S s-cis* TS for ketone **1**. In green **cat_OH_anthra**, in blue **cat_OH_cf3**, in purple **cat_OH_ph** and in orange **cat_OH_tbu**.

The enantioselectivity of all the catalysts studied has been maintained (**cat_OH_tbu** 90%*ee S*), slightly increased (**cat_OH_anthra** 94%*ee S* and **cat_OH_cf3** 97%*ee S*) and greatly increased (**cat_OH_ph** 99%*ee S*). However, the most remarkable effect of this modification is not this increase of the selectivity but the inversion of the favoured enantiomer for **cat_OH_anthra**, **cat_OH_cf3** and **cat_OH_tbu**. These catalysts previously favoured the *R* enantiomer, with $R_1 = OMe$, and now they exhibit a preference for the *S* product. Moreover, the addition of this second anchor point produces very consistent TS amongst all the studied R_1 substituents. This consistency emerges from the new HB established between the catalyst and the oxygen atom in the ketone, stabilising the *S s-cis* transition state in all the cases under investigation.

However, despite this agreement between the most stable TSs for all of the studied catalysts, their selectivity is not uniformly high across the board. Analysing the structures of the most stable TS in each system (*S s-cis*, Figure 3.19 and S3.9.2) it can be seen that both the catalyst and the reactants are in very similar positions, showing almost perfect superimposition in certain

areas. Nevertheless, **cat_OH_anthra** (green) presents a more distorted structure due to the size of the anthracene ring in the R_1 position, destabilising the system and contributing to the reduction in observed selectivity, as mentioned earlier. In the case of **cat_OH_tbu**, the structure closely resembles that of the highly stable **cat_OH_ph** and **cat_OH_cf3**. However, due to the bulkiness of the tBu substituents, the stabilising π - π interactions are replaced by weaker π - σ_{CH} interactions. This substitution results in a less stable transition state and a less enantioselective catalyst.

Characterisation of the Intermolecular Interactions

In order to characterise the different HBs present in each system, a NBO analysis of different TSs wave functions under study was performed. Examining the total NBO values (Eq. 10, Figure 3.20) across all investigated transition states, a notable finding emerges: the most stabilising (stronger) HBs are consistently associated with the *s-trans* TS leading to the *S* enantiomer.

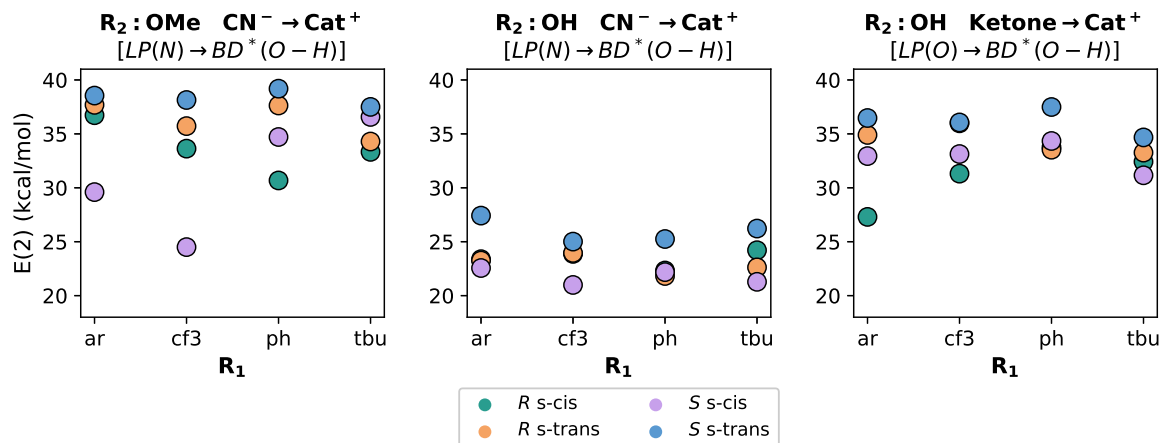


Figure 3.20. NBO E(2) stabilisation energies corresponding to the different HBs present in the transition state in reaction 1. Left $R_2 = OMe$, $CN^- \rightarrow Cat^+$ [$LP(N) \rightarrow BD^*(OH)$], Center: $R_2 = OH$, $CN^- \rightarrow Cat^+$ [$LP(N) \rightarrow BD^*(OH)$], Right: $R_2 = OH$, $Ketone \rightarrow Cat^+$ [$LP(O) \rightarrow BD^*(OH)$].

However, when $R_2 = OMe$, this consistency in the strength of the analysed HBs is not reflected in the global stability of the corresponding transition states. This discrepancy between the HB stabilisation and the binding energy reveals that the driving force of the selectivity in these catalysts is not the HBs established between the catalyst and the CN^- but different steric issues and NCIs of the ketone when approaching the system, as shown in the previous section.

On the other hand, when $R_2 = OH$, the observed NBO values for HB between cyanide and catalyst ($LP(N) \rightarrow \sigma^*(O-H)$) decreased drastically, while the new HB between ketone and catalyst ($LP(O) \rightarrow \sigma^*(O-H)$) shows higher values, similar to those between cyanide and catalyst with $R_2 = OMe$.

This decrease in the HB strength between the cyanide and the catalyst is in complete agreement with the position of the theoretical negative charge in the TS. When $R_2 = OMe$, only one strong HB could be established; consequently, the negative charge was partially located in the cyanide, restrained by this HB. However, the addition of this new anchor point drags this negative charge towards its final location in the oxygen atom within the ketone, decreasing

the charge located in the cyanide and consequently reducing the strength of the HB that it establishes with the catalyst.

Finally, a thorough QTAIM analysis was performed for all the studied TSs for the reaction. The values for all the BCPs corresponding to the different HBs and the reaction coordinates are collected in Table S3.6.2. Plotting these ρ_{BCP} values against bond distances (Figure 3.21) revealed robust exponential relationships between the densities ($\rho(BCP)$) and distances for all the examined interactions.

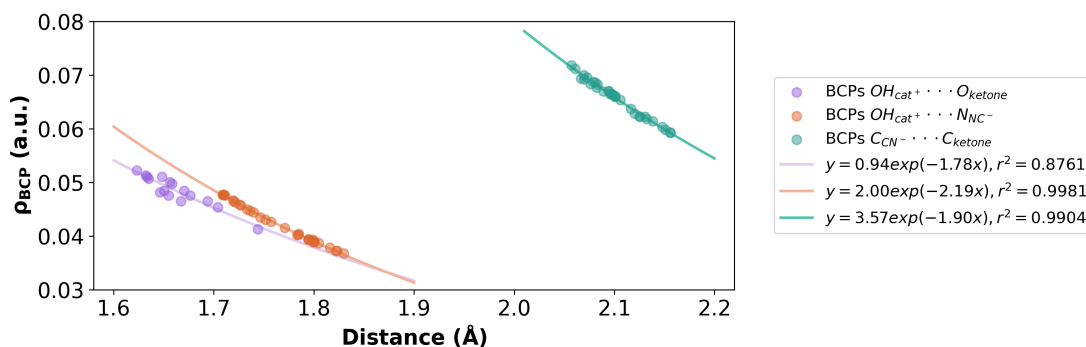


Figure 3.21. Exponential relationships of the intermolecular distance, in Å, and the corresponding $\rho(BCP)$ between the catalyst and the anion CN^- (orange), the catalyst and the ketone (blue and purple) and the CN^- anion and the CN^- and the ketone (green).

It would be expected that these data points would form three distinct groups, one for each type of interaction: $OH_{cat+} \cdots O_{ketone}$, $OH_{cat+} \cdots N_{CN^-}$, and $C_{CN^-} \cdots C_{ketone}$. However, $OH_{cat+} \cdots O_{ketone}$ presents two different trends and has to be divided into two different groups. The first of these trends (purple) contains those TSs leading to the *R* with the ketone in *s-cis* conformation and those TSs leading to the *S* enantiomer with the ketone in *s-trans*; and the latter (blue) containing the other two TSs (*R*-enantiomer & *s-trans* ketone and *S*-enantiomer & *s-cis* ketone).

This behaviour appears highly unusual, considering that the interactions analysed in the different systems are analogous, containing the same atoms and varying only their spacial position. However, the 3D structures of the different TS binding modes (Figure 3.22) reveal that, in this case, the first two groups, *R s-cis* and *S s-trans* enantiomers (Figure 3.22 top, purple in Figure 3.21), present the phenyl ring of the ketone close to the quinoline ring in the catalyst. On the other hand, the other two groups *S s-cis* and *R s-trans* (Figure 3.22 bottom, blue in Figure 3.21), position the phenyl ring of the ketone next to the catalyst group R_1 , in an almost vertical position.

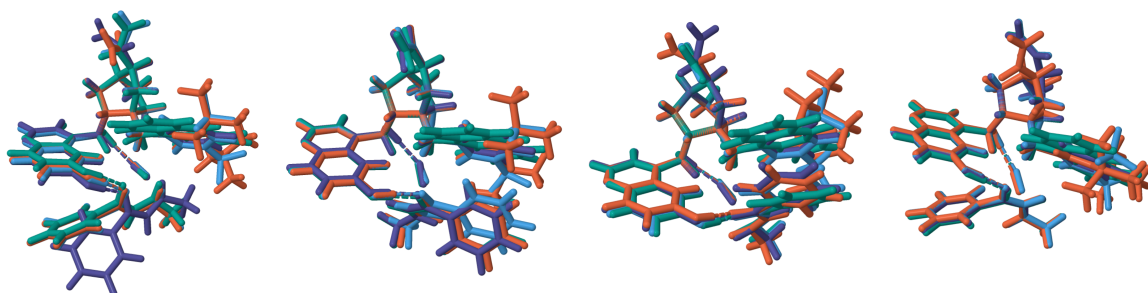


Figure 3.22. 3D structures of the optimised TSs under study for the cyanation of ketone **2**.

These two different placements of the ketone and the different alignments that they cause are able to distort the established HBs enough to produce changes that are noticeable in the QTAIM analysis.

(2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one cyanation

The second reaction under study was the conjugated cyanation of **(2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one**, **2**, (Figure 3.15-b) using the same set of catalysts and modifications as used above (**cat_OMe_ph**, **cat_OMe_anthra**, **cat_OMe_cf3**, **cat_OMe_tbu**, **cat_OH_ph**, **cat_OH_anthra**, **cat_OH_cf3** and **cat_OH_tbu**), and under the same conditions.

This new α - β -unsaturated ketone was selected to study how a different aromatic ring (pyrrole instead of phenyl) and different aliphatic chain length affect the mechanism, the binding modes and the selectivity of the reaction, studying the effect of possible new NCIs and steric hindrance within the system.

The PES of all the different binding modes in Figure 3.16 were studied (Table S3.9.3) and the selectivity of each catalyst was calculated based on the average free energy ($\Delta\overline{G}_i^\ddagger$) of the different transition states obtained (Figure 3.23). The calculated *R:S* ratio for the four studied transition states for each catalyst is shown in Figure 3.23.

Focusing on the first set of catalysts ($R_2 = OMe$) it can be seen that the *R* enantiomer is preferred in all cases. This situation is similar to the one found in the cyanation of ketone **1**, where this same enantiomer was the most favoured one in most of the cases. However, the selectivities achieved in this new reaction are, in general, greater than in the previous case (**cat_OMe_anthra**=98%*ee R*, **cat_OMe_cf3**=99%*ee R*, **cat_OMe_ph**=77%*ee R*, **cat_OMe_tbu**=88%*ee R*).

When analysing the 3D structure of the most stable transition state in each case (Figure 3.24), the situation is almost identical to the one found in the previous reaction where the ketone is located with the *O* atom as close as possible to the N^+ in the catalyst, forming a web of charge assisted HBs, and does not present any steric clashes with the substituents within the catalyst.

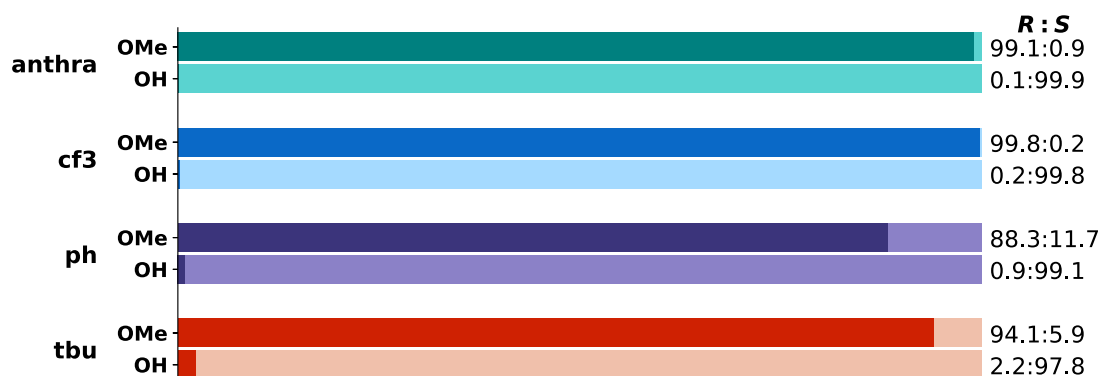


Figure 3.23. *R:S* ratio of the transition states under study for the conjugated cyanation of α - β -unsaturated ketone **2**.

These *R* leading, *s-cis*, transition states overlap almost perfectly independently of the catalyst, showing very clearly the consistency of their stability and the strength of the NCIs involved (Figure S3.4.2). However, they do not present any attractive NCI between the catalyst and the

ketone that favours them over the other three binding modes, like happened in the previous reaction with **cat_OMe_anthra** and **cat_OMe_ph**. The only attractive interaction present, the aforementioned HB web, is not exclusive of this binding mode and appears in all of them.

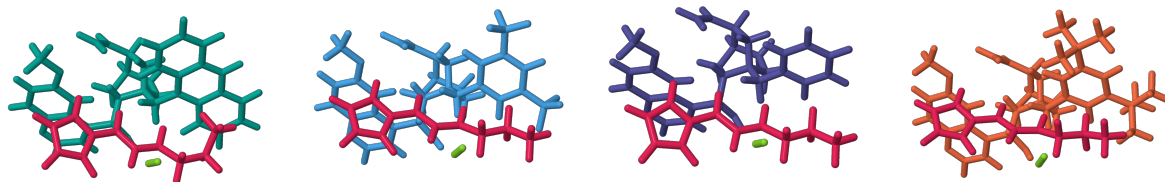


Figure 3.24. 3D structures of the TS leading to enantiomer *R* with ketone **2** in the *s-cis* conformation.

If the structures of the second most stable binding mode, *S s-trans*, is visualised (Figure S3.4.2), it can be seen how the longer aliphatic chain the ketone produces a bigger distortion of the structure, moving the ketone away from the catalyst. A similar situation can be also seen in the other two binding modes (Figure S3.4.2). Therefore, the increase of selectivity in this reaction seems to be caused by the destabilisation of some of the binding modes due to the bigger size of the ketone, rather than the effect of any attractive interaction.

In accordance with the preceding procedure employed in the previous reaction, the present investigation involves the examination of the reaction employing catalysts modified as $R_2 = OH$. This investigation aims to elucidate the impact of this modification on selectivity and to determine whether this modification exhibits consistent effects when different ketones are used.

As it can be seen in Figure 3.23 the *S* enantiomer is favoured in all the cases (**cat_OH_anthra**=99%*ee S*, **cat_OH_cf3**=99%*ee S*, **cat_OH_ph**=98%*ee S*, **cat_OH_tbu**=96%*ee S*). This preferred enantiomer is in complete agreement with what was found in the previous reaction (Figure 3.17-down), highlighting the consistency of this modification. When visualising the most stable binding mode for all the catalysts (Figure 3.25), a similar situation to what was found with ketone **1** is presented.

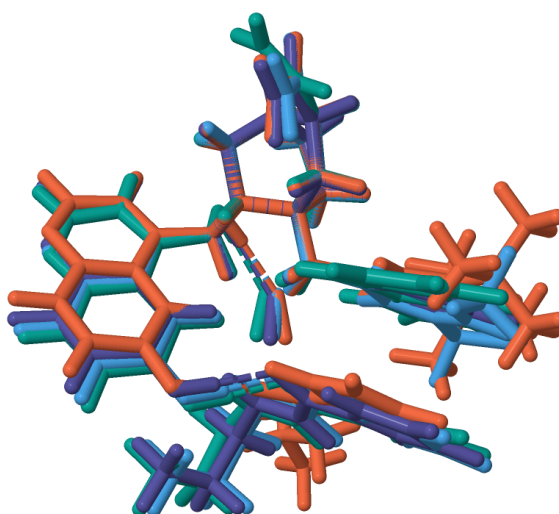


Figure 3.25. Superposition of the most abundant binding mode for ketone **2**. In green **cat_OH_anthra**, in blue **cat_OH_cf3**, in purple **cat_OH_ph** and in orange **cat_OH_tbu**.

The ketone and the new catalyst establish an extra hydrogen bond that anchors the ketone

in place, favouring one conformation over all the others. Besides, all these TSs show a very nice π - π stacking between the aromatic ring in ketone **2** and the aromatic R_1 in the catalyst, further stabilising the system.

Despite being the least populated binding mode, it is also worth mentioning that, in the case of the TSs leading to the *R* enantiomer (Figure S3.4.2), with ketone **2** in *s-scis* conformation, this second HB is not established between the $OH_{catalyst}$ and O_{ketone} and instead an interaction between the $OH_{catalyst}$ and the aromatic ring within the ketone is established.

Characterisation of the Intermolecular Interactions

Following the methodology established in the preceding section, the distinctive HBs present in the studied TSs have been systematically characterised by means of NBO, evaluating the stabilisation associated with the charge transfer (E(2)) between the different HB-donors and HB-acceptors ($LP(N_{CN^-}) \rightarrow BD^*(OH_{cat})$, $LP(O_{ketone}) \rightarrow BD^*(OH_{cat})$). All the extract values can be found in Table S3.5.3 and the sum of the individual charge transfer (Eq. 10) can be found in Figure 3.26.

Similarly to what happened in the previous reaction, the TS presenting the strongest HBs do not correspond to the major TS, showing that this interaction is not the driving force of the selectivity in this process. This observation is in agreement with the geometrical analysis in which it was observed that the most stable transition states were those with lower sterical hindrances.

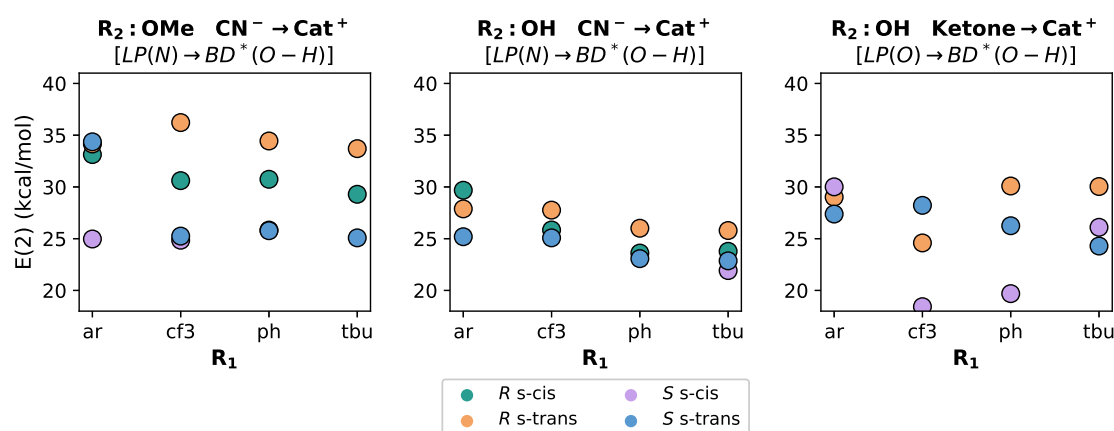


Figure 3.26. NBO E(2) stabilisation energies corresponding to the different HBs present in the transition state in reaction 1. Left $R_2 = OMe$, $CN^- \rightarrow Cat^+$ [$LP(N) \rightarrow BD^*(OH)$], Center: $R_2 = OH$ $CN^- \rightarrow Cat^+$ [$LP(N) \rightarrow BD^*(OH)$], Right: $R_2 = OH$ Ketone $\rightarrow Cat^+$ [$LP(O) \rightarrow BD^*(OH)$].

When comparing the values obtained for the HBs between the cyanide and the catalyst in the different versions of the catalyst ($R_2 = OMe$ and $R_2 = OH$), the same decrease in their strength as in the previous reaction is found. However, this difference is smaller and the new HB ($LP(O_{ketone}) \rightarrow BD^*(OH_{cat})$) is weaker than the one shown in the previous reaction. Besides, looking at the 3D structures of these transition states (Figure 3.25 and S3.4.2) and comparing them with the ones found in the previous reaction (Figure 3.22), it can be seen that the smaller aromatic ring is now better aligned with the π system of the ketone, allowing the

negative charge to be delocalised all over the system instead of locating it in the oxygen atom, consequently reducing the strength of the new HB.

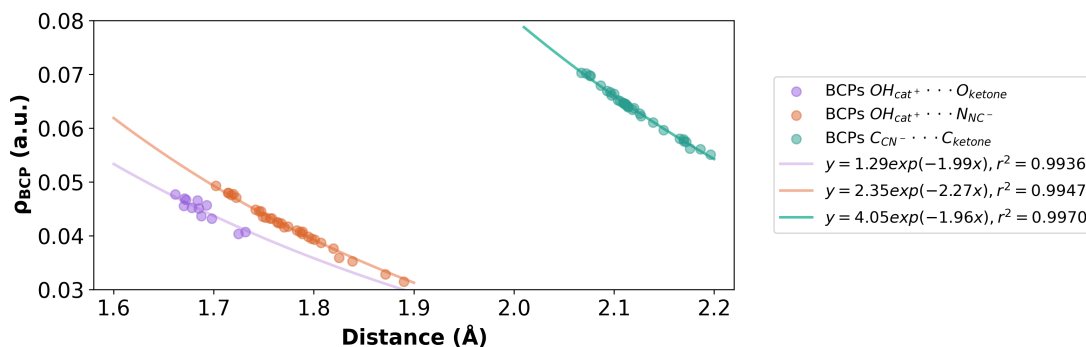


Figure 3.27. Exponential relationships of the intermolecular distance, in Å, and the corresponding $\rho(BCP)$ between the catalyst and the anion CN^- (blue), the catalyst and the ketone (red) and the CN^- anion and the CN^- and the ketone (green).

Subsequently, mirroring the procedure employed in the previous reaction, an exhaustive QTAIM analysis was systematically conducted for all transition states. This analysis includes the evaluation of the electron density at the BCPs (ρ_{BCP}) associated with the established hydrogen bonds ($OH_{cat+} \cdots N_{CN^-}$ in cases where $R_2 = OMe$ or OH , and $OH_{cat+} \cdots O_{ketone}$ when $R_2 = OH$). Additionally, the analysis encompassed an examination of the reaction coordinate ($C_{CN^-} \cdots C_{ketone}$). The resulting values for electron density at the BCP and the associated interaction distances are reported in Table S3.6.3.

Upon plotting the electron density at the BCP against the corresponding interaction distances (Figure 3.27), a robust exponential correlation emerges among these values. In contrast to observations made in the previous reaction, it is noteworthy that the values corresponding to $OH_{cat+} \cdots O_{ketone}$ interactions do not segregate into distinct clusters, showing a more uniform alignment of this interaction within these transition states. It is also worth mentioning that such interactions are absent in any of the transition states leading to the *R* enantiomer in the *s-cis* conformation, thereby limiting the number of transition states in which this hydrogen bond can be analysed.

3.4 Conclusions

This chapter was divided into two interconnected sections.

In the first section, a theoretical study of different interactions that can be established between a well-known alkaloid quinine-derived PTC catalyst and four different anions of interest in organocatalysis was performed: Cl^- , Br^- , $MeCO_2^-$ and CN^- . Competition between the ion pairing processes with the establishment of an intermolecular hydrogen bond between the catalyst counteranion and potential hydrogen bond donors was studied.

The results revealed that acetate anion complexes exhibit the strongest interactions, while the bromide ion complexes show the weakest interaction energy among of all them. Consequently, the complex establishing an intermolecular hydrogen bond presented the strongest interaction. Notably, the complex exhibiting a strict ion-pair interaction, where both charged atoms are closest (**pos_2**) (the shortest $X^- \cdots N^+$ distance), is 30 kJ/mol less stable than the HB-bound complex (**pos_1**).

A theoretical study of the free-energy profile corresponding to the enantioselective conjugate cyanation of α,β -unsaturated ketone, (E)-Chalcone (**1**), in the presence of the catalyst **cat_OMe_ph** was performed. The barrier for both enantiomers of the cat1-bound enolate product indicates that the C–C bond formation is very similar for both pathways ($\delta\Delta G = 0.6 \text{ kJ/mol}$). From the free-energy profile both enantiomers (*R* and *S*) are energetically accessible by cyanide, obtaining a theoretical ee value of 13%.

To enhance the enantioselectivity of this reaction, an additional anchor point was added, **cat_OH_ph**. The corresponding free-energy profile was studied. The calculated ee value when the reaction proceeds in the presence of cat2 increases up to 99%.

In the second section, the theoretical study of the cyanation reaction of two new α,β -unsaturated ketones was conducted under the presence of a set of four common cinchona catalyst modifications (**cat_OMe_anthra**, **cat_OMe_cf3**, **cat_OMe_ph**, **cat_OMe_tbu**) and a modified set of catalyst (**cat_OH_anthra**, **cat_OH_cf3**, **cat_OH_ph**, **cat_OH_tbu**).

The effect of the different substituents R_1 on the relative stability of the different binding modes under study was analysed, showing how each one of them can favour different binding modes and the effect that this has on the favoured enantiomer. Greater selectivity values than those seen in the first part were obtained with the first set of catalysts, $R_2 = OMe$, ranging 76.6-99.6%ee. Despite those better results, the modification of R_2 proposed in the first section was applied to these catalysts to analyse its effect and determine whether the effect was specific for the cyanation of (E)-chalcone or if it produces a similar effect independently of the α,β -unsaturated ketone used. The obtained selectivities with the new modified set of catalysts, $R_2 = OH$, ranged 90.0-99.6%ee, achieving greater selectivity values.

However, the most remarkable effect of this modification is that, in addition to the selectivity improvement, it favours the opposite enantiomer demonstrating how a simple modification can invert the selectivity of a catalytic process.

References

- (1) Z. G. Hajos and D. R. Parrish.
- (2) C. M. Starks, *Journal of the American Chemical Society*, 1971, **93**, 195–199.
- (3) S. Shirakawa and K. Maruoka, *Angewandte Chemie International Edition*, 2013, **52**, 4312–4348.
- (4) H. Wang, *Catalysts*, 2019, **9**, 244.
- (5) T. Ooi and K. Maruoka, *Angewandte Chemie International Edition*, 2007, **46**, 4222–4266.
- (6) D. Qian and J. Sun, *Chemistry–A European Journal*, 2019, **25**, 3740–3751.
- (7) S.-s. Jew and H.-g. Park, *Chemical Communications*, 2009, 7090–7103.
- (8) E. Corey, F. Xu and M. C. Noe, *Journal of the American Chemical Society*, 1997, **119**, 12414–12415.
- (9) K. Brak and E. N. Jacobsen, *Angewandte Chemie International Edition*, 2013, **52**, 534–561.
- (10) G. P. Petrova, H.-B. Li, K. Maruoka and K. Morokuma, *The Journal of Physical Chemistry B*, 2014, **118**, 5154–5167.
- (11) N. Kuroono and T. Ohkuma, *ACS Catalysis*, 2016, **6**, 989–1023.
- (12) S. J. Connon, *Angewandte Chemie International Edition*, 2008, **47**, 1176–1178.
- (13) L. Bini, C. Müller and D. Vogt, *ChemCatChem*, 2010, **2**, 590–608.
- (14) G. M. Sammis and E. N. Jacobsen, *Journal of the American Chemical Society*, 2003, **125**, 4442–4443.
- (15) G. M. Sammis, H. Danjo and E. N. Jacobsen, *Journal of the American Chemical Society*, 2004, **126**, 9928–9929.
- (16) Y. Tanaka, M. Kanai and M. Shibasaki, *Journal of the American Chemical Society*, 2008, **130**, 6072–6073.
- (17) Y. Tanaka, M. Kanai and M. Shibasaki, *Journal of the American Chemical Society*, 2010, **132**, 8862–8863.
- (18) L. Bernardi, F. Fini, M. Fochi and A. Ricci, *Synlett*, 2008, **2008**, 1857–1861.
- (19) B. A. Provencher, K. J. Bartelson, Y. Liu, B. M. Foxman and L. Deng, *Angewandte Chemie*, 2011, **123**, 10753–10757.
- (20) F. Buttard and P. A. Champagne, *ACS Catalysis*, 2022, **12**, 8185–8194.
- (21) C. Trujillo, I. Rozas, A. Botte and S. J. Connon, *Chemical Communications*, 2017, **53**, 8874–8877.
- (22) C. Trujillo, M. Litvajova, S. A. Cronin, R. Craig and S. J. Connon, *ChemCatChem*, 2019, **11**, 3776–3780.
- (23) B. Lockett-Walters, C. Trujillo, B. Twamley and S. Connon, *Chemical Communications*, 2019, **55**, 11283–11286.
- (24) A. G. Collar, C. Trujillo, B. Lockett-Walters, B. Twamley and S. J. Connon, *Chemistry–A European Journal*, 2019, **25**, 7275–7279.

- (25) Y.-h. Lam, M. N. Grayson, M. C. Holland, A. Simon and K. Houk, *Accounts of Chemical Research*, 2016, **49**, 750–762.
- (26) P. H.-Y. Cheong, C. Y. Legault, J. M. Um, N. Çelebi-Ölçüm and K. Houk, *Chemical reviews*, 2011, **111**, 5042–5137.
- (27) Y. Zhao and D. G. Truhlar, *Theoretical chemistry accounts*, 2008, **120**, 215–241.
- (28) T. H. Dunning Jr, *The Journal of chemical physics*, 1989, **90**, 1007–1023.
- (29) D. E. Woon and T. H. Dunning Jr, *The Journal of chemical physics*, 1995, **103**, 4572–4585.
- (30) a. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley and J. Mantzaris, *The Journal of chemical physics*, 1988, **89**, 2193–2218.
- (31) G. Petersson and M. A. Al-Laham, *The Journal of chemical physics*, 1991, **94**, 6081–6090.
- (32) A. V. Marenich, C. J. Cramer and D. G. Truhlar, *The Journal of Physical Chemistry B*, 2009, **113**, 6378–6396.
- (33) E. Engelage, N. Schulz, F. Heinen, S. M. Huber, D. G. Truhlar and C. J. Cramer, *Chemistry–A European Journal*, 2018, **24**, 15983–15987.
- (34) M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji et al., *Wallingford, CT*, 2016.
- (35) R. F. Bader, M. T. Carroll, J. R. Cheeseman and C. Chang, *Journal of the American Chemical Society*, 1987, **109**, 7968–7979.
- (36) T. Lu and F. Chen, *Journal of computational chemistry*, 2012, **33**, 580–592.
- (37) *Jmol: An Open-source Java Viewer for Chemical Structures in 3D*, <http://www.jmol.org/>.
- (38) L. J. Hauptert, J. C. Poutsma and P. G. Wenthold, *Accounts of chemical research*, 2009, **42**, 1480–1488.
- (39) S. T. Schneebeli, M. L. Hall, R. Breslow and R. Friesner, *Journal of the American Chemical Society*, 2009, **131**, 3965–3973.
- (40) Q. Peng, F. Duarte and R. S. Paton, *Chemical Society Reviews*, 2016, **45**, 6093–6107.
- (41) A. E. Reed, L. A. Curtiss and F. Weinhold, *Chemical Reviews*, 1988, **88**, 899–926.
- (42) R. Bader et al., *Clarendon: Oxford, UK*, 1990.
- (43) P. L. A. Popelier, F. Aicken and S. O'Brien, *Atoms in molecules*, Prentice Hall Manchester, 2000, vol. 188.
- (44) T. A. Keith, *AIMAll (Version 19.10.12)*, <http://aim.tkgristmill.com>, Overland Park, KS, USA, 2019.
- (45) G. Landrum et al., 2016.
- (46) I. Iribarren and C. Trujillo, *Journal of Chemical Information and Modeling*, 2022, **62**, PMID: 36271836, 5568–5580.
- (47) T. A. Halgren, *Journal of computational chemistry*, 1996, **17**, 490–519.
- (48) J.-D. Chai and M. Head-Gordon, *Physical Chemistry Chemical Physics*, 2008, **10**, 6615–6620.

- (49) S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher and M. Stahn, *The Journal of Physical Chemistry A*, 2021, **125**, PMID: 33688730, 4039–4054.
- (50) I. H. Williams, *Journal of Physical Organic Chemistry*, 2022, **35**, e4312.
- (51) T. Nakamura, K. Okuno, R. Nishiyori and S. Shirakawa, *Chemistry–An Asian Journal*, 2020, **15**, 463–472.
- (52) I. Iribarren and C. Trujillo, *Physical Chemistry Chemical Physics*, 2020, **22**, 21015–21021.

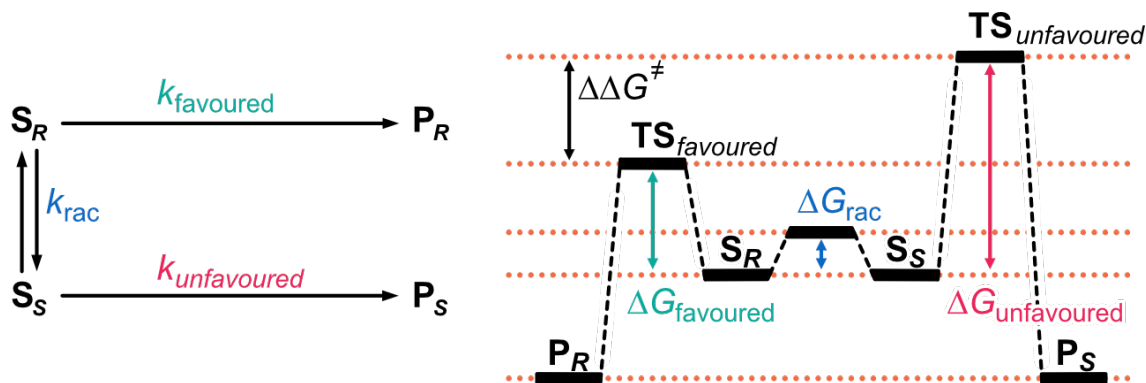
4 | Dynamic kinetic resolution of azlactones for the production of bis-protected amino acids

4.1 Introduction

Phase-Transfer Catalysis (PTC),¹ as previously discussed in this thesis, stands out as one of the most prolific catalytic methods within the organocatalysis family, demonstrating high enantioselectivities in various reactions.^{2–8}

Asymmetric organocatalysis can be applied to generate enantioenriched products in two different manners. Using an achiral reactant in which a new chiral centre is created, e.g. the asymmetric cyanation reactions explored in the previous chapter, or using a racemic mixture of a chiral compound as starting material and carrying out what is known as a *deracemisation* reaction, where the reaction of one of the enantiomers is favoured over the other. This way of producing an enantiomerically pure product using a racemic mixture of starting material is known as Kinetic Resolution (KR) and has been widely used in the past.⁹ However, the yield is limited to a maximum of 50% due to the racemic nature of the reactant. An elegant way to overcome this yield limitation can be achieved by developing a dynamic variant of these processes, obtaining what is known as Dynamic Kinetic Resolution (DKR).^{10,11} Reactions under DKR control have recently emerged as a powerful alternative that leads to complete conversion of both enantiomers, producing a high enantioenriched product.^{12–16}

However, several conditions must be fulfilled in order to consider that a reaction is under DKR control. In terms of reaction rates, as it can be seen in Scheme 4.1 left, the rate of racemisation of the reactant (k_{rac}) must be higher than the rate of any of the reaction rates, either favoured ($k_{favoured}$) or unfavoured ($k_{unfavoured}$). Likewise, the reaction rate of the favoured enantiomer ($k_{favoured}$) must be much higher than the reaction rate of the unfavoured enantiomer ($k_{unfavoured}$). Therefore, the DKR conditions can be summarised in terms of reaction rates as $k_{rac} > k_{favoured} \gg k_{unfavoured}$.

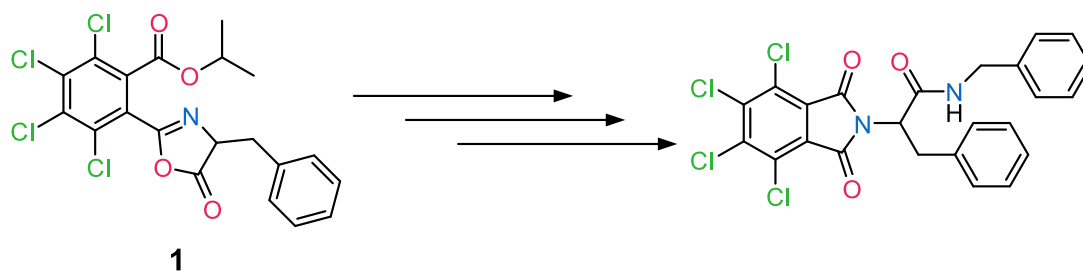


Scheme 4.1. General scheme of Dynamic Kinetic Resolution reaction rates (left) and the corresponding energy scheme of the different processes (right).

These conditions can be shown in terms of relative energies (Scheme 4.1 right), observing a small racemisation energy barrier (ΔG_{rac}) and two different reaction barriers; one smaller reaction barrier ($\Delta G_{favoured}$) and one much bigger ($\Delta G_{unfavoured}$). The selectivity of the global reaction will be determined by the energy difference between the two TSs ($\Delta\Delta G^\ddagger$).

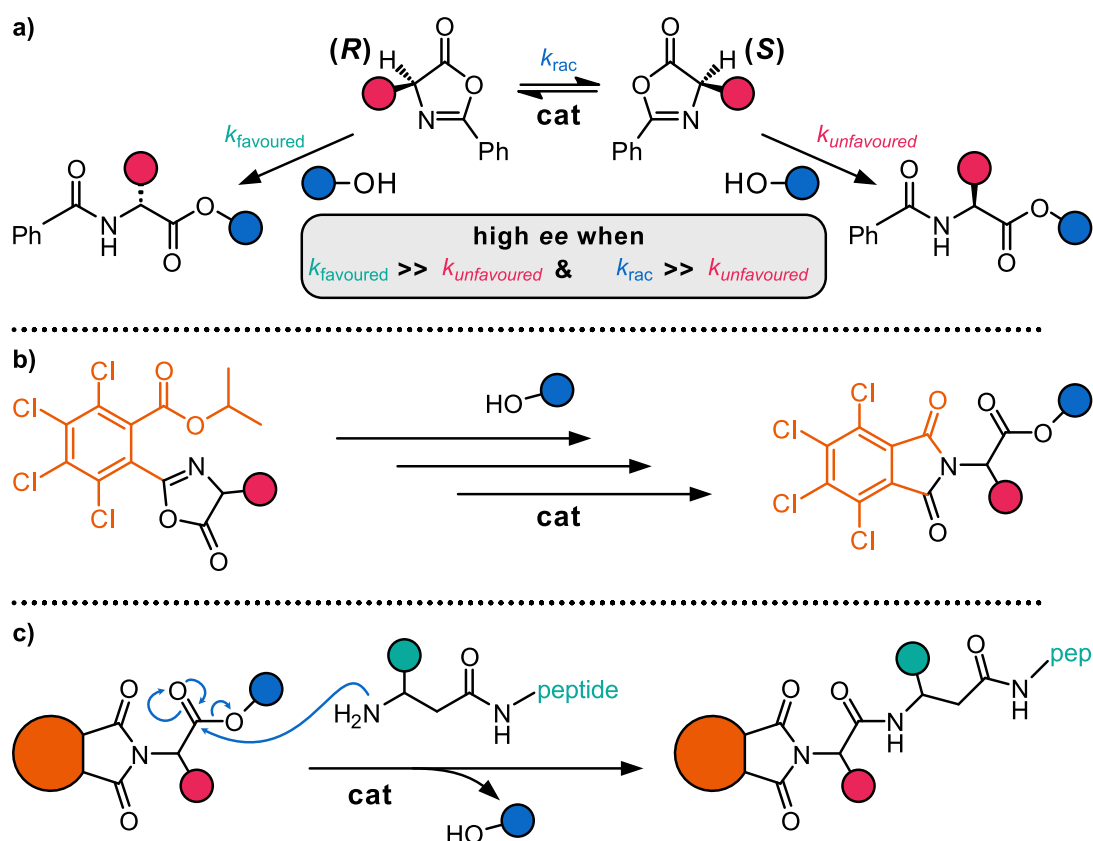
One very well-known and widely used example of these *deracemisation* reactions is the DKR of α -substituted azalactones via alcoholysis, a valuable route to obtain *N*-protected α -amino acid derivatives with high enantioselectivity and it has been extensively investigated over the last decade. In this chapter, the hydrolysis of an azalactone (*isopropyl 2-(4-benzyl-5-oxo-4,5-*

dihydrooxazol-2-yl)-3,4,5,6-tetrachlorobenzoate, **1**) has been study (Scheme 4.2).



Scheme 4.2. Hydrolysis of azalactone **1** to produce a bis-protected amino acid.

DKR conditions are ideal for this reaction because the acidity of the α -hydrogen of the azalactone allows the interconversion of both enantiomers, continuously regenerating the fast reacting enantiomer as it reacts producing the protected amino acid in an enantioselective manner. The *derasemisation* reaction of an α -substituted azalactone to produce an enantioenriched protected amino acid under DKR conditions is depicted in Scheme 4.3-a.



Scheme 4.3. a) Scheme of the DKR of an α -substituted azalactone. b) Scheme of the new protective group (tetrachloroisopropoxyester) and the thalamide product. c) Peptide growth mechanism and intermediate involved.

The chemistry of this process is well-known as several publications on the field have been published over the past few years.^{17–20} However, the potential uses of these enantioenriched amino acids present two problems: the two protecting groups are not orthogonal, so the product cannot be used for peptide chemistry,²¹ and also, it is not possible to use these products to grow peptides enantioselectively since their reaction with amines is too fast and happen even in the absence of an asymmetric catalyst, resulting in very poor enantioselectivities.

The problem with the non-orthogonal protective groups was solved by Cannon et al. with the addition of tetrachloroisopropoxyester substituent to the azalactone (Scheme 4.3-b). This substituent, after the azalactone ring is opened in the first step of the reaction upon the nucleophilic attack, allows the intermolecular ring closure of the intermediate, forming thalamide, that can be removed in the presence of other organic reactants, e.g. an ester.

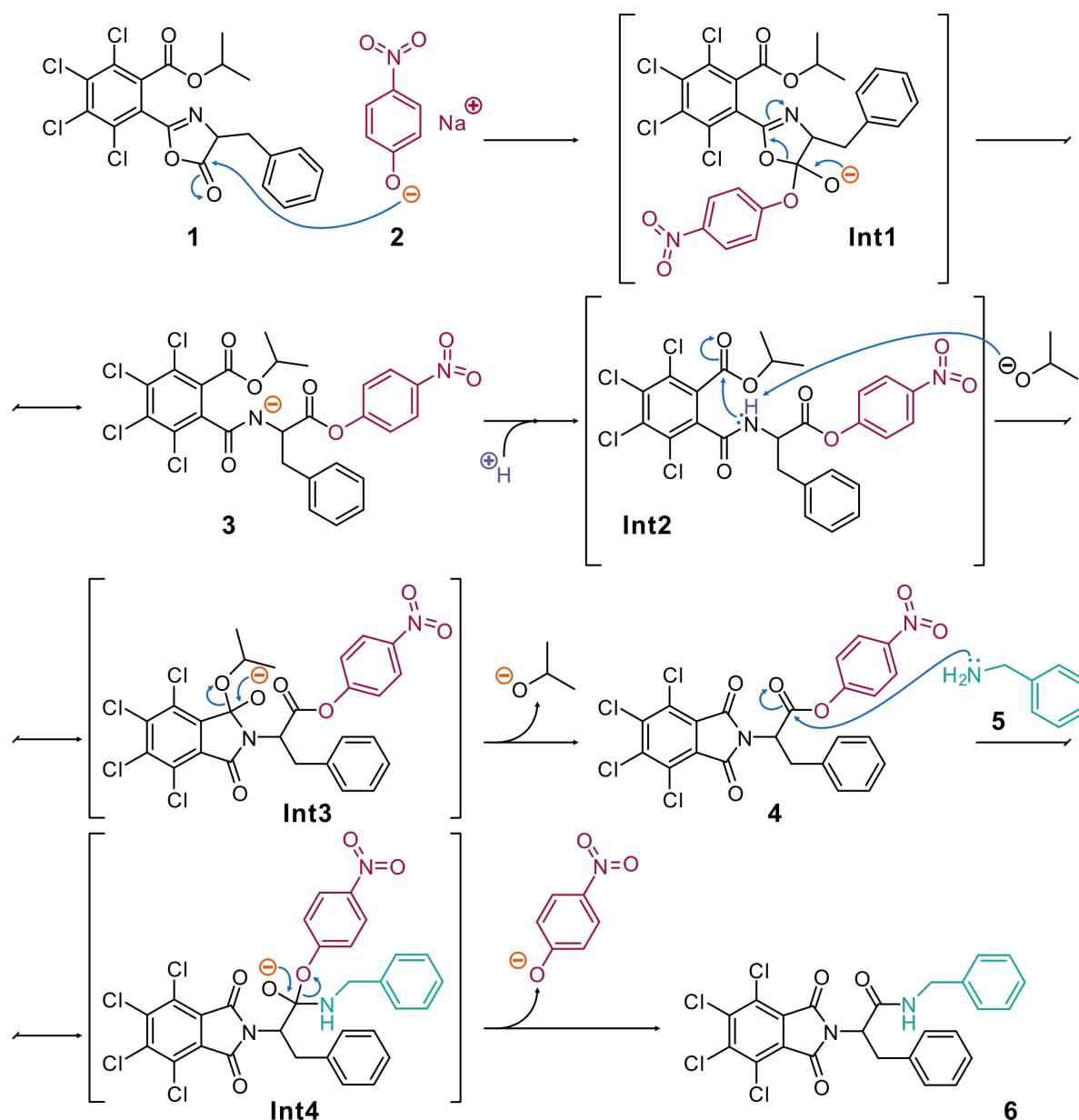
However, in order to take one step further and use these enantiomerically pure amino acids to grow peptides, the second problem must be also addressed. The novel approach to solve the rapid reaction of the amines was the use of a faster nucleophile that attacks the azalactone under PTC conditions, producing an intermediate stable enough to be isolated and that is, at the same time, active enough to react with the amine in one pot. The PTC conditions generate a charge separation between the catalyst, cationic, and the nucleophile, anionic, greatly improving the reaction rate of this nucleophilic attack over the amine reaction, theoretically solving the problem (Scheme 4.3-c).

This novel solution was tested by was experimentally tested by our experimental collaborators in Cannon's group and it was successfully proved that the intermediate is stable enough to be isolated and used to react with the selected amine, producing the desired peptide as a product. After several attempts, it was determined that the best nucleophile for this reaction was *p*-nitrophenolate as it is not as basic as other nucleophiles tested, e.g. fluoride, it is a decent leaving group for the second step of the reaction and, most importantly, it is insoluble in organic solvents, avoiding the nucleophilic attack in absence of the catalyst. However, the enantioselectivity values obtained with the catalysts selected for this preliminary study were either non-existent, reacting with both enantiomers of the azalactone, or very low; therefore, the rational design of new catalysts is required. The complete mechanism of the reaction under study can be seen in Scheme 4.4.

In this mechanism, it can be observed that the reaction is initiated by the nucleophilic attack of the phenolate **2** to the azalactone **1**, generating the tetrahedral intermediate **Int1**. This nucleophilic attack is followed by the ring opening of the unstable intermediate, resulting in **Int2** where the negatively charged amide N atom is immediately protonated, generating **3**. Subsequently, this amide is deprotonated again, leading to an intra-molecular cyclisation, forming the tetrahedral intermediate **Int3**. The unstable intermediate is followed by the elimination of the isopropylate anion, generating **4**. Finally, amide **5** attacks the ester generated in the first step of the reaction, producing the tetrahedral intermediate **Int4**, followed by the elimination of phenolate **1**, resulting in the formation of the protected bis-aminoacid **6**.

The novelty of this reaction lies in its comprehensive study under PTC, all in one pot. Additionally, a new nucleophilic agent initiates the process—the phenolate **1**, replacing the previously suggested F^- ion. This substitution helps avoid the difficulties and problems associated with the latter.

In the preceding chapters, different methodologies for studying cinchona-based catalysts were benchmarked, leading to the identification of the most efficient one for working with these organic catalysts. The nature of the interaction established between charged organo-catalysts and anionic reactants was revisited and a detailed analysis was conducted, providing a foundation for their examination in future projects. Finally, the topology of the cinchona-based catalyst was analysed, exploring several binding positions and revealing a definitive binding mode between the catalyst and the anionic reactants crucial in PTC catalysis reactions.



Scheme 4.4. Reaction mechanism of the production of protected amino acids using azalactone as starting material.

The preceding steps were undertaken to identify the most efficient and unbiased workflow and techniques for studying reactions under PTC conditions, specifically in the presence of cinchona-based catalysts.

In this chapter, all those techniques will be applied to the computational design of efficient cinchona-based organocatalysts for the production of enantiomerically pure protected bis-aminoacids under PTC and Dynamic Kinetic Resolution (DKR) conditions.

4.2 Computational Details

Analogously to what it was done in the previous chapters, the CREST^{22–24} package, part of the xTB²⁵ program, was used to perform a conformational search of the isolated catalysts. This conformational analysis was performed using the iMTD-GC algorithm at the GFN n -xTB²⁶ (GFN2-xTB) semiempirical level. The energy limit for the metadynamics simulation was set to 6.0 kcal/mol, the RMSD cutoff to 0.125 Å and the energy threshold between conformer pairs to 0.05 kcal/mol. All the calculations were performed using version 6.2 of the CREST package.

This RMSD calculations were performed using *calculate_rmsd*²⁷ package. The Hungarian algorithm²⁸ was used as the atomic reordering method and the Kabsch algorithm²⁹ was used as the system rotation method.

The RMSD value for all the possible pairs of conformers of the same catalyst is calculated,²⁷ using the Hungarian algorithm²⁸ and the Kabsch algorithm²⁹ as the reordering and system rotation methods, obtaining an $N \times N$ RMSD matrix. Then an RMSD threshold was iteratively calculated to filter the generated conformers, obtaining $N_{remaining} = 10$ that will be reoptimised using DFT. The structures of the optimised catalysts can be found in Table S3.2.1.

The structures of all the systems under study were optimised at the M062X/6-31g(d)^{30–32} computational level. Harmonic vibrational frequencies were computed at the same level used for geometry optimization in order to confirm that the stationary points are local minima.

Single point energies for the lowest energy small basis set calculations were computed using the M062X/6-311+g(d,p) computational level for the conformational analysis and all the studied reaction steps.

Solvent effects (toluene) were included in the optimization by means of a continuum method, the Solvation Model based on solute electron Density (SMD) approach.³³

The reported free energies ($\Delta G_{def2tzvp}$) in this document were computed by adding the thermal correction ($TC_{def2svp}$) from the small basis set calculations to the electronic energy obtained from the higher level single point calculations ($E_{def2tzvp}$), as shown in Equation 1.

$$G_{def2tzvp} = E_{def2tzvp} + TC_{def2svp} \quad (1)$$

The interaction free energies ($\Delta G_{complex}$) were calculated as the difference between the free energy of the optimised complex and the energy of each monomer in its optimised geometry, as expressed in Equation 2.

$$\Delta G_{complex} = G_{complex} - \sum G_{monomer} \quad (2)$$

All calculations were carried out using the Gaussian09 software package[34] and the optimised structures for all studied catalysts, complexes and reaction steps can be found in Table S4.1.

Based on the Curtin-Hammett principle³⁵ and according to the calculated Gibbs free interaction energies of the transition states, the theoretical enantiomeric excess (*ee*) value was determined based on Equation 3 where $[R]$ and $[S]$ represent the concentrations of the different enantiomers.

$$ee = \frac{k_{favoured} - k_{unfavoured}}{k_{favoured} + k_{unfavoured}} \quad (3)$$

The *ee* value was determined as the ratio of the rate constants at $T = 243$ K as described

in Equation 4),³⁶ where $\delta\Delta G^\ddagger$ is the difference of Gibbs free activation energy between the enantiomers, R is the gas constant, and T is the temperature.

$$ee = \frac{\exp\left(\frac{-\delta\Delta G^\ddagger}{RT}\right) - 1}{\exp\left(\frac{-\delta\Delta G^\ddagger}{RT}\right) + 1} \quad (4)$$

The energy values for the different species and the corresponding calculated %*ee* can be found in Table S4.2

The strength of the different HBs found in the cat-reactant complexes under study was evaluated by means of NBO and QTAIM analysis.

The NBO methodology³⁷ version 7 was used to assess the strength of various NCI present in the systems, measuring the charge transfer between the occupied *LP* from the HB-acceptor and the empty *BD** orbitals of the HB-donor.

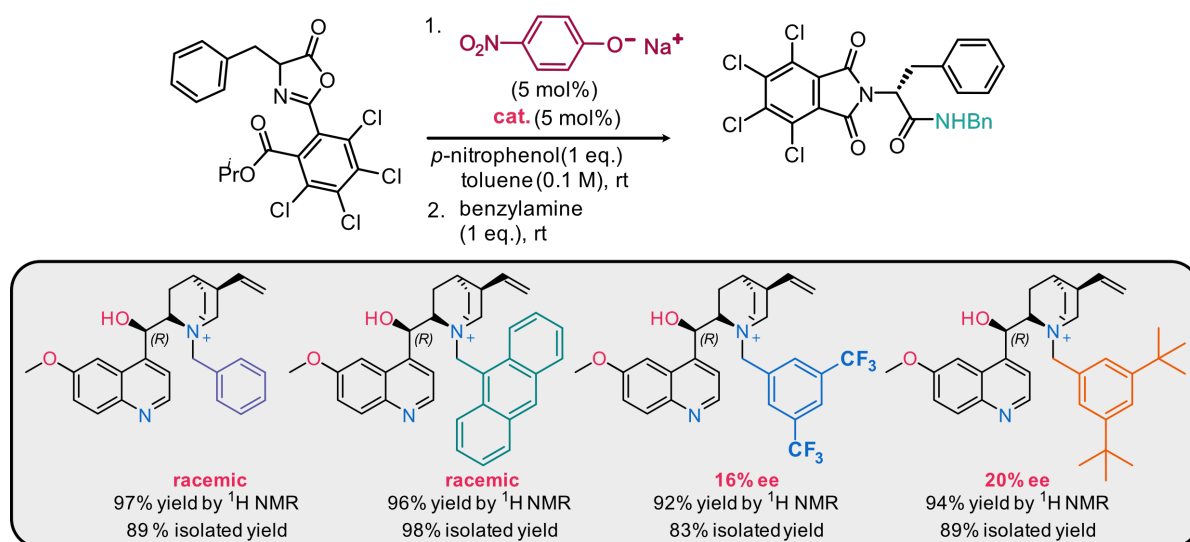
The QTAIM^{38,39} methodology was used to analyse the electron density of the systems, conducted within the AIMAll software.⁴⁰ The electron density at the BCP between the catalysts and the different anions was examined.

The NBO and QTAIM calculations were performed at the optimisation level of theory.

4.3 Results and Discussion

4.3.1 Preliminary Study

As previously highlighted in the Introduction section, this study marks a groundbreaking chapter since it is the first time that the production of bis-protected amino acids (Scheme 4.5) is performed under PTC conditions. Hence, the initial step in this collaborative project was to demonstrate the feasibility of this reaction under these conditions. This validation was conducted by our experimental collaborators, the research group led by Professor Connon, who tested four of the most commonly employed cinchona-based catalysts (**cat_ph**, **cat_anthra**, **cat_cf3**, **cat_tbu**). The modifications at position R_1 cover a wide range of sizes (**ph** < **cf3** < **tbu** $\approx$ **anthra**). This variation takes into consideration the steric effects of the substituents on the selectivity. Additionally, different electron withdrawing groups (EWG) (**cf3**) and electron donating groups (EDG) (**tbu**) were incorporated to assess how different electronic properties affect the selectivity of the catalyst.



Scheme 4.5. Reaction under study and catalyst tested experimentally to validate the experimental conditions.

This preliminary experimental study effectively served as a proof-of-concept for the reaction. It demonstrated that these catalysts were capable of producing high reaction yields consistently in all cases (^1H NMR yield > 92% and isolated yield > 83%). Additionally, it provided the optimal experimental conditions that should be replicated in computational studies. These conditions involve using toluene as a solvent and studying the reaction at room temperature (298K).

Despite the high yields achieved in this preliminary study, the enantioselectivity of all of the studied catalysts is absent or extremely low (%ee < 20%). This results in the production of both enantiomeric products at comparable rates, highlighting the inadequate design for this particular reaction.

In light of these findings, the initial objective of this project was to elucidate the factors behind the lack of selectivity exhibited by these catalysts. To achieve that, a conformational analysis of both the catalysts and the reactants was performed, to address their structural flexibility.

Conformational Analysis

This initial set of catalysts was already studied in the second part of the previous chapter; therefore, despite using different methodologies (see computational details), the same computational results will be considered.

The active conformer of each of the catalysts under study, along with the Boltzmann population of all of the different conformers studied, can be observed in Figure 4.1.

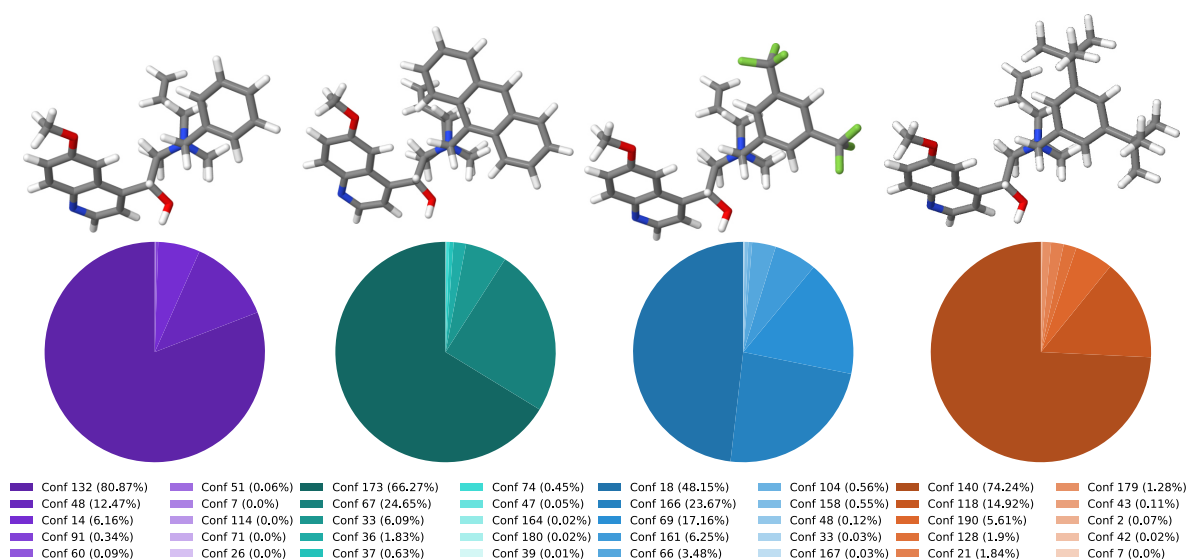


Figure 4.1. Conformational analysis catalyst and their active species. Left to right: **cat_ph**, **cat_anthra**, **cat_cf3**, **cat_tbu**.

It is clear that for **cat_ph** and **cat_anthra**, the active conformation is also the most stable. Furthermore, in cases where the most stable and active conformation do not align (e.g. **cat_cf3** and **cat_tbu**), the active remains significant, representing approximately 15% of the total population. Additionally, an examination of the preliminary experimental results, consistently reveals high yields in those reactions, demonstrating that all the catalysts present the active conformation that facilitates the transfer of the phenolate from the aqueous phase to the organic phase, thereby activating the reaction mechanism.

The catalysts are not the only flexible components involved in this reaction. The azlactone derivative **1** (*isopropyl 2-(2-benzyl-5-oxo-2, 5-dihydrooxazol-4-yl)-3, 4, 5, 6-tetrachlorobenzoate*, Figure 4.2 top left) also has several rotatable bonds that will greatly affect the stability of the system and that must be studied. Consequently, a conformational analysis of this azlactone was performed (Figure 4.2).

The conformational analysis revealed that the majority of the conformers were produced by the rotation of only two of the rotatable bonds within the structure, resulting in a myriad of structures with similar populations, rather than a single very stable conformer that distinguishes itself from the rest.

In view of these results, the most adequate conformer in each situation was used to build the different systems using chemical intuition and giving priority to those that are more stable and establish the best non-covalent interactions with the catalysts.

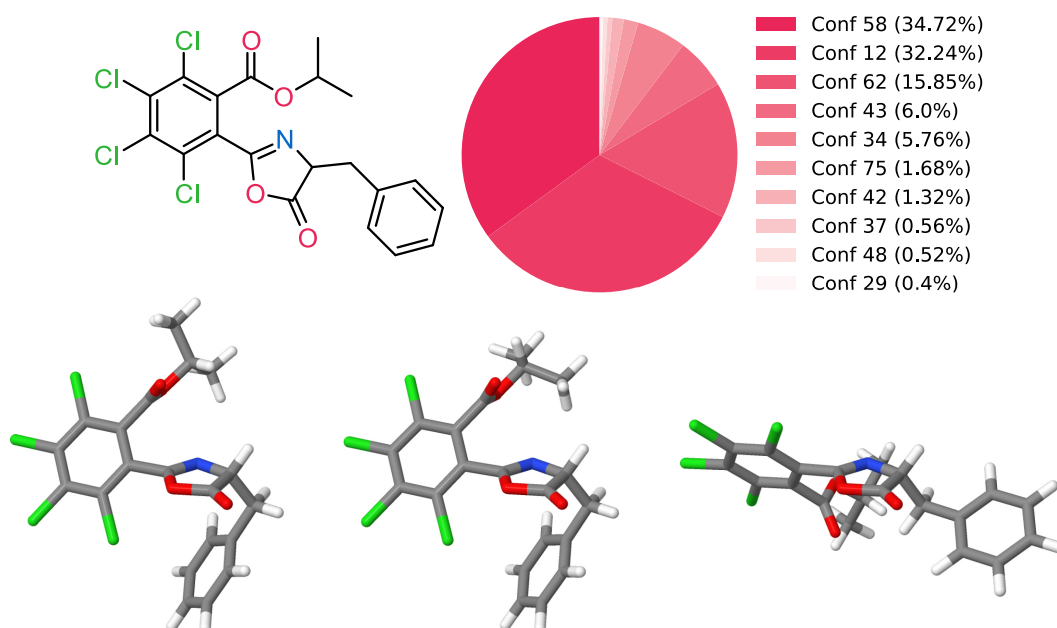


Figure 4.2. From left to right and top to bottom: 2D scheme of the azlactone derivative under study, Boltzmann population of the different conformers obtained, structure of the three most stable conformers.

Binding Mode Study

As in the previous chapter concerning the cyanation reaction, these catalysts exhibit different binding positions for accommodating the different reactants. However, given that all of them were thoroughly examined in the previous chapter, and the results consistently demonstrated that establishing a HB was always the most favoured interaction, only this specific position will be considered.

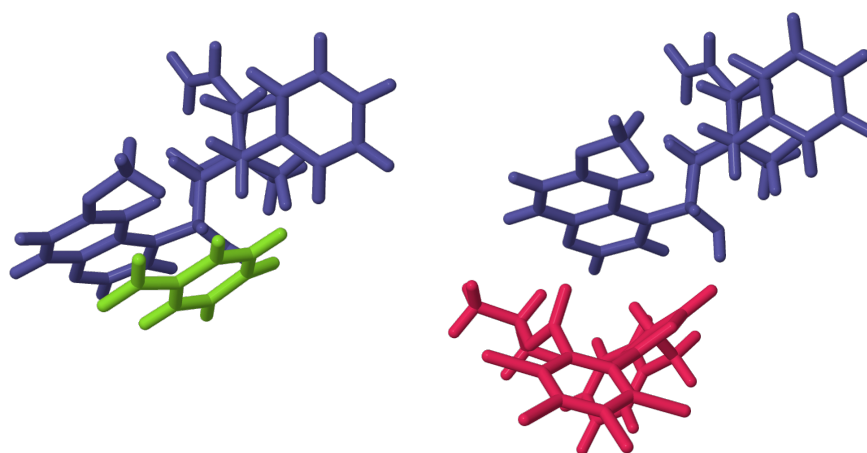


Figure 4.3. 3D structures of **cat-phenolate** (left) and **cat-azlactone** (right).

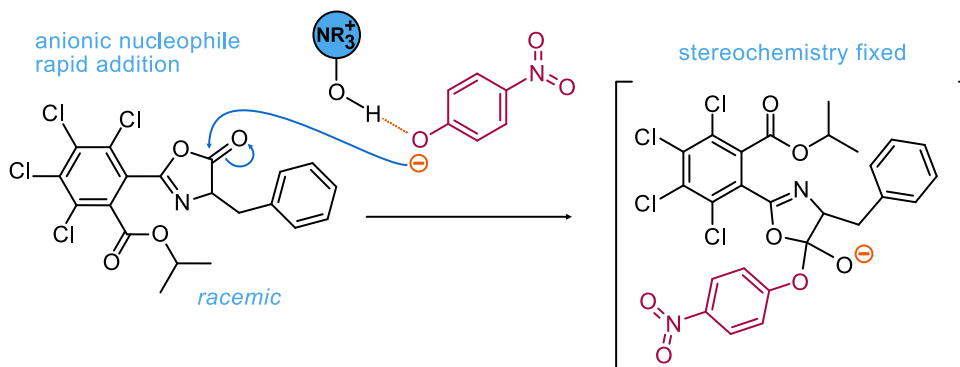
However, the interaction energy between the catalyst and the two reactants must be studied in this position to determine which establishes the strongest HB. Considering the mechanism of a PTC (Scheme 2), the catalyst must interact with the phenolate in the aqueous phase and transfer it to the organic phase for the reaction to proceed. Therefore, it is expected that the phenolate is the reactant interacting at this OH position in the catalyst, especially taking into account that a HB between a cation and an anion would be particularly strong. However, this

situation may change in the organic phase. Consequently, the interaction energies in the two potential **cat-reactant** complexes (**cat-phenolate** and **cat-azlactone**) were optimised and analysed (Figure 4.3).

As anticipated, considering that both the catalyst and the phenolate are charged species with opposite charges, the complex **cat-phenolate** is considerably more stable than the catalyst **cat-azlactone**. To illustrate, the **cat-phenolate** complex is 38.6 *kcal/mol* more stable than the **cat-azlactone** complex. Therefore, it can be confidently concluded that the only viable interaction in the organic phase is between the catalyst and the phenolate.

Decoding Chirality: Mechanism and Catalyst Role in Azlactone Nucleophilic Attack

After determining the definite binding mode of the catalyst with the different reactants, forming a HB between the catalyst and the phenolate, the mechanism of the reaction and the role of the catalyst was studied. The first step of the reaction involves the rapid nucleophilic attack of the *p*-nitrophenolate on the azlactone derivative, locking the chirality of the product (Scheme 4.6). This step is the focus of the current investigation.



Scheme 4.6. Nucleophilic attack of the phenolate to the azlactone; the first step of the reaction and responsible for the enantioselectivity of the product.

The different TSs leading to the *R* and *S* products were constructed on the binding mode determined in the previous section, where the phenolate forms a HB with the catalyst. The placement of the azlactone was guided by chemical intuition to enhance the NCIs with the catalyst while avoiding sterical clashes. The aim was to obtain the most stable and accessible TS for each of the chiral products.

Applying the Curtin-Hammett principle, the %*ee* induced by each catalyst was calculated, based on the energy of the different TS (see computational details). The resulting theoretical and experimental %*ee* values are provided in Figure 4.4.

Examining the results reveals two distinct situations. The computational and the experimental %*ee* values align closely for **cat_ph** (3.8%*ee*_{comp} vs. 0%*ee*_{exp}) and reasonably for **cat_anthra** (44.4%*ee*_{comp} vs 0%*ee*_{exp}). However, a notable disparity emerges for **cat_tbu** (87.0%*ee*_{comp} vs 20%*ee*_{exp}). The largest discrepancy is observed with **cat_cf3**, where the difference between the experimental and computational enantioselectivity exceeds 80% (96.6%*ee*_{comp} vs 16%*ee*_{exp}).

These discrepancies may have arisen because the most stable conformer of one of the two transition states in each case, *TS_S* for **cat_cf3** and *TS_R* for **cat_tbu**, was not successfully obtained. This highlights that relying solely on chemical intuition is insufficient for comprehending

the mechanism of a reaction and the role played by its various components involved.

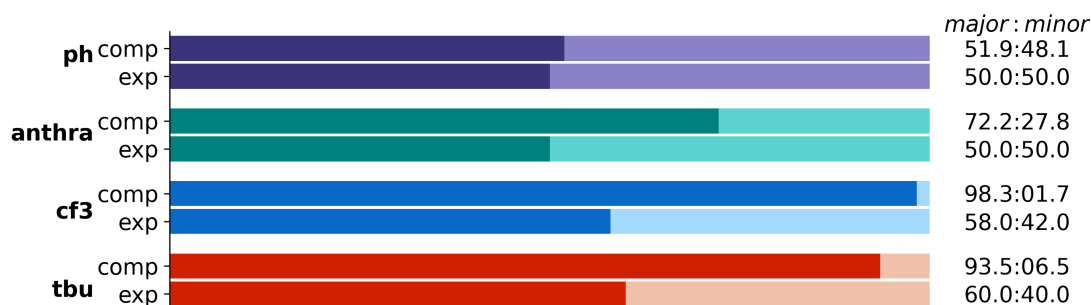


Figure 4.4. Computational and experimental *major:minor* ratios for, top to bottom, **cat_ph**, **cat_anthra**, **cat_cf3** and **cat_tbu**.

Furthermore, it is important to note that this disagreement between experimental and computational %*ee* results becomes significantly pronounced at low %*ee* values. This amplification is due to the exponential nature of the %*ee*_{comp} calculation (Figure 4.5). This exponential nature of this calculation means that even small energy differences between the obtained TS lead to substantial differences in the *ee* calculated values.

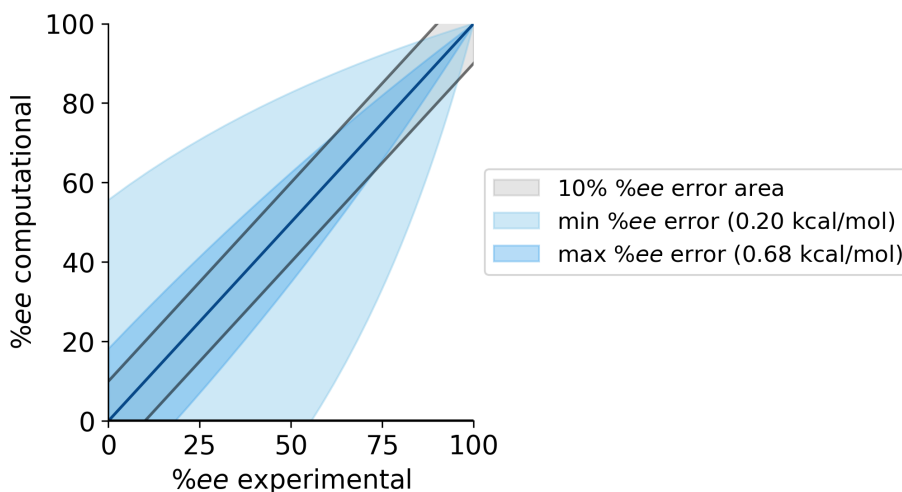


Figure 4.5. Experimental vs computational %*ee* error calculated with 0.20 kcal/mol (light blue) and 0.68 kcal/mol (dark blue) DFT free energy error and $\pm 10\%$ confidence area (grey).

Based on the minimum and maximum error values calculated by *Walker et al.*,⁴¹ using a methodology akin to that applied to similar NCIs, the minimum DFT error, 0.20 kcal/mol, lead to a %*ee*_{error} up to 18% at low %*ee* values (dark blue area in Figure 4.5); while the maximum DFT error, 0.68 kcal/mol, can lead to %*ee*_{error} bigger than 50% at the low %*ee* regime (light blue area in Figure 4.5). Within this maximum error range, predictions for selectivities below 50% can favour the opposite enantiomer, implying not only a different but also an opposite selectivity. In fact, it is necessary to reach *ee* values up to 75% to ensure that the selectivity error remains below 10% (grey area in Figure 4.5). Considering the previously discussed insights, it can be stated that computational results for *ee* maintain their quantitative accuracy primarily within the high selectivity ranges (< 80%). In contrast, they should be viewed as qualitative in the context of low selectivity regimes.

In light of these considerations, it becomes evident that the selectivity values for these catalysts must be regarded as qualitative. This is mainly because they tend to fall within the low selectivity regime. Consequently, our analysis will shift its focus towards a detailed examination of the different TS, aiming to uncover the origins of this limited selectivity.

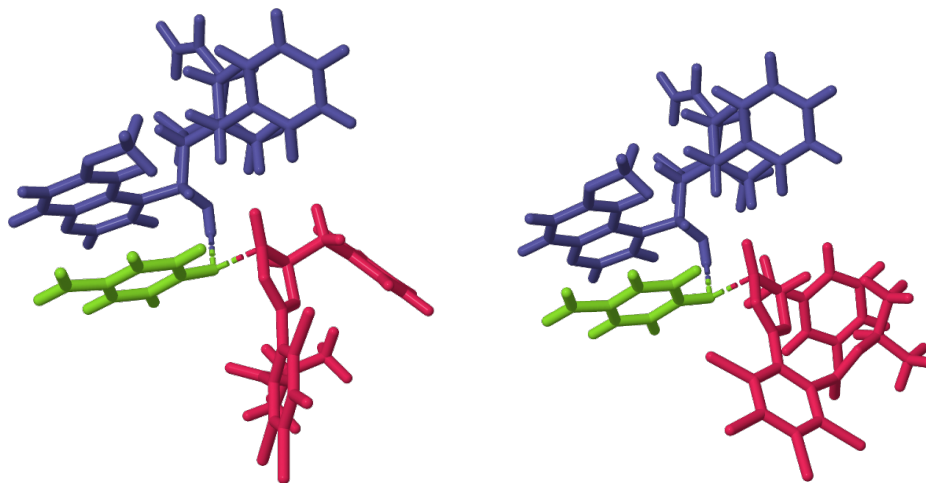


Figure 4.6. 3D structures for TS_R (left) and TS_S (right) under the presence of **cat_ph**.

Despite the suboptimal design of these catalysts for this specific reaction, the transition states must be analysed to gain insight into the factors contributing to this lack of success. Upon examining the 3D structures of the different TS obtained (Figure 4.6 and S4.1) it becomes apparent that in all the cases the azlactone is located in a very similar position for both TS_R and TS_S . Furthermore, it is noteworthy that the azlactone appears entirely unbound, displaying no interactions with the catalyst. This mobility and the open orientation necessary for the azlactone to facilitate the reaction explains the minimal variations in energy among the different TS, thus accounting for the process's limited selectivity.

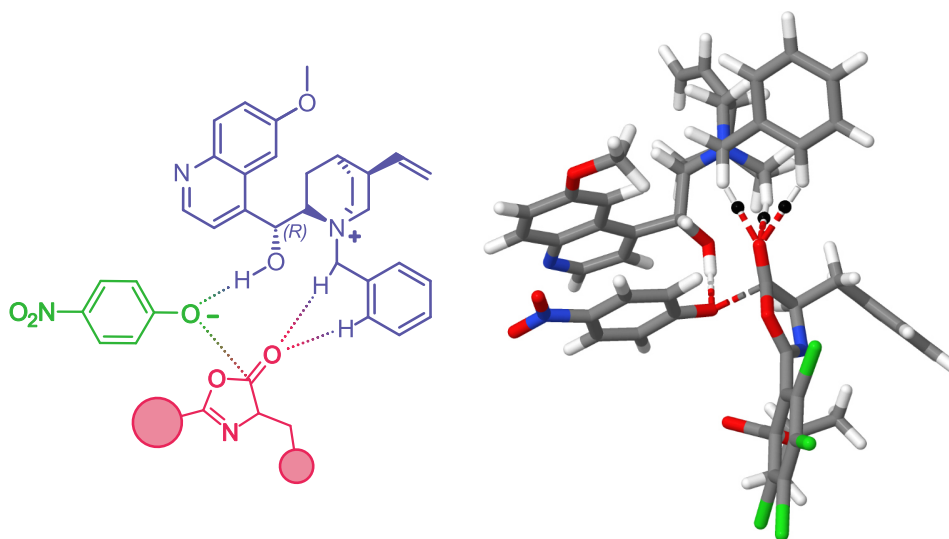


Figure 4.7. a) Schematic view of the NCIs established in the TS under study under the presence of **cat_ph**. b) 3D molecular graph of the TS showing the web of HBs established between the azlactone and the catalyst and their corresponding BCPs as black spheres.

Nonetheless, upon closer examination, it becomes apparent that the O atom in the azlactone is establishing a bifurcated HB with one of the H atoms in the CH_2 and another H atom in the aromatic ring in all the structures under study (Figure 4.7-a). In order to confirm the presence of these weak interactions, an QTAIM and a NBO calculations were performed on both TSs under the presence of **cat_ph**.

Three relevant HBs are identified by means of QTAIM analysis between the O atom in the azlactone and the different H atoms in α -position to the N^+ of the quinuclidine and one H_{BCP} of the phenyl ring (Figure 4.7-b) and the NBO charge transfer between the LP orbital of the O atom and the different anti-bonding sigma orbitals of the different HB-donors ($LP(O) \longleftrightarrow BD^*(X - H)$).

The results of the topological analysis, ρ_{BCP} and H_{BCP} , of these HBs and their corresponding second-order perturbation energy, $E(2)$, are presented in Table 4.1. This analysis reveals that, as anticipated, these interactions constitute exceptionally weak HBs ($\rho_{BCP} \leq 0.0153 a.u.$, $H < 0.001 a.u.$, $E(2) < 5.0 kcal/mol$). They give rise to an exceedingly flexible HB-web that only loosely confines azlactone in position. This feeble confinement fails to generate a substantial disparity between TS_R and TS_S , favouring the formation of both enantiomers equally.

Table 4.1. Density values at the BCP, ρ_{BCP} (a.u.), Energy density, H (a.u.), and Second order perturbation energy, $E(2)$ (kcal/mol), of the different HB in TS_R and TS_S between the azlactone and **cat_ph**.

	HB ₁			HB ₂			HB ₃		
	ρ_{BCP}	H_{BCP}	$E(2)$	ρ_{BCP}	H_{BCP}	$E(2)$	ρ_{BCP}	H_{BCP}	$E(2)$
TS_R	0.0153	0.000938	4.72	0.0150	0.000143	4.80	0.0113	0.000745	3.11
TS_S	0.0150	0.000966	4.55	0.0141	0.000221	4.49	0.0114	0.000724	1.49

Nonetheless, the presence of these weak HBs between the catalyst and the azlactone represents an auspicious initial step to fortify the interactions between the catalyst and the azlactone, aiming to stabilise one of the TSs and favour the production of one of the enantiomers over the other. Although the initial results may have been disappointing in terms of achieving high enantioselectivity and designing an optimal catalyst, this discovery paved the way for the following enhancements and modifications.

4.3.2 Additional Anchor Point

Considering the observed weak bifurcated HB between the azlactone and the aromatic substituents in the catalyst, a decision was made to modify the catalyst at this position. Different groups were chosen to introduce an additional HB-donor, with the goal of strengthening the interaction between the catalyst and the azlactone. This enhancement aimed to stabilize the azlactone and promote the preference of one of the TS over the other.

The selected groups must satisfy two requirements. First, they needed to closely resemble the original groups in terms of their structural arrangement, thus minimising any significant alterations to the positions of the hydrogen atoms. Additionally, these selected groups had to incorporate a strong HB-donor capable of forming stronger hydrogen bonds, thereby creating a second anchor point in the catalyst. Our investigation led us to consider two distinct groups as potential candidates: acetamide and n-phenyl acetamide (Figure 4.8).

The selection of these two groups allows us to study the role of the isolated anchor point

(**cat_conh2**) and study the combined effect of the HB-donor and the addition of an aromatic ring (**cat_conh2ph**), along with their effects.

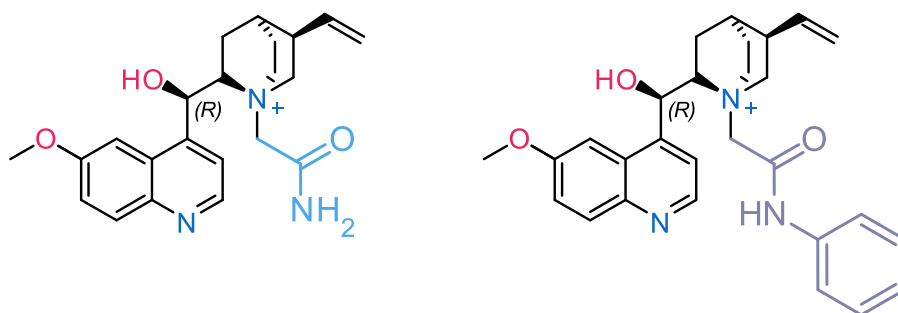


Figure 4.8. Modified catalysts under study after adding a second HB-donor: **cat_conh2** (left), **cat_conh2ph** (right).

The introduction of a second anchor point fundamentally transformed the catalyst's interaction with the phenolate before the reaction. Therefore, a detailed analysis of the catalyst, exploring the different binding modes that can be established was performed.

Binding Mode Study

As observed in the previous section, the preferred binding position for the anion to interact with the catalyst involves the formation of a HB with the hydroxyl group within the catalyst. However, with the introduction of this second amide HB-donor, a new potential binding site comes into play, and it requires consideration.

To initiate the comparison of the two distinct binding sites, a topological analysis of the Molecular Electrostatic Potential (MEP) of the catalyst (Figure 4.9) was performed. This analysis not only yielded a map of the high- and low-electron-density areas but also provided the electrostatic potential value at the critical points corresponding to the two different donors in each catalyst (OH and NH). This comparative analysis aimed to reveal valuable insights into the strength of the HBs that each HB-donor could potentially contribute to complexation. Consequently, it would help determine the most likely position for the anion to interact with the catalyst.

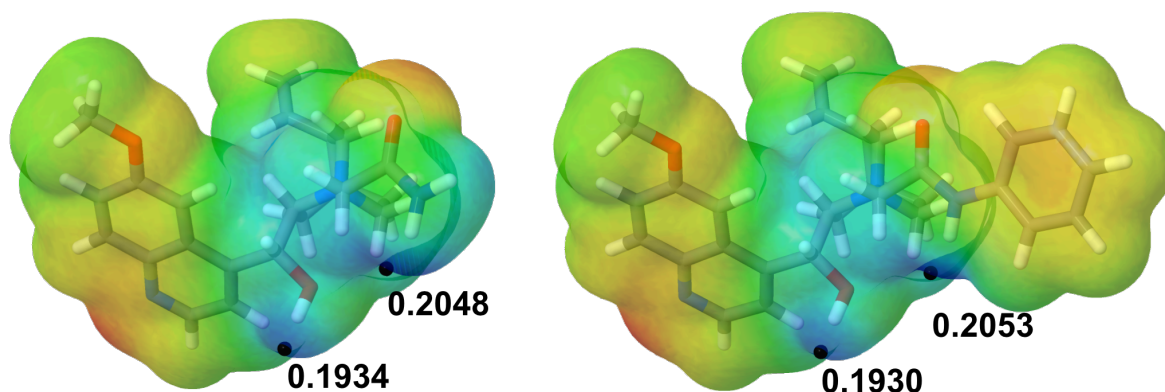


Figure 4.9. Molecular Electrostatic Potential (MEP) (a.u.) of **cat_conh2** and **cat_conh2ph** showing the critical points corresponding to the two different HB-donors.

The results showed that the NH HB-donor position is slightly more positive than the OH

binding site. Consequently, it is more likely for the phenolate to establish the interaction in this position. However, this difference is not significant enough to provide a definite answer. Furthermore, there are additional NCIs that can be established between the catalyst and the phenolate that stabilises one position over the other. For instance, a π - π can be observed between the quinoline ring within the catalyst and the phenolate in all the structures in the previous section. These supplementary interactions may contribute to the overall stability and preference for a particular binding site.

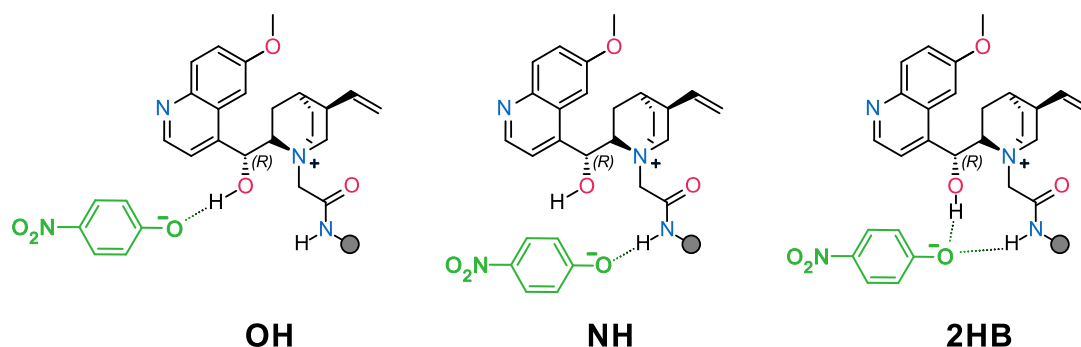


Figure 4.10. Different binding modes studied for the cat-phenolate complexes; from left to right **NH**, **OH** and **2HB**.

To comprehensively explore the diverse binding possibilities, our study examined three distinct binding modes (Figure 4.10). The first two binding modes are the most straightforward, involving the formation of HB with each of the HB-donors (**NH** and **OH**). However, a third binding mode (**2HB**) is essential to consider due to the nature of the phenolate. The O atom in phenolate possesses three *LP* orbitals, and the two HB-donors are attached to flexible groups that have the potential to rotate. This dynamic behaviour can bring these hydrogen bond donors closer in spatial proximity, resulting in the formation of a bifurcated HB.

The **cat-phenolate** complexes in the three different binding modes (**NH**, **OH** and **2HB**) were built and optimised with both **cat_conh2** and **cat_conhph** catalysts. However, it was observed that in the case of the HB systems, they invariably transitioned to the **2HB** binding mode, rendering the other modes elusive. Despite this challenge, three distinct positions of the phenolate name **back**, **front**, **bottom** were identified in both catalysts and their 3D structures can be found in Figure 4.11.

All the different complexes obtained exhibit identical interactions, two HBs between the phenolate and the catalyst. Remarkably, the 3D structures of these complexes remain highly consistent, regardless of the catalyst under study. Furthermore, the orientation of the phenolate of the **front** conformation is analogous to the binding mode found for the previous set of catalysts (Figure 4.3) establishing, in addition to the two HBs, a π - π interaction with the quinoline ring of the catalyst. To compare the importance of the different conformers of these complexes, their Boltzmann populations were calculated and shown in Figure 4.12.

As expected, the combination of the good alignment of the O atom within the phenolate and the two different HB-donors of the catalyst in the front conformation and the extra stabilising π - π interaction in the **front** conformation (Figure 4.11 middle) makes this conformation the most stable and, therefore, the most abundant. The second most abundant conformation, **down** (Figure 4.11 right), presents a similar structure, with the phenolate placed on the same side of the catalyst, and a parallel alignment between the different HB-donors and the HB-acceptor

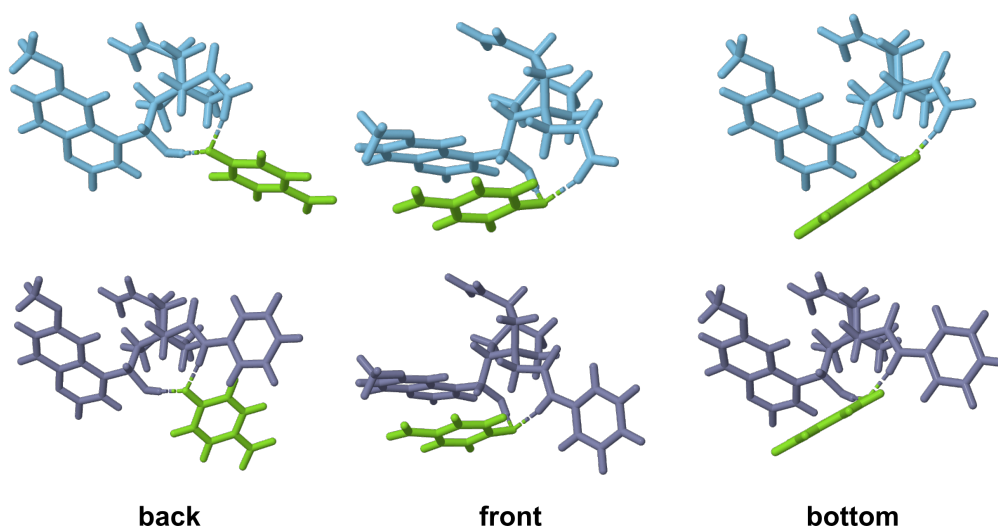


Figure 4.11. 3D structures of the different conformers of the cat-phenolate complexes between the phenolate and **cat_conh2** (up) and **cat_conhph** (down).

(O), minimising the strain of the structure of the catalyst. Conversely, the least favoured and most unstable configuration is the **back** conformer, constituting less than 1% of the Boltzmann population. This conformation exhibits a markedly different structure, with the phenolate positioned on the opposite side of the aromatic ring of the catalyst. Unfavourable alignment between HB donors and acceptors, coupled with the absence of other attractive interactions, renders the **back** conformation highly unstable. Because of its negligible population, this conformation is considered insignificant and will not be subjected to further scrutiny in the subsequent analysis.

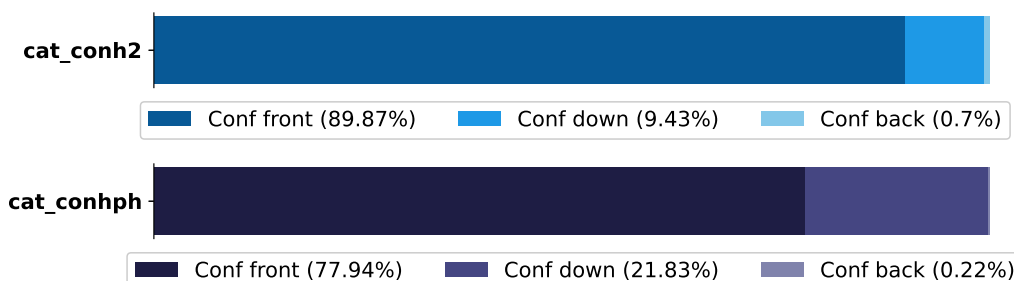


Figure 4.12. Boltzmann population of the different conformers of the cat-phenolate complexes between the phenolate and **cat_conh2** (left) and **cat_conhph** (right).

As aforementioned, the incorporation of this second anchor point into the catalyst serves the purpose of creating a robust and attractive interaction between the catalyst and the azlactone. To achieve this interaction, once the cat-phenolate complex is transferred to the organic phase, one of the two HBs established between the catalyst and the phenolate must be disrupted by the presence of the azlactone. Without this necessary structural adjustment, the reactants would be unable to approach each other closely enough, rendering the nucleophilic attack and subsequent reaction unattainable.

To assess the relative strength of both Hydrogen Bonds (OH and NH), as well as the additional HB found (CH) (Figure 4.13), an in-depth analysis by means of QTAIM and NBO was conducted. The primary goal was to identify which of these bonds would be more susceptible to breaking upon the approach of the azlactone, determining the binding mode essential for study-

ing the transition states. The density at the BCP (ρ_{BCP} , the energy density of the interaction (H) and the second-order perturbation energy ($E(2)$) were measured and compared for each of the interactions analysed (Table 4.2).

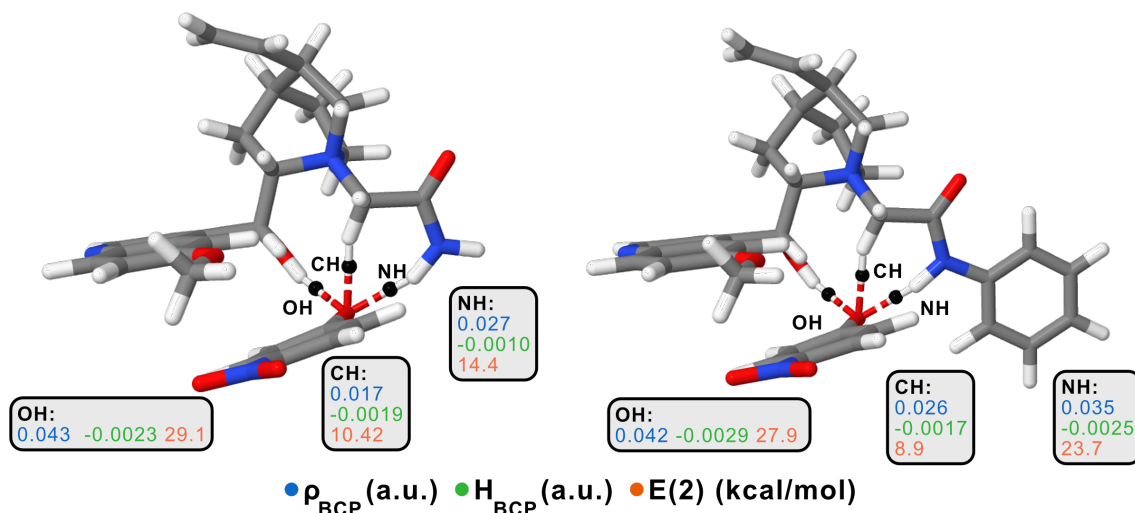


Figure 4.13. Molecular graphs of **cat_conh2-phenolate** and **cat_conhph-phenolate** complexes in **front** conformation and ρ_{BCP} , H and $E(2)$ values associated with each HB between them.

Upon studying the various HBs among the different conformers analysed, a surprising revelation emerges. Contrary to expectations, the strongest HBs do not correspond to the most stable conformation (**front**) but rather to the **bottom** conformation, which presents the optimal alignment between the donor and the acceptors. In the **front** conformation, the individual hydrogen bonds exhibit lower strength, highlighting the importance of the π - π interaction to the global stability of the system.

Remarkably, a consistent pattern emerges when examining the strength of individual HBs, regardless of the catalyst (**cat_conh2** or **cat_conhph**), the conformation of the complex analysed (**front** or **bottom**) or the measurement compared (ρ_{BCP} , H or $E(2)$).

Table 4.2. Density values at the BCP, ρ_{BCP} (a.u.), Energy density, H (a.u.), and Second order perturbation energy, $E(2)$ (kcal/mol), of the different HB in the different conformers under study.

(a) HB analysis in **cat_conh2**.

Position	OH			NH			CH		
	ρ_{BCP}	H_{BCP}	$E(2)$	ρ_{BCP}	H_{BCP}	$E(2)$	ρ_{BCP}	H_{BCP}	$E(2)$
Front	0.043	-0.0023	29.1	0.027	-0.0010	14.4	0.017	-0.0019	10.42
Bottom	0.063	-0.0073	55.1	0.033	-0.0020	17.4	0.026	-0.0016	10.47

(b) HB analysis in **cat_conhph**.

Position	OH			NH			CH		
	ρ_{BCP}	H_{BCP}	$E(2)$	ρ_{BCP}	H_{BCP}	$E(2)$	ρ_{BCP}	H_{BCP}	$E(2)$
Front	0.042	-0.0029	27.9	0.035	-0.0025	23.7	0.026	-0.0017	8.9
Bottom	0.062	-0.0068	53.3	0.042	-0.0026	27.1	0.024	-0.0013	7.9

The HB established with the OH HB-donor (HB_{OH}) is consistently the strongest interaction, often surpassing the combined strength of the other two HBs. As anticipated, HB_{NH} follows in strength, while the HB_{CH} exhibits the weakest one.

Despite the consistent superiority of the hydrogen bonds at the OH position (HB_{OH}) over those in NH position (HB_{NH}), the difference in strength is not pronounced enough to conclusively predict the breaking point upon the approach of the azlactone. Therefore, both scenarios will be explored under the presence of **cat_conh2** allowing for a comprehensive comparison. This investigation aims to ascertain whether establishing the new HB between the catalyst and the azlactone in the NH position is more stable than its counterpart in the OH position.

The 3D optimised structures of the two TSs obtained for the two different binding modes originating from the **front** conformation of the cat-phenolate complex are presented in Figure 4.14. In the first binding mode, **BM1**, the phenolate interacts with the catalyst in the OH position and the azlactone establishes the new HB in the NH position. In contrast, **BM2** reveals the opposite situation, where the phenolate establishes the HB in the NH position, while the azlactone is anchored at the OH position.

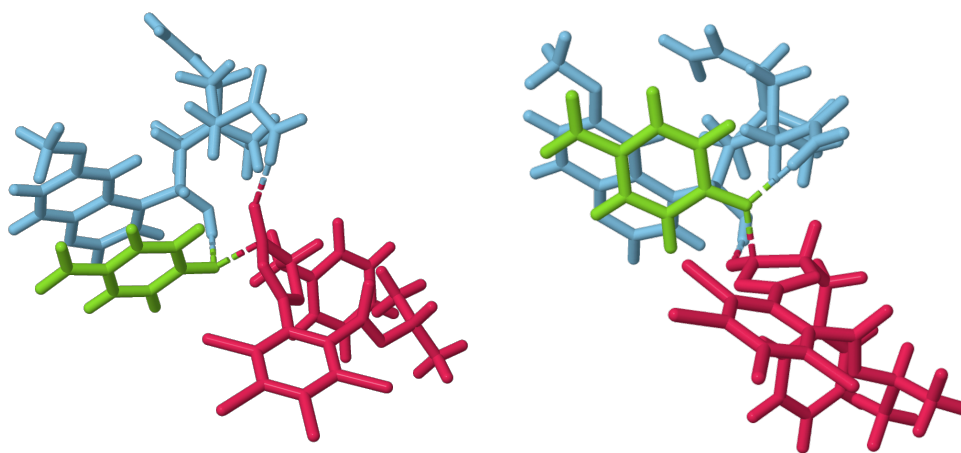


Figure 4.14. 3D structures of the two different binding modes under study for the TS under the presence of **cat_conh2**. **BM1** left, **BM2** right.

The energy difference between these two binding modes is $\Delta\Delta G = 4.88 \text{ kcal/mol}$. Consequently, the Boltzman population of the most stable one, **BM1**, is greater than 99.99%. This significant disparity unequivocally establishes **BM1** as the overwhelmingly more stable binding mode. Consequently, it is evident that, for the investigation of these catalysts, only **BM1** needs consideration, in where the azlactone interacts with the new anchor point of the catalyst.

Decoding Chirality: Mechanism and Catalyst Role in Azlactone Nucleophilic Attack

Upon establishing the definitive binding mode, the focus shifts towards studying the enantioselectivity of the modified catalysts, namely **cat_conh2** and **cat_conhph**. This investigation follows a systematic approach, akin to the methodology employed in the preliminary evaluation of catalysts in the initial round.

The optimisation of the two TSs, denoted as TS_R and TS_S , is carried out based on the previously determined binding mode (**BM1**). In this mode, the phenolate forms a HB interaction at the OH position, while the azlactone establishes it at the NH position of the catalyst. The 3D structures of these optimised TSs can be referenced in Figure 4.15 and S4.1.2.

In Figure Figure 4.15, both TS exhibit structural similarities, with the new anchor point securing the azlactone in a consistent manner for both *R* and *S* enantiomers. The primary

difference lies in the orientation of the phenyl ring within the azlactone, pointing in different directions in each of the TSs. In TS_R (Figure 4.15-left), the phenyl ring engages in σ - π interactions with several H atoms from the quinuclidine ring establishing. However, the rigidity of this new **hb** between the azlactone and the catalyst restricts the space for the interaction, forcing the phenyl ring to adopt a significantly altered conformation, thereby destabilising the structure.

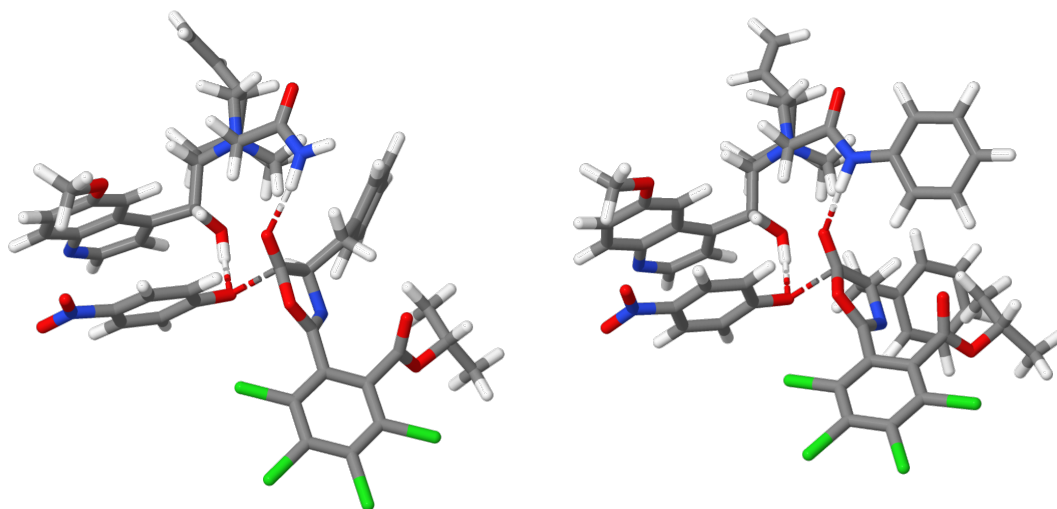


Figure 4.15. Optimised 3D structures for TS_R (right) and TS_S (left) with **cat_conh2**.

This difference between the two transition states and the rigidity of the interaction established in this new anchor point, contribute to an increased energy difference between the two TSs ($\Delta\Delta G = 0.49 \text{ kcal/mol}$), favouring the approach of the *R* enantiomer over the *S*, and resulting in a 42%*ee*.

A similar situation can be observed with **cat_conhph**, where the azlactone is located in an analogous position in the corresponding TSs. However, the addition of the phenyl ring to the amide introduces steric hindrance between the catalyst and the azlactone, marginally destabilising this position and decreasing the energy difference between the two TSs ($\Delta\Delta G = 0.06 \text{ kcal/mol}$). This slight destabilisation leads to a preference for the reaction with the *S* enantiomer, yielding a result of 6.2%*ee*, nearing a racemic mixture.

Despite the structural disparities and the modest enantioselectivities achieved, the potential of this new anchor point appears promising. Following discussions with our experimental collaborators, they express eagerness to conduct laboratory trials (Figure 4.16) to assess its impact on the reaction and viability.

Furthermore, two new catalysts based on **cat_conhph** were developed, to investigate how modifications to the phenyl ring impact catalyst selectivity. The first modification strengthens the new HB by introducing an EWG ($-NO_2$, **cat_conhph-pno2**) to the phenyl ring at the *para* position; while the second one introduces a bulky group to the phenyl ring ($-Br$, **cat_conhph-obr**) at the *ortho* position, increasing the steric hindrance between the catalyst and the azlactone in one of the TS. These modifications were collaboratively selected based on their chemical and structural properties, considering experimental feasibility.

The same computational procedure was followed with the two new catalysts, optimising the TSs with the different enantiomers. Simultaneously, the experimental group synthesised and tested the four catalysts are shown in Figure 4.8 and Figure 4.16.

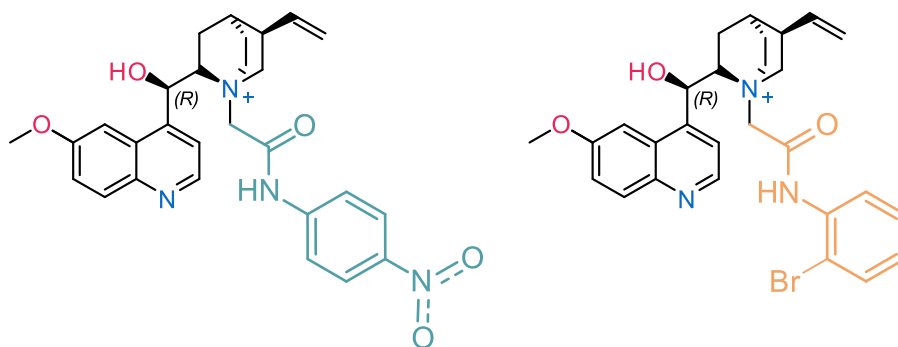


Figure 4.16. Modified catalysts under study after adding a second HB-donor: **cat_conh2** (left), **cat_conhph** (right).

The TS for each enantiomer with the different catalyst were optimised, and computational *ee* were calculated using the energy difference ($\Delta\Delta G$) between TS_R and TS_S . Notably, the two new catalysts exhibited higher and comparable selectivity values, 84.8%*ee* with **cat_conhph-pno2** and 85.0%*ee* with **cat_conhph-obr**, representing a significant enhancement in selectivity compared to the previous ones. These selectivity values, along with the corresponding experimental *ee* results, are presented as the *major:minor* ratio, in Figure 4.17.

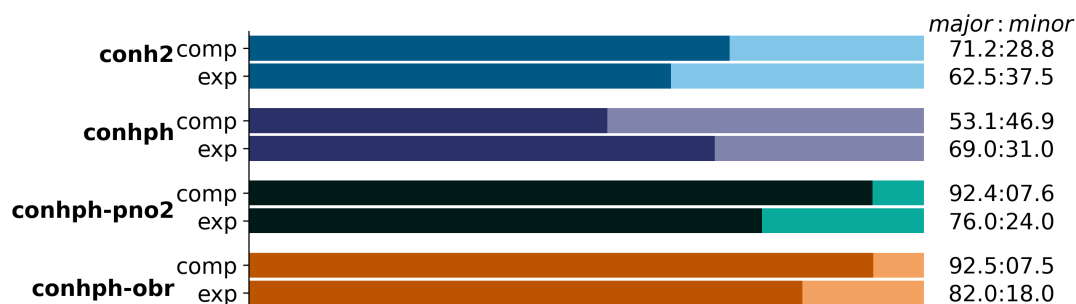


Figure 4.17. Computational and experimental *major:minor* ratios for, top to bottom, **cat_conh2**, **cat_conhph**, **cat_conhph-no2** and **cat_conhph-obr**.

Comparing the experimental and the computational data, a better agreement than in the first set of catalysts is observed, despite the low %*ee* values obtained for the first two catalysts, **cat_conh2** and **cat_conhph**. From the computational results, it can be seen how the most stable TS gradually changes from TS_R under the presence of **cat_conh2** (42.4%*ee* *R*) to TS_S when **cat_conhph-pno2** (84.8%*ee* *S*) or **cat_conhph-obr** (85.0%*ee* *S*) are considered, obtaining almost a racemic mixture with **cat_conhph** (6.2%*ee* *S*). This behaviour was also observed experimentally, presenting very low selectivities for **cat_conh2** (25.0%*ee*) and **cat_conhph** (38.0%*ee*). However, selectivity values increased up to 52%*ee* with **cat_conhph-pno2** and 64%*ee* with **cat_conhph-obr**, highlighting the importance of combining attractive NCIs formed with the new anchor point of these catalysts and bulkier groups that create sterical crashes, favouring one enantiomer over the other.

The 3D structures of the TSs obtained for **cat_conhph-pno2** (Figure S4.1.2), are observed to be almost identical to those obtained with **cat_conhph**, with the azlactone located in a very similar position and establishing the same interactions with the amide. The addition of the $-NO_2$ group to the phenyl ring has increased the acidity of the H atom in the amide and, therefore, increased the strength of the HBs established with the azlactone. Furthermore, it has

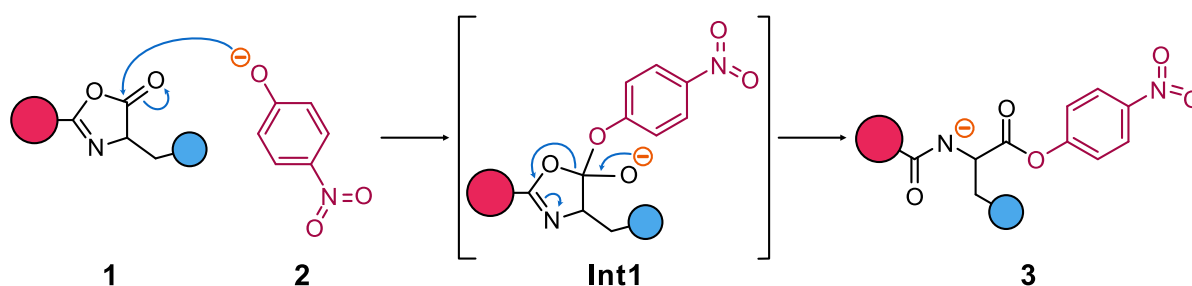
also increased the acidity of the H atoms in the phenyl ring, which are now acidic enough to establish a weak, bifurcated, HB with the O atom in the isopropyl ester present in the azlactone (Figure S4.1.2).

The situation, when looking at the TSs with **cat_conhph-obr**, remains similar but, in this case, due to the bigger size of the *Br* atom in the phenyl ring, the azlactone presents a greater displacement and the weak bifurcated HB between the ester and the aromatic ring, present in the previous catalyst, is not observed.

The results obtained from studying these modified catalysts greatly improved the selectivity of the reaction, starting from a racemic result in the first catalyst **cat_ph** and getting a 64%*ee* with **cat_conhph-obr**. This improvement is considered a great achievement for a single round of modifications and it reveals the huge potential of computational rational design of asymmetric catalysts; however, further improvements of this selectivity have been proven to be a computational and experimental challenge.

Additionally, when computationally studying these systems, several potential problems that might be the cause of the lack of selectivity were identified. The first one is the flexibility of these TSs and the myriad of NCIs that can be established between the catalyst and the reactants. This makes it particularly difficult to obtain the most stable conformation, representing a challenge for chemical intuition and resulting in biased results.

It was also observed that, as the reaction proceeds and the negative charge from O in the phenolate to O in the azlactone (Scheme 4.7) forming the intermediate, **Int1**, the binding mode changes and the interactions between the catalysts and the reactants change, establishing new NCIs that might affect the global stability of the reaction or even mean that the first step of the reaction is not the only one defining the selectivity of the global mechanism.



Scheme 4.7. Suggested movement of electrons and of charge over during the nucleophilic attack of the phenolate to the azlactone and the ring opening process.

Furthermore, our experimental collaborators observed that traces of the protonated version of the phenolate (*p* – *nitrophenol*) present in the organic phase of the reaction have an effect on global selectivity, making it necessary to include this species in the computational study of the reaction.

Finally, racemisation studies are currently being conducted, testing the stability of the intermediate and the rate of this process in order to check whether it is the cause of the low selectivity results obtained experimentally.

4.3.3 Future Work

In view of the current results and given the impossibility of improving the selectivity induced by these catalysts, the different problems aforementioned will be addressed in the future.

The most crucial problem, but easiest to address, is the obtention of the most stable conformer of each TS studied. A workflow that automates this search and explores different possibilities, along with different reaction paths, will be implemented, not only facilitating the process but eliminating most of the human input and the consequent bias.

It was also observed how, as the reaction proceeds, the negative charge, originally located in the phenolate, moves around the system at the different stages of the reaction (Scheme 4.7), modifying the strength of the different intermolecular NCIs between the reactants and the catalyst and even breaking the different HBs established, completely changing the binding mode. It is possible that this change in the binding mode has crucial effects on the global process and, therefore, the different possibilities must be studied. Additionally, further steps of the reaction will be studied, attending to these different BMs, aiming to analyse if any of them has a negative effect on the selectivity of the catalyst.

Finally, the ring-opening step of the reaction, after **Int1**, increases the flexibility of the molecule and, particularly, the chiral C atom of the azlactone; potentially allowing the partial racemisation of the structure and consequently decreasing the overall selectivity of the process. The racemisation barrier of this intermediate product will be computationally studied and experimental racemisation studies will be conducted, testing the stability of the intermediate and its racemisation rate.

In addition to the study of the aforementioned problems, a new set of catalysts was designed due to the observation of the new bifurcated HB between the phenyl ring of the phenylamide and the O atom in isopropyl ester within the azlactone. modifying the current catalyst and increasing the flexibility of the different phenylamide substituents, replacing the phenyl ring with a cyclohexane ring with an extra phenylamide in *ortho* position (Figure 4.18).

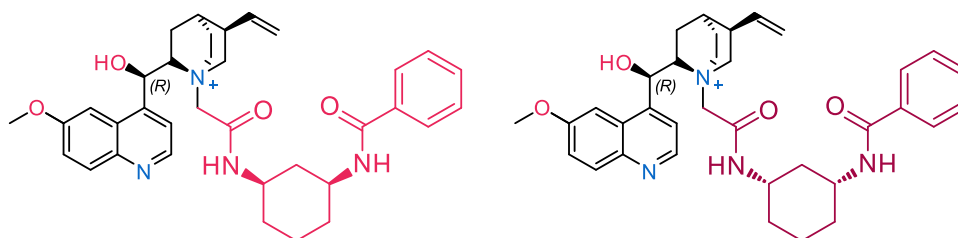


Figure 4.18. New modified catalysts for future studies: **cat_conhc6conhph-rr** and **cat_conhc6 conhph-ss**.

These catalysts will be computationally and experimentally studied, evaluating how a more flexible substituent will affect the stability of the different TSs and also the possibility of reinforcing that extra HB between the catalyst and the isopropyl ester of the azlactone by adding a third strong HB-donor to the system, like the new phenylamide.

However, the remarkable flexibility of these new catalysts makes imperative the development of the systematic and automatic workflow aforementioned and, consequently, these catalysts will be studied once the workflow is fully optimised.

4.4 Conclusions

A theoretical study of the generation of bis-protected amino acids in the presence of different cinchona-based catalysts was performed.

A conformational analysis was performed on all the reactants and catalysts under study to obtain the most stable conformers and their relative stabilities; followed by the study of the different complexes that can be formed between the catalyst and these reactants, observing that, as expected for a PTC system, the complex between the catalyst and the phenolate (**cat-phenolate**) is more stable than the complex between the catalyst and the azlactone (**cat-azlactone**).

Once the isolated catalysts and reactants were analysed and the most stable complex was determined, the first step of the reaction, the nucleophilic attack of p-nitrophenolate to azlactone **1**, was computationally studied in the presence of a first set of catalysts, and the results were compared with the experimental results obtained by our collaborators. Very low *ee* values were obtained both computationally and experimentally for **cat_ph** (3.8%*ee*_{comp} and 0%*ee*_{exp}) and **cat_anthra** (44.4%*ee*_{comp} and 0%*ee*_{exp}) and a considerable disagreement between the computational and the experimental selectivities was observed for **cat_cf3** (96.6%*ee*_{comp} and 16%*ee*_{exp}) and **cat_tbu** (87.0%*ee*_{comp} and 20.0%*ee*_{exp}).

The poor selectivity of these catalysts was found to be caused by the lack of interactions between the catalyst and azlactone, which did not establish any relevant NCI that creates a relevant difference between the TS of each enantiomer of the azlactone. However, a very weak but promising, bifurcated HB between the azlactone and the aromatic ring within the catalyst was found.

A second set of catalysts was designed based on this finding, aiming to enhance the interactions between the catalyst and the azlactone, substituting the aromatic rings with two different amides (**cat_conh2** and **cat_conhph**), introducing a second HB-donor and creating a second anchoring point in the structure.

Incorporating this additional binding site into the structure expands the range of potential binding modes, necessitating a comparative analysis of their characteristics, the strength of HBs they can form with each reactant and their respective relative stabilities to discern their importance in the reaction mechanism. The MEP of the catalysts, the different **cat-phenolate** complexes and the different possibilities for the TSs were studied, obtaining the definite binding mode with the phenolate establishing a HB in the **OH** position and the azlactone in the **NH** position.

The selectivity of these catalysts was computationally studied, highlighting the importance of this new anchor point in the holding of azlactone, creating structural differences between the two enantiomeric TSs. Two additional catalysts were designed based on these results to study the effect of increasing the acidity of the new anchor point by adding an EWG group (**cat_conhph-pno2**) and the effect of combining this attractive HB with sterical hindrance (**cat_conhph-obr**). The computational *ee* values obtained ranged from low (6.2%*ee*) to moderate (91%*ee*), without reaching a perfect selectivity but greatly improving the results obtained by the previous set of catalysts.

Our collaborators also studied this new set of four catalysts experimentally, obtaining similar results to those obtained computationally (25%*ee*-64%*ee*), highlighting the importance of

computational studies in the early stages of the catalyst design to perform efficient modifications based on the mechanistic and structural insight obtained.

References

- (1) C. M. Starks, *Journal of the American Chemical Society*, 1971, **93**, 195–199.
- (2) U. H. Dolling, P. Davis and E. J. J. Grabowski, *Journal of the American Chemical Society*, 1984, **106**, 446–447.
- (3) C. M. Starks and R. M. Owens, *Journal of the American Chemical Society*, 1973, **95**, 3613–3617.
- (4) S. Juliá, J. M. del Mazo, L. Avila and J. Elguero, *Organic Preparations and Procedures International*, 1984, **16**, 299–307.
- (5) E. V. Dehmlow, *Angewandte Chemie International Edition in English*, 1977, **16**, 493–505.
- (6) L. J. Mathias and R. A. Vaidya, *Journal of the American Chemical Society*, 1986, **108**, 1093–1094.
- (7) A. McKillop, J.-C. Fiaud and R. Hug, *Tetrahedron*, 1974, **30**, 1379–1382.
- (8) H. Freedman, *Pure and Applied Chemistry*, 1986, **58**, 857–868.
- (9) L. Pasteur, *Mémoire sur la fermentation de l'acide tartrique*, 1858.
- (10) H. Pellissier, *Tetrahedron*, 2008, **64**, 1563–1601.
- (11) I. Coldham, S. Dufour, T. F. Haxell, J. J. Patel and G. Sanchez-Jimenez, *Journal of the American Chemical Society*, 2006, **128**, 10943–10951.
- (12) K. Makino, M. Iwasaki and Y. Hamada, *Organic Letters*, 2006, **8**, 4573–4576.
- (13) S. Hoffmann, M. Nicoletti and B. List, *Journal of the American Chemical Society*, 2006, **128**, 13074–13075.
- (14) Y. Hayakawa, M. Hyodo, K. Kimura and M. Kataoka, *Chemical communications*, 2003, 1704–1705.
- (15) V. Jurkauskas and S. L. Buchwald, *Journal of the American Chemical Society*, 2002, **124**, 2892–2893.
- (16) M. Kitamura, T. Ohkuma, M. Tokunaga and R. Noyori, *Tetrahedron: Asymmetry*, 1990, **1**, 1–4.
- (17) S. Tallon, F. Manoni and S. J. Connon, *Angewandte Chemie*, 2015, **127**, 827–831.
- (18) Z. Rodriguez-Docampo, C. Quigley, S. Tallon and S. J. Connon, *The Journal of Organic Chemistry*, 2012, **77**, 2407–2414.
- (19) C. Trujillo, S. A. Cronin and S. J. Connon, *European Journal of Organic Chemistry*, 2022, **2022**, e202100818.
- (20) C. Palacio and S. J. Connon, *European Journal of Organic Chemistry*, 2013, **2013**, 5398–5413.
- (21) F. Albericio, *Peptide Science*, 2000, **55**, 123–139.
- (22) P. Pracht, F. Bohle and S. Grimme, *Physical Chemistry Chemical Physics*, 2020, **22**, 7169–7192.
- (23) S. Grimme, *Journal of chemical theory and computation*, 2019, **15**, 2847–2862.

- (24) P. Pracht and S. Grimme, *Chemical science*, 2021, **12**, 6551–6568.
- (25) C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2021, **11**, e1493.
- (26) S. Spicher and S. Grimme, *Angewandte Chemie International Edition*, 2020, **59**, 15665–15673.
- (27) J. C. Kromann, *Calculate Root-mean-square deviation (RMSD) of Two Molecules Using Rotation*, version 1.4, 2020.
- (28) H. W. Kuhn, *Naval research logistics quarterly*, 1955, **2**, 83–97.
- (29) W. Kabsch, *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 1976, **32**, 922–923.
- (30) Y. Zhao and D. G. Truhlar, *Theoretical chemistry accounts*, 2008, **120**, 215–241.
- (31) a. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley and J. Mantzaris, *The Journal of chemical physics*, 1988, **89**, 2193–2218.
- (32) G. Petersson and M. A. Al-Laham, *The Journal of chemical physics*, 1991, **94**, 6081–6090.
- (33) A. V. Marenich, C. J. Cramer and D. G. Truhlar, *The Journal of Physical Chemistry B*, 2009, **113**, 6378–6396.
- (34) M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji et al., *Wallingford, CT*, 2016.
- (35) L. J. Hauptert, J. C. Poutsma and P. G. Wenthold, *Accounts of chemical research*, 2009, **42**, 1480–1488.
- (36) S. T. Schneebeli, M. L. Hall, R. Breslow and R. Friesner, *Journal of the American Chemical Society*, 2009, **131**, 3965–3973.
- (37) A. E. Reed, L. A. Curtiss and F. Weinhold, *Chemical Reviews*, 1988, **88**, 899–926.
- (38) R. Bader et al., *Clarendon: Oxford, UK*, 1990.
- (39) P. L. A. Popelier, F. Aicken and S. O’Brien, *Atoms in molecules*, Prentice Hall Manchester, 2000, vol. 188.
- (40) T. A. Keith, *AIMAll (Version 19.10.12)*, <http://aim.tkgristmill.com>, Overland Park, KS, USA, 2019.
- (41) M. Walker, A. J. Harvey, A. Sen and C. E. Dessent, *The Journal of Physical Chemistry A*, 2013, **117**, 12590–12600.

5 | Computational Design: Transition State Studies, Enantioselectivity Calculations, and Workflow Automation.

5.1 Introduction

In the preceding chapter, the intricate nature of asymmetric catalysis emerged as a focal point in contemporary organic and computational chemistry.¹ As a result, the development of new, selective, and more efficient strategies for the design of chiral catalysts remains an important factor, to enable advancement and to drive this sector into the future.²

Although the field of asymmetric organocatalysis is continuously expanding and has grown significantly in recent times, the ability to elucidate the mechanistic details of these types of reactions has failed to keep pace. Traditionally, experimental screening has been one of the most common methods employed in catalyst design, prioritising brute force over rational design. Nevertheless, over the past two decades, it has become evident that computational methods represent a more economical approach toward the investigation of large chemical systems with satisfactory accuracy due to increasing calculation capacity and the development of new techniques. The advantages of the application of computational methods to challenging chemical transformations are clear; not only can the mechanism of different 'real-world' reactions be investigated, but predicting and guiding novel experimental discoveries has also become a possibility.³

Despite all these advantages, computational chemistry has been used as a **retrospective** tool⁴ during the last decade. It has supported experiments and explained the origin of the most selective catalyst, but it has often overlooked the predictive potential inherent in computational studies.^{5–12} However, this trend has gradually shifted in recent years, due to the development of computational power and the establishment of a better balance between cost and accuracy. This shift has placed computational studies at the forefront of the design process, harnessing the predictive capabilities of computational chemistry. This new paradigm has redefined the role of computational chemistry, making it a **prospective** tool. It now possesses the ability to accurately predict the activity of the catalyst, making the catalyst design process more cost-effective and environmentally friendly.

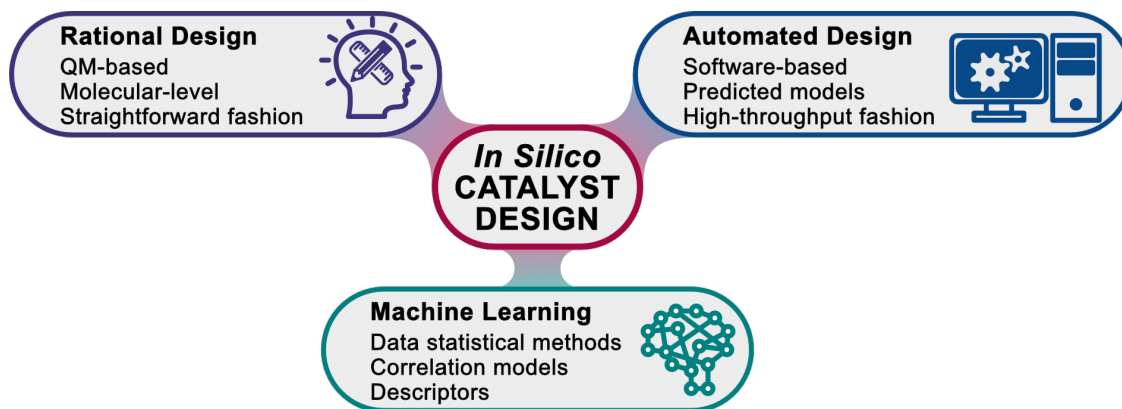


Figure 5.1. Three different approaches for computationally-led organocatalyst design.

For that reason, developing new techniques to improve the accuracy and produce near-instantaneous predictions is paramount in the present research. Furthermore, *de novo* design of effective and efficient catalysts for different types of reactions in silico remains a hopeful future objective, and, by extension, computational methods now not only predict but also design highly effective analogues of established catalysts.¹³

Several designing strategies, depicted in Figure 5.1, have been followed in the past aiming to

obtain an optimal catalyst with the ability to accelerate a desired organic reaction or the ability to control its selectivity: screening an existing catalyst library; producing modified derivatives of well-known catalysts, and testing the activity of the newly designed molecules; rationally designing a completely new catalyst. Computational chemistry emerges as a versatile means of achieving these objectives within reasonable time frames.

As computational chemistry assumes a leadership role in asymmetric organocatalyst design, three primary strategies have emerged: rational design, software-aided automated design, and ML-aided design (Figure 5.1).

Rational design of catalysts (Figure 5.2 a), involves an initial computational study of the reaction, obtaining detailed information on its mechanism in the presence of the selected catalyst. This is followed by the rational design of the catalyst, with the catalyst modified according to the information obtained from the previous study and with a clear objective in mind, improving the catalyst. Once a qualitative change has been achieved, the new catalyst is experimentally validated, starting an iterative process in which the catalysts are computationally refined and experimentally validated until good selectivities are achieved.

This rational design approach has been successfully used in numerous studies to optimise different catalytic processes in the past.^{14–21}

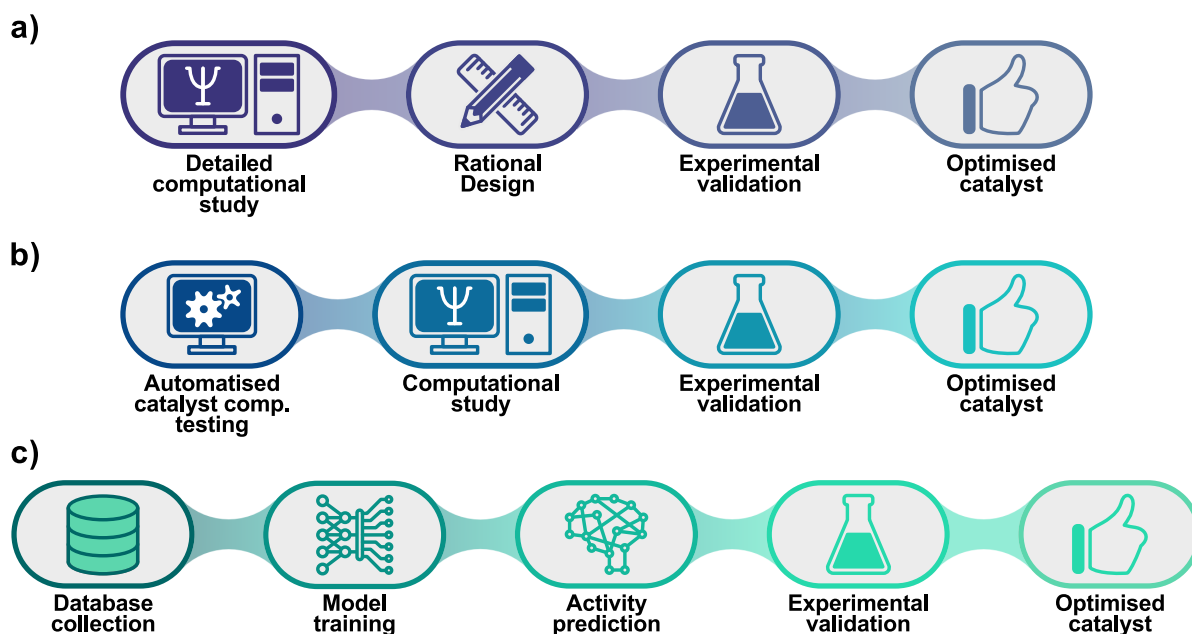


Figure 5.2. General workflow of the different design approaches. a) Rational design, b) automatic design and c) ML design.

In the framework of the automated in silico design (Figure 5.2 b), virtual screening methodologies are applied to an initial library of compounds, estimating their catalytic properties and obtaining the most promising scaffolds for the selected reaction. This approach involves an iterative cycle of computational optimisation and experimental validation to yield an optimal catalyst.

The primary distinction between these two approaches lies in the fact that, in the first case, the scaffold of the catalyst is known from the outset while, in the automatic design, it must be identified through initial screening. Additionally, the optimisation process in the automatic design is usually performed using the tools included in different software packages

like the Asymmetric Catalyst Evaluation (ACE) software,²² the Automated Reaction Optimiser for New Catalysts (AARON)²³ package, amongst many others.^{15,24–26} Several catalyst design processes have efficiently benefited from this approach, optimising different catalysts.^{13,27–31}

Finally, the last design approach is the ML-aided design and is extremely broad, since the design of catalysts can benefit from any of the different branches of ML methods and molecular representations. Given its breadth, this discussion will focus solely on the design of catalysts based on molecular descriptors, which is the most common approach. In this approach (Figure 5.2 c), an *in silico* library of molecules and potential catalysts is constructed, followed by the calculation of the corresponding, robust, chemical descriptors,^{32–39} along with their corresponding activities, either experimental or theoretical. All the collected data are separated into a training and a testing dataset and used to train and validate the selected model. Once the model is trained and optimised, the activities of different catalysts and their corresponding modifications can be predicted, followed by the optimisation and experimental validation of their results.

These workflows encounter a bottleneck that hinders efficiency and delays the entire process: the retention of the most stable TS. Traditionally, as demonstrated in the previous chapter, these TSs were manually constructed, using the chemical intuition and data from the NCIs established in the different complexes studied to place the reactants in the correct binding mode. However, this manual process can yield biased results, considering the flexibility of the systems and the myriad of NCIs that can be established, making it extremely challenging to find the most stable conformer for the TS leading to each enantiomer.

One potential solution to this problem is the manual addition of several conformations of those TSs, attempting to cover different relevant possibilities. However, due to the flexibility and complexity of these catalysts, creating different conformations that are physically accessible and different from each other can be intricate and laborious. Additionally, this is inefficient and may still result in a biased outcome, potentially failing to identify all the relevant conformations of the TS for each enantiomer.

Therefore, the objective of this chapter is to develop an efficient and automatic workflow to accurately describe the possible TSs of a given catalyst, taking advantage of different tools explored in previous chapters of this thesis to predict their selectivity in the most unbiased and efficient manner possible.

5.2 Computational Details

5.2.1 Workflow

The workflow designed for the automatic generation of different conformations of the TSs involved in the reaction and the calculation of the corresponding *ee* is presented in Figure 5.3. It is noteworthy to mention that, in parallel to the development of this study, a similar workflow was published by Grimme *et al.*, leveraging the various tools developed by their research group in order to efficiently calculate the free energies of ensembles of non-rigid molecules.⁴⁰

The workflow comprises four distinct stages: the construction of the initial 3D TS structure, the constrained conformational analysis of these TSs, a cyclic stage involving comparison, clusterisation and filtration and reoptimisation, and finally, the calculation of the *ee* of the process.

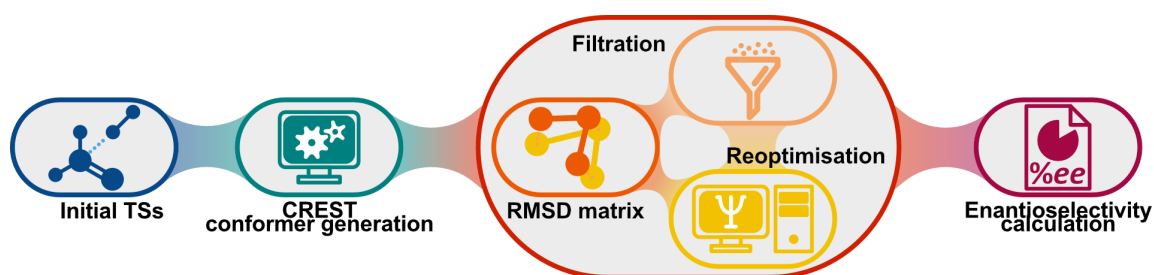


Figure 5.3. TS optimisation workflow scheme.

A set of non-interchangeable TSs, differing for each enantiomer of the azlactone, was constructed to encompass the nucleophilic attack (*p*-nitrophenolate) from both faces in each enantiomer (Figure 5.4 left). The binding mode was selected based on the results discussed in the previous chapter, locating the *p*-nitrophenolate anchored to the OH position in the catalyst and the azlactone establishing the HB in the NH position.

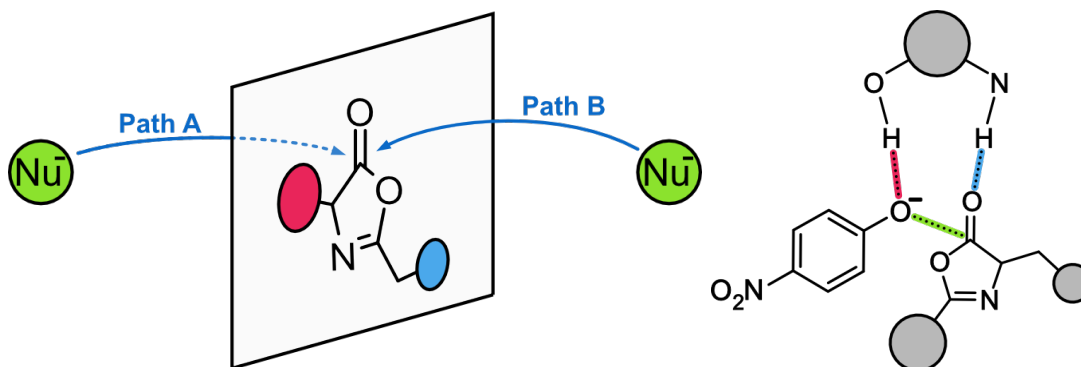


Figure 5.4. Nucleophilic attack to the azlactone from the two different faces, left, and interactions and distances fixed in the different orientations of the binding mode used as input, right.

A conformational analysis of these TSs was performed using CREST,^{41–43} at the semiempirical GFN2-xTB level, implemented in the xTB package.⁴⁴ The binding mode and the NCIs between the catalyst and the reactants was fixed by constraining the movement of the reaction coordinate and the HB established between them and the catalysts (Figure 5.4 right). The distance for the different HB and the reaction coordinate were fixed to the distance obtained from a previously optimised TS for each particular case.

The structures were optimised at the M062X/6-31g(d)^{45–47} computational level and solvent effects (toluene) were included in the optimization by means of a continuum method, the Solvation Model based on solute electron Density (SMD) approach.⁴⁸ The calculations were performed using Gaussian16.⁴⁹

The initial structures and all the generated conformers for each of them can be found in Table S5.1 and S5.2. The use of metadynamics to generate the TS ensemble ensures that all the generated conformers are physically accessible for that particular binding mode.

All the conformers obtained from the conformational analysis were extracted and used as the initial input for the filtration and reoptimisation iterative cycle. The first step of the cycle involves calculating the RMSD⁵⁰ between all the possible pairs of conformers originating from the same initial structure, creating the corresponding RMSD matrix. The RMSD threshold was automatically determined to obtain the desired $N_{remaining}$ most structurally different conformers in each case. The number of remaining conformers, $N_{remaining}$, was set to achieve a manageable number of conformers depending on the level of theory used in each cycle. Finally, the remaining conformers were reoptimised at a higher level of theory in each iteration, gradually increasing the accuracy of the calculations.

This stage of the workflow involves three of these filtration and reoptimisation cycles, each characterised by specific thresholds and computational details:

Cycle 1:

- $N_{remaining} = 15$ for each individual TS ($N_{remaining} = 60$ total).
- b3lyp/3-21g,^{46,47,51} gas phase
- Fixed reaction coordinate

Cycle 2:

- $N_{remaining} = 7$ for each individual TS ($N_{remaining} = 24$ total).
- m06-2x/6-31g(d),⁴⁵ Toluene, Polarizable continuum model (PCM)⁵² solvent model
- Fixed reaction coordinate

Cycle 3:

- $N_{remaining} = 4$ for each individual TS ($N_{remaining} = 16$ total).
- m06-2x/6-31g(d), Toluene, PCM solvent model
- Unfix reaction coordinate

Harmonic frequencies were calculated at all levels of theory to ensure that the obtained structure corresponds to the desired saddle point. All the optimisations were calculated using the Gaussian09 software.⁵³ The reoptimised structures at each level can be found in Table S5.3.

The obtained free energies (G) at the highest level of theory were used to calculate the selectivity of the catalyst, *ee*.

5.2.2 Enantioselectivity Calculation

Calculating the theoretical enantioselectivity of a reaction has been a widely discussed topic in the computational chemistry community. Various methods have been developed over the years, depending on the number of TSs and reaction paths taken into account,^{54–56} all of which operate under Curtin-Hammet⁵⁷ conditions.

The first option, used in the previous chapter and described by Schneebeli *et al.*,⁵⁴ calculates the selectivity of the process based on the difference in energy between the most stable TS leading to each enantiomer, expressed as $\Delta\Delta G^\ddagger = \Delta G_R^\ddagger - G_S^\ddagger$, with the *ee* calculated as the ratio between the rate constants (Equation 1).

$$\frac{K_R}{K_S} = e^{-\Delta\Delta G^\ddagger/RT} \quad (1)$$

When the reaction rate is significantly slower than the TS conformational interconversion rate, the obtained conformers represent an ensemble of physically accessible and interchangeable conformers. In such situations, the selectivity of the process must be calculated using the average free energy, $G_{average}$, of each TS, as described by Grimme *et al.*⁵⁵ (Eq. 2)

$$G_{average}^\ddagger = \sum_j^{N_{TS}} p_j G_j^\ddagger \quad (2)$$

where p_j (3) is the population of the different TS leading to each enantiomer

$$p_j = \frac{e^{-\Delta G_j^\ddagger/RT}}{\sum_j^{N_{TS}} e^{-\Delta G_j^\ddagger/RT}} \quad (3)$$

subsequently, the ratio between the constants can be calculated using the average free energy of each enantiomer (Eq. 4), similarly to the previous method.

$$\frac{K_R}{K_S} = e^{-\Delta\Delta G_{average}^\ddagger/RT} \quad (4)$$

However, when the different conformers of the TS leading to each enantiomer are not interconvertible, multiple reaction paths must be taken into account. In such a scenario, the different paths must be treated separately, as described by Williams in his work.⁵⁶

In this context, the individual rate constants, K_R and K_S , must be calculated individually as the sum of the individual rate constants leading to each enantiomer (Eq. 5).

$$\frac{K_R}{K_S} = \frac{\sum_j^{N_{TS}^R} e^{-\Delta\Delta G_{j,R}^\ddagger/RT}}{\sum_j^{N_{TS}^S} e^{-\Delta\Delta G_{j,S}^\ddagger/RT}} \quad (5)$$

The accurate description of this reaction requires considering the nucleophilic attack from both faces of the azlactone, resulting in two different reaction paths which TSs are non-interconvertible, similarly to the situation described earlier (Eq. 5). However, the flexibility of the reactants also makes it necessary to generate a conformational ensemble of each of the TS in order to obtain an accurate description of the process.

A schematic representation of the PES for the different reaction paths and TSs can be found in Figure 5.5. Two distinct paths are considered for each enantiomer of the azlactone,

corresponding to the nucleophilic attack in the different faces. A conformational ensemble for each of the TS is calculated using the workflow aforementioned.

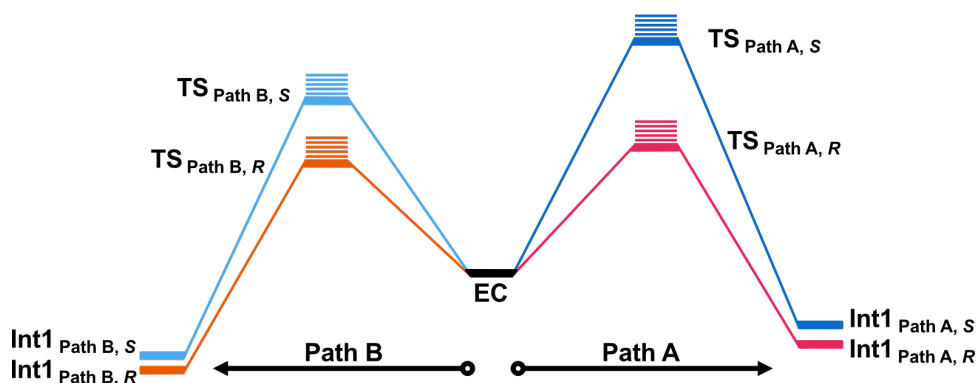


Figure 5.5. Schematic Potential Energy Surface representing the assumptions done for the study of these reaction. Two different reactions path for each enantiomer of the azlactone and a conformational ensemble for each TS under study.

Therefore, the enantioselectivity of the process is calculated as the sum of the individual reaction rates for each path (Eq. 5) using the average free energy, $\Delta G_{average}$ (Eq. 2, of the different TSs. This yields equation 6.

$$\frac{K_R}{K_S} = \frac{\sum_j^{N_{TS}^R} e^{-\Delta\Delta G_{average,j,R}^\ddagger/RT}}{\sum_j^{N_{TS}^R} e^{-\Delta\Delta G_{average,j,S}^\ddagger/RT}} \quad (6)$$

5.3 Results and Discussion

5.3.1 Conformer Generation

Developing an efficient workflow to automatically obtain a high-level conformational ensemble that accurately describes the possible TSs involved in a reaction and finds all the representative conformations of those TSs requires, as expected, fine-tuning and testing to make the most of it.

This testing must be conducted using systems that have been analysed in detail, using their results as a reference for evaluating the different steps of the workflow. The selected system should be as complete and complex as possible, pushing the limits of the workflow.

In the previous chapter, several catalysts with different functional groups that established a myriad of NCIs were detailed studied. Among all the catalysts investigated, **cat_conhph** not only presents some of the most interesting NCIs but also has been thoroughly studied from both the structural and energetic perspectives. For these reasons, **cat_conhph** is the perfect candidate to tune and test the workflow developed in this project.

As it was shown in Figure 5.4, the nucleophilic attack of the enolate can occur from both faces of the azlactone. The two different faces are named O_{back} and O_{front} based on the relative position of the O atom within the azlactone cycle with respect to the catalyst. Considering the chiral nature of azlactone, which is racemic before the reaction, four situations must be considered: Azlactone R enantiomer attacked from face O_{back} (**R** O_{back}), Azlactone R enantiomer attacked from face O_{front} (**R** O_{front}), azlactone S enantiomer attacked from face O_{back} (**S** O_{back}) and azlactone S enantiomer attacked from face O_{front} (**S** O_{front}) (Figure 5.6).

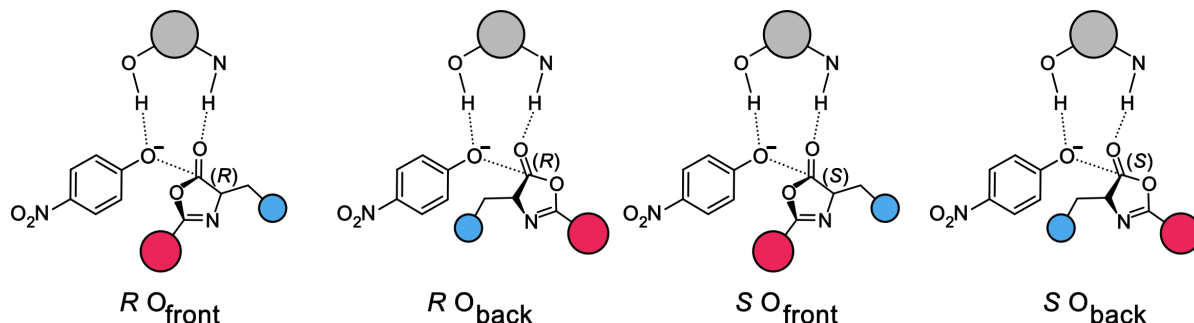


Figure 5.6. Binding modes considered for the study. From left to right $R O_{back}$, $R O_{front}$, $S O_{back}$, $S O_{front}$.

The conformational analysis of the optimised TS structures was performed, generating an ensemble for each of them (see computational details). The aligned 3D structures of each of the TS ensembles as well as the number of generated conformers can be found in Figure 5.7.

When examining the results, it is surprising to observe the significant disparity in the number of generated conformers among the different TSs. The largest $TS_{ensemble}$ is $S O_{front}$, formed by 2105 conformers; followed by the $R O_{front}$ ensemble, formed by 194 conformers. Finally, the two smallest ensembles are $R O_{back}$ and $S O_{back}$, formed by 86 and 85 conformers, respectively.

Given that the only difference among the four TSs under study is the chirality and orientation of the azlactone, one would expect to obtain similar-sized conformational ensembles in all the cases. However, upon closer inspection, it becomes evident that the smallest ensembles correspond to the TS with the azlactone in **back** position; leaving the biggest group,

o-tetrachloromethoxycarbonyl, located between the quinuclidine ring and the benzamide substituent, significantly limiting its flexibility. In contrast, the TSs generated with the azlactone in **front** position produce many more conformers due to reduced steric hindrance between the aforementioned group of the azlactone and the catalyst. This absence of repulsive interactions is particularly noticeable in the $S O_{back}$ TS_{ensemble}, where the azlactone can adopt almost any possible conformation.

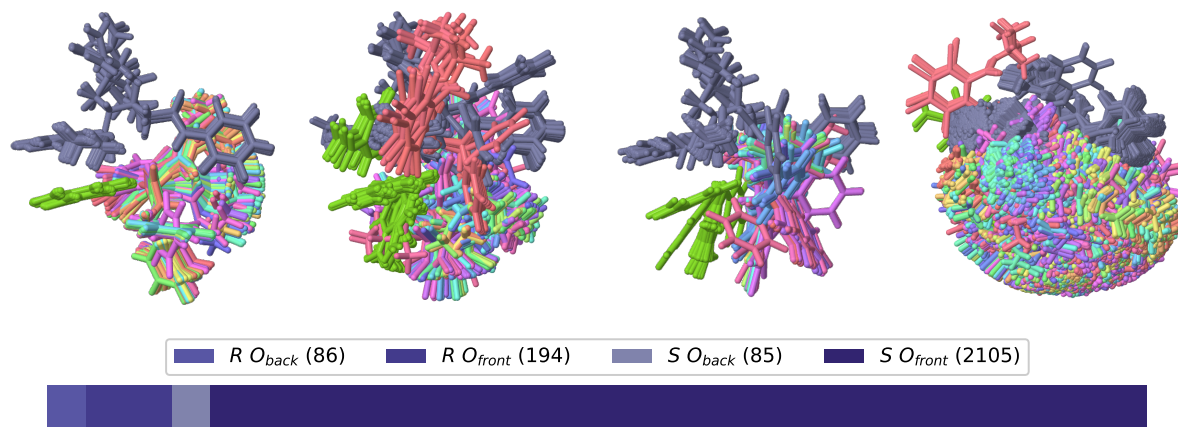


Figure 5.7. Conformational ensembles generated by CREST. From left to right $R O_{back}$, $R O_{front}$, $S O_{back}$, $S O_{front}$.

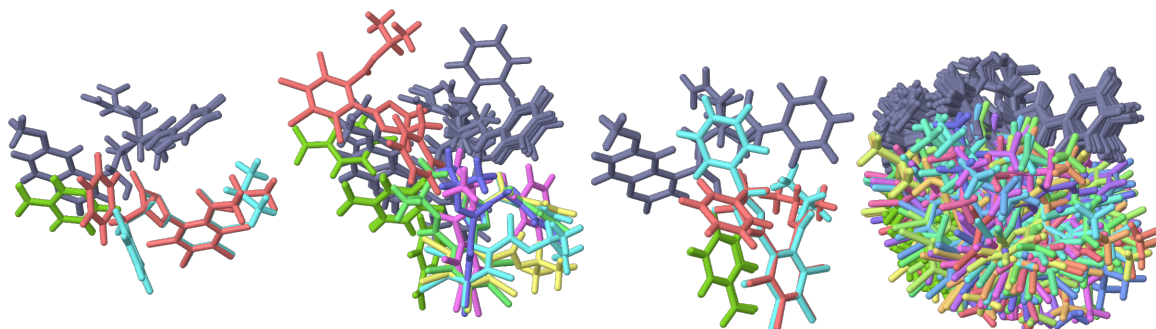
The flexibility of the generated ensembles is also reflected in the number of positions where the enolate can be found and the different conformations that the catalyst adopts within them. When visualising the same aligned 3D structures with the azlactone hidden (Figure S5.1), it can be seen that the $R O_{back}$ ensemble presents an almost perfect catalyst and enolate overlap in all the conformers, keeping the enolate establishing a π - π interaction with the catalyst throughout; suggesting a very rigid ensemble. As the flexibility of the system increases, the catalyst can adopt different conformations and the enolate can be found in a few different positions. For example, in the $S O_{back}$ ensemble, the enolate establishes a π - π interaction not only with the catalysts but also with the phenyl ring of the azlactone when it is pointing downward. In both ensembles with the azlactone in **front** position, both the catalyst and the enolate can be found in numerous conformations and locations. Particularly in the $S O_{front}$ ensemble, both the catalyst and the enolate are located in dozens of intermediate structures, contrasting with the previous ensembles where they were grouped in a reduced number and well-aligned positions.

5.3.2 Conformer Filtration and Selection

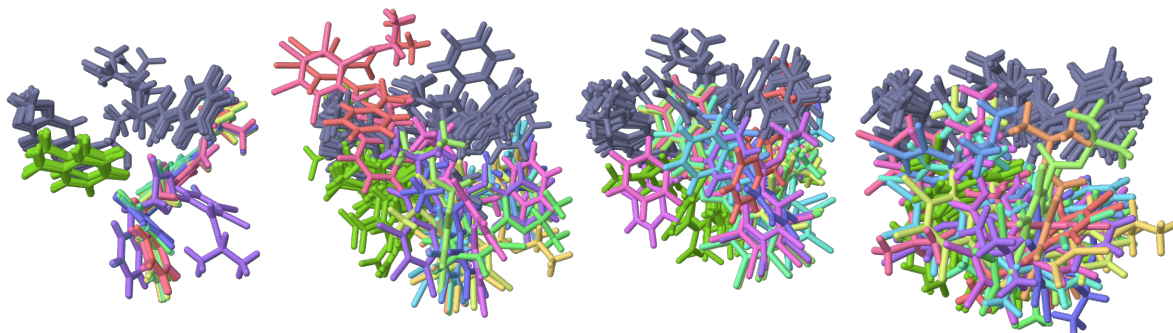
After generating the conformational ensembles, they need to be filtered to obtain a manageable sample that will be reoptimised at a higher theory level, since reoptimising 2470 TS conformers at any DFT level would not be feasible. As explained in the computational details section, the RMSD between each pair of conformers was calculated, and then a threshold was applied in order to retain a certain number of conformers. However, two different approaches can be taken for this filtration: the four different TS ensembles can be grouped in a bigger ensemble and then calculate the RMSD of all the possible pairs, obtaining a larger RMSD matrix to which apply the threshold; or the four different ensembles can be treated individually, resulting in four RMSD matrices.

The first approach selects the N most different structures among all the generated conformers, prioritising geometry diversity over equal representation of the different TSs. On the other hand, the second approach generates an even distribution of the four different TSs, assuring that each of them will be adequately described. Both approaches were tested in order to evaluate which one more accurately describes the reaction.

As described in the computational section, the number of remaining conformers in both cases was set to $N_{remaining} = 60$ for this first selection. The two filtering processes produce completely different outcomes, and the remaining conformers of each procedure, grouped by orientation can be found in Figure 5.8.



(a) Remaining TS conformers after RMSD filtration of all the generated conformers together. Grouped by binding mode, left to right: $R O_{back}$ (2), $R O_{front}$ (6), $S O_{back}$ (2) and $S O_{front}$ (50).



(b) Remaining TS conformers after RMSD filtration of the individual conformer ensembles generated. Grouped by binding mode, left to right: $R O_{back}$ (15), $R O_{front}$ (15), $S O_{back}$ (15) and $S O_{front}$ (15).

Figure 5.8. Remaining conformers after RMSD filtration using different procedures. a) Filtered together, b) filtered separately.

When all the generated conformers are compared together (5.8a), the number of remaining conformers is completely uneven, obtaining only 2 conformers for $R O_{back}$ and $S O_{back}$, 10 for $R O_{front}$ and 50 for $S O_{front}$. This filtering method maintains a similar ratio between the number of conformers in each binding mode and the total of conformers before and after the filtration. Before filtering, the conformers $R O_{back}$ and $S O_{back}$ represent 3.5% of the total conformers each, $R O_{front}$ represents a 7.9% and $S O_{front}$ represents the remaining 85% of the conformers. After the filtration process, the first two binding modes represent a 3.3% of the total conformers each, the ratio of $R O_{front}$ slightly increases to 10% and the largest ensemble now represents a 83.0%.

This uneven distribution could pose several problems. When the four are combined and placed one after the other, the conformers are no longer sorted by energy in ascending order. Consequently, it cannot be ensured that the most stable conformer of each binding mode re-

mains after the filtration. Even if the conformers forming the large ensemble were sorted by energy to ensure the inclusion of the most stable conformer generated by CREST, regardless of the ensemble size, two conformers would not be enough for describing the TS accurately. Additionally, as determined in the first chapter, CREST is not particularly adept at identifying the most stable conformer of a system, as it tends to overestimate certain NCIs, such as π - π interactions, which are crucial in these systems. Therefore, larger and more diverse conformer ensembles are preferred for an accurate analysis of these systems.

When the four $\text{TS}_{\text{ensemble}}$ are filtered separately instead of all together, the same number of remaining conformers is obtained for each of the binding modes (5.8b). This even distribution of conformers allows a more detailed study of the different binding modes as fewer potentially important conformers of each TS are lost along the way. The difference in the ensembles of remaining conformers is particularly important in the RO_{back} and SO_{back} since they are formed by 15 conformers, compared to only 2 in the previous example. Despite being much larger, the observed ensemble in RO_{back} does not change excessively due to the rigidity of the system and the similarity of the generated conformers. However, the observed change in the SO_{front} ensemble is significant because the generated conformers are much more diverse, and only two of them remained in the previous situation. The situation of the RO_{front} ensemble remains similar to the previous one, with the addition of some extra conformers that will help describe the TS more accurately. Finally, the most substantial change in terms of the absolute number of conformers is the SO_{front} ensemble, which was reduced from 50 conformers to 15 in this second filtration. However, as mentioned earlier, many of these conformers represent intermediate situations between stable conformers. Therefore, when reoptimised at a higher level, they will collapse in the same conformation.

Additionally, since the individual ensembles are sorted by energy in ascending order, this method ensures that the most stable one at this optimisation level is included. This is not particularly important during the first filtration process due to the limited prediction capacity of CREST, but it becomes crucial in the subsequent steps to avoid losing the most stable TS conformer in each case.

In conclusion, the analysis conducted underscores the advantages of separate filtering of the individual TS ensembles rather than adopting a grouped filtration approach.

5.3.3 Reoptimisation Cycles

After determining that the best way to filter the different $\text{TS}_{\text{ensembles}}$ is by treating each one of them separately, these ensembles can be introduced in the iterative cycle of reoptimisation and filtration (Figure 5.3) to reduce their size and increase the level of theory at which they are optimised. This is the step of the workflow in which more parameters can be tweaked to optimise the process: the number of cycles, the number of conformers in each cycle, the level of theory of the different cycles and where the calculations are relaxed, fixed of single points.

In an ideal situation with unlimited resources and instantaneous calculations, all the conformers would be studied at the highest level of theory, describing the ensembles as accurately as possible and minimising the error at calculating the ee . However, since computational power is not limitless and the cost of quantum calculations grows exponentially with the level of theory and size of the basis set, these parameters must be optimised to obtain a good description of the $\text{TS}_{\text{ensemble}}$ at the highest level of theory using a reasonable amount of computational resources

- the limiting factor.

Considering that each round of this cycle implies the reoptimisation of different conformers, a large number of cycles would considerably increase the computational time required. However, if the number of cycles is too low, some of the most stable conformers of the TS- key for an accurate calculation of the *ee*- could be lost due to the incorrect description of their relative stabilities.

Based on our previous experience working with conformational ensembles, as demonstrated in Chapter 1, and optimising similar TSs, as illustrated in Chapter 4, it was decided to establish three reoptimisation cycles, starting with $N_{conformers} = 60$ at the lowest DFT level and finishing with $N_{conformers} = 16$ at the highest one. The level of theory was increased from b3lyp/3-21g in the gas phase for the first cycle to m06-2x/6-31(d) including a solvent model (PCM, toluene), which is a standard level of theory for this kind of system and similar to the one used in the previous chapter. Additionally, the reaction distance was fixed for the first two rounds of the cycle, greatly decreasing the computational time of the optimisation.

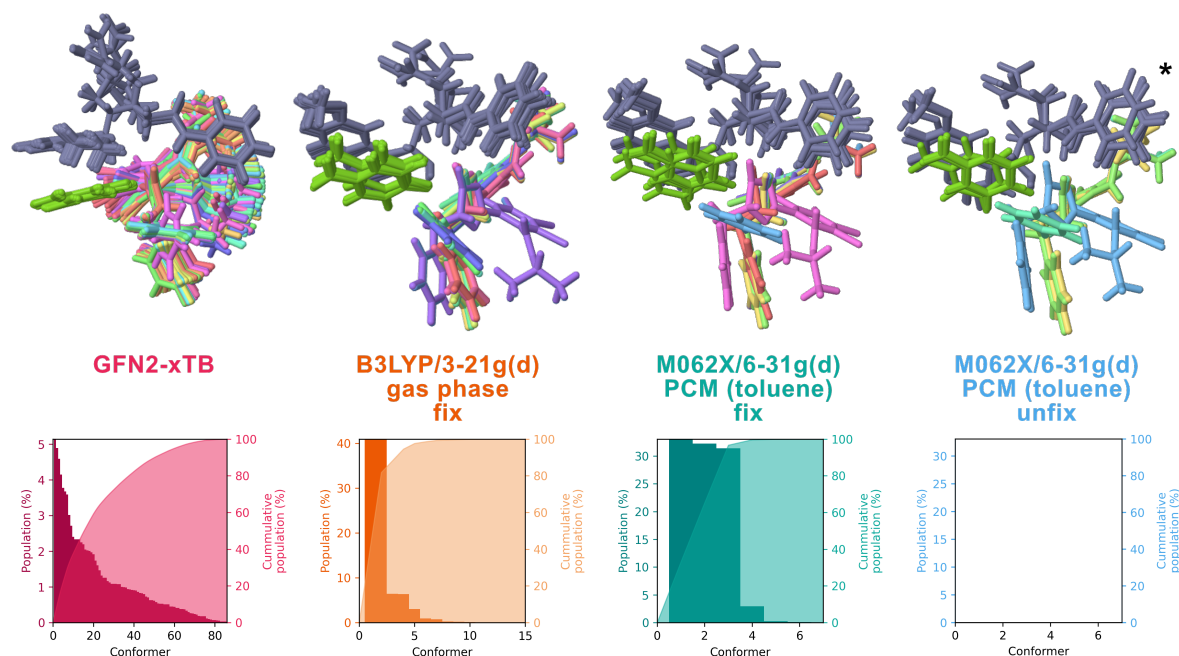


Figure 5.9. Top: aligned 3D conformers of **cat_conhph** generated from $R O_{back}$ starting point at the different theory levels of each reoptimisation cycle. Bottom: Population and cumulative population of the different conformers. Structures marked with * are not optimised, only filtered.

These selected parameters were applied to all four $TS_{ensembles}$ generated for **cat_conhph**, resulting in a smaller and higher-quality ensemble with each iteration. The aligned 3D structures can be seen in Figure 5.9-top and S5.2. In addition to the 3D structures, the individual population of each conformer and their cumulative population were calculated to assess the relative importance of each conformer. The results are plotted and can be found in Figure 5.9-bottom and S5.3.

In terms of conformational possibilities covered at each level, from the 3D structures (Figure 5.9 top and S5.2), it is noticeable that, since the filtration method used maximises the RMSD between the conformer pairs, the most different structures remain after each cycle. This ensures the best and most diverse representation of the conformational space at each step. However, it does not ensure that all studied conformers are relevant to the accurate description of $TS_{ensemble}$.

or even their energetic accessibility.

For instance, when examining the populations of each conformer and their growth at each round of the cycle (Figure 5.9 bottom and S5.3), one can observe significant changes in the situation. At the CREST level, the population grows smoothly with the number of conformers, providing a good and even representation of the conformational space. However, this situation changes dramatically with the first round of the reoptimisation cycle, where most of the conformers become irrelevant (*Population* < 1%). The population growth is particularly pronounced for the *R O_{front}* and *S O_{back}* binding modes, where the first conformer already represents the 99.7% and 99.9%, respectively.

Interestingly, the inclusion of the PCM solvent model in the second round of the cycle, resulted in a change of this situation for some of the binding modes, revealing a smoother growth at the lower-energy conformers (*R O_{back}* and *S O_{back}*).

Unfortunately, the unfixed optimisation of the different TSs is not finished yet and only the selected, not reoptimised, structures can be shown.

These changes in the relative stability of different conformers might also modify which one is the most stable conformer at each stage. This would not be surprising for the first round since, as observed in the first chapter, CREST does not excel at predicting the most stable conformers in these types of systems.

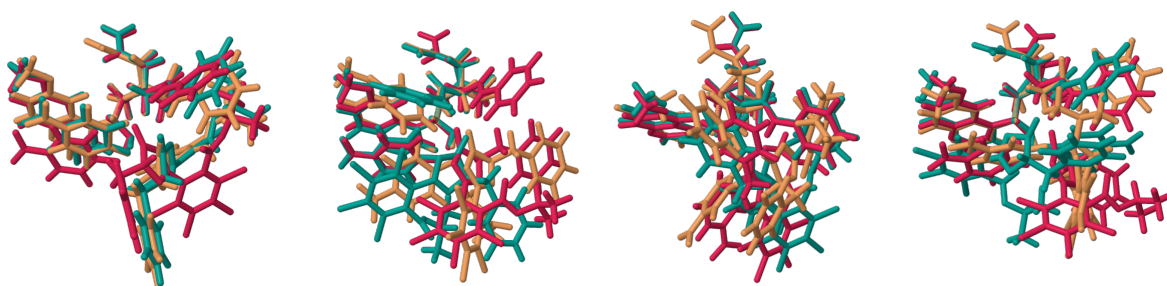


Figure 5.10. Most stable conformer at each round of the reoptimisation cycle. Left to right *R O_{back}*, *R O_{front}*, *S O_{back}* and *S O_{front}*. Each round is represented in a different colour: CREST in **red**, fixed b3lyp/3-21g in the gas phase in **orange** and fixed m06-2x/6-31g(d) in toluene in **green**.

When examining the aligned 3D structures of the most stable conformers at each level (Figure 5.10 and S5.4), it can be seen that, as expected, the most stable conformer at the CREST level does not correspond to the most stable conformer at any of the DFT levels in most cases. Surprisingly, significant changes can be observed between the most stable conformers at different DFT levels of theory. In the binding mode *R O_{front}* it is evident how the conformation of the catalyst undergoes a complete transformation when transitioning from CREST to fixed DFT structures, leading to a significant alteration in the orientation of the phenylamide substituent.

Another observed effect is the changes induced by the solvent model. In the *S O_{back}* and *S O_{front}* binding modes, it can be observed that the most stable conformer in the gas phase (**orange**) does not align with the most stable one when the solvent model is considered (**green**).

These significant changes in the global stability of the system and the relative population of each conformer are expected to have a dramatic impact on the calculated *ee* at each of the theoretical levels. When the selectivity of the catalyst is calculated at each of these levels (Figure 5.11) using equation 6, it can be observed how the results completely differ from those

found in the previous chapter, where only one TS was considered for each enantiomer.



Figure 5.11. Experimental and theoretical *major:minor* ratio and *ee* values at the different levels of theory.

The differences between the experimental selectivity and the calculated one decrease as the level of theory increases, obtaining a perfect match between them already at the $DFT_{fix, high}$ level of theory.

It can also be observed how the agreement between the experimental and the computational *ee* is significantly better than the one obtained in the previous chapter, where only one TS was considered for each of the enantiomers. Furthermore, upon closer examination, it becomes apparent that the major enantiomer obtained in the previous chapter (DFT_{chap4}) corresponds to the *S* enantiomer, whereas the one obtained at the highest level of theory using the new workflow is the *R* enantiomer. These contrasting results highlight the importance of exploring the conformational space of the TS and considering all the different possibilities.

It also highlights how computational results may exhibit bias when constructing TS based solely on chemical intuition, overlooking alternative possibilities and yielding inaccurate results that may coincidentally align, to varying degrees, with experimental findings by pure chance.

5.3.4 Future Work

The obtained results from the initial parameterization of the automated workflow for studying flexible organocatalyst transition states (TSs) suggest several promising directions for future exploration and refinement.

This investigation has laid the foundation for a more in-depth exploration and refinement of all the parameters involved in each level, as well as various algorithms for the filtration and selection of the relevant conformers from the $TS_{ensemble}$ at each level of theory.

One notable observation was that filtering and selecting the remaining conformers by means of RMSD alone leads to smaller and very diverse ensembles, in which only one or two of the conformers are mathematically relevant for the final average energy $G_{average}$ of the ensemble. To address this issue two different approaches will be considered:

1. Incorporating an energy threshold (δG_{max}) into the filtration algorithm before calculating the RMSD matrix. This step would exclude conformers from the ensemble whose population is smaller than 1%.
2. Implementing a dimensionality reduction and clustering algorithm that considers both the RMSD and δG between each pair of conformers, similar to the one presented in Chapter 1. This approach involves selecting conformers from the different clusters to avoid redundant conformers, especially in those TSs with a more rigid structure.

However, as was observed in this chapter and Chapter 1, the relative energies extracted from the CREST conformational analysis are not trustworthy, as some of the present NCIs in these systems are overestimated. This leads to a failure in predicting the most stable conformers and their relative importance to the ensemble. This lack of confidence in the preliminary results is particularly important if an energy threshold, δG_{max} is implemented to the first conformed filtration. This implementation could result in the loss of some conformers that become very relevant to the ensemble at higher levels of theory. To address this issue, an intermediate stage will be added to the workflow, performing a single point DFT calculation at the lowest DFT level, prior to the filtration. This addition aims to increase the accuracy of the relative energies of the conformers, preventing issues during their consideration in subsequent filtration processes.

Finally, it was observed that, with the current number of reoptimisation cycles and the number of remaining conformers in each cycle, the change in the total number of conformers between the first and the second stage was very abrupt. It descended from 2470 to 60 conformers in one step, thereby losing many conformers that might be relevant at higher levels of theory. A different number of cycles will be tested, with a different number of conformers at each one in order to obtain a smoother narrowing process of the different $TS_{ensemble}$ under study. This adjustment will be made according to the ensemble size in each case and the corresponding level of theory.

5.4 Conclusions

A new workflow for the automatic and accurate calculation of computational enantioselectivities in organocatalytic reactions was developed. The workflow thoroughly analyses the different TSs responsible for the selectivity by performing a conformational analysis of the initial structure used as input, representing the different reaction paths and binding modes involved.

Despite not being fully optimised yet, this workflow has demonstrated promising potential in calculating the *ee* of a reaction, matching with experimental results even without reaching the highest level of theory employed. It has also underscored the potential bias in results when relying solely on chemical intuition in flexible systems where numerous NCIs contribute to stability. As exemplified in the optimisation case study, **cat_conhph**, discussed in the previous chapter, this workflow not only achieved higher agreement between *ee_{experimental}* and *ee_{computational}*, but also revealed that the most favoured enantiomer in this reaction is *R* enantiomer, contrary to previous predictions

However, the accurate description of the individual TSs is not the sole advantage of this workflow. The new method for calculating the *ee* (Eq. 6), combining the average free energy of a TS_{ensemble}, $G_{average}$,⁵⁵ and accounting for different reaction paths, accessible from the same EC,⁵⁶ allows for an accurate and physically meaningful calculation of the enantioselectivity of the reaction.

In summary, despite the lack of fine optimisation of the different parameters, the developed workflow has proven to be efficient at calculating the *ee* of a given reaction and describing the TSs involved. This provides an ideal foundation to further study the different NCIs involved in the process as explored in the previous chapter. The ability to explain the mechanism behind catalyst selectivity contributes to rationalising the design process for future catalysts.

References

- (1) H. Gröger and A. Berkessel, *Asymmetric Organocatalysis: From Biomimetic Concepts to Applications in Asymmetric Synthesis*, 2005.
- (2) A. Calcaterra and I. D'Acquarica, *Journal of pharmaceutical and biomedical analysis*, 2018, **147**, 323–340.
- (3) K. N. Houk and P. H.-Y. Cheong, *Nature*, 2008, **455**, 309–313.
- (4) Z. Tu, T. Stuyver and C. W. Coley, *Chemical Science*, 2023.
- (5) S. Grimme, C. Bannwarth and P. Shushkov, *Journal of chemical theory and computation*, 2017, **13**, 1989–2009.
- (6) C. Trujillo, I. Rozas, A. Botte and S. J. Connon, *Chemical Communications*, 2017, **53**, 8874–8877.
- (7) A. G. Collar, C. Trujillo, B. Lockett-Walters, B. Twamley and S. J. Connon, *Chemistry–A European Journal*, 2019, **25**, 7275–7279.
- (8) B. Kótai, G. Kardos, A. Hamza, V. Farkas, I. Pápai and T. Soós, *Chemistry–A European Journal*, 2014, **20**, 5631–5639.
- (9) S. E. Wheeler, T. J. Seguin, Y. Guan and A. C. Doney, *Accounts of Chemical Research*, 2016, **49**, 1061–1069.
- (10) Y.-h. Lam, M. N. Grayson, M. C. Holland, A. Simon and K. Houk, *Accounts of Chemical Research*, 2016, **49**, 750–762.
- (11) M. N. Grayson and K. Houk, *Journal of the American Chemical Society*, 2016, **138**, 9041–9044.
- (12) P. H.-Y. Cheong, C. Y. Legault, J. M. Um, N. Çelebi-Ölçüm and K. Houk, *Chemical reviews*, 2011, **111**, 5042–5137.
- (13) M. Foscato and V. R. Jensen, *ACS catalysis*, 2020, **10**, 2354–2377.
- (14) G. G. Gerosa, M. O. Marcarino, R. A. Spanevello, A. G. Suárez and A. M. Sarotti, *The Journal of Organic Chemistry*, 2020, **85**, 9969–9978.
- (15) M. Dittner and B. Hartke, *Journal of chemical theory and computation*, 2018, **14**, 3547–3564.
- (16) L. Dell'Amico, V. M. Fernández-Alvarez, F. Maseras and P. Melchiorre, *Angewandte Chemie International Edition*, 2017, **56**, 3304–3308.
- (17) R. Mitra, H. Zhu, S. Grimme and J. Niemeyer, *Angewandte Chemie International Edition*, 2017, **56**, 11456–11459.
- (18) H. Yang and M. W. Wong, *Molecules*, 2020, **25**, 1045.
- (19) P. Yingcharoen, C. Kongtes, S. Arayachukiat, K. Suvarnapunya, S. V. Vummaleti, S. Wannakao, L. Cavallo, A. Poater and V. D'Elia, *Advanced Synthesis & Catalysis*, 2019, **361**, 366–373.
- (20) G.-J. Cheng, X. Zhang, L. W. Chung, L. Xu and Y.-D. Wu, *Journal of the American Chemical Society*, 2015, **137**, 1706–1725.

- (21) S. M. Banik, A. Levina, A. M. Hyde and E. N. Jacobsen, *Science*, 2017, **358**, 761–764.
- (22) C. Corbeil, S. Thielges, J. Schwartzentruber and N. Moitessier, *Angewandte Chemie-German Edition*, 2008, **120**, 2675–2678.
- (23) Y. Guan, V. M. Ingman, B. J. Rooks and S. E. Wheeler, *Journal of chemical theory and computation*, 2018, **14**, 5249–5261.
- (24) M. Foscatto, V. Venkatraman and V. R. Jensen, *Journal of chemical information and modeling*, 2019, **59**, 4077–4082.
- (25) A. R. Rosales, J. Wahlers, E. Limé, R. E. Meadows, K. W. Leslie, R. Savin, F. Bell, E. Hansen, P. Helquist, R. H. Munday et al., *Nature Catalysis*, 2019, **2**, 41–45.
- (26) A. R. Rosales, T. R. Quinn, J. Wahlers, A. Tomberg, X. Zhang, P. Helquist, O. Wiest and P.-O. Norrby, *Chemical Communications*, 2018, **54**, 8294–8311.
- (27) R. Laplaza, S. Gallarati and C. Corminboeuf, *Chemistry-Methods*, 2022, **2**, e202100107.
- (28) G. G. Gerosa, R. A. Spanevello, A. G. Suárez and A. M. Sarotti, *The Journal of Organic Chemistry*, 2015, **80**, 7626–7634.
- (29) B. J. Rooks, M. R. Haas, D. Sepúlveda, T. Lu and S. E. Wheeler, *ACS Catalysis*, 2015, **5**, 272–280.
- (30) A. C. Doney, B. J. Rooks, T. Lu and S. E. Wheeler, *ACS Catalysis*, 2016, **6**, 7948–7955.
- (31) E. M. Fleming, C. Quigley, I. Rozas and S. J. Connon, *The Journal of Organic Chemistry*, 2008, **73**, 948–956.
- (32) C. A. Tolman, *Chemical reviews*, 1977, **77**, 313–348.
- (33) O. Diebolt, G. C. Fortman, H. Clavier, A. M. Slawin, E. C. Escudero-Adan, J. Benet-Buchholz and S. P. Nolan, *Organometallics*, 2011, **30**, 1668–1676.
- (34) L. Perrin, E. Clot, O. Eisenstein, J. Loch and R. H. Crabtree, *Inorganic chemistry*, 2001, **40**, 5806–5811.
- (35) K. D. Cooney, T. R. Cundari, N. W. Hoffman, K. A. Pittard, M. D. Temple and Y. Zhao, *Journal of the American Chemical Society*, 2003, **125**, 4318–4324.
- (36) H. V. Huynh, *Chemical reviews*, 2018, **118**, 9457–9492.
- (37) A. Verloop, W. Hoogenstraaten and J. Tipker, *Drug design*, 1976, **7**, 165–207.
- (38) T. Piou, F. Romanov-Michailidis, M. Romanova-Michaelides, K. E. Jackson, N. Semakul, T. D. Taggart, B. S. Newell, C. D. Rithner, R. S. Paton and T. Rovis, *Journal of the American Chemical Society*, 2017, **139**, 1296–1310.
- (39) A. V. Brethome, S. P. Fletcher and R. S. Paton, *ACS Catalysis*, 2019, **9**, 2313–2323.
- (40) S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher and M. Stahn, *The Journal of Physical Chemistry A*, 2021, **125**, 4039–4054.
- (41) P. Pracht, F. Bohle and S. Grimme, *Physical Chemistry Chemical Physics*, 2020, **22**, 7169–7192.
- (42) S. Grimme, *Journal of chemical theory and computation*, 2019, **15**, 2847–2862.
- (43) P. Pracht and S. Grimme, *Chemical science*, 2021, **12**, 6551–6568.

- (44) C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2021, **11**, e1493.
- (45) Y. Zhao and D. G. Truhlar, *Theoretical chemistry accounts*, 2008, **120**, 215–241.
- (46) a. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley and J. Mantzaris, *The Journal of chemical physics*, 1988, **89**, 2193–2218.
- (47) G. Petersson and M. A. Al-Laham, *The Journal of chemical physics*, 1991, **94**, 6081–6090.
- (48) A. V. Marenich, C. J. Cramer and D. G. Truhlar, *The Journal of Physical Chemistry B*, 2009, **113**, 6378–6396.
- (49) M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji et al., *Wallingford, CT*, 2016.
- (50) J. C. Kromann, *Calculate Root-mean-square deviation (RMSD) of Two Molecules Using Rotation*, version 1.4, 2020.
- (51) A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 1372–1377.
- (52) S. Miertuš, E. Scrocco and J. Tomasi, *Chemical Physics*, 1981, **55**, 117–129.
- (53) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 09, Revision A.02*, Wallingford, CT: Gaussian, Inc., 2016.
- (54) S. T. Schneebeli, M. L. Hall, R. Breslow and R. Friesner, *Journal of the American Chemical Society*, 2009, **131**, 3965–3973.
- (55) S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher and M. Stahn, *The Journal of Physical Chemistry A*, 2021, **125**, 4039–4054.
- (56) I. H. Williams, *Journal of Physical Organic Chemistry*, 2022, **35**, e4312.
- (57) L. J. Hauptert, J. C. Poutsma and P. G. Wenthold, *Accounts of chemical research*, 2009, **42**, 1480–1488.

Conclusions

In this thesis, various projects related to the rational design of organocatalysts were shown. The focus was on a shift in paradigm from a **retrospective** use of computational and theoretical chemistry to a **prospective** approach. This new approach involves conducting computational studies at the outset, unravelling the mechanism of action of a catalyst, and allowing its rational modification. This modification is based on facts and not mere chemical intuition, resulting in improved selectivity.

In the first chapter, an extensive benchmark on different conformational analysis software packages. These were evaluated based on five criteria to assess different aspects of the conformational analysis process. The criteria include the accuracy of each program in finding the most stable conformer; **tunability**, accounting for the flexibility of the package in terms of adjustable parameters and conditions; **space exploration**, assessing how much of the conformational space of a molecule is explored by each software; and finally, **structure accuracy** and **energy accuracy**, which evaluate the precision of the structures and energies calculated by each program.

While this project did not yield a definitive answer to the question of the best conformational package, it does offer valuable insights. There is no universally perfect package, but the project provides a user guide detailing the strengths and weaknesses of each software. Additionally, it includes a list of recommendations to assist users in selecting the most suitable package for their specific needs and workflow.

The following chapter provides a detailed study of the interactions between the most common cationic PTC catalyst and different anions. Despite being commonly referred to as ion-pair over the last decades in many studies, these interactions and their nature lacked conclusive evidence to support such categorisation. Through a comprehensive investigation, it was revealed that the interactions between different catalysts and anions exist along a spectrum ranging from a pure HB to a pure ion pair. The different interactions can be divided into **charge-assisted HB**, **HB assisted ion pair** and **long-range ion pair** and a detailed description of each one of them is provided, analysing the different NCIs involved and the NBO involved in the interaction. The primary objective of this project is to establish an appropriate nomenclature for researchers to use, allowing them to accurately label interactions based on their true nature.

The preliminary studies and concepts explored in the first two chapters found practical application in chapter three, where they were applied to a real PTC reaction. The study focused on exploring the various interactions between different cinchona-based catalysts and commonly used anions in PTC, along with their interactions with different neutral reactants. The investigation revealed the definite binding mode between the catalysts and the reactants was determined. Contrary to previous studies that contemplated a strict ion-pair, it was determined that the catalyst forms a HB with the CN^- anion. In addition to this discovery, three different cyanation reactions were studied as a proof of concept for the rational design of new and improved catalysts. Various modifications involving both electron-withdrawing (EWG) and electron-donating (EDG) groups on the catalyst were tested. Based on the studied mechanism and the structural and energetic data extracted from the different catalysts under study, a modification was rationally proposed. This modification introduced a second anchor point to the catalyst, resulting in a significant enhancement of the theoretical enantioselectivities obtained for the first reaction. Additionally, it led to the inversion of enantioselectivities in the other two reactions. This highlights the remarkable **prospective** potential of computational chemistry.

Once the new **prospective** approach was tested with the three previous reactions, and the workflow to study reactions under PTC from a computational perspective was established, a collaboration with an experimental group was initiated to optimise the production of enantiomerically pure bis-protected amino acids using racemic azlactone as the starting point; a novel reaction that has never been performed before under PTC conditions using cinchona-based catalysts.

Our computational studies led the investigation, unravelling the mechanism of the reaction and the origin of the selectivity of the process. Similar to the approach in the previous chapter, the problem of the lack of selectivity was identified. This mechanistic knowledge was used to propose different modifications to the catalysts, adding a second anchor point that interacts with the azlactone, holding it in a specific position and favouring the reaction of one of the enantiomers. These modifications were tested experimentally, resulting in a remarkable improvement from a racemic mixture of products with the first catalysts to a 64%*ee*, a stunning advancement for a single round of modifications.

Despite this impressive improvement, the selectivity is not perfect, and further modifications are currently under study to enhance it even more. Additionally, various aspects of the reaction, such as the racemisation rate of the product, will be studied to determine whether they affect the overall selectivity of the reaction.

Another issue identified in the previous chapter is that manually studying different systems, particularly finding the most stable TS of each enantiomer, can lead to biased and, therefore, incorrect results. Considering the structural flexibility of these systems and the myriad of NCIs that they can establish between the different reactants, finding the most stable combination presents a challenge. To address this problem, the last chapter of this thesis focuses on the development of an automatic workflow that minimises human input and automatically explores the different possible conformers of the TSs. This approach incrementally increases the level of theory and refines the TS ensemble until the most relevant ones are obtained. It not only reproduces the experimental results better than the manually constructed TSs but also proves to be more efficient, using fewer computational resources due to its incremental nature.

In conclusion, this thesis provides several pieces of evidence highlighting the importance of using computational chemistry prospectively instead of a mere retrospective tool. This contribution supports the paradigm shift, advocating for the incorporation of computational studies at the beginning of every project. The thesis establishes foundations for future collaborations with experimental groups based on the rational design of catalysts, led by these computational studies, and proposes a new and efficient automatic workflow for the study of enantioselective reactions from a computational perspective.

A | Appendixes

A.1 Supporting Information

The supporting information for this manuscript can be found in the following Zenodo repository:
DOI 10.5281/zenodo.10407604.

Table of Contents

Chapter 1 - Efficiency and Suitability when Exploring the Conformational Space of Phase-Transfer Catalysts

Figures

- Figure S1.1 - Population of the possible conformations of Thiourea and Squaramide substituted models with $R = EWG$ and EDG in H_2O and THF .
- Figure S1.2 - 3D conformational ensembles generated for each unconstrained catalyst.
- Figure S1.3 - 2D graph of the clusters obtained for each unconstrained catalyst.
- Figure S1.4 - Number of conformers generated for each unconstrained catalyst.
- Figure S1.5 - RMSD statistic values obtained for each unconstrained catalyst.
- Figure S1.6 - Sankey diagrams for the different clusters obtained for each unconstrained catalyst.
- Figure S1.7 - Energy and RMSD comparison between the studied conformers optimised with the different software packages and the DFT-optimised structures.
- Figure S1.8 - Most stable conformer obtained for each unconstrained catalyst with the different software packages under study.
- Figure S1.9 - Relative energy of the generated conformers for each unconstrained catalyst.
- Figure S1.10 - Population and cumulative population of the generated conformers for each unconstrained catalyst.
- Figure S1.11 - 3D conformational ensembles generated for each constrained catalyst.
- Figure S1.12 - 2D graph of the clusters obtained for each constrained catalyst.
- Figure S1.13 - Number of conformers generated for each constrained catalyst.

- Figure S1.14 - RMSD statistic values obtained for each constrained catalyst.
- Figure S1.15 - Most stable conformer obtained for each constrained catalyst with the different software packages under study.
- Figure S1.16 - Relative energy of the generated conformers for each constrained catalyst.
- Figure S1.17 - Population and cumulative population of the generated conformers for each constrained catalyst.

Tables

- Table S1.1 - SMILE codes of the catalysts under study.
- Table S1.2 - Atomic coordinates (XYZ) for all the conformers generated in the unconstrained conformational analysis.
- Table S1.3 - Atomic coordinates (XYZ) for all the conformers generated in the constrained conformational analysis.
- Table S1.4 - Dimension reduction and clusterisation parameters for the unconstrained conformational analysis.
- Table S1.5 - Dimension reduction and clusterisation parameters for the constrained conformational analysis.
- Table S1.6 - RMSD and dihedral matrix results for each unconstrained catalyst.
- Table S1.7 - Clusterisation results for each unconstrained catalyst.
- Table S1.8 - RMSD and dihedral matrix results for each constrained catalyst.
- Table S1.9 - Clusterisation results for each constrained catalyst.

Scripts

- Script S1 - *catalyst_generation* - Script for generating the 3D structure used as input in the different conformational analyses performed.
- Script S1 - *clusterisation_UMAP* - Script for the clusterisation of the obtained conformers for each catalyst.
- Script S1 - *rmsd_filtration* - Script for the filtration of the generated conformers by means of an RMSD matrix.

Chapter 2 - Revisiting Ion-Pair Interactions in Phase Transfer Catalysis: From Pure Ionic Compounds to Real Catalyst Systems

Figures

- Figure S2.1 - NBO charges of the monomers of the different complexes under study.
- Figure S2.2 - LMO-EDA energy partition terms for the different complexes under study.

- Figure S2.3 - NBO E(2) values corresponding to the different complexes under study.
- Figure S2.4 - QTAIM 3D molecular graphs of the complexes under study.
- Figure S2.5 - EDB-EDA energy partition terms for the different complexes under study in different solvents.

Tables

- Table S2.1 - Atomic coordinates (XYZ) for all the complexes under study.
- Table S2.2 - Interaction energies, ΔE , for all the complexes under study.
- Table S2.3 - LMO-EDA energy partition terms values for all the complexes under study.
- Table S2.4 - NBO E(2) values corresponding to the different complexes under study.
- Table S2.5 - QTAIM $\rho(r)_{BCP}$, $\nabla^2\rho(r)_{BCP}$, H_{BCP} and d_{bond} corresponding to the different complexes under study.
- Table S2.6 - NBO charges of the monomers of the different complexes under study.

Chapter 3 - Improving Phase-Transfer Catalysis by enhancing Non-Covalent Interactions.

Figures

- Figure S3.1 - Conformational analysis figures for the different catalysts under study.
- Figure S3.2 - 3D molecular graphs for the different structures under study.
- Figure S3.3 - 2D NCI plots for the different cat-anion complexes under study with cat_ome_ph (part 1).
- Figure S3.4 - 3D TS structure comparison for reaction 2 and 3 with all the catalysts under study.

Tables

- Table S3.1 - Basis set energy benchmark results for catalyst complexes and cat-anion complexes.
- Table S3.2 - Atomic coordinates (XYZ) for all the conformers, complexes and reaction steps under study for the cyanation of (E)-Chalcone.
- Table S3.3 - Atomic coordinates (XYZ) for all the conformers, complexes and reaction steps under study for the cyanation of (2E)-1-Phenyl-2-buten-1-one and (2E)-1-(1H-Pyrrol-1-yl)-2-hexen-1-one.
- Table S3.4 - Comparison of the different methods for *ee* calculation.
- Table S3.5 - NBO results for the different reactions under study.
- Table S3.6 - QTAIM results for the different reactions under study.

- Table S3.7 - Conformational analysis energy and population data of the different catalysts and reactants under study.
- Table S3.8 - Energy values and N^+-X^- distances of the different cat-anion complexes studied with cat_ome_ph (part1).
- Table S3.9 - PES of the different reactions studied.

Chapter 4 - Dynamic kinetic resolution of azlactones for the production of bis-protected aminoacids: a theoretical approach.

Figures

- Figure S4.1 - 3D structures of the studied TSs with the different catalysts under study.

Tables

- Table S4.1 - XYZ coordinates of the optimised structures. Isolated catalysts and reactants, cat-reactant complexes and different reaction steps.
- Table S4.2 - Energy values of the optimised TSs and %ee theoretical values obtained.

Chapter 5 - Computational Design: Transition State Studies, enantioselectivity Calculations, and Workflow automatisation

Figures

- Figure S5.1 - Aligned 3D conformational generated by CREST without hiding the azlactone.
- Figure S5.2 - Aligned 3D conformational ensembles reoptimised at the different levels of theory.
- Figure S5.3 - Population of the different conformational ensembles at the different levels of theory.
- Figure S5.4 - 3D most stable conformer at each level of theory.

Tables

- Table S5.1 - Initial XYZ structures used as input for CREST conformational ensembles generation.
- Table S5.2 - Atomic coordinates of the conformational ensembles generated by CREST.
- Table S5.3 - Atomic coordinates of the reoptimised structures at the different levels of theory.

A.2 Published papers

- (14) M. Ferrer, I. Iribaren, T. Renningholtz, I. Alkorta and T. Cristina, *BJOC - Beilstein Journal of Organic Chemistry- Thematic Issue "Adaptive experimentation and optimization in organic chemistry"*, In preparation.
- (13) I. Iribarren, M. Ferrer, I. Alkorta and C. Trujillo, *EuroJOC - Special Collection Physical Organic Chemistry*, In preparation.
- (12) I. Iribarren, E. Mates and C. Trujillo, *Dalton Trans.*, 2024.
- (11) I. Iribarren, C. Trujillo, G. Sánchez-Sanz, E. Hénon, J. Elguero and I. Alkorta, *Theoretical Chemistry Accounts*, 2023, **142**, 81.
- (10) M. R. Garcia, I. Iribarren, I. Rozas and C. Trujillo, *Chemistry–A European Journal*, 2023, e202203577.
- (9) N. Melnyk, M. R. Garcia, I. Iribarren and C. Trujillo, *Tetrahedron Chem*, 2023, 100035.
- (8) I. Iribarren and C. Trujillo, *Journal of Chemical Information and Modeling*, 2022, **62**, 5568–5580.
- (7) I. Iribarren, M. R. Garcia and C. Trujillo, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2022, **12**, e1616.
- (6) N. Melnyk, I. Iribarren, E. Mates-Torres and C. Trujillo, *Chemistry–A European Journal*, 2022, **28**, e202201570.
- (5) I. Iribarren, G. Sánchez-Sanz, I. Alkorta, J. Elguero and C. Trujillo, *The Journal of Physical Chemistry A*, 2021, **125**, 4741–4749.
- (4) I. Iribarren, G. Sánchez-Sanz, J. Elguero, I. Alkorta and C. Trujillo, *ChemistryOpen*, 2021, **10**, 724–730.
- (3) I. Iribarren and C. Trujillo, *Physical Chemistry Chemical Physics*, 2020, **22**, 21015–21021.
- (2) I. Iribarren, G. Sánchez-Sanz and C. Trujillo, *Molecules*, 2020, **25**, 798.
- (1) I. Iribarren, M. M. Montero-Campillo, I. Alkorta, J. Elguero and D. Quiñonero, *Physical Chemistry Chemical Physics*, 2019, **21**, 5796–5802.