MODELING OF HOMOGENEOUS CATALYSIS: FROM DFT TO QSPR APPROACHES

Sonia Aguado Ullate

Dipòsit Legal: T. 279-2012

Sonia Aguado Ullate

Modeling of homogeneous catalysis:

from DFT to QSPR approaches

PhD THESIS

Supervised by Dr. Jorge J. Carbó Martín

Departament de Química Física i Inorgànica

UNIVERSITAT ROVIRA I VIRGILI

Tarragona

2012

UNIVERSITAT ROVIRA I VIRGILI
MODELING OF HOMOGENEOUS CATALYSIS: FROM DFT TO QSPR APPROACHES
Sonia Aguado Ullate
DL:T. 279-2012

UNIVERSITAT ROVIRA I VIRGILI

Departament de Química Física i Inorgànica
C/ Marcel·lí Domingo s/n
43007 - Tarragona

Jorge Juan Carbó Martín, Professor Agregat del Departament de Química Física i Inorgànica de la Universitat Rovira i Virgili,

FAIG CONSTAR que la present memòria, titulada:

"Modeling of homogeneous catalysis: from DFT to QSPR approaches"

ha estat realitzada sota la meva direcció al Departament de Química Física i Inorgànica de la Universitat Rovira i Virgili per Sonia Aguado Ullate per a l'obtenció del títol de Doctor.

Tarragona, Febrer de 2012

El director de la tesi doctoral

Dr. Jorge J. Carbó Martín

UNIVERSITAT ROVIRA I VIRGILI
MODELING OF HOMOGENEOUS CATALYSIS: FROM DFT TO QSPR APPROACHES
Sonia Aguado Ullate
DL:T. 279-2012

Agradecimientos

Me gustaría comenzar haciendo referencia a la memoria del Dr. Pedro Sarasa por haberme iniciado en la química teórica, dándome la oportunidad de realizar un trabajo experimental durante mi último año de carrera. Gracias a él, se me despertó el interés por el mundo de la química computacional. Quiero agradecer a mi director de tesis, Dr. Jordi Carbó por su amabilidad, profesionalidad y esfuerzo durante estos años. El desarrollo y aprendizaje de las diferentes metodologías utilizadas en la tesis no hubiera sido posible sin su apoyo.

Quiero agradecer al Dr. Carles Bo, Dr. Manolo Urbano, Dr. José Ignacio García y Dr. Miguel Mena su colaboración en los diferentes estudios realizados en esta tesis doctoral. Agradezco al Prof. Jonathan Hirst haberme brindado la oportunidad de trabajar en su grupo de investigación durante mi estancia en Nottingham y al PhD Abrar Hussain por su ayuda durante el proyecto. A Dolores, Marta, Mara, Eduardo, Pilar, Yolanda, Óscar, Luis y Lee por las cervezas de los viernes, la bolera, la despedida y todos los momentos divertidos que pasamos en Nottingham.

La experiencia vivida en el doctorado ha sido muy especial tanto a nivel profesional como a nivel personal, que ha sido posible gracias a los todos los miembros del grupo de Química Cuántica del Departamento de Química Física. Destacando a mis compañeros del doctorado que han sido mi familia en Tarragona. Gracias por las terapias de grupo en la Plaça la Font, las cenas y celebraciones, y por supuesto, las charlas para resolver el mundo en la cafetería de la Universidad. No sólo han sido mis compañeros de trabajo, son unos amigos a los que nunca olvidaré.

Mi familia es mi gran apoyo, sin ellos no podría conseguir mis metas. Quiero agradecer a toda mi familia, desde mis primos, tíos a mis abuelos, por todos los momentos que hemos vivido juntos. En especial a mis padres, Mª Jesús y José Mª, y a mi hermana, Edurne, porque siempre puedo contar con ellos, me dan un apoyo incondicional, son los que me animan, me aconsejan y me hacen mirar hacia adelante. No tengo palabras para agradecerles todo lo que me aportan. Gracias a Héctor, mi pareja, por estar a mi lado, por ser mi confidente, mi mejor amigo, es decir, por ser como es y compartirlo conmigo.

También quiero agradecer a mis amigas; Rosa, Susana, Eva, Charo, Judith y María con las que he crecido y he madurado. Por todas las experiencias vividas, desde la niñez hemos estado juntas y nunca me habéis fallado, gracias.

Por último, me gustaría dedicar la tesis doctoral a la memoria de mi tía Maite, una de las personas más importantes en mi vida. Gracias por cuidarme de niña, por los consejos, los disfraces del cole y carnavales, las vacaciones en Noain, en otras palabras, por haber sido parte de mi vida.

UNIVERSITAT ROVIRA I VIRGILI
MODELING OF HOMOGENEOUS CATALYSIS: FROM DFT TO QSPR APPROACHES
Sonia Aguado Ullate
DL:T. 279-2012

A Maite

Contents

Chapter 1

Introduction

Chapter 1

Introduction

The discovery of new materials and chemical processes is increasingly important for the development of a modern society, in which environmental protection is a major concern. Technological evolution constantly demands the synthesis of new materials and chemicals products which maintain the sustainability of production and the environment. At industrial level, chemical processes must be specifically designed and manufacturing methods, efficient, clean and ecological. In this context, the catalytic processes can play a crucial role, since they are able to minimize materials, energy and resources.

Catalysis is a field of science that explores solutions to environmental problems such as pollution, elimination of waste generated in the process of materials synthesis or regeneration of natural resources. Homogeneous catalysis refers to a catalytic system, in which the substrates for a reaction and the catalyst components are brought together in one phase, most often the liquid phase. Homogeneous catalysis usually involves organometallic complexes as the catalysts and the formation of C-C, C-H, C-O or C-N bonds. An emerging area within this field is asymmetric catalysis, in which enantiomerically enriched products are obtained using chiral catalysts. Enormous efforts are being dedicated to the development, discovery and optimization of active chiral catalysts, due to the increasing demand of enantiomerically pure compounds. One of the advantages of asymmetric catalysis is to avoid the product of optically inactive species, and as result, to improve the efficiency of processes.

Frequently, catalysts and catalytic processes have been discovered through a trial-error approach, rather than by rational design. Computational methods are rapidly becoming powerful tools for changing this scenario. In fact, catalysis is identified as one of the areas in which computational chemistry has achieved the most success. Currently, it is possible to determine reaction mechanisms with sufficient accuracy in a reasonable time, and therefore understand the behaviour of a catalyst and propose changes to alter its function. The primary objective of any modeling technique is to reproduce the experimental values, and then to identify and rationalize the catalytic outcome. The ultimate goal is to predict the behaviour of the catalysts and establish rules for the design of new catalysts. These goals are even more challenging in asymmetric catalytic process due to the small energy differences between the diastereomeric reaction paths. Computational approaches have received attention at academic and industrial level. However, despite significant advances, the *de novo* design of catalysts has been restricted to compounds structurally related to previous ones.

The development of computers and the use of density functional theory (DFT) and hybrid quantum mechanics / molecular mechanics (QM/MM) methods makes it possible to calculate

geometries, electronic structures, bond energies, spectroscopic properties, reaction energies, and activation energies of chemical reactions with high accuracy. Thus, in order to computationally assess the key factors governing the activity and the selectivity of the catalytic processes, it is common practice to locate and analyze the minima (reactants, intermediates and products) and the saddle points (transition states, TS) on the *Potential Energy Surface* (PES).

As an alternative to TS-based approaches, other computational methodologies define descriptors for ground-state properties to look for quantitative relationships with the activity or selectivity. They are usually referred to as Quantitative Structure-Activity Relationship (QSAR) or -Property Relationship (QSPR) approaches, in which it is assumed that it is possible to correlate the mathematical features of the catalyst molecular structure (molecular descriptors) with its activity or selectivity. In principle, these mathematical models enable the prediction of the activity and selectivity without prior experimentation synthesis and lead to the design of new compounds. Nowadays, QSAR approaches are widely used in the process of drug discovery; one could even say that they are an essential part of it. Despite acceptance in the pharmaceutical world and their potential utility, the use of QSAR approaches in the design of catalysts is still rare. Its detractors argue that this type of models based on ground-state structures a step back compared to models based on transition state structures.

The PhD Thesis is divided into six chapters. The first part of *Chapter 1* includes the catalytic processes studied in the Thesis and summarizes the most representative experimental and theoretical findings on these topics. The second part of the *Chapter 1* gives a general overview of the computational methodologies and strategies for the study of homogeneous catalysis ranging from TS-based to QSPR approaches; and the last part is reserved for the objectives of this work. *Chapter 2* reports a DFT study on the N-H σ-bond activation of ammonia by the μ_3-alkylidyne titanium species using the [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-CH)] model complex. In *Chapter 3*, we combine the TS-based approach and qualitative analysis through a newly defined molecular descriptor (*distance-weighted volume,* V_W), in order to analyze the asymmetric hydroformylation of styrene catalyzed by Rh-binaphos complexes. Using our previous mechanistic knowledge, *Chapter 4* presents a QSPR study to predict the activity and the enantioselectivity in the hydroformylation of styrene catalyzed by Rh-diphosphane complexes. In *Chapter 5,* we develop a new methodology to predict enantioselectivity based on the quantitative quadrant-diagram representation of the catalysts and 3D-QSSR modeling; and we apply it in the asymmetric cyclopropanation of alkenes catalyzed by copper complexes. Finally, *Chapter 6* contains the concluding remarks of this PhD Thesis.

UNIVERSITAT ROVIRA I VIRGILI
MODELING OF HOMOGENEOUS CATALYSIS: FROM DFT TO QSPR APPROACHES
Sonia Aguado Ullate
DL:T. 279-2012

1.1. Catalytic processes

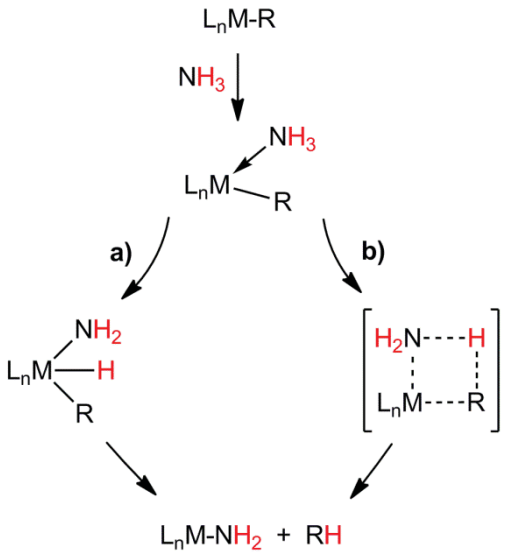
As mentioned in the above paragraphs, in this section, we describe the reactions and catalytic processes studied in this PhD Thesis: ammonia activation, hydroformylation of alkenes catalyzed by rhodium complexes and asymmetric cyclopropanation of alkenes.

1.1.1. Ammonia activation

The activation of the ammonia molecule has attracted considerable interest because it is a synthetic precursor for a wide range of commercially useful products. Most of the chemistry of ammonia is dominated by its nucleophilicity due to the presence of a lone pair at nitrogen, the high N-H bond dissociation energy (~ 104±2 kcal mol^{-1})[1] and the scarce tendency to form N-H σ complexes.[2] In consequence, the majority of the reactions of ammonia with transition metal complexes generate Lewis acid-base adducts,[2a-3] the activation of its N-H bonds still being unusual.[4-5]

Relatively few examples of ammonia N-H bond activation processes by direct use of ammonia have been reported to synthesize amido complexes (M-NH_2). Most of them have been prepared via the oxidative addition of ammonia to a transition-metal complex (Scheme 1.1, path a).

Scheme 1.1.



Only recently have Hartwig and co-workers reported the first stable mononuclear amido hydride species via cleavage of one of the N-H bonds of ammonia molecule by a Ir(I) complex with an aliphatic PCP ligand.[5] Interesting, the same authors showed that in the case of an

aromatic PCP ligand the reverse reaction to give a σ-bonded NH_3 complex via reductive elimination takes place.[6]

More recently have Turculet and co-workers detailed the formation of an isolable complex of the type [{CyPSiP}Ir(NH_2)H] (Cy-PSiP = κ^3-(2-$Cy_2PC_6H_4)_2$SiMe),[7] significantly more resistant to N-H bond reductive elimination than the [(PCP)Ir] pincer systems previously mentioned.[5,7] The groups of Mindiola and Ozerov have achieved the activation of ammonia via a binuclear oxidative addition using a dinuclear palladium (PNP) pincer species to give monomeric species.[8] Alternatively, the H_2N-H bond-breaking processes can also lead to the formation of dinuclear bridging amido species such as [{Ir(μ-NH_2)(PEt_3)}$_2$][3a] and *cis*-[{Ir(PEt_3)$_2$(NH_3)H(μ-NH_2)}$_2$][9] reported by Milstein, and [Ir(4-C_5NF_4)H(μ-NH_2)(NH_3){P^iPr_3)}]$_2$ reported by Braun.[10] There have been examples of oxidative addition of ammonia to trimetallic systems with a cooperative action of the metal centres. The osmium trinuclear carbonyl cluster, [$Os_3(CO)_{11}$L] (L = *c*-C_6H_8 CH_3CN), reacts with ammonia to produce the complex [$Os_3(CO)_{10}$(μ-H)(μ-NH_2)], where two osmium centres support a μ-amido moiety.[11]

The oxidative addition can be also achieved by main group elements. Bertrand and co-workers[12] published a facile splitting of ammonia using cyclic and acyclic [(alky)(amino)carbenes] and their strong nucleophilic character. Research on carbene type species, involving oxidative addition of NH_3 and creating an amide fragment and a terminal hydride ligand, has been further expanded to other group 14 elements by the groups of Power,[13] Roesky,[14] and Stephan.[15] The latter reported a N-H bond cleavage by a frustrated Lewis pairs (FLP) as N-heterocyclic carbene/B(C_6F_5)$_3$.

As an alternative to oxidative addition, the N-H bond can be also activated via metal-ligand cooperation without involving oxidative addition of the metal centre (Scheme 1.1, path b).[16] Some complexes containing d^0 early transition metal centres, such as [(η^5-$C_5Me_5)_2MH_2$] (M = Zr, Hf)[17] and [(η^5-$C_5Me_5)_2$ScR] (R = H, Me)[16c] or [{$(Me_3Si)_2N$}$_3$M-Cl],[18] react with ammonia to yield amide complexes and H_2, methane, or hydrogen halide elimination, respectively. Carbene type derivatives [Ar_2M] (M = Sn,[19] Mn, Fe;[20] Ar = bulky terphenyl ligands) afford bridging μ-amide complexes with arene elimination. Milstein and co-workers have reported a novel type of reversible N-H bond activation by ruthenium complexes modified with pyridine-based PNP pincer ligand that involves hydrogen transfer to an unsaturated bond of the ligand, leading to aromaticity of the central pyridine ring.[21] Very recently have Driess and co-workers reported an ammonia N-H bond activation by a silylene ligand in a [LSi(II)]$\rightarrow$[Ni(CO)$_3$] complex.[22]

Apart from single N-H activation, further N-H bond cleavage to yield imide and nitride complexes has been also reported; however, these examples are still rare. Some tantalum complexes promote a second N-H bond cleavage of ammonia to yield imide complexes of the type [(η^5-$C_5Me_5)_2$Ta(=NH)H][23] and [{$(Me_3SiNCH_2CH_2)_3$N}Ta=NH].[24] Another interesting example, which constitutes a bridge of knowledge between the organometallic chemistry and the surface chemistry, was reported by Basset, Emsley, Quadrelli and co-workers.[25] The authors showed that a solid-supported tantalum hydrido [($\equiv$SiO)$_2$TaH] converts into the imido amido surface complex [($\equiv$SiO)$_2$Ta(NH)(NH_2)] by reaction with ammonia. Similar to the osmium

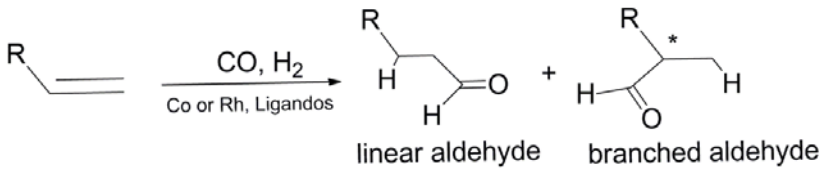
trinuclear cluster mentioned above, Suzuki and co-workers reported the ammonia oxidative addition over the triruthenium complex [{(η^5-C_5Me_5)Ru(μ-H)}$_3$(μ_3-H)$_2$] to form the μ_3-imido cluster [{(η^5-C_5Me_5)Ru(μ-H)}$_3$(μ_3-NH)].[26]

Finally, the activation of ammonia via cleavage of the three N-H bonds leading to nitrido-species is rather scarce.[27] The first example of such a process was reported by Roesky and co-workers, who synthesized the titanium imido nitrido [{(η^5-C_5Me_5)Ti(μ-NH)}$_3$(μ_3-N)], by reaction of [(η^5-C_5Me_5)TiMe_3] with excess of ammonia and loss of methane at room temperature.[28] Ammonia N-H bonds activation was further expanded by Wolczanski to other tantalum[29] and zirconium[30] nitrido derivatives. Some linear nitrido complexes of the type [X(NH_3)$_4$OsNOs(NH_3)$_4$X]Cl_n (X = NH_3, Cl)[31] were prepared by reaction of $(NH_4)_2[OsCl_6]$ with hydrazine hydrate involving disproportionation of hydrazine into N_2 and NH_3. A similar result is obtained in the thermal treatment of $(NH_4)_2[OsCl_6]$ with concentrated aqueous ammonia.[32] In the nineties, Mena and co-workers succeeded in synthesizing the first nitrido organometallic with metalocubane structure [{Ti(η^5-C_5Me_5)(μ_3-N)}$_4$] by ammonolysis of the amido derivative [Ti(η^5-C_5Me_5)(NMe_2)$_3$].[33] Similar ammonolysis reaction from [Ti(η^5-C_5Me_5)Cl_3] or [Ti(η^5-C_5Me_5)Cl_2(NMe_2)] let to the isolate of the species [Ti_2(η^5-C_5Me_5)$_2Cl_3$(μ-N)(NH_3)].[3c] Also, ammonoysis of homoleptic metal amide complexes[34] [M(NR_2)$_n$] provides access to nitride solids with ammonia as coprecursor. On collaboration with Professor Mena's group, we have performed a study on the reactions of ammonia with [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-CR)] [R = H (**1**), Me (**2**)] to give rise to the first example of ammonia N-H bonds activation by μ_3-alkylidyne transition metal derivatives.

1.1.2. Hydroformylation of alkenes catalyzed by rhodium complexes

Hydroformylation is the oldest and largest homogeneous catalytic reaction of alkenes, which are converted into aldehydes by the addition of CO and H_2 (synthesis gas). The reaction can be catalyzed by several transition metal complexes, with low-valent cobalt and rhodium catalysts being the most popular in industry. Scheme 1.2 shows that hydroformylation of terminal alkenes generates a mixture of linear and branched aldehydes. Usually, the linear aldehyde is more desirable than the branched one in achiral hydroformylation, but asymmetric hydroformylation generates branched aldehydes that contain a chiral centre.

Scheme 1.2.



Achiral hydroformylation is one of the most important homogeneous catalytic processes in the chemical industry of 'bulk materials'. A clear example is the hydroformylation of propene which provides around 75% of all oxo chemicals consumed in the world.[35] Much effort has

been devoted to discovering and designing active and regioselective catalysts in hydroformylation. Wilkinson and co-workers reported that [Rh(I)-PPh_3] catalysts were active and highly regioselective in hydroformylation for 1-alkenes under ambient conditions.[36] The first highly *n*-selective Rh-catalyst was patented by the Eastman Kodak Company.[37] Nowadays, the best catalysts for achieving high regioselectivity control in the hydroformylation of terminal alkenes are the rhodium catalysts derived from the bidentate P-donor ligands such as the bisbi,[38] biphephos[39] and xantphos ligands[40] (Figure 1.1).

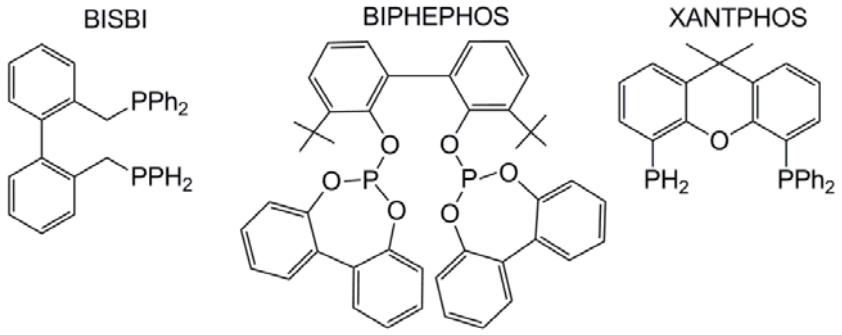


Figure 1.1. Examples of ligands used in hydroformylation of terminal alkenes.

In the late sixties, Wilkinson and co-workers proposed the *dissociative mechanism* for ethylene hydroformylation catalyzed by [Rh(I)-PPh_3]complexes (Figure 1.2).[41] Although the complexes shown in Figure 1.2 contain at least two coordinated PPh_3 ligands, a large body of indirect evidence suggests the influence of equilibria involving catalytically active species containing a single coordinated phosphane ligand.[42] Figure 1.2 shows the seven fundamental steps involved in the catalytic cycle: i) ligand dissociation from trigonal-bipyramidal rhodium hydride species **1** to generate the coordinatively unsaturated species **2**; ii) alkene coordination to yield complex **3**; iii) alkene insertion into the Rh-H bond to form alkyl complex **4**; iv) CO addition to generate **5**; v) CO insertion to form the acyl complex **6**; vi) H_2 oxidative addition; and vii) aldehyde reductive elimination.

In literature, there has been considerable controversy about the rate-determining step of catalytic hydroformylation. The relative rates of individual reaction steps are influenced by the steric and electronic properties of the ligands. Therefore, different reaction kinetics have been observed experimentally depending on these properties. In general, the reaction kinetic fits into one of two extreme cases. For electron-poor ligands such as bulky monophosphites and the unmodified rhodium-carbonyl catalysts, the hydrogenolysis of acyl specie **6** controls the overall rate of hydroformylation.[43] Therefore, the rate is independent of the alkene concentration, showing a negative order in CO concentration and a first-order rate dependency of the hydrogen concentration (usually referred to as Type II kinetics). Under these conditions, the resting state is complex **7**, which has been observed spectroscopically.[43a-44] On the contrary, the kinetics for electron-rich ligands are consistent with a rate-determining step early in the catalytic cycle (usually referred to as Type I kinetics). For these ligands, its

catalyst resting state is pentacoordinated rhodium hydride species **1**. Recently, a combination of isotope effects study and computational analysis by our group demonstrated that the overall process from the resting state specie **1** to hydride migration (from complex **3** to complex **4**) governs the overall activity in 1-octene hydroformylation catalyzed by the rhodium-xantphos complexes.[45]

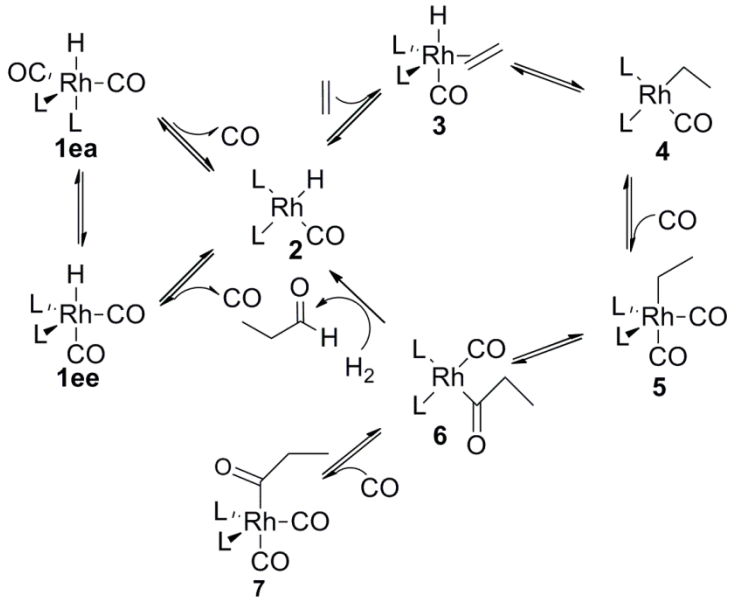


Figure 1.2. Dissociative hydroformylation catalytic cycle of ethylene as proposed by Wilkinson and co-workers (L = PPh_3).

Catalytic activity has been one of the most studied aspects of hydroformylation. Recent clarification of the rate determining step[45] indicates that the ligands that promote CO dissociation, alkene coordination, or hydride migration might yield higher catalytic turnovers. Besides, the understanding of the factors influencing the individual steps of the catalytic cycle, some systematic correlation studies have attempted to establish relationships between ligand structure and catalytic activity. Early experimental studies on monodentated ligands showed that a relationship exists between ligand basicity and catalysts activity, thus the least basic phosphanes enhance the activity.[42] In addition, van Leeuwen and co-workers showed that phenoxophosphane (phosphacyclic) moieties are less basic than diphenylphosphino moieties through the determination of carbonyl stretching frequencies of [(P-P)RhH$(CO)_2$] complex; they observed an increase in the rate of hydroformylation upon introduction of phosphacyclic moieties.[46] This was attributed to the decreased electron density at the metal centre, which should facilitate CO/alkene exchange and hydride migration. Theoretical studies have shown that electron-withdrawing ligands facilitate the hydride migration step.[47] The contribution of back-donation to the metal-alkene bond is small, leading to facile rotation of the alkene moiety to reach the transition state. Similarly, for bidentate cyclic phosphanes, Pringle and co-

workers found a correlation between C-P-C angles and catalyst activity for a series of isoelectronic and isosteric ligands.[48] The smaller ring size imposes small valence angles around the phosphorus atom, which seems to become a worse σ donor and a better π acceptor.

Apart from these studies, in which differences in catalytic activity have been related to the electronic features of the ligands, the steric effects have been also considered. For a series of phosphabenzenes, the variation in catalytic performance was attributed to the steric properties of the ligand.[49] According to the authors, the bulky phosphabenzene ligand favours the formation of a monoligand rhodium species, which should have a larger accessible space compared with diligand rhodium species; and hence the energy barrier for coordination or hydride metallation of alkene should be reduced. Kamer, van Leeuwen and co-workers studied the effect of the *natural bite* angle on the catalytic performance by synthesizing a series of diphosphane xantphos-type ligands.[50] Xantphos ligands were especially designed to ensure that mutual variation in electronic properties and in steric size within the series is minimal, and that the bite angle is the only characteristic that has a significant variation within the series.[40] For the hydroformylation of 1-octene and styrene, the authors observed an increase in the rate with increasing the bite angle, suggesting that the bite angle effect on activity is dominated by the reaction of [(diphosphane)Rh(CO)H] intermediate with CO and alkene.[40] However, how the bite angle affects the activity is still not understood in detail.[51] On the contrary, the effect of the bite angle on regioselectivity was unravelled with the help of QM/MM calculations.[52] By determining the transition states for alkene insertion, the authors reproduced the experimentally observed trends and showed that regioselectivity is governed by the nonbonding interactions between the diphenylphosphino substituents and the substrate. Thus, wider bite angles increase the steric congestion around the metal centre, disfavouring the most sterically demanding paths, which yield the branched product.

The asymmetric hydroformylation (AHF) of alkenes offers a potentially very useful synthetic route to enantiomerically pure aldehydes, which can be used as precursors of biologically active compounds and fine chemicals.[53] Thus, much effort has been devoted during recent years to designing efficient catalysts for asymmetric catalysis (Figure 1.3). A real breakthrough in this area was the publication of Rh/(*R,S*)-binaphos catalysts by Takaya and co-workers.[54a] Binaphos is an unsymmetric bidentate phosphane phosphite ligand, which has been successfully employed in the hydroformylation of several substrates to achieve high activity and stereoselectivity.[54b,55] However, it is not an ideal catalyst; its functional group tolerance is limited, making it difficult to apply in the real world to the synthesis of pharmaceutical intermediates, which often have complex substitution patterns. Another catalyst that had some success was the Rh/(*2R,4R*)-chiraphite from Union Carbide, which exhibited high enantioselectivity (76% yield) with high regioselectivity control for styrene.[56] Thus, the broad applicability of the reaction was, and still is, a goal to pursue due to the intrinsic difficulties of controlling simultaneously the regio- and the enantio-selectivity, and obtaining highly active catalysts for a broad range of substrates. In consequence, the asymmetric hydroformylation was not used industrially until 2003, when some companies start offering AHF in their technological portfolio.[57] Wills and co-workers reported the first esphos-type ligand[58] that was found to be effective for AHF of vinyl acetate but, interestingly, quite unselective in the

hydroformylation of styrene. Dowpharma company described the use of a new ligand, called (*S,S*)-kelliphite.[59]

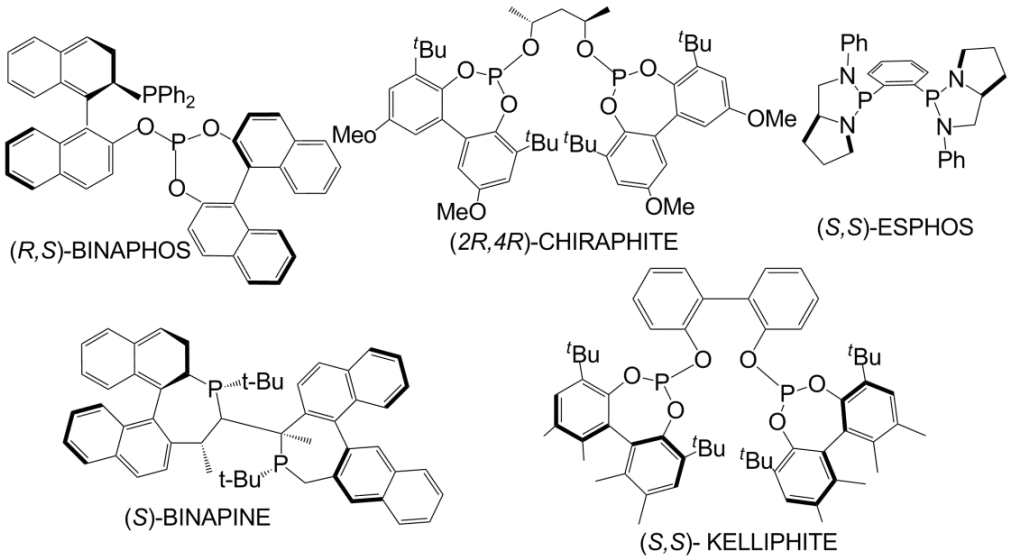


Figure 1.3. Examples of ligands used in asymmetric hydroformylation reactions.

Most of the selective catalysts are based on rhodium and chelating P-donor ligands, including C_1- and C_2-symmetric ligands, and diverse phosphorus groups such as diphosphites,[56,59,60] diphosphonites,[61] diphosphanes,[62] diazaphospholanes,[58,63] unsymmetric phosphane-phosphite,[60,64] or phosphane-phosphoramidites.[65] The (*S,S*)-kelliphite (diphosphite) developed by Whiteker and co-workers provide high enantioselectivities with very high regioselectivity control for allyl cyanide and vinyl acetate substrates.[59] Unfortunately, the reaction of styrene occurs with poor enantioselectivity, although the regioselectivity control is high. Other outstanding examples are (*R,R*)-Ph-BPE[62b] (diphospholanes), yanphos[65-66] (phosphane-phosphoamidite), (*S,S*)-binaphine and (*S,S,R,R*)-tangphos (diphosphanes).[67] The latter ligands were originally developed by the Zhang group for asymmetric hydrogenation,[67a] and then tested by Klosin and co-workers in AHF.[67b] Finally, mention should be made of the P-stereogenic aminophosphane phosphinite ligands (AMPP) reported by Vogt and co-workers that provide relative high enantioselectivity with high regioselectivity control for styrene, but medium enantioselectivity with poor regioselectivity control for vinyl acetate.[68]

For establishing empirical relationships between catalyst structure and selectivity in hydroformylation, some researches carried out systematic studies in small ligand families that ideally vary in a single ligand variable.[69] For a series of biaryl-bridged diphosphite ligands, Klosin, Whiteker and co-workers found a linear relationship between the dihedral angle of bridging aryl moieties and the enantioselectivity for vinyl acetate and allyl cyanide.[69a] A similar correlation was also observed between the bridging dihedral angle and the regioselectivity of branched aldehyde. In both cases, decreasing the dihedral angle increases the

enantioselectivity and regioselectivity of the branched isomer. Based on the DFT calculation of reaction intermediates, it was observed that the dihedral angle of bridging aryl moieties in turns correlates with the computed P-Rh-P bite angle: the narrower the dihedral angle, the narrower the bite angle is. However, for a family of diphospholane ligands, the same type of correlation was unclear.[69b] Klosin and co-workers explored more varied set diphosphacycle structures, which were identified as a promising ligand class.[69a] Pàmies, Diéguez and co-workers have screened a library of 64 diphosphite ligands, using parameters such as the ligand backbone, the length of the bridge and the substituents of the biphenyl moieties for the hydroformylation of heterocyclic alkenes.[69c] Their findings can be summarized as follows: i) diphosphites with 3 carbons in the bridge provide higher regio- and enantio-selectivities than ligands with 2 and 4 carbons; ii) increasing the rigidity of the ligand is beneficial; iii) substitution at positions adjacent to phosphite moieties is also beneficial; iv) bulky substituents in the *ortho* and *para* positions of the biphenyl moieties are needed for high enantioselectivities. However, these studies do not provide a clear insight into the factors governing enantioselectivity.

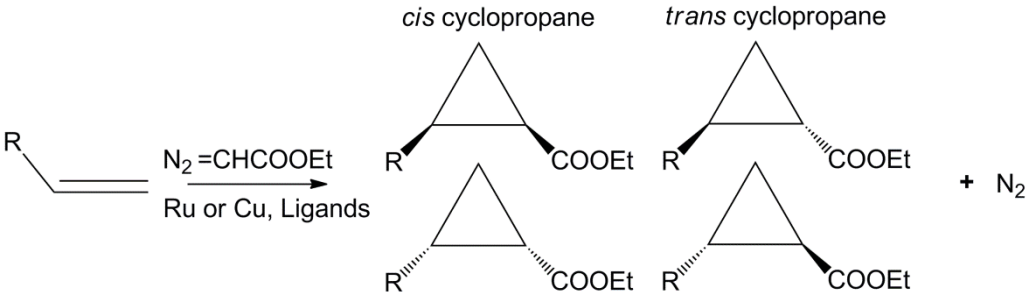
Some computational studies on asymmetric hydroformylation have been performed using different levels of calculation.[70] Hermann and co-workers made a first contribution to the theoretical description for binaphos[70a] and a series of C_2-symmetric diphosphane ligands,[70b] using semiquantitative *quantum mechanics and then molecular mechanics*. Then, our research group provided the first quantitative theoretical assessment of enantioselectivity for AMPP ligands, using hybrid quantum mechanics / molecular mechanics (QM/MM) methods.[70c] Based on that finding, Pizzano and co-workers rationalized the observed enantioselectivity in phosphane-phosphite ligands.[70d] They proposed that π-π stacking interactions between the phosphane group in the apical position and vinyl arene substrates govern the enantioselectivity. More recently, Zhang and co-workers have also performed semiquantitative MM calculations on the yanphos system for rationalizing the asymmetric hydroformylation of allylic compounds.[66b] All these theoretical studies use transition state-based approaches, assuming that the alkene insertion into the Rh-H bond is the enantioselectivity-determining step. This hypothesis should be applied with caution since alkene insertion can be reversible or irreversible depending on the catalyst system, and/or the experimental conditions. Also, purely qualitative models have been proposed from structures mimicking the alkene insertion transition state (TS).[54b] On the basis of these models and kinetic data, Landis and co-workers proposed a two-dimensional quadrant steric diagram for the qualitative rationalization of enantioselectivity for the Rh-(*S,S,S*)-bisdiazaphos catalyst.[71]

1.1.3. Asymmetric cyclopropanation of alkenes

The asymmetric cyclopropanation of alkenes with diazoesters catalyzed by chiral transition-metal complexes enables cyclopropanes compounds to be obtained enantioselectively (see Scheme 1.3).[72] The reaction generates two diasteroisomers which have two stereogenic centres. Thus, cyclopropanation gives rise to the question about control of the diastereoselectivity and enantioselectivity, depending on the structure of the carbene and the alkene. The most often used diazo compounds are those containing a carbonyl group such as

diazo acetates (N_2=C(H)COOEt) because they are stable up to 100 °C, and the catalyzed reaction occurs at considerably low temperatures.

Scheme 1.3.



Catalysts based on copper complexes are interesting because of their high efficiency,[73] and their low cost compared with other catalysts such as those based on rhodium or ruthenium complexes.[74] In fact, the use of chiral copper salicylaldimine complexes as catalysts in the reaction was reported as this first example of enantioselective reaction under homogenous conditions.[75] Since then, other chiral ligands have been used, including the widespread family of oxazoline-based ligands, which showed the best results. Figure 1.4 displays several types of skeletons which form the family of oxazoline-based ligands; bis(oxazoline) (Box), azabis(oxazoline) (AzaBox), quinolinyl(oxazoline) (Quinox), pyridyl(oxazolines) (Pyox).[76]

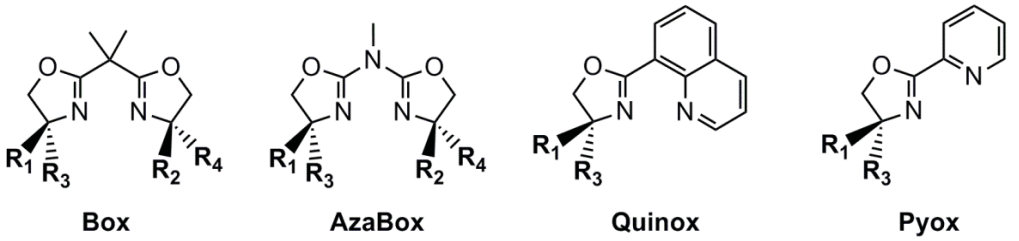


Figure 1.4. Ligands used for enantioselective C-H carbene insertion.

There are several experimental[73,77] and theoretical studies[74,78-79] on the mechanism of asymmetric cyclopropanation of alkenes with diazo compounds catalyzed by copper complexes. The theoretical studies included systems such as bipyridines ligands in neutral copper-complexes[77e] or bis(oxazoline) ligands in cationic copper-complexes.[78a] Thus, the catalytic cycle and the origin of enantioselectivity are now reasonably well established. The reaction proceeds via a Cu(I)-carbene species, which is formed from the diazo compound with extrusion of nitrogen. Some copper-carbene complexes have been detected experimentally and even isolated; and their role as reaction intermediates in cyclopropanation reactions has been unequivocally established.[80] The copper-complex formation is the rate-limiting step of the reaction; however, the step that controls the stereochemistry occurs later in the catalytic

cycle. Two different mechanisms for the addition of alkenes copper-carbene complexes to give cyclopropanes were originally proposed,[81] namely the *concerted* and the *stepwise pathway* involving a metallacyclobutane intermediate (Figure 1.5). Isotope effects[77c] and Hammett-type studies on cyclopolypyrazolylborate and bis(oxazoline)-copper(I)[78a] catalysts are consistent with a concerted but asynchronous mechanism for the cyclopropanation step. Likewise, the computational results show that the *concerted pathway* is favoured over the stepwise *pathway*; the direct addition of the copper-carbene species to the alkene being the step that controls the stereoselectivity.[74,78a]

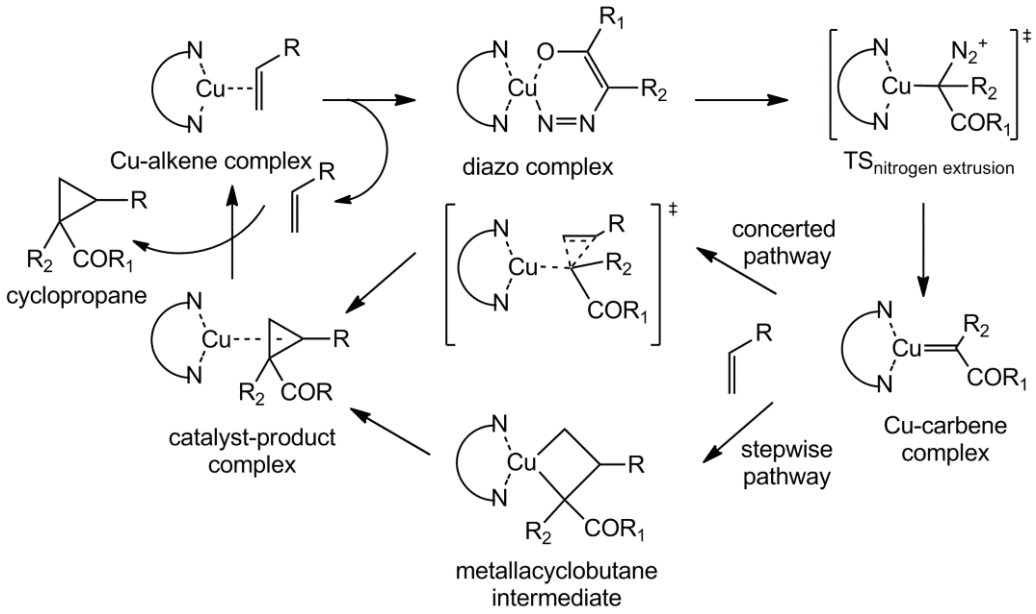


Figure 1.5. Proposed mechanisms for the asymmetric cyclopropanation reaction of alkene with diazo compound.

Modeling this reaction can become very difficult because the PES is rather flat in the area surrounding the TS of the alkene addition to Cu-carbene. In fact, some authors have found that there is not enthalpic barrier on the corresponding PES, and that the activation barrier can only be detected on the estimated *Gibbs Free-Energy Surface* (GFES), which is extremely difficult to calculate for complex systems.[78a-b] Maseras, Moreno-Mañas and co-workers did not find the transitions states in the *Potential Energy Surface* on copper-catalyzed cyclopropanation of $Ph_2C{=}CH_2$ alkene.[78b] Therefore, the authors used a procedure based on the definition of reaction paths in an approximate *Gibbs Free-Energy Surface*. Nevertheless, this procedure is time consuming and less accurate than the usual approaches based on electronic energy surfaces.[78b] Later, García and co-workers have shown that the existence or absence of the TS is highly dependent on the substrate type and on the theory level employed.[78c]

1.2. Tools for the computational modeling of homogeneous catalysis

In this section, the computational strategies frequently used in the modeling of homogenous catalysis are discussed. We have divided these strategies into two types. The first approach involves determining the key transition states (TS) of the mechanism using electronic structure methods, and it is frequently referred to as *TS-based* approach. The second approach tries to account for catalytic activity and selectivity from ground-state structures, defining *molecular descriptors* and setting *qualitative* or *quantitative* relationships with the catalytic properties. The latter approach comprises the so-called *Quantitative Structure-Property Relationship* (QSPR) methods. As mentioned above, the sequence of objectives of a modeling technique is to reproduce the experimental results, and to rationalize the behaviour of the catalysts to obtain the criteria for design and predict the new catalysts. In principle, it is possible to use both strategies to achieve these objectives. For QSPR approaches, the time-consuming determination of the catalytic mechanism as for TS-based approaches is not necessary, but a set of uniform experimental values are required.

1.2.1. Transition state-based approaches

The homogeneous catalysis occurs through a sequence of reactions which constitutes the catalytic cycle. Thus, within the TS-based approach, the theoretical study of this field requires the determination of the geometries and the energies of the reactants, products, and all relevant intermediates and transition states. In other words, it is concerned with the localization of the stationary points of the *Potential Energy Surface* (PES) and their energy evaluation. Optimization of geometries is usually done through gradient techniques. In the procedure, all degrees of freedom are varied simultaneously until the gradient (first derivates) of the energy is zero. The characterization of the stationary points involves the differentiation between the local minima (intermediates, reactant, and products) and the saddle points (transition states), by computing the matrix second derivatives of the energy with respect to the molecular coordinates (Hessian matrix). In the case of local minima all the eigenvalues or the Hessian matrix are positive, whereas in the saddle points, there is one and only one negative eigenvalue. Based on this information, it is possible to identify the activity- and the selectivity-determining step of the catalytic cycle by evaluating the different relative energies and the energy barriers.

The analysis of the rate-determining transition state can provide information about the factors governing the catalytic activity. However, in multistep processes with species of comparable energy, several transition states and intermediates could contribute to the catalytic turnover. In such cases, it is not straightforward to calculate the efficiency of a catalytic cycle, that is, its turnover frequency (TOF) from a theoretically obtained energy profile. To obtain the catalytic TOF from the overall energy profile, Kozuch and Shaik proposed a more sophisticated methodology, the *Energetic Span Model*, which is based on Eyring's Transition State Theory (TST) and corresponds to a steady-state regime.[82] To computationally assess the key factors governing selectivity, it is common practice to compute the relative energies of the different transition states for each of the pathways, and then to analyze the ligand-substrate

interactions at the TS. Figure 1.6 shows an example of a hypothetical process of asymmetric catalysis with two competing diastereomeric pathways. Computing the relative energies of the highest-energy transition states of the pro-*R* and the pro-*S* path provides access to the enantioselectivity values of the catalytic process. This procedure assumes that the intermediates interconvert rapidly, fitting in with the Curtin-Hammett principle. The criterion needs to be modified when the reaction does not fit in with the Curtin-Hammett conditions. Finally, it is interesting to note that the rate-determining step (usually associated with the highest energy TS) does not need to be the same step as the selectivity-determining step, but they frequently coincide.

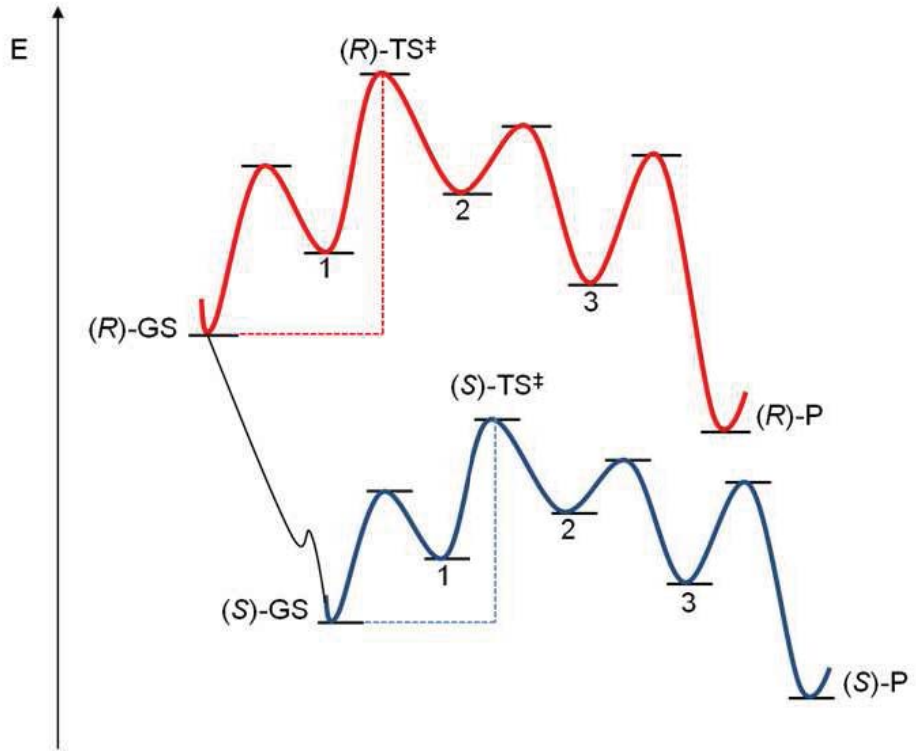


Figure 1.6. Energy profiles of a typical process of asymmetric catalysis with one pathway leading to the *R* product and the other to the *S* product.

The study of reactivity involves bond-breaking/bond-forming processes, and therefore in principle, it requires the use of quantum mechanics (QM) methods, in which the electrons are treated explicitly. The development of density functional theory (DFT) methods,[83] introducing electron correlation with a low computational cost, has made it possible to obtain reliable structures and energies for medium-sized molecular systems. Nowadays, DFT methods have become the standard choice for the study of transition-metal reactivity and homogenous catalysis.[84] Larger systems are amenable to the hybrid quantum mechanics/molecular mechanics (QM/MM) methods.[85] The QM/MM method divides the system into two regions.

The active site region is described by the QM method and the bulky conformationally flexible regions are described classically by the less computationally demanding MM method. The most popular implementations of QM/MM treatment for organometallic-type catalysts are the IMOMM method[85d] and its extension, the ONIOM method.[85e] Within this partition, there is no charge transfer from the MM part, thereby resulting in a ligand with symmetric electronic properties but its steric properties are accounted for. In addition to these QM methods, some specific MM-based approaches have been employed to determine transition states and computationally access selectivity, such as the QM-guided molecular mechanics (Q2MM) method proposed by Norrby.[86] In Q2MM method, the TS is modeled as a minimum, allowing for the application of standard MM tools such as minimization and conformational searching. The negative eigenvalue corresponding to the TS vector is assigned into an arbitrarily large positive value which turns the "TS" optimization into a minimization (see Figure 1.7). Definition of the force field is based on model QM calculations, which is the time bottle-neck of the procedure. Another interesting MM-based approach is asymmetric catalyst evaluation (ACE) which treats the approximate transition state as a "real" species, with partial bonding between the reacting centres.[87] ACE is a molecular-mechanics-based independent program that has been developed from simple organic chemistry principles. ACE constructs TSs from a linear combination of reactants and products, including a factor λ describing the position of the TS on the *Potential Energy Surface* (TS = (1-λ) reactant + λ product).

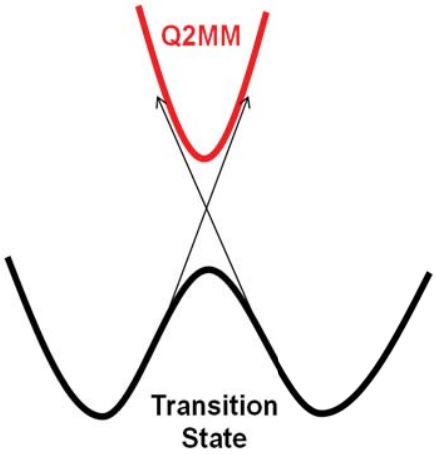


Figure 1.7. Schematic representation of the Q2MM TS force-field approach.

Today, these computational approaches can obtain results at accuracies that make it possible to understand and predict the roles of homogeneous catalysts. Even in the challenging asymmetric catalysis, where the energy differences can be as low as 1-3 kcal.mol^{-1}, many reports have demonstrated the accuracy of stereochemical predictions using Quantum Methods (QM).[88] However, there have been only moderate successes in the *a priori* computational design of effective catalyst either in achiral or asymmetric catalysis.[89] In addition, the use of TS-based approaches has some disadvantages and it can be of difficult

application in some situations. Most of the popular functionals, such as BP86, B3LYP or PBEh, are unable to describe noncovalent π-π interactions in their attractive regime.[90] For example, in the previous contribution of our research group to asymmetric hydroformylation based on TS determination, full DFT calculation were unable to describe the arene-arene interactions involved in the process, whereas QM/MM calculations overestimated the interactions.[91] Recently, Truhlar and co-workers recommended the M06-class of functionals for the study of noncovalent interactions.[92] Also, explicit dispersion corrections have been empirically added to conventional functionals (DFT-D) such as in B97D.[93]

The location of a stationary point in the *Potential Energy Surface* (PES) with the topology of TS presents intrinsic difficulties because the point is a maximum in the reaction direction and a minimum in all the other coordinates. It is not straightforward for implementing the double criteria in a mathematical algorism. The lack of experimental and sometimes computational references makes it difficult to generate an initial guess for the search relying in chemical intuition and knowledge of the process. Additionally in some cases, the PES is rather flat in the area surrounding the TS, making location very difficult. Even so, we can face the extreme case that the TS exists in the Gibbs free energy surface and it does not in the electronic energy surface, for which analytical gradients and second derivatives are available.

In general for finding the catalyst structure, an additional difficulty is the large number of conformations that the ligand can geometrically arrange. The conformational variability problem has not been yet completely solved. One possible approach is the use of molecular dynamic (MD) simulations, as a method for exploring conformational space. Another limitation of the TS-based approach is that the relative rates of individual reaction steps are influenced by ligand properties employed as catalysts in the reaction. A small change in the ligand can alter the behaviour of the catalyst and consequently the rate- or the selectivity-determining step can change. Thus, each specific catalyst can have its own mechanism; and sometimes, we are forced to handle multiple kinetic scenarios simultaneously.

1.2.2. Molecular descriptors

In homogeneous catalysis by transition metal complexes, modifications of the coordinated ligands are often used to achieve an improvement in the activity and selectivity of the process.[94] Thus, even before the use of computational methods on the study of homogenous catalysis, researchers attempted to quantify the effects of the ligands in order to rationalize and predict the catalytic outcome. Ligand properties are roughly classified into electronic and steric effects, which are considered as the standard concepts in the chemical language. The qualitative and quantitative understanding of these ligand parameters is crucial for rational ligand design. The electronic effects arise from the variations in electron density through bonding or spatial interactions and include inductive effects, conjugative effects, or orbital symmetry considerations. Steric effects usually refer to repulsions originated as atoms or molecular fragments which approach each other, although the approaches can also lead to attractive interactions. Commonly, steric effects are modelled by the Lennard-Jones potential, which includes a repulsive and an attractive term for nonbonding interactions. At this point, it

is important to note that the separation between steric and electronic effects is not always straightforward, and any proposal for separation between both effects is as arbitrary as any other. Experimentally, it is difficult to vary one ligand parameter without affecting the others. Interestingly, computational tools allow the separation, identification and quantification of the different effects of the ligands on the structure and energies.

Many efforts have been devoted to quantifying electronic and steric effects through defining what are known as molecular descriptors.[95] A *descriptor* is a mathematical formula which quantifies the different parameters of the ligand or catalyst structure. Todeschini and Consonni defined molecular descriptors as *the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment.*[96] Descriptors can be structural such as bond lengths and Tolman's cone angle[97] (see bellow), related to energy such as interaction energies and proton affinities,[98] or obtained from spectroscopic measures such as infrared stretching frequencies and NMR chemical shifts. They can be experimentally measured or calculated with computational methods, the latter being useful for experimentally inaccessible or novel structures. In many cases, a univariate or a multivariate mathematical relationship can be found between the experimental variables and 1D/2D or 3D type descriptors that promoted the use of the Quantitative Structure-Property Relationships (QSPR) modeling.

The following sections include a summary of the most popular and significant steric and electronic descriptors proposed within the field of homogeneous catalysis, as well as their transferability to different ligand types, and their potential application for the design of new catalysts though the definition of qualitative and quantitative relationships.

Steric descriptors

One of the first steric descriptors was proposed by Taft.[99] This experimentally obtained descriptor is a geometrical parameter which measures the bulk of several functional groups. In the original studies, the relative rates of methyl ester hydrolysis were measured for a wide range of aliphatic esters in acid or basic conditions (Figure 1.8). The Taft equation shown in Figure 1.8 measures the steric influence of substituents on the rate of hydrolysis, where k_s is the rate of ester hydrolysis for a given substrate, k_{CH3} is the methyl acetate hydrolysis rate, which is used as the standard, δ is a proportionality constant which is a measure of sensibility, and E_s is the *Taft steric* parameter. If the hydrolysis is conducted under acidic conditions, the reaction involves a positively charged intermediate formed from a positively charged precursor (Figure 1.8). Therefore, the electronic inductive and conjugative effects should be minimal, and the rates should mostly be influenced by the steric effects of the substituents. Nevertheless, this hypothesis was controversial. In order to improve the *Taft steric* parameter as a pure measure of steric effects, Hancock[100a] and Palm[100b] introduced hyperconjugation corrections, and Charton[101] correlated the *Taft steric* parameters with the van der Waals radius of the corresponding groups, ruling out both inductive and resonance effects and defining new values.[102] Charton presented the equation $log(k/k_0)=\psi\nu$, where the *log* of the

relative rate (k/k_0) is proportional to the product of ψ (sensibility factor) and ν is the adjusted Taft value based on the van der Waals radius of the given substituent. These types of descriptors are easily accessible from tabulated values, but they are only defined for some specific functional groups and cannot be applied to more complex ligand architectures.

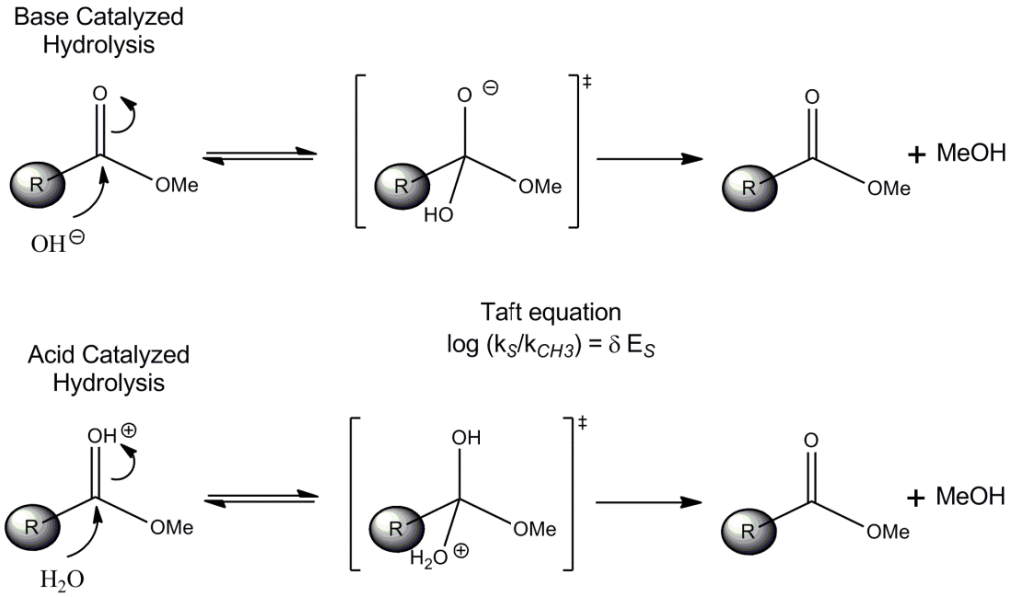


Figure 1.8. Correlations of methyl ester substituents to the rate of hydrolysis in the development of *Taft steric* parameter.

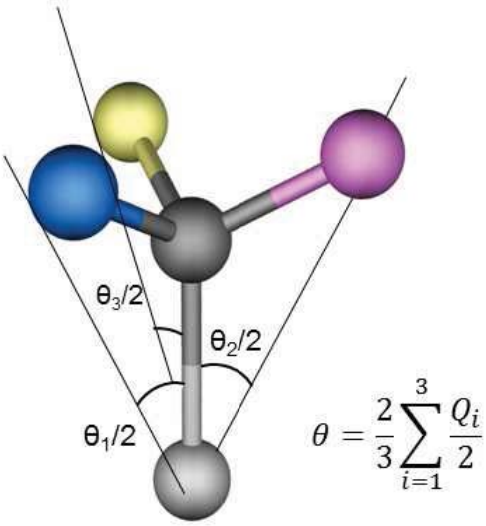


Figure 1.9. Representation of *cone angle* for monophosphane ligand.

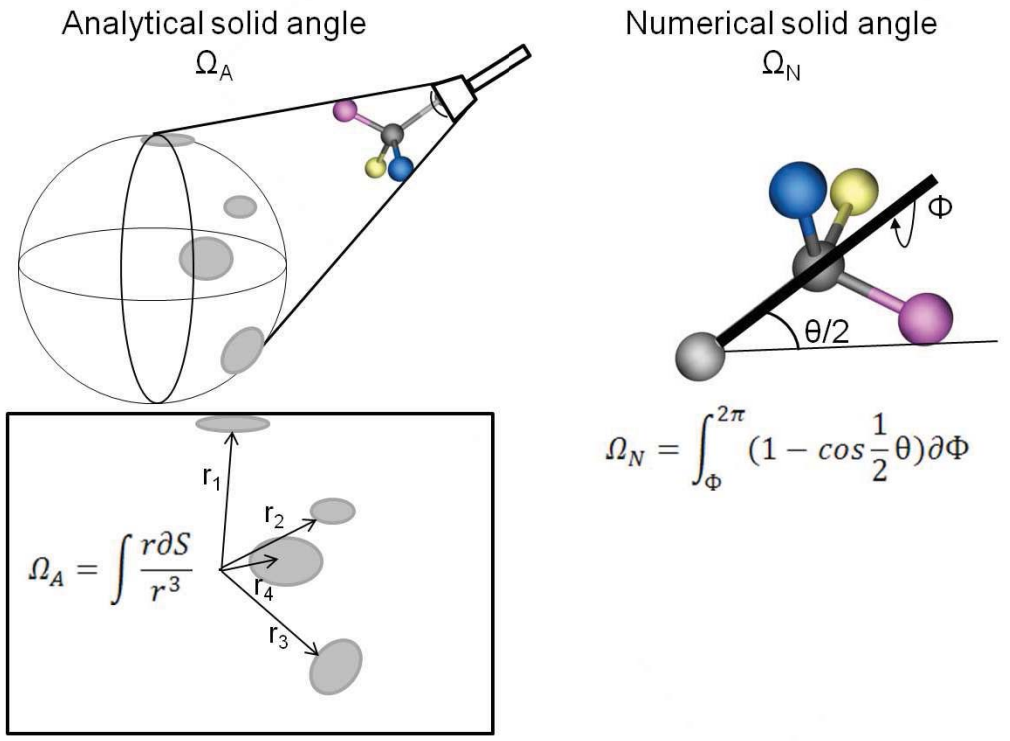


Figure 1.10. Representation of analytical (Ω_A) and numerical (Ω_N) *solid angle* for the phosphane ligand.

In organometallic homogeneous catalysis, descriptors have been developed for different ligand classes including mono- and bidentate P-donor ligands, carbenes, substituted metallocenes and mixed bidentate P,N-donor ligands. One of the most popular steric parameters for monophosphane ligands is the *cone angle* (θ), originally proposed by Tolman.[97] Tolman simply sought to measure the smallest possible cone angle of the ligand using the simple formula which consists of the sum of the half-vertex angle for *3* unsymmetrical groups in a ligand, multiplied by two (Figure 1.9). The θ has been measured from ligand structures determined by X-ray crystallography,[103a] crystal data in the Cambridge Structural Database (CSD)[103b] or calculated geometries.[103c-d] Although the θ was developed especially for phosphorus ligands, the values of θ have been successfully used for ligands such as cyclopentadienyl.[104] A related development for evaluating ligand size is the *solid angle* (Ω) proposed by White and co-workers, which can be measured using crystallographic data or calculated structures.[105] The *solid angle* distinguishes methods between two of calculation, the *analytical solid angle* (Ω_A) and the *numerical solid angle* (Ω_N). The Ω_A is defined as the surface area of that part of a solid projected onto the surface of a unit sphere (Figure 1.10). The analytical method correctly quantifies the shadow of individual atoms but it does not take into account the possible overlap between the atoms. The Ω_N is the angle formed by non-circular cone traced around the outer van der Waals radius of the ligand being measured (Figure

1.10).[106] The cone is measured by rotating a fixed axis through ϕ degrees, by way of measuring the semi-cone angle, θ/2, around the group. Unfortunately, the θ and Ω parameters may suffer from conformational noise in different coordination environments leading to a range of values when different crystal structures are considered. Computationally, it may be necessary to explore multiple conformers before using calculated geometries of ligands for the derivation of θ and Ω, and it is worth bearing in mind that conformational preferences can change in response to different coordination environments, and therefore a range of *cone* and *solid angles* may be more useful than a single value.

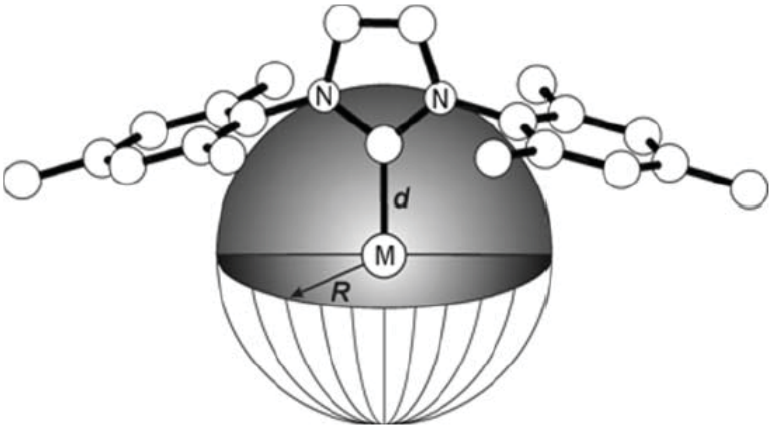
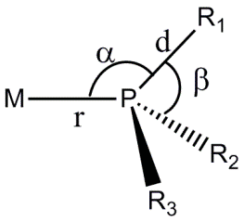


Figure 1.11. Representation of *percent of buried volume*, $\%V_{Bur}$, for N-heterocyclic carbene (NHC) ligands. Figure was reproduced from reference 108b.

Conformational issues aside, the θ and Ω described reasonably well when the ligand was close to spherical symmetry such as C_3-symmetrical P-donor and C_5-symmetrical metallonece ligands,[97,103-106] but are less suitable for ligands with symmetry such as C_2-symmetrical carbene ligands.[107] For C_2-symmetrical carbene ligands, the *percent of buried volume* ($\%V_{Bur}$) has been described.[107-108] The $\%V_{Bur}$ descriptor measures the volume amount of a sphere centered on the metal that is occupied by the atoms of the ligands (see Figure 1.11). Apart from N-heterocyclic carbene (NHC) ligands,[109] the $\%V_{Bur}$ has been used to qualitatively rationalize trends in the phosphane ligands,[108] metallocenes, phenoxy-amine and phenoxy-imine.[110] The $\%V_{Bur}$ value is very sensitive and the choice of the sphere radius and metal-ligand distances makes it difficult to determine a single set of values and extension to other ligands. To solve this problem, the authors have tried to standardize calculations by implementing a simple web interface.[108b]

The S_1', S_4'and S_7' descriptors, called the *symmetric deformation coordination*, are other simple structural parameters either derived from X-ray crystallography databases[111] or from computational chemistry-based geometry optimizations[112] (see Figure 1.12). For example, the

S_4' is calculated as the difference between M-P-R and R-P-R angles in coordinated phosphane ligands. The value of S_4' is inversely correlated with the ligand size as measured by the *cone angle*.[111] This descriptor can again change in response to conformational flexibility of ligands.[112] Unfortunately, S_4' is unsuitable for use with C-donor ligands and the S_4' parameter has only been used with phosphane ligands.[98a,111-112]



$$S_1' = r$$

$$S_4' = \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \beta_i$$

$$S_7' = \sum_{i=1}^{n} d_i$$

Figure 1.12. Representation of *symmetric deformation coordinate* descriptors (S_1', $S_4'$and S_7') for phosphane ligands.

The steric descriptors described thus far are often insufficient for estimating the steric effects of bidentated ligands. Casey and Whiteker introduced the concept of the *natural bite angle*, as a characteristic of chelate ligands (see Figure 1.13).[113] The *bite angle* is the angle formed by the ligand atoms bonded to the metal (P-M-P angle).[114] It can be determined from suitable X-ray structures,[115] but originally,[113] was derived from mechanics molecular calculations. The calculations fixed the phosphane donor atoms at a constant distance from a dummy atom, ensuring that the P-M-P angle is a function of the ligand backbone preferences and not a consequence of metal-ligand interactions. The van Leeuwen´s group has shown its predictive utility for ligand design in a broad range of reactions.[115-116] Actually, the bite angle generates two kinds of effects: firstly, steric effects related with the ligand-ligand and ligand-substrate nonbonding interactions, and secondly, electronic effects induced at the metal centre (see below). Also for bidentate ligands, other approaches have been defined such as the *pocket angle* concept[117] or the *accessible molecular surface.*[118]

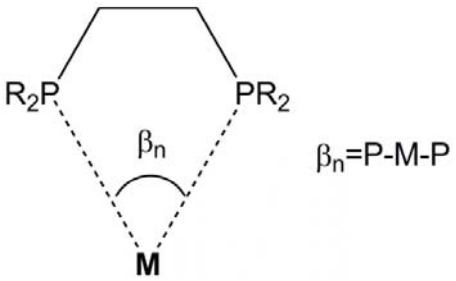


Figure 1.13. *Natural bite angle* (β_n).

Besides the steric descriptors based on structural measures, other steric parameters derive from electronic and energetic calculations by computational methods. The *Molecular Electrostatic Potential*[119] (MESP) approach combines QM calculations in model systems (*e.g.*: PH_3) and QM/MM calculations in real systems (*e.g.*: PR_3), separating the steric (S_{eff}) and electronic (E_{eff}) effects. Both parameters were described by the *values of the molecular electrostatic potential minimum* (V_{min}), corresponding to the lone pair region of several substituted phosphane ligands. The V_{min} is calculated from electron density distribution $\rho(r)$ by employing the equation 1.1, where Z_A is the charge on nucleus A, located at R_A. The S_{eff} is measured as the difference between $V_{min}(PH_3)$ and $V_{min}(QM/MM\_PR_3)$. In principle, the S_{eff} can be applied to any ligands, but it has only been used with P-donor ligands.

$$V(r) = \sum_{A}^{N} \frac{Z_A}{|r - R_A|} - \int \frac{\rho(r')d^3r'}{|r - r'|} \qquad (1.1)$$

The quantification of the steric effects of a ligand can also be performed from the ligand repulsive energies with sterically demanding fragments. The *ligand steric repulsive energy*, E_R, is based on molecular mechanics calculation of the $[M(CO)_5(L)]$ complex.[120] The energy-minimized structures employ the van der Waals repulsive force between the ligand and [M(CO)] fragments (equation 1.2).

$$E_R = \frac{\partial E_{VdW}(repulsive)}{\partial r(M - L)} r_e \qquad (1.2)$$

E_R values correlate well with the *cone angle* for all phosphane, phosphite and arsenic ligands investigated.The E_R allows for comparison of the ligand size across different classes on a single scale. The E_R parameters have been applied to mainly P-donor ligands[120a-b] and extended to other monodentate ligands.[120c-d] Other steric parameters which are calculated from interaction energies are the *interaction energy between ligand and ring of 8 helium atoms*[98a-b] (He_8_steric) and the *interaction energy between ligands and wedge of 8 helium atoms*[121] (He_8_wedge). The He_8_steric have been applied to monodentate ligands such as P-donor and C-donor ligands,[98a-b] and the He_8_wedge for the chelating ligands such as P,P-donor and P,N-donor ligands.[121] In general, the parameters derived from energy calculations capture detailed ligand steric properties, but such calculations may be time-consuming and too complicated for widespread acceptance by non-specialists.

Electronic descriptors

One of the most popular electronic parameters is the *Tolman Electronic Parameter* (TEP).[97] This parameter provides information about the net donor/acceptor character of the ligand. Tolman defined this electronic parameter for phosphorus ligands based on the highest CO stretching of A_1 symmetric in $[Ni(PR_3)(CO)_3]$ complex using the $P(^tBu)_3$ ligand as reference. When the ligands have mixed substitutions, it is also possible to assess their individual contributions (χ_i) using a simple formula (equation 1.3), where 2056.1 is the frequency of

$P(^tBu)_3$ and χ_i is the individual contributions, provided the frequencies of symmetrically substituted ligands are known.

$$TEP = 2056.1 + \sum_{i=1}^{N} \chi_i \quad (1.3)$$

The better the donor power of PR_3 ligand, the lower the CO stretching frequency is, as a result of the larger back-donation into CO π* orbitals. For popular NHC ligands, the TEP values in Ni complexes showed little responsiveness to changes in the substitution pattern. In this case, the $[IrCl(CO)_2L]$complexes were identified as the suitable carbonyl complexes for the determination of electronic properties of carbene ligands.[109c] Similarly, bidentate ligands have been compared experimentally in molybdenum complexes.[122] However, the extension of the TEP concept for bidentate ligands computationally is still missing.

The TEP has a few limitations, such as the toxicity of a common precursor $[Ni(CO)_4]$ or the instability of $[Ni(CO)_3L]$ complex with important ligands such as pyridine derivates, NO, H_2O, or H_2S. As an alternative, DFT frequency calculations can be used to derive a *Computed Electronic Parameters* (CEP),[123] showing a linear relationship between the calculated and the experimental CO frequencies. Similarly, Cundari and co-workers have used semi-empirical methods to derive related electronic parameters (SEP).[112] Both computational approaches extended the ligand family beyond the P-donor set including N-donor, anionic and cationic ligands.[112,123] Recently, Gusev compared and ranked a large group of two-electron ligands with the help of DFT calculations.[124] The author proposed as parameters the CO and C=C frequencies in iridium and osmium complexes, as well as bond and exchange enthalpies.

An alternative to TEP the *Lever Electronic Parameters* (LEP), which is based on the electrochemical values (E_0) for various redox couples, notably Ru(III)/Ru(II), in series of complexes containing the ligands investigated.[125] It assumes that the M(n)/M(n-1) redox potentials are additive for all ligand contributions. LEP is believed to measure the net donor power (both σ and π effects) of ligand to metal fragment. For example, in the $[Ru(bipy)_nL_{6-2n}]^{m+}$ (n = 0-3; bipy = 2,2'-bipyridyl) complex and using the potential of the Ru(III)/Ru(II) redox couple (E_{obs}), the LEP value for the ligand L derived from the equation 1.4. One advantage of LEP compared with TEP is that the set of ligands can be greater, including anionic ligands and O-donor and N-donor ligands. However, one limitation is that the measurement requires the use of an electrochemical apparatus, which is less accessible for experimentalists than IR spectroscopy. In addition, the electrochemical process is irreversible and can degrade the quality of the data for certain ligands; and the solvation and ion-pairing effects might interfere to some extent. From a computational point of view, the calculation of a TEP is easier than the computation of an LEP, because the solvent effects and open shell system are difficult to model.

$$E_{obs}(Ru^{III}/Ru^{II}) = 0.51n + (6 - 20)LEP \quad (1.4)$$

As mentioned in the previous section, the MESP approach[119] can be used to measure both the *steric effects* (S_{eff}) and *total electronic* effects (E_{eff}). The E_{eff} is the difference between V_{min}(QM/MM_PR_3) and V_{min}(QM_PR_3), the R groups being included in the part MM. Electronic properties can also captured by other calculated parameters such as the *energy of the highest occupied molecular orbital* (E_{HOMO}), the *lowest unoccupied molecular orbital* (E_{LUMO}) and ligand *proton affinity* (PA).[98a-b,121] The energy of the HOMO in the free ligand usually corresponds to a lone pair and can be considered an indication of ligand donor ability. The PA is calculated as the difference between the energy of the neutral and the protonated single ligand and it was also related to the *σ*-bonding properties of the ligands. These descriptors have been applied to P-donor and C-donor ligands, giving good results.[98a-b,121]

The quantitative separation of electronic effects in terms of σ and π contributions is difficult, as only the net effect on metal-ligand interactions can be measured directly. There are indirect techniques, in which the interaction of interest is isolated from any others as part of the quantum chemical treatment. In general, these techniques are based on stepwise processes to calculate the interaction energy between the isolated ligand and the metal fragments. The *Energy Decomposition Analysis*[126] (EDA) method, the *Extended Transition State*[127] (ETS) model and the *Constrained Space Orbital Variation*[128] (CSOV) method and the *Natural Orbitals for Chemical Valence*[129] (NOCV) method have attempted to decompose the metal-ligand interaction in terms of σ/π and donation/back-donation contributions. The EDA, ETS, and CSOV methods decompose the interaction energy, whereas the NOCV method provides an approximate valency associated with each type of orbital interaction. Another approach is *Natural Bonding Orbital*[130] (NBO) analysis, and in particular the use of the perturbation theory to quantify secondary orbital interactions. The roles of σ and π bonding in the variations of physicochemical properties have been also evaluated empirically in terms of the *quantitative analysis of ligand effects*[131] (QALE) model. In *Chapter 3*, we have used a modified version of EDA method based on orbital deletion[132] to separate the σ donation and π back-donation effect of the phosphane-phosphite binaphos ligand in Rh catalysts.

Finally, it is to be noted that in many cases, the parameters capture both the steric and electronic contributions and it is difficult to classify them as steric or electronic descriptors. One example is the *bite angle*, which is one of the most frequently used parameters for chelating ligands (see Figure 1.13).[114] Apart from nonbonding steric effects and ligand bulkiness, the bite angle determines the metal hybridization and this in turn determines the metal orbital energy.

1.2.3. Qualitative analysis for ligand design

Once the steric and electronic descriptors have been presented, the next question to be answered is how the molecular descriptors can help us understand the catalyst behaviour and assist ligand design. This analysis can be classified into two main approaches; the qualitative and the quantitative analysis. The qualitative analysis can be used to explore ligand properties, in terms of a single parameter or by deriving ligand maps from suitable descriptor scatter plots. Bivariate correlations and simple scatter plots between descriptors and suitable

response data such as structural features, thermodynamic, kinetic or spectroscopic results can be used to establish the importance of the steric and electronic contributions to observed chemistry.[95b-97]

The combination of steric and electronic descriptors can be used to map ligand space, and experimental data can be projected onto such maps, as originally shown by Tolman[97] and developed further by Cundari[112] and Suresh.[119] The modeling by these two-dimensional *stereo-electronic maps* assumes that proximity between two ligands implies a similar catalytic behaviour. This can be useful in the design of ligands by targeting a specific area, which is known to contain active catalysts, by designing a broad and varied ligand set to sample ligand space comprehensively, or by identifying depopulated areas, in which novel ligands with new properties may exist. Suresh and co-workers proposed stereo-electronic maps based on MESP[119] type descriptors, which can be divided into steric (S_{eff}) and electronic parameters (E_{eff}) (see above), leading to a general classification of the ligands into categories. For example, PR_3 ligands were classified into four categories; i) ($+E_{eff}$, $+S_{eff}$), ii) ($+E_{eff}$, $-S_{eff}$), iii) ($-E_{eff}$, $+S_{eff}$), iv) ($-E_{eff}$, $-S_{eff}$). Cundari and co-workers developed stereo-electronic maps for characterizing phosphanes and phosphates using semi-empirical quantum mechanics (SEQM) methods.[112]

Using DFT calculated ligand and ligand-complex descriptors, Harvey and co-workers have developed *ligand knowledge bases* (LKBs) or *maps of the ligand space*.[95b,98] The descriptors include electronic parameters such as E_{HOMO}, E_{LUMO} and PA,[98] TEP[97] and CEP,[123] or steric parameters such as the bite angle or the He_8-wedge,[98c] and chemically relevant information such as metal-ligand bond lengths or changes in ligand geometry upon complexation.[95b,98] The multiple-computed descriptors generate *maps of ligand space* via *Principal Component Analysis*[133] (PCA) method.

The PCA approach is a projection technique, which makes it possible to reduce the number of variables in a correlated dataset by deriving new variables (principal components) as linear combinations of the original descriptors, optimized to maximize the information content in fewer dimensions and uncorrelated themselves. This reduction is performed by identifying the directions (called principal components) along which the variation in the data is maximal. Theoretically, the number of the observed variables is equal to the number of components extracted in PCA. However, in most cases, the first few components alone already account for the majority of the variance, so only these first few components are retained, interpreted, and used (one example is illustrated in Figure 1.14). The PCA components possess two very important characteristics: i) each component accounts for the maximal amount of variance in the observed variables that was not accounted for by the preceding components; and ii) no component is correlated with any preceding component. PCA is an exploratory technique that allows clustering of the compounds related to their chemical nature.

These maps can prove very useful in the qualitative analysis of ligand clustering, identifying ligand similarities and perhaps in intuitive ligand design. Currently the limitation of such an approach is in the availability of experimental data to test and challenge the calculated models and maps.[98] Moreover, these types of works fit perfectly well into a combinatorial design,

since they rely on fast and reliable calculations of parameters that are known to affect the catalytic process of transition metal catalyzed reactions. Using a simple 2D plot, the catalyst designer can achieve the maximum degree of chemical diversity among libraries and therefore enrich the ligand set with new diverse structures. In this way "empty holes" in the experimental space are filled by virtual ligands, and highly populated areas can be reduced by discarding duplicate molecules.

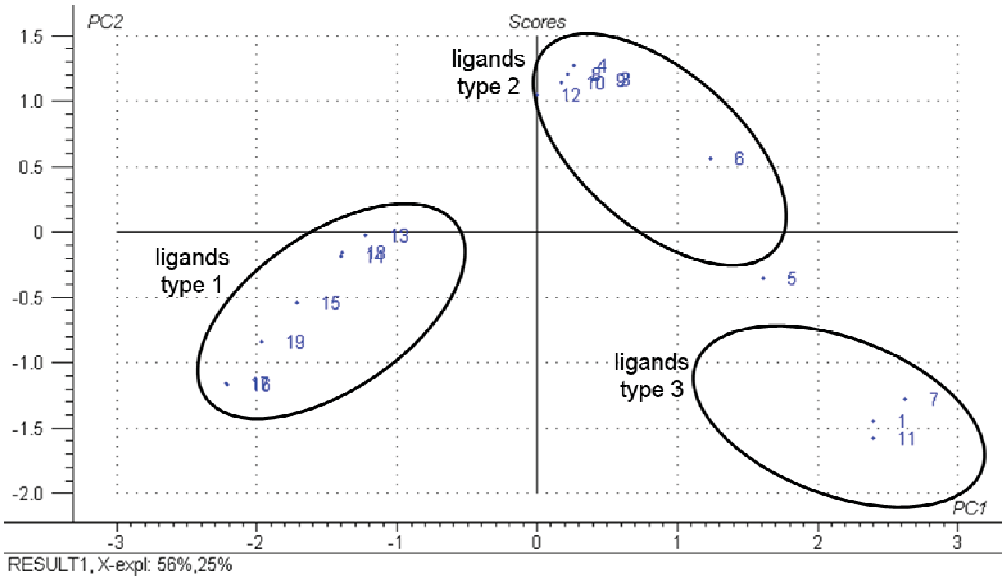


Figure 1.14. Example of PCA analysis. The ligands were classified into three categories; i) ligands type 1; ii) ligands type 2; and iii) ligands type 3.

1.2.4. Linear correlation and QSPR modeling

The establishment of quantitative correlations between diverse molecular properties and the chemical structures is of great technological importance for the development of environmental and medical aspects, and is an emerging tool in homogeneous catalysis. These approaches are referred to as Quantitative Structure-Property Relationships (QSPR) which relate the physicochemical properties of compounds to their structures. The main goal of the QSPR studies is to find a mathematical relationship between the property of interest and one or more descriptors derived from the structure of the molecule. For biological systems, these approaches are usually referred to as Quantitative Structure-Activity Relationships (QSAR) because the biological activity is the main focus; and nowadays, they have become an essential part of drug design. Although QSAR techniques are not commonly used in homogenous catalysis, there are some outstanding examples that correlate catalytic activity and selectivity (see below). Frequently, these studies not only pursue a reliable predictive ability, but also chemically meaningful interpretation. The procedure begins by posing a hypothesis which relates some chemical features of the catalytic system to the experimental

activity or selectivity. In general, the hypothesis is refined by examining structural similarities and differences for active/inactive and selective/non-selective catalysts. Ideally, we should find an optimum quantitative relationship, which can be used for making predictions; and we should end with the proposal of a new catalyst, which maximizes the presence of functional groups or features believed to be responsible for experimental value.

The simplest quantitative models are the univariate linear relationships between the descriptor and the experimental variable. The linear regression model assumes that the response variable is a straight-line function of a single dependent variable. The pioneering work appeared in the field of classical physical organic chemistry, where Hammett[134] and Taft[99] developed the linear free energy relationships (LFER) to get an insight into respectively, the electronic and the steric effects of substituents on chemical reactivity. Since then, many examples of Hammett analysis[135] and approaches to LFER and steric effects[136] have emerged. The development of LFER in inorganic chemistry focused initially on kinetic parameters for reactions such as the nucleophilic attack to *trans*-[$Pt(py)_2Cl_2$] complex.[137] In organometallic chemistry, it was generally admitted that both steric and electronic factors are important in determining the reactivity effects of the ligands. Thus, the most successful approaches actually consisted of multivariate models combining several ligand descriptors such as the QALE[131] approach and the ECW[138] model.

The LFER approaches have also been applied to the study of asymmetric catalysis for finding correlations between molecular descriptors and enantioselectivity. *Taft steric* parameters[99] have been used to correlate enantioselectivity as a function of substrate[139] and ligand substituents size.[140] However, the resulting LFERs showed some limitations in their use as tools for optimizing and designing *de novo* catalysts. To improve the prediction ability of these models, very recently, Sigman and co-workers have reported more sophisticated approaches to the problem.[141] The authors were able to generate multi-dimensional relationships of free energy to substituent size and electronic parameters, including crossed terms that account for interplays between substituents. In *Chapter 5*, we have extended this approach to obtain one of the quantitative quadrant-diagram representations of the catalysts used to predict the enantioselectivity in Cu-catalyzed cyclopropanation of alkenes.

The univariate relationships do not necessarily imply causation. As anticipated above, various property descriptors may contribute to the response experimental variable. Therefore, in many cases, multivariate approaches are required. They usually employ multiple descriptors which are reasonably easy to compute, but often have limited chemical meaning and transferability. Nowadays, the development of software technology enables the rapid calculation of thousands of structural descriptors for a large set of compounds. Building a QSPR model is an inductive process that depends on the set of compound; therefore there is no set of descriptors which can be applied to any type of compounds.[142] In practice, the QSPR methodology can be divided into two main stages: i) representation of the chemical object which is related with the chemical knowledge; and ii) statistical treatment to develop the mathematical equations. The first stage does comprise the selection of the dataset, the generation of molecular geometries, and the calculation and selection of significant

descriptors. The dataset must be informative and representative. In many cases it is not easy to find an available set of experimentally measured properties of sufficient size and diversity, measured under similar conditions with reproducibility and accuracy. Also, the normal distribution of the experimental data is important, thus in some cases, it may be desirable to mathematically transform the original values (*e.g.* transformation of %*ee* in $\Delta\Delta G^{\neq}$ by a logarithmic function). To obtain the molecular structures, one option is to optimize the geometry with MM, semi-empirical, or QM methods, dealing with mixtures of conformational or geometrical isomers. However, depending on the type of descriptors to be calculated, only atomic composition and connection is required. Once the molecular structures are determined, the next step in the QSPR modeling is the generation of descriptors, which can be empirical indices, quantum chemical, electronic, geometrical, constitutional, topological, or three-dimensional in nature.

In general, the QSPR approach generates a multivariate linear equation 1.5, where C_i are the regression coefficients, P_i are the independent variables (the values of descriptor), and *exp.value* is the dependent variable (experimental value). Multiple linear regression is an extension of simple linear regression including more than one dependent variable.

$$exp.value = C_0 + \sum_{i}^{N} C_i P_i \qquad (1.5)$$

Although the linear regression equation is very useful, the "real" correlations are rarely "linear" because most of the QSPR relationships are nonlinear in nature. Nonlinear approaches such as artificial neural networks (ANN) or support vector machines (SVM) have also been employed in order to derive flexible correlation models between the molecular structures and properties. In principle, they are able to "catch" hidden nonlinearities between the property and the descriptors. On the other hand, these nonlinear models are not as intuitively easy to interpret as the linear models.[142]

The most frequently used multivariate statistical methods are the *Principal Component Analysis*[133] (PCA, described above) and the *Partial Least Squares*[143] (PLS) regression. In order to obtain the optimal mathematical equation, a statistical method such as the PLS regression is needed. In the PLS the data analysis is simplified by projecting the data into a low dimensional "latent variable" space (the acronym PLS has also been taken to mean "projection to latent structure"). The components are constructed to maximize the covariance between the dependent variable *y* and the original independent variables *x*. Unlike the PCA analysis, the PLS analysis also simultaneously calculates the latent variables for the two matrices; the matrix of independent (*x*) and the matrix of dependent (*y*) variables, together with the relationship between them. Thus, for PLS the new set of "latent variables" is a set of conjugant gradient vectors to the correlation matrix rather than a set of successive orthogonal directions that explain the largest variance in the data as in PCA.

The QSPR models should be statistically significant and robust, be validated by making accurate predictions for external datasets not used in the model development, and have a defined domain of application. There are techniques for evaluating the predictive power of models such as internal validation, external validation and scramble sets. The internal validation is to predict the independent variable for a compound or a compound group using the regression equation calculated from the QSPR model. The most frequently used internal validation is the Leave-One-Out[144] (LOO) cross-validation, in which the cross-validated correlation coefficient (q^2) defines the predicting ability. In equation 1.6, $\hat{y}_n$ is the observed activity, y_n is the predicted activity of molecule *n*, and $\tilde{y}_n$ is the mean activity value of the dataset. [145]

$$q^2 = 1 - \frac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{\sum_{n=1}^{N}(y_n - \tilde{y}_n)^2} \qquad (1.6)$$

In general, it is considered that the QSAR model has predictive capacity if q^2 is greater than 0.5, but the use of only this criterion is often inadequate.[146] On analyzing such models, chance correlations, noisy variables, and excessively predictive collinearity are frequently the cause of their lack of predictivity. Other statistical parameters are employed for evaluating the predictive ability of models during the fitting and prediction stages, namely: the correlation coefficient (r^2), the standard error, and the slope and intercept.

The external validation is one of the most widely used methods of correlation testing. It divides the dataset into two subsets called training and test subsets. The test subset should be sufficiently large to give a reasonable estimation of the model quality, especially when a random selection was used for its construction. The purpose of the external validation is to evaluate how well the equation generalizes the data. Another validation is the scramble set, which is used to provide a measure of the risk of chance correlations occurring. First, the descriptor values are generated for each compound and properties are assigned at random, giving rise to a new set. The new QSPR model is built taking into consideration the new set; the entire process is repeated 1000 times, and the statistical parameters averages are calculated over the 1000 runs. The purpose of the scramble set is to compare the statistical parameters of the real QSPR model of "true" data with the QSPR models of scramble set and confirm that the best model is the real QSAR model of "true" data. This would indicate that the real QSAR model is not generated at random and the risk of change correlations occurring does not exist.

1D/2D-QSPR models. Applications in homogeneous catalysis

The QSPR approaches can be classified according to the dimensionality of molecular descriptors. The 1D/2D-descriptors do not require the explicit 3D geometry of the molecule. Initially, empirical molecular descriptors were obtained from experimental data. However, many empirical descriptors reflect a complex combination of different physical interactions and, in addition, are not available for compounds yet to be synthesized. Thus, many efforts

were made to develop descriptors which only require the information encoded in the chemical structure, as for example, the constitutional or the topological descriptors. The constitutional descriptors are related with the chemical composition of the molecule and are independent of molecular connectivity and geometry. Some examples are the total number of atoms in the molecule, the absolute and relative numbers of atoms of certain chemical identity (C, H, O, N, F, etc.), the absolute and the relative numbers of certain chemical groups and functionalities, the total number of bonds in the molecule, etc. Topological descriptors are calculated using data on the connectivity of atoms within a molecule and represent the size, the shape, and branching of the molecule. Some examples of topological indices are the Wiener index,[147a] the Kier shape index[147b] and the Kier flexibility index.[147c] Other descriptors are derived from 3D geometries but are 1D/2D in nature as for example those calculated with quantum mechanics methods[148] such as HOMO and LUMO orbital energies,[148] total electronic energy[148a] and protonation energy.[148c]

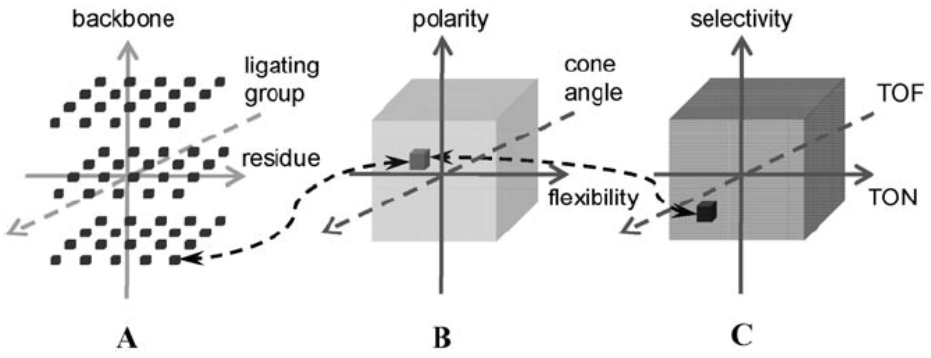


Figure 1.15. Simplified representation of three multidimensional spaces **A**, **B**, and **C**. Figure was reproduced from Ref. 150e.

These 1D- or 2D-QSPR approaches are normally used to set quantitative correlations of physical and chemical properties of organic and biological molecules; and also, there are some outstanding examples that correlate activity and selectivity in homogeneous catalysis. Jensen and co-workers have constructed QSPR/QSAR models based on DFT-generated molecular descriptors for alkene metathesis by ruthenium catalysts.[149] They found a multivariate correlation between the calculated measure of the activity and the properties of ruthenium complexes. The authors used a large number of descriptors (constitutional, topological, electronic, geometric, etc.), and those specific with chemical meaning gave insight into the factors governing the activity.[149] Rothenberg and co-workers proposed a computer-based method to screen virtual catalyst libraries mainly based on topological descriptors.[150] In this multidimensional approach, the authors divided the system into three spaces (Figure 1.15); i) space A, or catalyst space, contains a library of catalysts that have been tested; ii) space B, or descriptor space, contains the catalyst and reaction descriptors such as backbone flexibility, temperature, etc.; and iii) space C contains figures of merit, measures of catalyst performance

such as TOF and enantiomeric excess, and also those related with real-world concerns such as economic cost. Combining the descriptors and chemical principles and QSAR-type modeling, the authors predicted several characteristics of the catalysts and their performance. Examples are the prediction of ligand bite angle and backbone flexibility in bidentate ligand-metal complexes,[150a] and the catalytic activity of furfural hydrogenation by NHC-Ru complexes.[150e] Similarly, Chavali and co-workers demonstrated that 2D connectivity indexes are useful for setting structure-property correlations in reactions catalyzed by transition metal complexes such as the epoxidation by molybdenum catalyst.[151] These studies were performed with the *Computer Aided Molecular Design* (CAMD) computational tool.

Another 2D-QSPR approach is the *Topological Maximum Cross Correction*[152] (TMACC) method, which uses alignment-independent 2D-descriptors. These descriptors are generated using atomic properties determined by molecular topology: electrostatics, solubility, steric effect and lipophilicity. The electrostatic properties are represented by the *Gasteiger partial charge*,[153] which is calculated using the method of partial equalization of orbital electronegativity. This procedure calculates atomic charges in σ-bonded and non-conjugated π-system using only the topology of the catalysts. *Crippen-Wildman molar refractivity*[154] (MR) is used as a measure of the steric effects that it is determined through the classification of atoms based on neighbour atoms. *Crippen-Wildman partition coefficients*[154] (*logP*) are assigned to each atom as a measure of atomic lipophilicity, determined in the same way as *Crippen-Wildman molar refractivity* (MR). The solubility properties are described by *logS* parameters, representing solubility and solvation phenomena.[155] The property parameter is calculated (P_{calc}) using the equation 1.7, where n_i is the number of atoms of type i, a_i is the value of the property of type i and the sum is extended for all the atoms types. The property types produce positive and negative values which are considered as two separate properties. The molar refractivity is the exception that only takes positive values. The values of properties are rescaled by the largest absolute value, resulting in a range of +1 and -1.

$$P_{calc} = \sum_{i=1}^{N} n_i a_i \qquad (1.7)$$

The TMACC descriptor is derived from the autocorrelation descriptor (X_{ac}),[156] which considers a property (p), and a topological distance between atoms i and j (equation 1.8). They are calculated by computing the products of the properties for every pair of the atoms within a given molecule. Thus, only the maximum value of these products for each bond distance is selected. In TMACC descriptors, the different properties can be combined to generate cross-correlation descriptors as summarized in an equation 1.9, in which the products combine two different type of properties (p, q) for two atoms (i, j) separated by the topological distance (d). In this Thesis, we have performed TMACC-based study on asymmetric hydroformylation of styrene by Rh complexes in collaboration with the Prof. Hirst group (School of Chemistry, University of Nottingham, Nottingham, England) and obtained models that have predictive capacity for the activity (see *Chapter 4*).

$$X_{ac}(p,d) = \sum_{i=1,j\neq i}^{N} p_i p_j \tag{1.8}$$

$$X_{ac}(p,q,d) = \max\left(p_i q_i, p_j q_j\right) \tag{1.9}$$

Ideally, the descriptors used for model development should be quick to calculate and easy to interpret. The calculation of 2D and 1D descriptors are three orders of magnitude faster than the calculation of interaction fields based on MM calculations used to obtain 3D descriptors (see below). The time advantage of using 2D descriptors, however, is offset by several limitations; the conformational information is neglected and the chirality cannot be straightforwardly treated. Moreover, although 2D descriptors account for specific physicochemical properties of molecules, there is no mechanistic interpretation for them. 3D descriptors offer the most realistic way of representing a chemical system but their computational time depends on the geometry optimization step, and therefore on the size and degrees of freedom of the chemical system.

3D-QSPR models. Applications in asymmetric catalysis

Considering the difficulties of the synthesis and evaluation of chiral ligands, the use of QSPR modeling in asymmetric catalysis could be especially relevant. To identify the quantitative structure-enantioselectivity relationships, in principle, it is necessary to account for the 3D features of the chiral catalyst, in order to reflect its enantiotopic differentiation capacity. It is possible to use geometrical descriptors based on steric size,[140] molecular topology[157] and continuous chirality quantification methods[158] that indirectly account for 3-dimensionality, or 3D *molecular interaction field* (MIF) descriptors. Geometrical descriptors include easily accessible parameters from classical physical organic chemistry such as the above-mentioned *Taft-Charton steric* parameters.[99,102] The MIF-based approaches became the standard 3D-QSAR methods used in drug design, with the *Comparative Molecular Field Analysis*[159] (CoMFA) approach being the most popular. The generation of the CoMFA descriptors involves three stages (see Figure 1.16): i) generation of the molecular structures; ii) alignment of structures to obtain reproducible predictive spaces; and iii) computation of the MIFs at each point of a 3D grid surrounding the molecule.[159] Initially, these methods used MM-based interactions sampling the steric and electrostatic interactions by Lennard-Jones and Coulomb potentials,[159] but also methods that incorporate semiempirical[160] and QM[161-162] interaction energies have been developed. CoMFA and CoMFA-like approaches have produced fairly good results in some asymmetric catalytic processes.[161-165] Examples are the alkylation of aldehydes by Zn-aminoalcohols,[161a,162-163] the Diels-Alder reaction catalyzed by Cu-(bis-oxazoline) and Cu-phosphano(oxazoline) complexes,[164a] the ketone reduction by oxazaborolidines,[164b] the lithiation-substitution of N-Boc-pyrrolidine by diamines,[161b] the alkylation of the glycine imine by amines[165] and the enantioselective Nozaki-Hiyama-Kishi propargylation of ketones.[164c] Despite its undoubted popularity, CoMFA analysis can be highly sensitive to the parameters used to construct the matrix of 3D descriptor such us the methods for calculating the interaction fields and the grid spacing and dimensions. As results, these parameters need to be

optimized for each catalytic system. In particular, sensibility of the results to the translation of the grid relative to the aligned molecular structures has been shown in numerous investigations.[161-165]

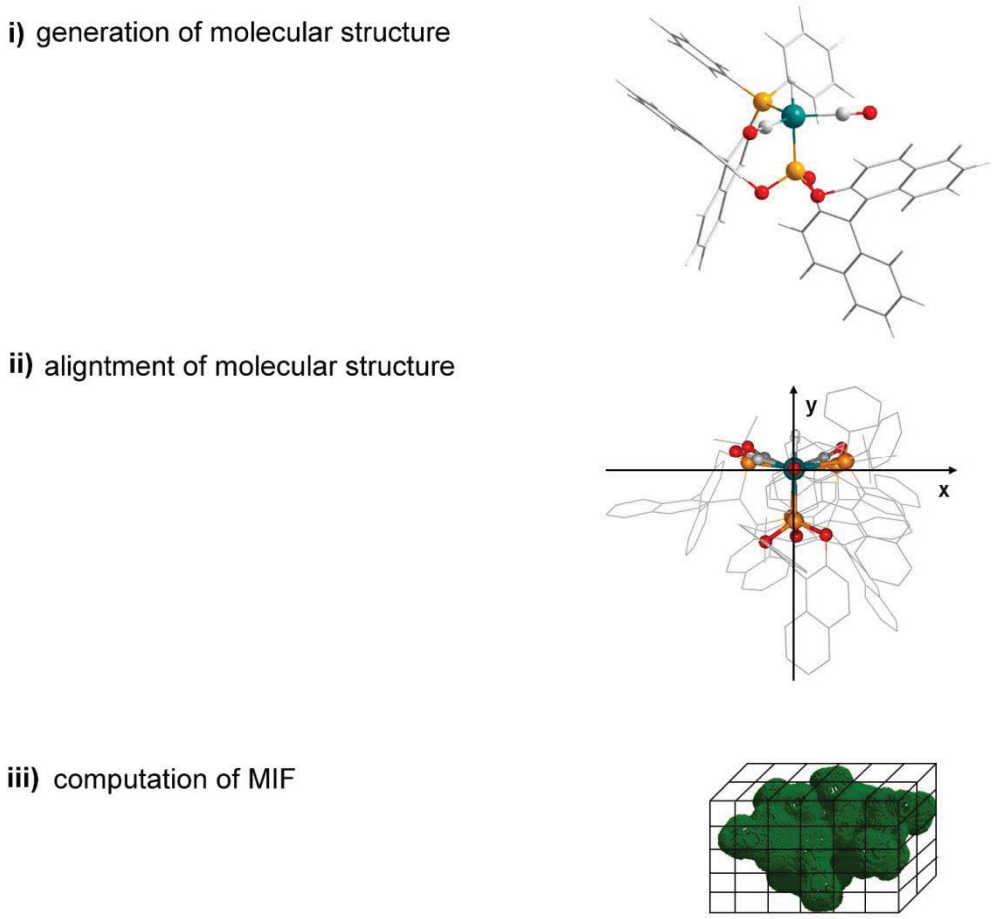


Figure 1.16. Schematic representation of the different steps of a prototypical CoMFA approach.

An alternative CoMFA-like approach was reported to compute property fields based on similarity indices of molecules which have been brought into common alignment. The *Comparative Molecular Similarity Indices Analysis*[166] (CoMSIA) is similar to the CoMFA approach; but the descriptors are different in their calculation form and in nature. In CoMSIA, it is also required to calculate the value of the descriptor at regularly spaced grid points. However, CoMSIA approach uses the similarity indices calculated via common probe atom which is localized at the intersections of a regularly spaced lattice using a distance-dependent

Gaussian-type function, $G_{pj}(c)$ (equation 1.10). The $G_{pj}(c)$ term is the similarity index of property *p* at grid point *j* for compound *c*; *i* is the atom index; P_i is the atomic property of atom *i*; α is the attenuation factor and r_{ij} is the distance between atom *i* and grid point *j*. Originally, the physicochemical properties evaluated were the steric properties using the third power of the atomic radii, the electronic properties using atomic charges, and the hydrophobic contributions using the parameterization developed by Viswahadhan.[167] Recently, several studies have investigated the same reaction with both the CoMFA and CoMSIA approaches, giving similar results which were acceptable in both cases.[168,169]

$$G_{pj}(c) = \sum_{i=1}^{N} p_i e^{-\alpha r_{ij}^2} \qquad (1.10)$$

The alignment step is usually time-consuming for large date sets. Also it is always subjective, and it is not necessarily valid for all the types of structures, limiting the application of 3D-QSPR methodologies. To avoid this, alignment-independent molecular descriptors have been introduced. The first alignment-independent molecular descriptors emerged from autocorrelation functions proposed by Broto.[170] The Broto approach consisted of representing the atomic properties of 2D and 3D structures by autocorrelation vectors which are used as independent variables in QSPR modeling. Pastor and co-workers have introduced the *GRid-Independent descriptors* (GRIND), which *aims to extract the information enclosed in the MIF and to encode it into new types of variables whose values are independent of the spatial position of the molecule studied.*[171] The alignment independence is achieved by computing the products of MIF values for each pair of points on the grid, and then selecting the highest value of these products for each distance range. This enables its representation in the original 3D space as a line linking two specific MIF nodes grouped into the graphics called *correlograms* (Figure 1.17). The GRIND protocol involves also the filtering of the number of surface data points to reduce the dataset size and select the most relevant regions of molecular structure. The GRIND methods combined with empirical molecular interaction fields have been proved effective in several asymmetric organometallic catalytic processes.[163]

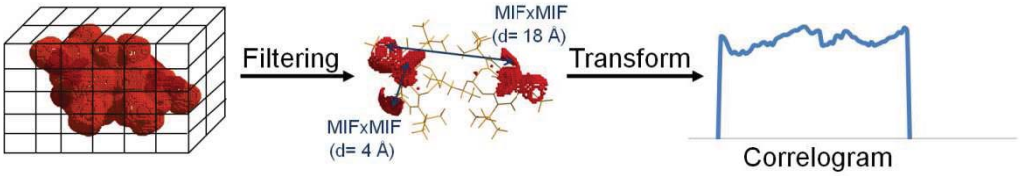


Figure 1.17. Schematic representation of the GRIND methodology.

Similarly, Bo and co-workers have developed a CoMFA-like alignment independent (GRIND) protocol based on 3D molecular descriptors derived from DFT calculations, directly called the *Q-QSSR* approach.[162] This approach was successfully applied to study the asymmetry alkylation

of aldehydes by Zn-aminoalcohols.[162] To assess the enantioinduction, it was assumed that it is necessary to account for the catalyst shape. First, the method generates the molecular surface from the DFT-derived electronic density isosurface. At each point on the isosurface the *Molecular Shape Field* (MSF) is computed. The MSF field is a geometrical parameter based on the local curvature of the isodensity surface that uses a cosine expression.[171b] The MSF values range between 0 and -1 for convex areas and 0 and +1 for concave ones. The filtering criterion selected the most convex areas as being the most descriptive of the molecular surface, accounting for the steric effects of the catalyst. In addition, the *Molecular Electrostatic Potential* (MEP) field was also computed at the grid points and provides a second level of steric effects.[162] The PhD Thesis applies an extended version of this *Q-QSSR* approach to the study of two asymmetric processes: the asymmetric hydroformylation of styrene by Rh-diphosphane (*Chapter 4*), and the asymmetric cyclopropanation of styrene by copper catalysts modified with oxazoline-type ligands (*Chapter 5*).

4D-QSPR models. Application to conformational flexible catalysts

For catalysts with loose structures, it may be necessary to account for their conformational flexibility. Hirst and co-workers have developed 3.5D- and 4D-QSPR approaches based on molecular dynamics (MD) simulations.[165] These methodologies employ the results of molecular dynamics simulations in order to find suitable low-energy conformations for deriving 3D-QSPR models and more directly, to use the trajectories in a 4D-QSSR approach. Additionally, the 3.5D-QSPR approach considers the different conformations sampled by selecting one conformer for each energy value. Then the contribution of each conformation is weighted with Boltzmann factor. The 3.5D-QSPR approaches can be considered a hybrid between the 3D- and 4D-QSPR. An example is the study on the alkylation of the glycine imine by amines in which standard CoMFA parameters were used.[159]

1.3. Objectives of the Thesis

The general objective of this Thesis is to explore and to develop computational tools based on QSPR modeling for their use in the study of homogeneous catalysis as an alternative to the traditional DFT approaches involving transition-state (TS) determination. Also, the intention is to gain an insight into the chemistry of outstanding catalytic (or potentially catalytic) processes in order to derive guidelines for *de novo* catalyst design. The selected examples are ammonia activation, asymmetric hydroformylation of alkenes by Rh-diphosphanes and asymmetric cyclopropanation of alkenes by Cu complexes. First, the Thesis attempts to determine the mechanism of the N-H σ-bond activation of ammonia by the μ_3-alkylidyne titanium species using DFT methods. After recognizing some limitations in the TS-based approaches with respect to the hydroformylation and cyclopropanation of alkenes, other computational methodologies based on ground-state properties were evaluated and developed, in order to seek quantitative relationships with the activity and selectivity. These approaches aim to define and use chemically meaningful descriptors, to combine 2D and 3D alignment-independent protocols with our previous mechanistic knowledge, and to test whether it is possible to build 3D-QSPR models using quantitative quadrant-diagram representations of the

catalysts. Thus, the specific goals of this Thesis can be classified according to the processes studied.

Ammonia activation by μ_3-alkylidyne fragments supported on a Ti molecular oxide

- To investigate the observed N-H σ-bond activation of ammonia by μ_3-alkylidyne titanium oxide complexes.
- To analyze the structural, electronic and energetic properties of the spectroscopically detected amido, imido and nitride titanium intermediates.
- To identify and characterize the mechanism for the three N-H bond activation of ammonia.

Asymmetric hydroformylation of alkenes by Rh-diphosphane

- To identify and quantify the different effects governing the coordination preferences of the phosphane-phosphite binaphos ligand in [$HRh(CO)_2${(*R,S*)-binaphos}] complex as a key factor in the catalytic performance of asymmetric hydroformylation of styrene.
- To identify and quantify the key ligand-substrate interactions responsible for enantioinduction.
- To define a new molecular descriptor to evaluate the steric hindrance of the different quadrants.
- To propose a stereochemical model based on a quadrant-diagram representation of the catalyst.
- To generate 2D- and 3D-QSPR models for the catalytic enantioselectivity and activity asymmetric hydroformylation of styrene by Rh-diphosphane for a dataset of 21 ligands.
- To use chemically meaningful descriptors based on steric and electronic concepts in QSPR modeling in order to provide insight into the factors governing catalytic activity and selectivity.
- To combine the analysis of the equations based on alignment-independent descriptors with our previous mechanistic knowledge in order to orient the ligand design.

Asymmetric cyclopropanation of alkenes by Cu complexes

- To generate 3D-QSSR models for enantioselectivity in asymmetric cyclopropanation of alkenes by Cu complexes, testing alignment-independent and -dependent descriptors for a dataset of 30 ligands.
- To test two steric descriptors; the *Taft-Charton steric* parameter and the *distance-weighted volume* (V_W) to evaluate the steric hindrance in quadrant-diagram representation of the catalyst.
- To develop a new methodology for predicting enantioselectivity based on a quantitative quadrant-diagram representation of the catalysts and QSSR-type modeling.
- To propose a simple sterochemical model to predict the outcome of new catalysts and assist the design of *de novo* catalysts.

1.4. References and notes

1 N. A. Lange, J. G. Speight, *Lange's Handbook of Chemistry*, *16th ed*. Mc Graw-Hill, New York, **2005**.

2 a) M. Shultz, D. Milstein, *J. Chem. Soc. Chem. Commun.* **1993**, 318. b) W. Yao, O. Eisenstein, R. H. Crabtree, *Inorg. Chim. Acta* **1997**, *254*, 105.

3 a) R. Koelliker, D. Milstein, *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 707. b) R. Koelliker, D. Milstein, *J. Am. Chem. Soc.* **1991**, *113*, 8524. c) A. Abarca, P. Gómez-Sal, A. Martín, M. Mena, J. M. Poblet, C. Yélamos, *Inorg. Chem.* **2000**, *39*, 642. d) D. Conner, K. N. Jayaprakash, T. R. Cundari, T. B. Gunnoe, *Organometallics* **2004**, *23*, 2724. e) D. Rais, R. G. Bergman, *Chem. Eur. J*. **2004**, *10*, 3970. f) M. García-Castro, A. Martín, M. Mena, J. M. Poblet, C. Yélamos, *Eur. J. Inorg. Chem.* **2006**, 1155.

4 a) A. L. Casalnuovo, J. C. Calabrese, D. Milstein, *J. Am. Chem. Soc.* **1988**, *110*, 6738. b) D. M. Roundhill, *Chem. Rev.* **1992**, *92*, 1. c) T. Braun, *Angew. Chem. Int. Ed.* **2005**, *44*, 5012. d) I. Asiv, Z. Gross, *Chem. Commun*. **2006**, 4477. e) J. I. van der Vlugt, *Chem. Soc. Rev*. **2010**, *39*, 2302.

5 J. Zhao, A. S. Goldman, J. F. Hartwig, *Science* **2005**, *307*, 1080.

6 M. Kanzelberger, X. Zhang, T. J. Emge, A. S. Goldman, J. Zhao, C. Incarvito, J. F. Hartwig, *J. Am. Chem. Soc.* **2003**, *125*, 13644.

7 E. Morgan, D. F. MacLean, R. McDonald, L. Turculet, *J. Am. Chem. Soc*. **2009**, *131*, 14234.

8 C. M. Fafard, D. Adhikari, B. M. Foxman, D. J. Mindiola, O. V. Ozerov, *J. Am. Chem. Soc.* **2007**, *129*, 10318.

9 A. L. Casalnuovo, J. C. Calabrese, D. Milstein, *Inorg. Chem.* **1987**, *26*, 971.

10 M. A. Salomon, A.-K. Jungton, T. Braun, *Dalton Trans.* **2009**, 7669.

11 E. G. Bryan, B. F. G. Johnson, J. Lewis, *J. Chem. Soc. Dalton Trans.* **1977**, 1328.

12 G. D. Frey, V. Lavallo, B. Donnadieu, W. W. Schoeller, G. Bertrand, *Science* **2007**, *316*, 439.

13 Y. Peng, J.-D. Guo, B. D. Ellis, Z. Zhu, J. C. Fettinger, E. Rivard, P. P. Power, *J. Am. Chem. Soc*. **2009**, *131*, 16272.

14 a) A. Jana, I. Objartel, H. W. Roesky, D. Stalke, *Inorg. Chem.* **2009**, *48*, 798. b) A. Jana, C. Shulzke, H. W. Roesky, *J. Am. Chem. Soc*. **2009**, *131*, 4600.

15 P. A. Chase, D. W. Stephan, *Angew. Chem. Int. Ed*. **2008**, *47*, 7433.

16 a) M. L. H. Green, C. R. Lucas, *J. Chem. Soc. Dalton Trans*. **1972**, 1000. b) J. N. Armor, *Inorg. Chem.* **1978**, *17*, 203. c) J. E. Bercaw, D. L. Davies, P. T. Wolczanski, *Organometallics* **1986**, *5*, 443.

17 G. L. Hillhouse, J. E. Bercaw, *J. Am. Chem. Soc.* **1984**, *106*, 5472.

18 X. Yu, Z.-L. Xue, *Inorg. Chem*. **2005**, *44*, 1505.

19 Y. Peng, B. D. Ellis, X. Wang, P. P. Power, *J. Am. Chem. Soc*. **2008**, *130*, 12268.

20 C. Ni, H. Lei, P. P. Power, *Organometallics* **2010**, *29*, 1988.

21 E. Khaskin, M. A. Iron, L. J. W. Shimon, J. Zhang, D. Milstein, *J. Am. Chem. Soc.* **2010**, *132*, 8542.

22 A. Meltzer, I. Shigeyoshi, C. Präsang, M. Driess, *J. Am. Chem. Soc.* **2010**, *132*, 3038.

23 G. Parkin, A. van Asselt, D. J. Leathy, L. Whinnery, N. G. Hua, R. W. Quan, L. M. Henling, W. P. Schaefer, B. D. Santarsiero, J. E. Bercaw, *Inorg. Chem.* **1992**, *21*, 82.

24 J. S. Freundlich, R. R. Schrock, W. M. Davis, *J. Am. Chem. Soc.* **1996**, *118*, 3643.

25 P. Avenier, A. Lesage, M. Taoufik, A. Baudouin, A. De Mallmann, S. Fiddy, M. Vautier, L. Veyre, J.-M. Basset, L. Emsley, E. A. Quadrelli, *J. Am. Chem. Soc.* **2007**, *129*, 176.

26 Y. Nakajima, H. Kameo, H. Suzuki, *Angew. Chem. Int. Ed.* **2006**, *45*, 950.

27 K. Dehnicke, J. Strähle, *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 955.

28 H. W. Roesky, Y. Bai, M. Noltenmeyer, *Angew. Chem. Int. Ed. Engl*. **1989**, *28*, 754.

29 a) M. M. B. Holl, M. Kersting, B. D. Pendley, P. T. Wolczanski, *Inorg. Chem.* **1990**, *29*, 1518. b) M. M. B. Holl, P. T. Wolczanski, G. D. Van Duyne, *J. Am. Chem. Soc.* **1990**, *112*, 7989.

30 M. M. B. Holl, P. T. Wolczanski, *J. Am. Chem. Soc.* **1992**, *114*, 3854.

31 P. A. Lay, H. Taube, *Inorg. Chem.* **1989**, *28*, 3561.

32 S.-H. Kim, B. A. Moyer, S. Azan, G. M. Brown, A. L. Olins, D. P. Alison, *Inorg. Chem*. **1989**, *28*, 4648.

33 a) P. Gómez-Sal, A. Martín, M. Mena, C. Yélamos, *J. Chem. Soc. Chem. Commun*. **1995**, 2185. b) C. Yélamos, Thesis Doctoral, Universidad de Alcalá, Madrid, **1996**.

34 a) R. M. Fix, R. G. Gordon, D. M. Hoffman, *J. Am. Chem. Soc.* **1990**, *112*, 7833. b) D. M. Hoffman, *Polyhedron* **1994**, *13*, 1169. c) L. H. Dubois, *Polyhedron* **1994**, *13*, 1329. d) D. V. Baxter, M. H. Chisholm, G. J. Gama, V. F. DiStasi, A. L. Hector, I. P. Parkin, *Chem. Mater.* **1996**, *8*, 1222.

35 K. Weissermel, H.-J. Arpe, *Industrielle Organische Chemie*, *5th ed*. VCH: Weinheim, Germany: **2000**.

36 J. A. Osborn, F. H. Jardine, J. F. Young, G. Wilkinson, *J. Chem. Soc. A* **1966**, 1711.

37 a) T. J. Devon, G. W. Phillips, T. A. Puckette, J. L. Stavinoha, J. J. Vanderbilt, *Chem*. *Abs*. **1988**, *108*, 7890 (US Pat., 4694109, **1987**). b) T. J. Devon, G. W. Phillips, T. A. Puckette, J. L. Stavinoha, J. J. Vanderbilt, *Chem*. *Abs*. **1994**, *121*, 280879 (US Pat., 5332846 **1994**).

38 a) T. J. Devon, G. W. Philips, T. A, Puckette, J. L. Stavinoha, J. J. Vanderbilt, US Pat., 4694109, **1989**. b) C. P. Casey, G. T. Whiteker, M. G. Melville, L. M. Petrovich, J. A. Gavey, D. R. Powell, *J. Am. Chem. Soc.* **1992,** *114*, 5535.

39 a) E. Billig, A. G. Abatjoglou, D. R. Bryant, US Pat., 4769498, **1988**. b) G. D. Cunny, S. L. Buchwald, *J. Am. Chem. Soc.* **1993,** *115*, 2066.

40 a) P. C. J. Kamer, M. Kranenburg, P. W. N. M. van Leeuwen, J. G. de Vries, *Pat. Appl.* 9400470, **1994**, WO95/ 30680. b) M. Kranenburg, Y. E. M. van der Burgt, P. C. J. Kamer, W. N. M. P. van Leeuwen, K. Goubitz, J. Fraanje, *Organometallics* **1995,** *14*, 3081. c) S. Hillebrand, J. Bruckmann, C. Krüger, M. W. Haenel, *Tetrahedron Lett.* **1995**, *36*, 75. d) P. C. J. Kamer, W. N. M. P. van Leeuwen, J. N. H. Reek, *Acc. Chem. Res.* **2001,** *34*, 895.

41 D. Evans, J. A. Osborn, G. Wilkinson, *J. Chem. Soc. A* **1968**, 3133.

42 a) P. W. N. M. van Leeuwen, C. F. Roobeek, *J. Organomet*. *Chem*. **1983**, *258*, 343. b) W. R. Moser, C. J. Papite, D. A. Brannon, R. A. Duwell, *J. Mol. Catal*. **1987**, *41*, 271.

43 a) M. Garland, P. Pino, *Organometallics* **1991**, *10*, 1693. b) V. S. Nair, S. P. Mathew, R. V. Chaudhari, *J. Mol. Catal. A. Chem.* **1999**, *143*, 99. c) J. Feng, M. Garland, *Organometallics* **1999**, *18*, 417.

44 P. C. J. Kamer, A. Van Rooy, G. C. Schoemaker, P. W. N. M. van Leeuwen, *Coord. Chem. Rev.* **2004**, *248*, 2409.

45 E. Zuidema, L. Escorihuela, T. Eichelsheim, J. J. Carbó, C. Bo, P. C. J. Kamer, P. W. N. M. van Leeuwen, *Chem. Eur. J.* **2008**, *14*, 1843.

46 E. Zuidema, P. E. Goudriaan, B. H. G. Swennenhuis, P. C. J. Kamer, P. W. N. M. van Leeuwen, M. Lutz, A. L. Spek, *Organometallics* **2010**, *29*, 1210.

47 E. Zuidema, E. Daura-Oller, J. J. Carbo´, C. Bo, P. W. N. M. van Leeuwen, *Organometallics* **2007**, *26*, 2234.

48 M. F. Haddow, A. J. Middleton, A. G. Orpen, P. G. Pringle, R. Papp, *Dalton Trans*. **2009**, 202.

49 B. Breit, R. Winde, T. Mackewitz, R. Paciello, K. Harms, *Chem. Eur. J.* **2001**, *7*, 3106.

50 L. A. van der Veen, P. H. Keeven, G. C. Schoemaker, J. N. H. Reek, P. C. J. Kamer, P.W. N. M. van Leeuwen, M. Lutz, A. L. Spek, *Organometallics* **2000**, *19*, 872.

51 C. P. Casey, E. L. Paulsen, E. W. Beuttenmueller, B. R. Proft, L. M. Petrovich, B. A. Matter, D. R. Powell, *J. Am. Chem. Soc.* **1997**, *119*, 11817.

52 J. J. Carbo´, F. Maseras, C. Bo, P. W. N. M. van Leeuwen, *J. Am. Chem. Soc.* **2001**, *123*, 7630.

53 a) I. Ojima, *Catalytic Asymmetric Synthesis*, Wiley-VCH, New York, **2000**. b) E. N. Jacobsen, A. Pfaltz, H. Yamamoto, *Comprehensive Asymmetric Catalysis*, Springer-Verlag, New York, **1999**.

54 a) N. Sakai, S. Mano, K. Nozaki, H. Takaya, *J. Am. Chem. Soc*. **1993**, *115*, 7033. b) K. Nozaki, N. Sakai, T. Nanno, T. Higashijima, S. Mano, T. Horiuchi, H. Takaya, *J. Am. Chem. Soc*. **1997**, *119*, 4413.

55 K. Nozaki, *Chem. Rec.* **2005**, *5*, 376.

56 a) J. E. Babin, G. T. Whiteker, US Pat., 911518, **1992** (WO 9303839). b) G. T. Whiteker, J. R. Briggs, J. E. Babin, B. A. Barner*, Chemical Industries*, Vol. 89, Marcel Dekker, New York, **2003**, 359.

57 M. Rouhi, *Chem. Eng. News* **2003**, *81*, 52.

58 a) S. Breeden, D. J. Cole-Hamilton, D. F. Foster, G. J. Schwarz, M. Wills, *Angew. Chem.* **2000**, *112*, 4272. b) S. Breeden, D. J. Cole-Hamilton, D. F. Foster, G. J. Schwarz, M. Wills, *Angew. Chem. Int. Ed.* **2000**, *39*, 4106.

59 a) C. J. Cobley, K. Gardner, J. Klosin, C. Praquin, C. Hill, G. T. Whiteker, A. Zanotti-Gerosa, J. L. Petersen, K. A. Abboud, *J. Org. Chem.* **2004**, *69*, 4031. b) C. J. Cobley, J. Klosin, C. Qin, G. T. Whiteker, *Org. Lett*. **2004**, *6*, 3277.

60 a) G. J. H. Buisman, L. A. van der Veen, A. Klootwijk, W. G. J. der Lange, P. C. J. Kamer, P. W. N. M. van Leeuwen, D. Vogt, *Organometallics* **1997**, *16*, 2929. b) M. Diéguez, O. Pàmies, A. Ruiz, S. Castillón, C. Claver, *Chem. Eur. J.* **2001**, *7*, 3086.

61 B. Zhao, X. Peng, Z. Wang, C. Xia, K. Ding, *Chem. Eur. J.* **2008**, *14*, 7847.

62 a) A. T. Axtell, C. J. Cobley, J. Klosin, G. T. Whiteker, A. Zanotti-Gerosa, K. A. Abboud, *Angew. Chem. Int. Ed*. **2005**, *44*, 5834. b) A. T. Axtell, C. J. Cobley, J. Klosin, G. T. Whiteker, A. Zanotti-Gerosa, K. A. Abboud, *Angew. Chem*. **2005**, 117, 5984.

63 T. P. Clark, C. R. Landis, S. L. Freed, J. Klosin, K. A. Abboud, *J. Am*. *Chem*. *Soc*. **2005***, 127*, 5040.

64 a) A. Kless, J. Holz, D. Heller, R. Kadyrov, R. Selke, C. Fisher, A. Börner, *Tetrahedron: Asymmetry* **1996**, *7*, 33. b) S. Deerenberg, P. C. J. Kamer, P.W. N. M. van Leeuween, *Organometallics* **2000**, *19*, 2065. c) O. Pámies, G. Net, A. Ruiz, C. Claver, *Tetrahedron: Asymmetry* **2002**, *12*, 3441. d) T. Robert, Z. Abiri, J. Wassenaar, A. J. Sandee, S. Romanski, J. M. Neudçrfl, H. G. Schmalz, J. N. H. Reek, *Organometallics* **2010**, *29*, 478. e) S. H. Chikkali, R. Bellini, G. Berthon- Gelloz, J. I. van der Vlugt, B. de Bruin, J. N. H. Reek, *Chem. Commun.* **2010**, *46*, 1244. f) F. Doro, J. N. H. Reek, P. W. N. M. van Leeuwen, *Organometallics* **2010**, *29*, 4440. g) I. Arribas, S. Vargas, M. Rubio, A. Surez, C. Domene, E. Alvarez, A. Pizzano, *Organometallics* **2010**, 29, 5791.

65 Y. Yan, X. Zhang, *J. Am. Chem. Soc*. **2006**, *128*, 7198.

66 X. Zhang, B. Cao, Y. Yan, S. Yu, B. Ji, X. Zhang, *Chem. Eur. J.* **2010**, *16*, 871.

67 a) W. Tang, X. Zhang, *Angew. Chem. Int. Ed*. **2002**, *41*, 1612. b) J. Klosin, C. R. Landis, *Acc. Chem. Res.* **2007,** *40,* 1251.

68 R. Ewalds, E. B. Eggeling, A. C. Hewat, P. C. J. Kamer, P. W. N. M. van Leeuween, D. Vogt, *Chem. Eur. J.* **2000**, *6*, 1496.

69 a) C. J. Cobley, R. D. J. Froese, J. Klosin, C. Quin, G. T. Whiteker *Organometallics* **2007**, *26*, 2986. b) A. T. Axtell, J. Klosin, G. T. Whiteker, C. J. Cobley, M. E. Fox, M. Jackson, K. A. Abboud, *Organometallics* **2009,** *28,* 2993. c) J. Mazuela, M. Coll, O. Pàmies, M. Diéguez, *J. Org. Chem.* **2009**, *74*, 54440.

70 a) D. Gleich, R. Schmid, W. A. Herrmann, *Organometallics* **1998,** *17,* 2141. b) D. Gleich, W. A. Herrmann, *Organometallics* **1999,** *18,* 4354. c) J. J. Carbó, A. Lledós, D. Vogt, C. Bo, *Chem. Eur. J.* **2006**, *12*, 1457. d) M. Rubio, A. Suárez, E. Álvarez, C. Bianchini, W. Oberhauser, M. Peruzzini, A. Pizzano, *Organometallics* **2007,** *26,* 6428.

71 A. L. Watkins, C. R. Landis, *J. Am. Chem. Soc.* **2010**, *132*, 10306.

72 P. W. N. M. van Leeuwen, *Homogeneous Catalysis*, Published by Kluwer Academic Publishers, **2004**.

73 H. Fritschi, U. Leutenegger, A. Pfaltz, *Helv. Chim. Acta* **1988**, *71*, 1553.

74 a) A. Cornejo, J. M. Fraile, J. I. Garcia, M. J. Gil, V. Martínez-Merino, J. A. Mayoral, L. Salvatella, *Organometallics* **2005**, *24*, 3448. b) A. Cornejo, J. M. Fraile, J. I. Garcia, M. J. Gil, V. Martínez-Merino, J. A. Mayoral, L. Salvatella, *Angew. Chem. Int. Ed.* **2005**, *44*, 458.

75 H. Nozaki, S. Moriuti, H. Takaya, R. Noyori, *Tetrahedron Lett.* **1966**, 5239.

76 a) A. K. Glosh, M. Packiarajan, J. Capiello, *Tetrahedron: Asymmetry* **1998**, *9*, 1. b) J. S. Johnson, D. A. Evans, *Acc. Chem. Res*. **2000**, *33*, 325. c) G. Desimoni, G. Faita, P. Quadrelli, *Chem. Rev*. **2003**, *103*, 3199. d) H. Lebel, J. F. Marcoux, C. Molinaro, A. B. Charette, *Chem. Rev*. **2003**, *103*, 977.

77 a) R. G. Salomon, J. K. Kochi, *J. Am. Chem. Soc.* **1973**, *95*, 3300. b) J. Green, E. Sinn, S. Woodward, R. Butcher, *Polyhedron* **1993**, *12*, 991. c) M. M. Díaz-Requejo, P. J. Pérez, M. Brookhart, J. L. Templeton, *Organometallics* **1997**, *16*, 4399. d) M. M. Díaz-Requejo, T. R. Belderrain, M. C. Nicasio, F. Prieto, P. J. Pérez *Organometallics* **1999**, *18*, 2601. e) M. Bühl, F. Terstegen, F. Löffler, B. Meynhardt, S. Kierse, M. Müller, C. Näther, U. Lüning, *Eur. J. Org. Chem.* **2001**, 2151.

78 a) T. Rasmussen, J. F. Jensen, N. Østergaard, D. Tanner, T. Ziegler, O.-O. Norrby, *Chem. Eur. J.* **2002**, *8*, 177. b) G. Drudis-Solé, F. Maseras, A. Lledós, A. Vallribera, M. Moreno-Mañas, *Eur. J. Inorg. Chem.* **2008**, 5614. c) J. I. García, G. Jiménez-Osés, J. A. Mayoral, *Chem. Eur. J.* **2011**, *17*, 529.

79 a) J. M. Fraile, J. I. García, M. J. Gil, V. Martínez-Merino, J. A. Mayoral, L. Salvatella, *Chem. Eur. J.* **2004**, *10*, 758. b) J. I. García, G. Jiménez-Osés, V. Martínez-Merino, J. A. Mayoral, E. Pires, I. Villalba, *Chem. Eur. J.* **2007**, *13*, 4064. c) B. F. Straub, I Gruber, F. Rominger, P. Hofmann, *J. Organomet. Chem.* **2003**, *684*, 124.

80 a) B. F. Straub, P. Hofmann, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 1288. b) X. Dai, T. H. Warren, *J. Am. Chem. Soc.* **2004**, *126*, 10085. c) Y. M. Badiei, T. H. Warren, *J. Organomet. Chem.* **2005**, *690*, 5989. d) P. Hofmann, I. V. Shishkov, F. Rominger, *Inorg. Chem.* **2008**, *47*, 11755. e) I. V. Shishkov, F. Rominger, P. Hofmann, *Organometallics* **2009**, *28*, 1049.

81 M. Brookhart, W. B. Studabaker, *Chem. Rev.* **1987**, *87*, 411.

82 S. Kozuch, S. Shaik, *Acc. Chem. Res.* **2011**, *44*, 101.

83 a) R. G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules*. New York, Oxford University Press, **1989**. b) W. Koch, M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Weinheim, Wiley-VCH **2000**.

84 a) A. Ghosh, *Coor. Chem. Rev.* **2009**, *253*, 523. b) C. J. Cramer, D. G. Truhlar, *Phys. Chem. Chem. Phys*. **2009**, *11*, 10757.

85 a) H. Lin, D. G. Truhlar, *Theor*. *Chem*. *Acc*. **2007**, *117*, 185. b) C. Bo, F. Maseras, *Dalton Trans.* **2008**, 2911. c) F. Maseras, *Chem. Comm.* **2000**, 1821. d) F. Maseras, K. Morokuma, *J. Comput. Chem*. **1995**, *16*, 1170. e) T. Vreven, K. S. Byun, I. Komaromi, S. Dapprich, J. A. Montgomery, K. Morokuma, M. J. Frisch, *J. Chem. Theory Comput.* **2006**, *2*, 815.

86 a) P.-O. Norrby, *J. Mol. Struct.* (*Theochem)* **2000**, *506*, 9. b) P. J. Donoghue, E. Kieken, P. Helquist, O. Wiest, *Adv. Synth. Catal.* **2007**, *349*, 2647. c) P. Rydberg, L. Olsen, P.-O. Norrby, U.Ryde, *J. Chem. Theory Comput.* **2007**, *3*, 1765.

87 a) C. R. Corbeil, S. Thielges, J. A. Schwartzentruber, N. Moitessier, *Angew. Chem.* **2008**, *120*, 2675. b) R. Corbeil, S. Thielges, J. A. Schwartzentruber, N. Moitessier, *Angew. Chem. Int. Ed.* **2008**, *47*, 2635.

88 a) S. Bahmanyar, K. N. Houk, H. J. Martin, B. List, *J. Am*. *Chem*. *Soc*. **2003**, *125*, 2475. b) F. R. Clemente, K. N. Houk, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 5766. c) S. Mitsumori, H. Zhang, P. H.-Y. Cheong, K. N. Houk, F. Tanaka, C. F. Barbas, *J. Am. Chem. Soc.* **2006**, *128*, 1040. d) Y. Wang, J. Wang, J. Su, F. Huang, J. Jiao, Y. Liang, D. Yang, S. Zhang, P. A. Wender, Z.-X. Yu, *J. Am*. *Chem*. *Soc*. **2007**, *129,* 10060. e) W. H. Jr. Pitcock, R. L. Lord, M.-H. Baik, *J. Am*. *Chem*. *Soc*. **2008**, *130*, 5821.

89 S. T. Schneebeli, M. L. Hall, R. Breslow, R. Friesner, *J. Am*. *Chem*. *Soc*. **2009**, *131*, 3965.

90 a) Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2007**, *3*, 289. b) S. Tsuzuki, H. P. Lüthi, *J. Chem. Phys*. **2001**, *114*, 3949. c) E. J. Meijer, M. Sprik, *J. Chem. Phys.* **1996**, *105*, 8684.

91 T. Horiuchi, E. Shirakawa, K. Nozaki, H. Takaya, *Organometallics* **1997**, *16*, 2981.

92 Y. Zhao, D. G. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 157.

93 S. Grimme, *J. Comp. Chem.* **2006**, *27*, 1787.

94 a) N. Fey, *Dalton Trans.* **2010**, *39*, 296. b) J. A. Gillespie, D. L. Dodds, P. C. J. Kamer, *Dalton Trans.* **2010**, *39*, 2751.

95 a) D. White, N. J. Coville, *Advances in Organometallic Chemistry*, vol. 36, 1994, p.95. b) N. Fey, A. G. Orpen, J. N. Harvey, *Coord. Chem. Rev.* **2009**, *253*, 704.

96 R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics* (2 volumes), Wiley-VCH, **2009**.

97 C. A. Tolman, *Chem. Rev*. **1977**, *77*, 313.

98 a) N. Fey, A. C. Tsipi, S. E. Harris, J. N. Harvey, A. G. Orpen, R. A. Mansson, *Chem. Eur. J.* **2006**, *12*, 291. b) N. Fey, M. F. Haddow, J. N. Harvey, C. L. McMullin, A. G. Orpen, *Dalton Trans.* **2009**, 8183. c) N. Fey, J. N. Harvey, G. C. Lloyd-Jones, P. Murray, A. G. Orpen, R. Osborne, M. Purdie, *Organometallics* **2008**, *27*, 1372.

99 a) R. W. Jr. Taft, *J. Am. Chem. Soc*. **1952**, *74*, 2729. b) R. W. Jr. Taft, *J. Am. Chem. Soc*. **1953**, *75*, 4538. c) R. W. Jr. Taft, *Org. Chem.* **1956**, 556.

100 a) C. K. Hancock, E. A. Meyers, B. J. Yagar, *J. Am. Chem. Soc.* **1961**, *83*, 4211. b) V. A. Palm, *"Fundamentals of the Quantitative Theory of Organic Reactions."* Khimiya, Leningrad, **1967**, Chapter 10.

101 A. Bondi, *J. Phys. Chem.* **1964**, *68*, 441.

102 a) M. Charton, *J. Am. Chem. Soc*. **1969**, *91*, 615. b) M. Charton, *J. Am. Chem. Soc*. **1975**, *97*, 1552. c) M. Charton, *J. Am. Chem. Soc*. **1975**, *97*, 3691. d) M. Charton, *J. Am. Chem. Soc*. **1975**, *97*, 3694. e) M. Charton, *J. Am. Chem. Soc*. **1975**, *97*, 6159. f) M. Charton, *J. Am. Chem. Soc*. **1975**, *97*, 6472.

103 a) R. P. Hughes, J. M. Smith, L. M. Liable-Sands, T. E. Concolino, K.-C. Lam, C. Incarvito, A. L. Rheingold, *J. Chem. Soc. Dalton Trans.* **2000**, 873. b) T. E. Müller, D. M. P. Mingos, *Transition Met. Chem.* **1995**, *20*, 533. c) J. T. DeSanto, J. A. Mosbo, B. N. Storhoff, P. L. Bock, R. E. Bloss, *Inorg. Chem.* **1980**, *19*, 3086. d) M. Chin, G. L. Durst, S. R. Head, P. L. Bock, J. A. Mosbo, *J. Organomet. Chem.* **1994**, *470*, 73.

104 P. C. Möhring, N. J. Coville, *Coord. Chem. Rev.* **2006**, *250*, 18.

105 a) D. White, B. C. Taverner, P. G. L. Leach, N. J. Coville, *J. Comput. Chem.* **1993**, *14*, 1042. b) D. White, B.C. Taverner, N. J. Coville, P.W. Wade, *J. Organomet. Chem.* **1995**, *495,* 41. c) B. C. Taverner, *J. Comput. Chem.* **1996**, *17,* 1612.

106 A. Immirzi, A. Musco, *Inorg. Chim. Acta.* **1997**, *25*, L41.

107 a) L. Cavallo, A. Correa, C. Costabile, H. Jacobsen, *J. Organomet. Chem.* **2005**, *690*, 5407. b) R. H. Crabtree, *J. Organomet. Chem.* **2005**, *690*, 5451.

108 a) R. Dorta, E. D. Stevens, N. M. Scott, C. Costabile, L. Cavallo, C. D. Hoff, S. P. Nolan, *J. Am. Chem. Soc.* **2005**, *127*, 2485. b) A. Poater, B. Cosenza, A. Correa, S. Giudice, F. Ragone, V. Scarano, L. Cavallo, *Eur. J. Inorg. Chem.* **2009**, 1759. (URL http://www.molnac.unisa.it/OMtools.php).

109 a) H. Jacobsen, A. Correa, A. Poater, C. Costabile and L. Cavallo, *Coord. Chem. Rev.* **2008**, *253*, 687. b) A. C. Hillier, W. J. Sommer, B. S. Yong, J. L. Petersen, L. Cavallo, S. P. Nolan, *Organometallics* **2003**, *22*, 4322. c) R. A. Kelly, III, H. Clavier, S. Giudice, N. M. Scott, E. D. Stevens, J. Bordner, I. Samardjiev, C. D. Hoff, L. Cavallo, S. P. Nolan, *Organometallics* **2008**, *27*, 202. d) A. Poater, F. Ragone, S. Giudice, C. Costabile, R. Dorta, S. P. Nolan, L. Cavallo, *Organometallics* **2008**, *27*, 2679. e) M. S. Viciu, O. Navarro, R. F. Germaneau, R. A. Kelly, W. Sommer, N. Marion, E. D. Stevens, L. Cavallo, S. P. Nolan, *Organometallics* **2004**, *23*, 1629.

110 A. Poater, L. Cavallo, *Dalton Trans.* **2009**, 8878.

111 B. J. Dunne, R. B. Morris, A. G. Orpen, *J. Chem. Soc. Dalton Trans.* **1991**, 653.

112 K. D. Cooney, T. R. Cundari, N. W. Hoffman, K. A. Pittard, M. D. Temple, Y. Zhao, *J. Am. Chem. Soc.* **2003**, *125*, 4318.

113 C. P. Casey, G. T. Whiteker, *Isr. J. Chem.* **1990**, *30*, 299.

114 a) Z. Freixa, P. W. N. M. van Leeuwen, *Dalton Trans.* **2003**, 1890. b) M. N. Birkholz, Z. Freixa, P. W. N. M. van Leeuwen, *Chem. Soc. Rev.* **2009**, *38*, 1099. c) O. Kühl, *Can. J. Chem.* **2007**, *85*, 230.

115 P. Dierkes, P. W. N. M. van Leeuwen, *J. Chem. Soc. Dalton Trans.* **1999**, 1519.

116 P.W.N. M. van Leeuwen, P. C. J. Kamer, J.N.H. Reek and P. Dierkes, *Chem. Rev.* **2000**, *100*, 2741.

117 S. G. Koide, A. R. Barron, *Organometallics* **1996**, *15*, 2213.

118 K. Angermund, W. Baumann, E. Dinjus, R. Fornika, H. Gorls, M. Kessler, C. Kruger, W. Leitner, F. Lutz, *Chem. Eur. J.* **1997**, *3,* 755.

119 a) C. H. Suresh, *Inorg. Chem.* **2006**, *45*, 4982. b) J. Mathew, T. Tinto, C. H. Suresh, *Inorg. Chem.* **2007**, *46*, 10800.

120 a) T. L. Brown, *Inorg. Chem.* **1992**, *31*, 1286. b) T. L. Brown, K. J. Lee, *Coord. Chem. Rev.* **1993**, *128*, 89. c) D. P. White, T. L. Brown, *Inorg. Chem.* **1995**, *34*, 2718. d) R. J. Bubel, W. Douglass, D. P. White, *J. Comput. Chem.* **2000**, *21*, 239.

121 H. M. Senn, D.V. Deubel, P.E. Blöchl, A. Togni, G. Frenking, *J. Mol. Struct. (Theochem)*, **2000**, *506*, 233.

122 N. M. Scott, S. P. Nolan, *Eur. J. Inorg. Chem.* **2005**, 1815.

123 L. Perrin, E. Clot, O. Eisenstein, J. Loch, R. H. Crabtree, *Inorg. Chem.* **2001**, *40,* 5806.

124 D. G. Gusev, *Organometallics* **2009**, *28*, 763.

125 a) A. B. P. Lever, *Inorg. Chem.* **1990**, *29*, 1271. b) A. B. P. Lever, *Inorg. Chem.* **1991**, *30*, 1980.

126 a) K. Morokuma, *J. Chem. Phys.* **1971**, *55*, 1236. b) T. Ziegler, A. Rauk, *Inorg. Chem*. **1979**, *18*, 1755. c) T. Ziegler, A. Rauk, *Inorg. Chem*. **1979**, *18*, 1558.

127 T. Ziegler, A. Rauk, *Theor. Chim. Acta*, **1977**, *46*, 1.

128 a) P. S. Bagus, K. Hermann, C.W. Bauschlicher, *J. Chem. Phys.* **1984**, *80*, 4378. b) K. Hermann, P. S. Bagus, C. W. Bauschlicher, *Phys. Rev. B*, **1985**, *31*, 6371.

129 a) M. Mitoraj, A. Michalak, *J. Mol. Model.* **2007**, *13*, 347. b) A. Michalak, M. Mitoraj, T. Ziegler, *J. Phys. Chem. A* **2008**, *112*, 1933.

130 a) A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* **1988**, *88*, 899. b) F. Weinhold, C. R. Landis, V*alency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*, Cambridge University Press, Cambridge, **2005**.

131 a) M. R. Wilson, A. Prock, W. P. Giering, A. L. Fernandez, C. M. Haar, S. P. Nolan, B. M. Foxman, *Organometallics* **2002**, *21*, 2758. b) A. L. Fernandez, M. R. Wilson, A. Prock, W. P. Giering, *Organometallics* **2001**, *20*, 3429. c) A. J. Pöe, *Dalton Trans.* **2009**, 1999. d) C. Babij, A. J. Pöe, *J. Phys. Org. Chem.* **2004**, *17*, 162. e) K. A. Bunten, A. J. Pöe, *New J. Chem.* **2006**, *30*, 1638.

132 a) N. S. Antonova, J. J. Carbó, J. M. Poblet, *Organometallics* **2009**, *28*, 4283. b) N. S. Antonova, J. J. Carbó, J. M. Poblet, *Dalton Trans.* **2011**, *40*, 2975.

133 a) D. Livingstone, *Data Analysis for Chemists*, Oxford University Press, Oxford, **1995**. b) *Handbook of Chemoinformatics From Data to Knowledge*, ed. J. Gasteiger, Wiley-VCH, Weinheim, **2003**.

134 L. P. Hammett, *Chem. Rev.* **1935**, *17*, 125.

135 C. Hansch, A, Leo, R. W. Taft, *Chem*. *Rev*. **1991**, 91165.

136 K. A. Bunten, L. Chen, A. L. Fernandez, A. J. Poe, *Coord. Chem. Rev.* **2002**, *233-234*, 41.

137 R. G. Pearson, H. R. Sobel, J. Songstad, *J. Am. Chem. Soc.* **1968**, *90*, 319.

138 a) R. S. Drago, S. Joerg, *J. Am. Chem. Soc.* **1996**, *118*, 2654. b) S. Joerg, R. S. Drago, J. Sales, *Organometallics* **1998**, 17, 589.

139 J. H. Wu, G. Zhang, N. A. Porter, *Tetrahedron Lett.* **1997**, *38*, 2067.

140 a) J. J. Miller, M. S. Sigman, *Angew. Chem.* **2008**, *120*, 783. b) M. S. Sigman, J. J. Miller, *J. Org. Chem.* **2009**, *74*, 7633.

141 a) K. C. Harper, M. S. Sigman, *Proc. Natl. Acoad.Sci. U.S.A.* **2001**, *108*, 2179. b) K. C. Harper, M. S. Sigman, *Science* **2011**, *333*, 1875.

142 A. R. Katritzky, M. Kuanar, S. Slavov, C. D. Hall, M. Karelson, I. Kahn, D. A. Dobchev, *Chem. Rev.* **2010,** *110,* 5714.

143 a) S. Wold, M. Sjostrom, L. Eriksson, *Chemom. Intell. Lab. Syst*. **2001**, *58*, 109. b) P. Geladi, B. Kowlaski, *Anal. Chim. Acta* **1986**, *185*, 1.

144 L. Eriksson, E. Johansson, M. Muller, S. Wold, *J. Chemometrics* **2000**, *14*, 599.

145 D. M. Hawkins, S. C. Basak, D. Mills, *J. Chem. Inf. Comput. Sci.* **2003**, *23*, 579.

146 A. Tropsha, P. Gramatica, V. Gombar, *QSAR Comb. Sci.* **2003**, *22*, 69.

147 a) H. Wiener, *J. Am. Chem. Soc.* **1947**, *69*, 17. b) L. B. Kier, *Quant. Struct.-Act. Relat.* **1985**, *4*, 109. c) L. B. Kier, *Computational Chemical Graph Theory*, D. H. Rouvray (Ed.), Nova Science Publishers, New York **1990**.

148 a) I. G. Csizmadia, *Theory and Practice of MO Calculations on Organic Molecules*, Elsevier, Amsterdam, **1976**. b) B.W. Clare, *Theoret. Chim. Acta* **1994,** *87*, 415. c) G. Trapani, A. Carotti, M. Franco, A. Latrofa, G. Genchi, G. Liso, *Eur. J. Med. Chem.* **1993**, *28*, 13.

149 G. Occhipinti, H.-R. Bjørsvik, V. R. Jensen, *J. Am. Chem. Soc.* **2006**, *128*, 6952.

150 a) E. Burello, G. Rothenberg, *Adv. Synth. Catal.* **2005**, *347*, 1969. b) J. A. Hageman, J. A. Westerhuis, H.-W. Frühauf, G. Rothenberg, *Adv. Synth. Catal.* **2006**, *348*, 361. c) E. Burello, G. Rothenberg, *Int. J. Mol. Sci.* **2006**, *7*, 375. d) A. G. Maldonado, J. A. Hageman, S. Mastroianni, G. Rothenberg, *Adv. Synth. Catal.* **2009**, *351*, 387. e) Z. Strassberger, M. Mooijman, E. Ruijter, A.H. Alberts, A.G. Maldonado, R.V.A. Orru, G. Rothenberg, Adv. Synth. Catal. **2010**, *352*, 2201.

151 a) S. Chavali, B. Lin, D. C. Miller, K. V. Camarda, *Comp. Chem. Eng.* **2004**, *28*, 605. b) B. Lin, S. Chavali, K. Camarda, D. C. Miller, *Comp. Chem. Eng.* **2005**, *29*, 337.

152 a) J. L. Melville, J. D. Hirst, *J. Chem. Inf. Model*. **2007**, *47*, 626. b) B. M. Spowage, C. L. Bruce, J. D. Hirst, *J. Cheminf*. **2009**, *1*, 22.

153 J. Gasteiger, M. Marsili, *Tetrahedron* **1980**, *36*, 3219.

154 S. A. Wildman, G. M. Crippen, *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 868.

155 T. J. Hou, K. Xia, W. Zhang, X. J. Xu, *J. Chem. Inf. Comput. Sci*. **2004**, *44*, 266.

156 G. Moreau, P. Broto, *Nouv. J. Chim*. **1980**, *4*, 359.

157 C. Jiang, Y. Li, Q. Tian, T. You, *J. Chem. Inf. Comp. Sci.* **2003**, *43*, 1876.

158 a) K. B. Lipkowitz, S. Schefzick, D. Avnir, *J. Am. Chem. Soc.* **2001**, *123*, 6710. b) S. Alvarez, S. Schefzick, K. B. Lipkowitz, D. Avnir, *Chem. Eur. J.* **2003**, *9*, 5832.

159 R. D. III. Cramer, D. E. Patterson, J. D. Jeffrey, *J. Am. Chem. Soc.* **1988**, *110*, 5959.

160 S. Dixon, K. M. Jr. Merz, G. Lauri, J. C. Ianni, *J. Comp. Chem*. **2004**, *26*, 23.

161 a) M. C. Kozlowski, S. L. Dixon, M. Panda, G. Lauri, *J. Am. Chem. Soc.* **2003**, *125*, 6614. b) P.-W. Phuan, J. C. Ianni, M. C. Kozlowski, *J. Am. Chem. Soc.* **2004**, *126*, 15473.

162 M. Urbano-Cuadrado, J. J. Carbó, A. G. Maldonado, C. Bó, *J. Chem. Inf. Model.* **2007**, *47*, 2228.

163 a) S. Sciabola, A. Alex, P. D. Higginson, J. C. Mitchell, M. J. Snowden, I. Morao, *J. Org. Chem*. **2005**, *70*, 9025. b) J. C. Ianni, V. Annamalai, P.-W. Phuan, M. C. Kozlowski, *Angew. Chem. Int. Ed.* **2006**, *45*, 5502.

164 a) K. B. Lipkowitz, M. Pradhan, *J. Org. Chem.* **2003**, *68*, 4648. b) M. Hoogenraad, G. M. Klaus, N. Elders, S. M. Hooijschuur, B. McKay, A. A. Smith, E. W. P. Damen, *Tetrahedron: Asymmetry* **2004**, *15*, 519. c) S. E. Denmark, N. D. Gould, L. M. Wolf, *J. Org. Chem.* **2011**, *76*, 4337.

165 J. L. Melville, K. R. J. Lovelock, C. Wilson, B. Allbutt, E. K. Burke, B. Lygo, J. D. Hirst, *J. Chem. Inf. Model.* **2005**, *45*, 971.

166 G. Klebe, U. Abraham, T. Mietzner, *J. Med. Chem.* **1994**, *37*, 4130.

167 V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, R. K. Robins, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163.

168 a) V. Ravichandran, B. R. P. Kumar, S. Sankar, R. K. Agrawal, *Med. Chem. Res.* **2008**, *17*, 1. b) S. J. Bang, S. J. Cho, Bull. Korean, *Chem. Soc.* **2004**, *25*, 1525. c) L. V. Yuyin, Y. Chunsheng, L. Hongyan, Y. Zhongsheng, W. Yang, *J. Environ. Sci.* **2008**, *20*, 1433.

169 a) V. Murugesan, Y. S. Prabhakar S. B. Katti, *J. Mol. Graph. Model.* **2009**, *27*, 735. b) C. Kunick, K. Lauenroth, K. Wieking, X. Xie, C. Schultz, R. Gussio, D. Zaharevitz, M. Leost, L. Meijer, A. Weber, F. S. Jørgensen, T. Lemcke, *J. Med. Chem.* **2004**, *47*, 22. c) R. P. Verma, C. Hansch, *Chem. Rev.* **2009**, *109*, 213.

170 a) P. Broto, G. Moreau, C. Vandycke, *Eur. J. Med. Chem.* **1984**, *19*, 66. b) P. Broto, G. Moreau, C. Vandycke, *Eur. J. Med. Chem.* **1984**, *19*, 79.

171 a) M. Pastor, G. Cruciani, I. McLay, S. Pickett, S. Clementi, *J. Med. Chem*. **2000**, *43*, 3233. b) F. Fontaine, M. Pastor, F. Sanz, *J. Med. Chem*. **2004**, *47*, 2805. c) F. Fontaine, M. Pastor, I. Zamora, F. Sanz, *J. Med. Chem.* **2005**, *48*, 2687.

Chapter 2

DFT study on Ammonia Activation by μ_3-Alkylidyne Fragments Supported on a Titanium Molecular Oxide Model

Chapter 2

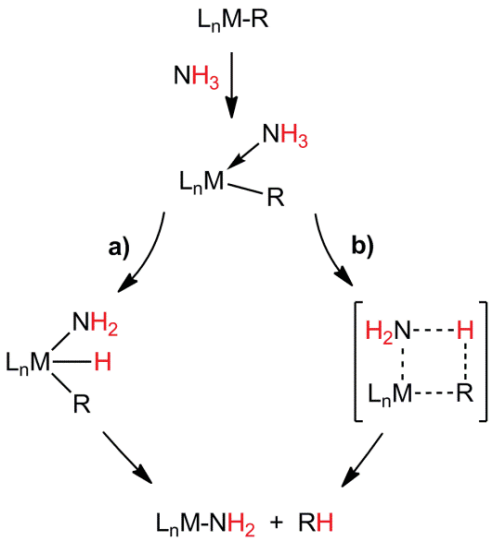
DFT study on Ammonia Activation by μ_3-Alkylidyne Fragments Supported on a Titanium Molecular Oxide Model

This Chapter reports a DFT study on the N-H σ-bond activation of ammonia by the μ_3-alkylidyne titanium species using the [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-CH)] model complex. Ammonolysis of the μ_3-alkylidyne derivatives [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-CR)] [R = H (**1**), Me(**2**)] produces a trinuclear oxonitride species, [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-N)] (**3**), via methane or ethane elimination, respectively. The calculations show the multiple bond character of the terminal amido and the bridging nature of imido ligand in the reaction intermediates. For the product [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-N)], calculations allow understanding its electronic structure, and rationalizing its chemical behaviour by comparison with electronic structures of the isoelectronic species [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-CH)] and [{Ti(η^5-C_5H_5)(μ-NH)}$_3$(μ_3-N)]. They also indicate that the sequential ammonia N-H bonds activation process goes successively downhill in energy and occurs via direct hydron transfer to the alkylidyne group on organometallic oxides **1** and **2**. The mechanism can be divided in three stages: i) coordination of ammonia to a titanium centre, in a *trans* disposition with respect to the alkylidyne group, and then the isomerization to adopt the *cis* arrangement, allowing the direct hydron migration to the μ_3-alkylidyne group to yield the amido μ-alkylidene complexes **4** and **5**; ii) hydron migration from the amido moiety to the alkylidene group; and finally iii) hydron migration from the μ-imido complex to the alkyl group to afford the oxo μ_3-nitrido titanium complex **3** with alkane elimination.

2.1. Introduction

As mentioned in the *Chapter 1*, the activation of the ammonia molecule is of great interest because it is a synthetic precursor for a wide range of commercially useful products. Due to the high N-H bond dissociation energy (~104±2 kcal mol^{-1})[1] and the scarce tendency to form N-H σ complexes,[2] there are relatively few examples of ammonia N-H bond activation processes by direct use of ammonia. Most of them have been prepared via the oxidative addition of ammonia to a transition-metal complex[3] (see Scheme 2.1, path a), but also the N-H bond can activated via metal-ligand without involving oxidative addition of the metal centre[4] (see Scheme 2.1, path b). We have recapitulated the most representative experimental findings on the *Chapter 1*.

Scheme 2.1.



Besides the studies on synthetic chemistry, quantum mechanical methods have provided some insight into this complex reactivity. The interactions of bare metals and bare metal cations with ammonia, where the N-H activation occurs via oxidative addition to the metal, have been the subject of a number of theoretical investigations.[5-6] To extrapolate the results of ion chemistry in gas phase to organometallic chemistry in solution, one has to analyze the trends from the variation of transition metal centre. Sicilia and co-workers have systematically analyzed the species expected in the dehydrogenation process of ammonia by bare first-row transition-metal cations (M^+-NH_3, H-M^+-NH_2, M^+-NH_2, M^+-NH).[5] For early transition metals, such as Ti^+, the metal-amido and imido binding energies are larger, which can be explained in terms of multiple bond character; that is, the NH_2 and the NH groups are doubly and triply bonded to the metal.[5c] Moreover, the relative energy of the transition states for N-H bond-breaking correlates with the stability of intermediate products, and, consequently in the case of Ti^+, the process does not require energy in excess of reactants. Analogously for bare second row transition metals, Sigbhan and co-workers found that for the metal atoms on the left the

ammonia activation has lower barriers than for metals on the right.[6a] However, in both cases, the authors considered gas-phase collisionless conditions, in which the energy of reactants is considered to be the threshold energy. Consequently, the formation of low-energy lying intermediates via ammonia coordination does not affect the kinetics of the system because the excess of energy cannot dissipate in the media. This is not the case in solution chemistry, where the stability of ammonia complex can prevent its N-H activation. Morokuma and co-workers have analyzed the activation of N-H bonds by the organometallic complex [CpRh(CO)(H)(R)] [R = H_2, CH_4, NH_3, H_2O, SiH_4], focusing on the trends among different H-R bonds.[7] For this late transition metal complex, the oxidative addition of N-H bond is significantly more difficult than for other model H-R bonds such as $HSiH_3$, H-H, $H-CH_3$, or H-OH, because of the depth energy level of the NH_3 complex. Thus, in solution conditions, it will be very difficult to activate the N-H bond of ammonia, explaining why the reported examples of oxidative addition of N-H bond of ammonia are relatively scarce.

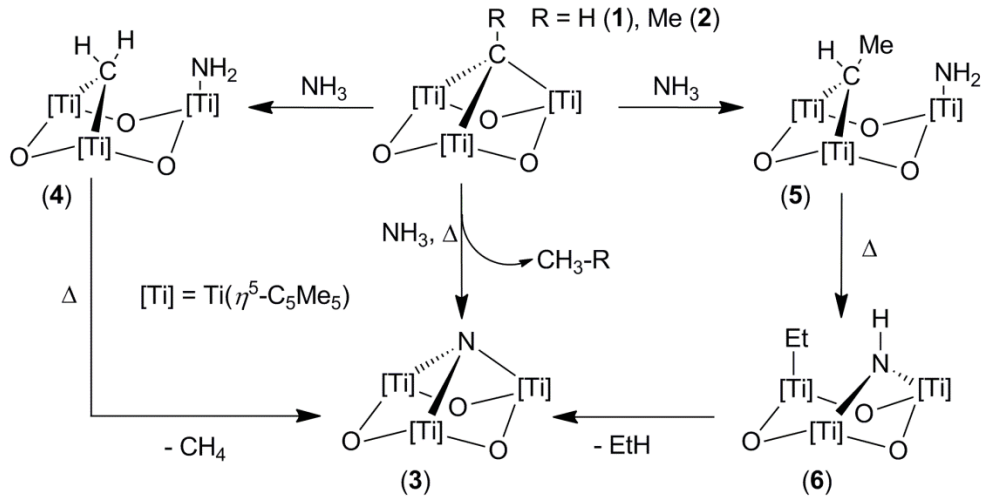
Some theoretical studies have alternatively focused on the N-H bond activation via addition across M=X multiple bonds. Matsubara investigated the hypothetical activation of the N-H bond of ammonia at the double bond between palladium and the ligand, Pd=X (X = Sn, Si, C).[8] Sakaki and co-workers also investigated a hypothetical process, in which the N-H bond is activated by Ti-imido and Ti-alkylidyne complexes, comparing them with the activation of C-H bond of methane.[9] Both groups proposed mechanisms involving coordination of NH_3 by donation of its lone pair, followed by heterolytic breaking of the N-H bond and migration of the hydrogen as hydron. Very interestedly, Sakaki and co-workers predicted that the N-H bond activation of ammonia can be achieved by terminal alkylidyne-titanium complexes. To the best of our knowledge, such process had been never observed until now. The ammonia activation is accomplished by a related bridging μ_3-alkylidyne titanium complex, further supporting previous theoretical prediction. This type of chemistry can also be related to the above-mentioned N-H bond activation by PNP-Ru complexes,[10] with the hydrogen transfer to a multiple bond of the pyridine ligand. The authors complemented their study with DFT calculations to support their claims.[10] Finally, previous theoretical work on group 4 molecular oxide models has provided us with some insight into the activation of σ bonds such as O-H.[11-12] For the same [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-CR)] complexes reported here, calculations showed that the activation of O-H bonds of silanols and alcohols occurs easily via heterolytic bond-breaking with the assistance of titanium bridging oxygen atoms.[12]

The work in this Chapter has been performed in collaboration with the experimental group of Prof. Mena (Departamento de Química Inorgánica, Universidad de Alcalá, Spain). First, in the Section 2.2, we summarize the results on the reactions of ammonia with [{Ti (η^5-C_5Me_5)(μ-O)}$_3$(μ_3-CR)] [R = H (**1**), Me (**2**)] to give rise to the first example of ammonia N-H bonds activation by μ_3-alkylidyne transition metal derivatives. In the next sections we show that, according to the DFT calculations performed on the [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-CH)] (**A**) model, the ammonia degradation occurs via direct hydron transfer to the alkylidyne group on the organometallic titanium oxide. To our knowledge, there is no other reported process in which the sequential activation of the three N-H bonds in ammonia has been studied within the same molecular system.

2.2. Experimental background

The group of Prof. M. Mena has studied the ammonolysis of the μ_3-alkylidyne derivatives [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-CR)] [R = H (**1**), Me (**2**)] producing a trinuclear oxonitride species, [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-N)] (**3**), via methane or ethane elimination, respectively (see Scheme 2.2). During the course of the reaction, the intermediates amido μ-alkylidene [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ-CHR)(NH_2)] [(R = H (**4**), Me (**5**)] and μ-imido ethyl species [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ-NH)Et] (**6**) were characterized and/or isolated.

Scheme 2.2.



Upon monitoring the ammonolysis of [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-CR)] [R = H (**1**), Me (**2**)] by ^{1}H NMR in benzene-d_6 solutions at room temperature, the presence of the intermediate amido μ-alkylidene species [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ-CHR)(NH_2)] [R = H (**4**), Me (**5**)], μ-imido ethyl [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ-NH)Et] (**6**), and the final μ_3-nitrido product [{Ti(η^5-C_5Me_5)(μ-O)}3(μ_3-N)] (**3**) was revealed as a result of the sequential ammonia N-H bond activation process. A later thermal treatment of the mixture reaction at ~75 °C for 3 days afforded [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-N)] (**3**) as the unique product (see Scheme 2.2).

The μ_3-nitrido complex **3** was isolated as green crystals by thermal treatment of **1** or **2** in toluene with an excess of ammonia at 75 °C for 72 h in 83% yield. Complex **3** probed to be stable under argon atmosphere at room temperature and practically soluble in most common solvents (e.g., toluene, hexane, tetrahydrofurane, chloroform). It is noteworthy the single set of signals corresponding to the η^5-C_5Me_5 ligands indicating the equivalence of all of them in the ^{1}H and ^{13}C NMR spectra of complex **3**, retaining the C_{3v} symmetry of the starting materials **1** and **2**.

During the course of the reaction, the species corresponding to the single and double ammonia N-H bond activation were detected by NMR spectroscopy. In an attempt to isolate them, the reaction conditions (temperature, time of reaction, ammonia exposure, etc.) were modified. Toluene solutions of **1** or **2** placed in an amber-stained Carius tube were treated with ammonia while maintaining the solution at 0 °C. After the excess of NH_3 was removed and the mixture was stirred at room temperature, complete conversion was achieved overnight to afford the amido μ-alkylidene complexes [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ-CHR)(NH_2)] [(R = H (**4**), Me (**5**)], as a result of the activation of one N-H bond. Compound **4** was isolated as a red solid in 34% yield, probed to be soluble in the habitual solvents (e.g., toluene, hexane), but it is unstable both in solution and in the solid state evolving to provide **3** and other uncharacterized species. The analogous compound **5** was only detected by ^{1}H NMR.

The NMR spectra of complexes **4** and **5** display two types of η^5-C_5Me_5 groups in a 2:1 ratio, consistent with a C_s symmetry and in contrast with the C_{3v} symmetry of the starting materials **1** and **2**. In the ^{1}H NMR spectra, **4** exhibits an AB spin system at δ = 5.99 and 6.38 ($^2J_{H-H}$ = 9.3 Hz) assigned to a μ-methylene moiety, while complex **5** shows a doublet at δ = 2.27 and a quartet at δ = 6.27 ($^3J_{H-H}$ = 7.8 Hz) corresponding to a μ-ethylidene group. Also, narrow signals at δ = 5.11 and 5.26 attributed to the amido fragment Ti-NH_2 were observed for complexes **4** and **5**, respectively.

The ^{13}C NMR spectrum of **4** reveals as an important feature the resonance of the μ-methylidene group as a triplet at δ = 185.5, demonstrating a strong high field shift with respect to the starting μ_3-methylidyne product **1** [μ_3-CH, δ = 383.8]. The chemical shift of the signal due to the μ-alkylidene carbon in **4** follows a trend similar to those observed for homodinuclear[12,13b,14,15] and heterodinuclear[16] μ-methylene titanium systems reported in the literature.

In an attempt to isolate the compounds derived of the double ammonia N-H bonds activation, a synthetic approach similar to that of **4** was used, increasing the reaction time. The reaction of **2** with ammonia allows to obtain the μ-imido ethyl compound [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ-NH)Et] (**6**) as a violet solid in 83% yield. The preparation of the analogous μ-imido methyl compound from the corresponding amounts of **1** and ammonia affords a mixture of products in which the μ-imido methyl compound together with **3** and other ammonolysis process derivatives were identified. Unfortunately, the analogous to complex **6** was not isolated, and no further characterization was possible.

In the ^{1}H NMR spectrum of **6**, two signals corresponding to the η^5-C_5Me_5 ligands in a 2:1 ratio indicate a C_s symmetry for this species. In addition, a quartet at δ = 0.98 and a triplet at δ = 1.55, assigned to an ethyl bonded to a titanium atom, together with a broad signal at δ = 10.77, attributed to a μ-NH group, were detected. The ^{13}C NMR spectra display as the most remarkable feature the resonance of the alkyl (C_α) moiety as a triplet at δ = 58.7, strong high field shifted with respect to the μ_3-ethylidyne starting material **2** [δ(μ_3-CMe) = 401.7].[17]

Green crystals of [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-N)] (**3**) were obtained in a glovebox at room temperature from a concentrated toluene solution after slow evaporation. The molecular structure of **3** is shown in Figure 2.1, the starting compound **2**, and the isoelectronic species [{Ti(η^5-C_5Me_5)(μ-NH)}$_3$(μ_3-N)].[18] Complex **3** has a crystallographic mirror plane, which contains Ti(1), N(1), and O(22). The three pseudotetrahedral titanium atoms are the apexes of a hypothetical equilateral triangle capped by a μ_3-nitride moiety and bridged by the three oxygen atoms to form a six-membered [Ti_3O_3] ring in a chair disposition. The Ti-N bond distances [*av.* 2.066(4) Å] in **3** are clearly longer than those found for the isoelectronic azaderivative [*av.* 1.912(1) Å][18] The steric restrictions imposed by the [Ti_3O_3] ring make the Ti-N-Ti angles clearly narrower than those found for the tertiary organic amines (~115°) and also than those of the mentioned azaderivative. The nitrogen atom is located 1.27(1) Å above the plane formed by the titanium atoms. This distance is slightly shorter than that observed in the starting μ_3-alkylidine product **2** [1.35(1) Å][13] but longer than that presented by the apical nitrido fragment in [({Ti{η^5-C_5Me_5)-(μ-NH)}$_3$(μ_3-N)] (1.03(1) Å).[18]

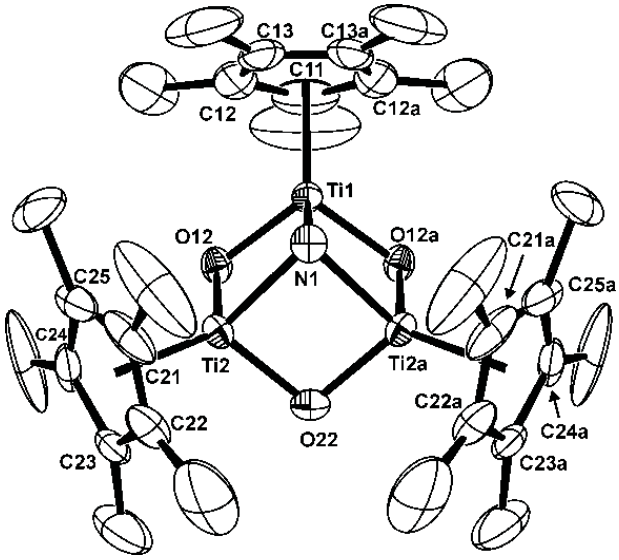


Figure 2.1. Molecular structure of complex **3** depicting thermal ellipsoids at the 50% probability level. Symmetry transformations used to generate equivalent atoms: a *x*, *y*, *-z*, +1/2.

On the other hand, the Ti-O bond distances [*av.* 1.853(1) Å] in **3** are very similar to those observed in other not only trinuclear,[13a,15] but also dinuclear, [{Ti(η^5-C_5Me_5)Me(η^2-MeNNC-Ph_2)}(μ-O)(Ti(η^5-C_5Me_5)Me_2)] (1.83 Å),[19] and tetranuclear oxocomplexes, [{Ti(η^5-C_5Me_5)}$_4$(μ-$O)_6$}] (1.84 Å).[20] Additionally, Ti-O-Ti angles [*av.* 99.0(1)°] are similar to those found in the precubane **2** [*av.* 100.8°][13a] or in the cube-type derivative [{$Mo(CO)_3$}(μ_3-O)$_3${Ti(η^5-C_5Me_5)}$_3$(μ_3-CMe)] (100.7),[21] but remarkably narrower than those observed in other titanium trinuclear

systems where the metal centres are only bridged by oxygen atoms, as, for example, in [{Ti(η^5-C_5Me_5) (μ-O)X}$_3$] (X = Cl 133.9°,[22] Br 134.0°)[23] and [{Ti(η^5-C_5Me_5)(μ-O)Me}$_3$] (132.9°).[24]

2.3. Computational details

All calculations were performed with the Gaussian 03 series of programs[25] within the framework of the Density Functional Theory (DFT)[26] using the B3LYP functional.[27] A quasi-relativistic effective core potential operator was used to represent the 10 innermost electrons of the Ti atom.[28] The basis set for Ti atoms was that associated with the pseudopotential,[28] with a standard double-ξ LANL2DZ contraction.[25] The 6-31G(d) basis set was used for C, N and O atoms,[29] the 6-31G(p) basis set was used for the hydrogen of ammonia and the alkylidyne group,[29] whereas the 6-31G basis set was used for the other hydrogens.[29] Geometry optimizations were carried out without any symmetry restrictions and all stationary points were optimized with analytical first derivatives. Transition states were characterized by single imaginary frequency, whose normal mode corresponded to the expected motion.

2.4. Results and discussion

2.4.1. Structural, electronic and energetic analysis of the amido-, imido- and nitrido-titanium species

To investigate the successive N-H σ-bond activation of ammonia by the molecular oxides **1** and **2** at a mechanistic level, a theoretical analysis was carried out on the model complex [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-CH)] (**A**). The study of the reaction was started with the search for minimum energy configurations at different intermediates of the ammonolysis process of complex **A**. Thus, we found three minima **D**, **E**, and **F**, which correspond respectively to the three experimentally characterized species, that is, the amido-titanium complex **4**, the μ-imido-titanium complex **6**, and the μ_3-nitrido-titanium complex **3**. Figure 2.2 shows the molecular structures and the main geometric parameters of **D,** and **E**, optimized without any symmetry constrains.

The first N-H bond activation via hydron transfer to the alkylidyne group leads to a μ-alkylidene amido complex. As commented above, the N-H bond activation of ammonia by Ti-alkylidyne complexes was predicted by theoretical calculations,[9] but it had not been observed until now. In structure **D**, we have considered a ligand arrangement analogous to that obtained by reaction of **1** with diphenylamine to give the X-ray characterized μ-ethylidene diphenylamido complex [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ-CHMe)(NPh_2)].[15] Both, the alkylidene group bridging two titanium atoms and the amido ligand located on the third titanium centre are situated on the same side of the plane formed by the three titanium atoms. The nature as a minimum of this proposed species **D** was confirmed by vibrational analysis showing no imaginary frequencies. Moreover, the formation of **D** from **A** and NH_3 is computed to be energetically favorable by 62 kJ·mol^{-1}.

The $[Ti_3O_3]$ core in **D** resembles that of the diphenylamido complex $[\{Ti\{\eta^5\text{-}C_5Me_5\}(\mu\text{-}O)\}_3(\mu_3\text{-}CHMe)NPh_2]$,[15] the calculated Ti-O-Ti angle (100.2°) corresponding to the titanium atoms bridged by the μ-alkylidene group is smaller than the other Ti-O-Ti angles (*av.* 128.4°), the values being very close to those of the diphenylamido complex (99.4(1)° and *av.* 127.4°, respectively). Also, the computed average Ti-O (1.83 Å) and Ti-μ-C (2.10 Å) bond lengths in **D** are close to those of the diphenylamido complex (1.84 and 2.11 Å, respectively). On the other hand, **D** exhibits a Ti-N bond length of 1.894 Å, which is somewhat shorter than that for the diphenylamido ligand (2.007 (4) Å). The Ti-N bond elongation in the latter complex can be related to the steric repulsion between the phenyl substituents of the amido ligand and the pentamethylcyclopentadienyl ligands; this steric repulsion is clearly weaker in the ammonia-derived complex. The nitrogen atom in **D** shows a planar environment, allowing a strong p_π-d_π bonding interaction between a vacant d orbital at the Ti centre and the amido N-atom lone pair, perpendicular to the NH_2 plane. Indeed, we found a molecular orbital (HOMO-4) formally corresponding to this interaction (see Figure 2.3). Thus, the NH_2 moiety acts as a four electron donor ligand, the amido group being doubly bonded to titanium as previously suggested by calculations on bare early transition-metal cations.[5c]

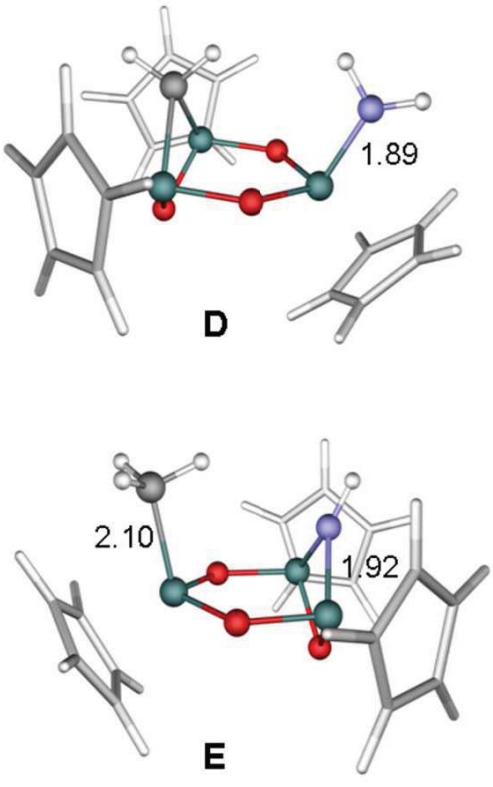


Figure 2.2. Computed molecular structures (Å) of the amido (**D**) and μ-imido (**E**) intermediates involved in successive N-H bonds activation of ammonia.

The second N-H bond activation leads to an alkyl μ-imido complex via migration of a hydron from the amido ligand to the μ-alkylidene group. In **E**, the methyl group and the imido moiety bridging the other titanium atoms are in a *cis* disposition (see Figure 2.2). The nitrogen atom shows a planar environment with Ti-N distances of 1.92 Å (*av.*), close to the value of 1.95 Å (*av.*) found in the phenyl imido moiety of the complex [{Ti{η^5-C_5Me_5)(μ-O)}$_3$(CH=CHPh)(μ-NPh)].[30] The analysis of frontier molecular orbitals of **E** shows that the first set of unoccupied orbitals are combinations of titanium d atomic orbitals, indicating that the titanium centres remain formally fully oxidized. The HOMO-1 and HOMO-2 orbitals are mixtures between π orbitals of Cp ligands and a nonbonding nitrogen p-type orbital perpendicular to the plane of μ-imido group. This orbital assessment indicates the presence of a lone-pair at the imido group, which is not involved in the Ti-N bonds. On going from **D** to **E**, calculations predict a stabilization energy of 48 kJ·mol^{-1} and confirm the nature as a minimum of **E** via frequency analysis.

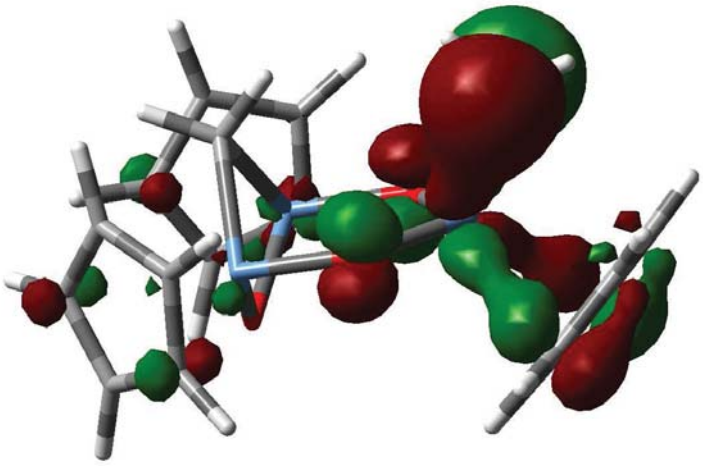


Figure 2.3. Representation of the p_π-d_π-type orbital HOMO-4 for the amido group in intermediate **D**.

Third, the final step in the sequential N-H bonds activation process leads to methane elimination and the μ_3-nitrido oxo model **F**, already characterized by X-ray analysis as complex **3**. The computed geometry for **F** compares reasonably well with the crystallographic data. The computed Ti-O, Ti-N and Ti···Ti distances in **F** (*av.* 1.84, 1.92 and 2.75 Å), as well as the Ti-O-Ti and Ti-N-Ti angles (*av.* 96.6° and 91.2°), follow the same trend as the experimental values (*av.* 1.854, 2.066 and 2.820 Å; and 99.1° and 86.5°, respectively). The formation of the final complex **F** from **E** is also computed to be exothermic (-62 kJ·mol^{-1}). We can conclude that, starting at **A**, the successive N-H activation of ammonia proceed through different stages (**A** → **D** → **E** → **F**), which go downhill in energy (0.0, -62, -110 and -172 kJ·mol^{-1}) providing the thermodynamic driving force for the reaction.

To understand the electronic structure of compound **3**, we compared the molecular orbital structure of model complex [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-N)] (**F**) with those of the previously studied oxoligand [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-CH)] (**A**) and imido-nitrido derivative [{Ti(η^5-C_5H_5)(μ-NH)}$_3$(μ_3-N)] (**G**) (Figure 2.4). To allow a more clear interpretation of frontier orbitals, the pentamethylcyclopentadienyl ligands on the real system were modeled by cyclopentadienyl groups and C_{3v} symmetry was assumed in calculations. Figure 2.4 shows the important molecular orbitals (MO) for these complexes.

The formation of oxonitrido complex **F** from the μ_3-alkylidyne oxocomplex **A** is not accompanied by changes in the oxidation state of the titanium metal centres, which remains +4, the first set of unoccupied orbitals being a combination of titanium d orbitals. The orbitals formally corresponding to the lone-pair electrons of the basal oxo groups (2e and 2a_1) remain qualitatively similar on going from **A** to **F**. We only notice that the 2e orbitals, which are mainly formed by p-type orbitals of the basal oxygen, stabilize 0.5 eV due to some mixing with the apical p-type nitrido orbitals. Thus, we expect a similar coordination behavior through basal oxygens for complexes **F** and **A**, as we will describe below.

The replacement of the alkylidyne group in **A** by the nitrido group in **F** generates a new Lewis basic centre through the available molecular orbitals 1e and 1a_1 (see Figure 2.4). These MOs are similar to those calculated for **G**,[31] but they show an inversion on the relative order of the energy levels. In **F**, the calculated MOs labeled 1e (HOMO-6 and -7), being nitrogen p-type orbitals perpendicular to the C_3 axis with some bonding contribution of d-type titanium orbitals, lay higher in energy than that MO labeled 1a_1(HOMO-8), formally corresponding to the lone-pair electrons of the apical nitrogen atom. Moreover, the molecular orbitals available for coordination through the apical nitride group are higher in energy than those through the basal oxo groups. This contrasts with the situation in the parent imido-nitrido titanium complex **G**, containing imido groups at the basal position instead of oxo groups. This could have significant implications on the chemical behavior of **3**, pointing to a more active role for the apical nitrogen atom in the chemical reactivity and preferred apical coordination site provided that repulsive steric effects are not into play.

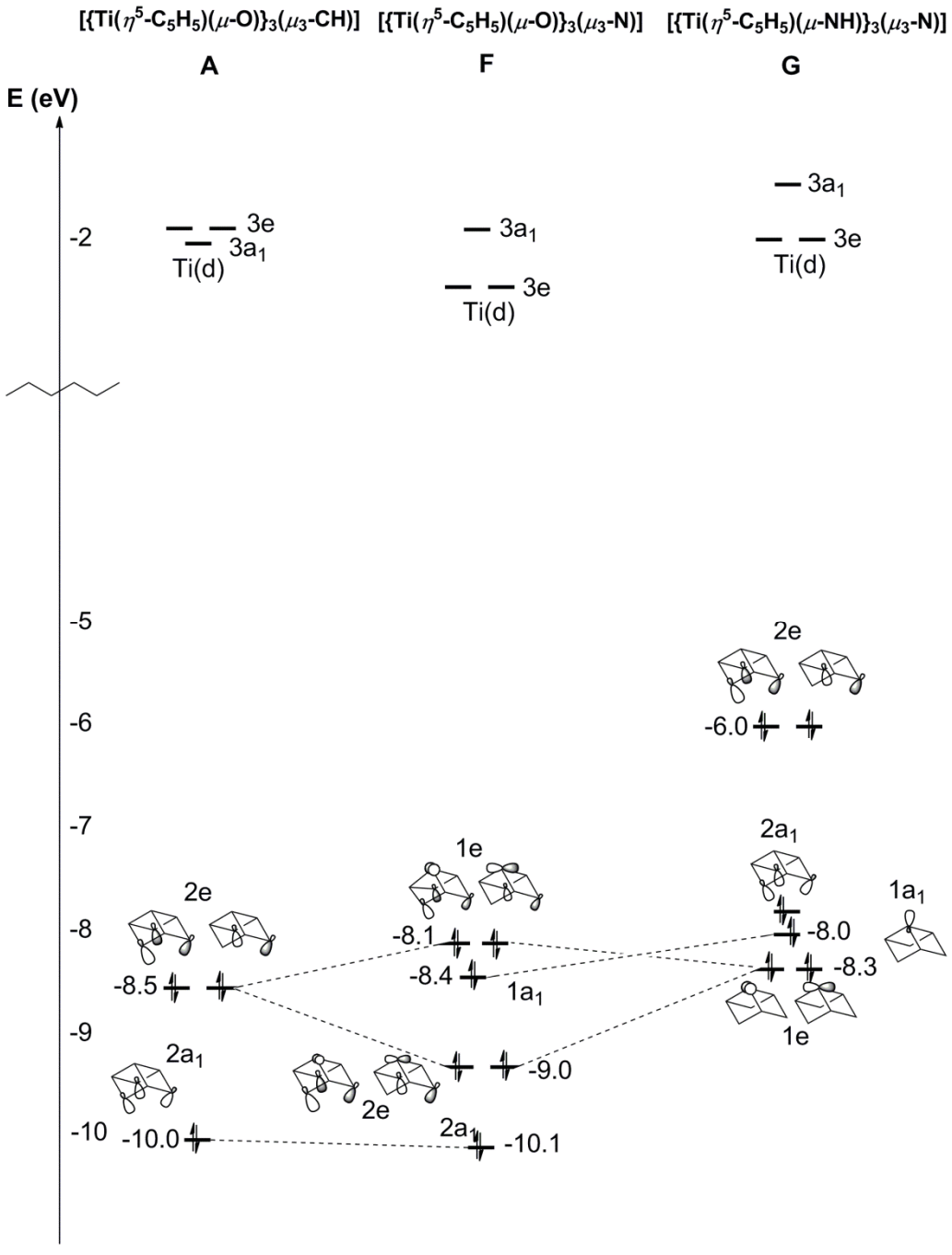


Figure 2.4. Frontier molecular orbitals diagram for complexes **A**, **F** and **G**.

2.4.2. Mechanism for the three N-H bonds activation of ammonia

The energy profiles of the different studied mechanisms are summarized in Figure 2.5. For the activation of the first N-H bond, we initially examined a mechanism analogous to that previously reported for the hydron transfer processes between silanols (or alcohols) and the μ_3-methylidyne complex **A**.[12] The first step of this mechanism would involve the hydron transfer from the substrate to one of the oxygen atoms of the Ti_3O_3 ring. Yet the results show qualitative differences with respect to those previously found in the analogous processes involving silanols and alcohols. The approach of NH_3 to the complex **A** does not form any adduct by hydrogen bonding N-H···O, but a complex **B** (see Figure 2.5) with the ammonia coordinated to titanium (Ti-N bond distance 2.39 Å). The short Ti-NH_3 distance indicates that a typical coordinative bond has been formed, in contrast with the weak initial interactions between the titanium atoms and the silanols[12a] or alcohols.[12b] The coordination of NH_3 stabilizes the complex by 49 kJ·mol^{-1}.

From **B** (see Figure 2.5), the computed energy barrier for the ammonia hydron transfer to one oxo group is relatively high (111 kJ·mol^{-1}), 2-folds higher than those computed for the CH_3OH (56 kJ·mol^{-1}) or SiH_3OH (42 kJ·mol^{-1}). Also, the process from **B** to the amido intermediate **C** is computed to be quite endothermic +107 kJ·mol^{-1}, leading to a very low inverse energy barrier of about 4 kJ·mol^{-1}. All these data indicate that, in this case, the hydron transfer to one of the oxygen atoms is unlikely. The differences found can be explained from the lower acidity of the hydrogen atoms in ammonia in comparison with those of alcohols and silanols.

Alternatively, we considered a mechanism, in which the hydron transfer to the alkylidyne group occurs directly from the coordinated ammonia. Actually, we located two ammonia titanium intermediates, **B** and **B'**, which result from the approach of the ammonia to two different sides of the titanium centre. The ammonia ligand in **B** is located *trans* with respect to the alkylidyne group, whereas in **B'** it is *cis*. The *trans* disposition of the ligand is quite stable, but the *cis* is not; their energies relative to the reactants are -49 and +15 kJ·mol^{-1}, respectively (see Figure 2.5). Moreover, the NH_3 approach through the basal region of [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-CH)] (**A**) complex yielding the *trans* isomer would be less sterically demanding than the approach through the apical one yielding the *cis* isomer. Thus, although the hydron migration to the μ_3-alkylidyne group should occur from **B'**, the interaction of NH_3 and **A** most probably forms the *trans* ammonia intermediate **B**. Next, the **B** complex can undergo an intramolecular rearrangement of the ammonia and the cyclopentadienyl ligands that are not involved in the Ti_3 skeleton to yield **B'**. In both complexes, the ammonia titanium centre can be regarded as a trigonal bipyramid structure, with the NH_3 and Cp ligands in axial-equatorial positions for **B** and equatorial-axial positions for **B'**. The transition state connecting **B** and **B'** species (**TS$_{rot}$**) is a square-based pyramid-type structure, in which the NH_3 and Cp ligands occupy basal positions. The computed energy barrier for the NH_3 rearrangement is modest, amounting 75 kJ·mol^{-1} (see Figure 2.5). Note that an analogous process was proposed for the rearrangement of alkynyl and cyclopentadienyl ligands in a trinuclear imido-nitrido titanium complex, the energy barrier being of the same order of magnitude (62 kJ·mol^{-1}).[32] Once **B'** is formed, the hydrogens of ammonia are facing to its neighboring alkylidyne group, allowing the hydron migration to

form the amido μ-alkylidene-titanium complex **D**. The estimated migration energy barrier is only 23 kJ·mol^{-1}. Thus, the estimated energy barrier for the overall process involving ligand rearrangement and direct hydron transfer from NH_3 amounts 87 kJ·mol^{-1}. This value is significantly lower than that computed for hydron transfer to the oxo groups (111 kJ·mol^{-1}), and consequently in this case, we can discard the participation of the oxo groups in the N-H activation.

The interaction of NH_3 with the titanium centre in **B'** elongates the Ti-C_{apical} bond length with respect to that in the reactant **A** by 0.28 Å. On going from **A** to **B'**, we also observed that the energy of the HOMO, which is mainly formed by a bonding combination of p orbitals of the bridging carbon atom and d orbitals of the titanium atoms in **B'**, rises significantly from -6.5 to -5.5 eV. This suggests an increase of the basicity of the alkylidyne carbon atom, enough to trap the relatively low acidic hydrogen of ammonia. Recently, Sakaki and cowokers have computationally investigated the hypothetical N-H σ-bond activation by the titanium alkylidyne complex [(PNP)Ti ≡ $CSiMe_3$] (PNP = N-[2-$(PH_2)_2$-phenyl]$_2^-$), reporting an activation barrier of 31 kJ·mol^{-1} at DFT level,[9] similar to that computed here for the transformation of **B'** to **D** (23 kJ·mol^{-1}). The authors concluded that the N-H activation can be easily achieved by the titanium alkylidyne complex, although to the best of our knowledge it had not been yet experimentally observed. The experimental and computational evidences provide a further proof that the Ti(IV) alkylidyne complexes activate the N-H bonds of ammonia.

In the second N-H activation, the hydrogen atom at the amido group of **D** could directly migrate to the hydrocarbon moiety to afford the μ-imido-titanium complex **E** (Scheme 2.3i and dashed lines in Figure 2.5). The energy barrier (**TS_{DE}**) for the step from **D** to **E** is calculated to be 151 kJ·mol^{-1}, which seems to be too high for a reaction occurring under mild experimental conditions. To achieve this transformation, several processes should occur simultaneously: breaking of the N-H and Ti-$C_{alkylidene}$ bonds, the formation of new C-H and Ti-N bonds, and the rotation of the μ-alkylidene group. In the transition state **TS_{DE}**, the breaking N-H and Ti-C bond distances lengthen 0.38 and 0.44 Å, respectively, and the formed C-H (1.40 Å) and Ti-N (2.21 Å) bond distances are significantly shorter than those of the reactants. The μ-alkylidene rotation process is required to allow the hydron migration from the NH_2 ligand. Note that in the reactant **D**, the μ-alkylidene CH_2 ligand is within the same plane of NH_2 ligand with two hydrogen atoms from the two different ligands facing each other (see Figure 2.2). We have already evaluated the full rotation of the μ-ethylidene ligand in a closely related system amounting about 129 kJ·mol^{-1}.[12a] The latter value together with the other simultaneous processes could explain the high computed energy barrier.

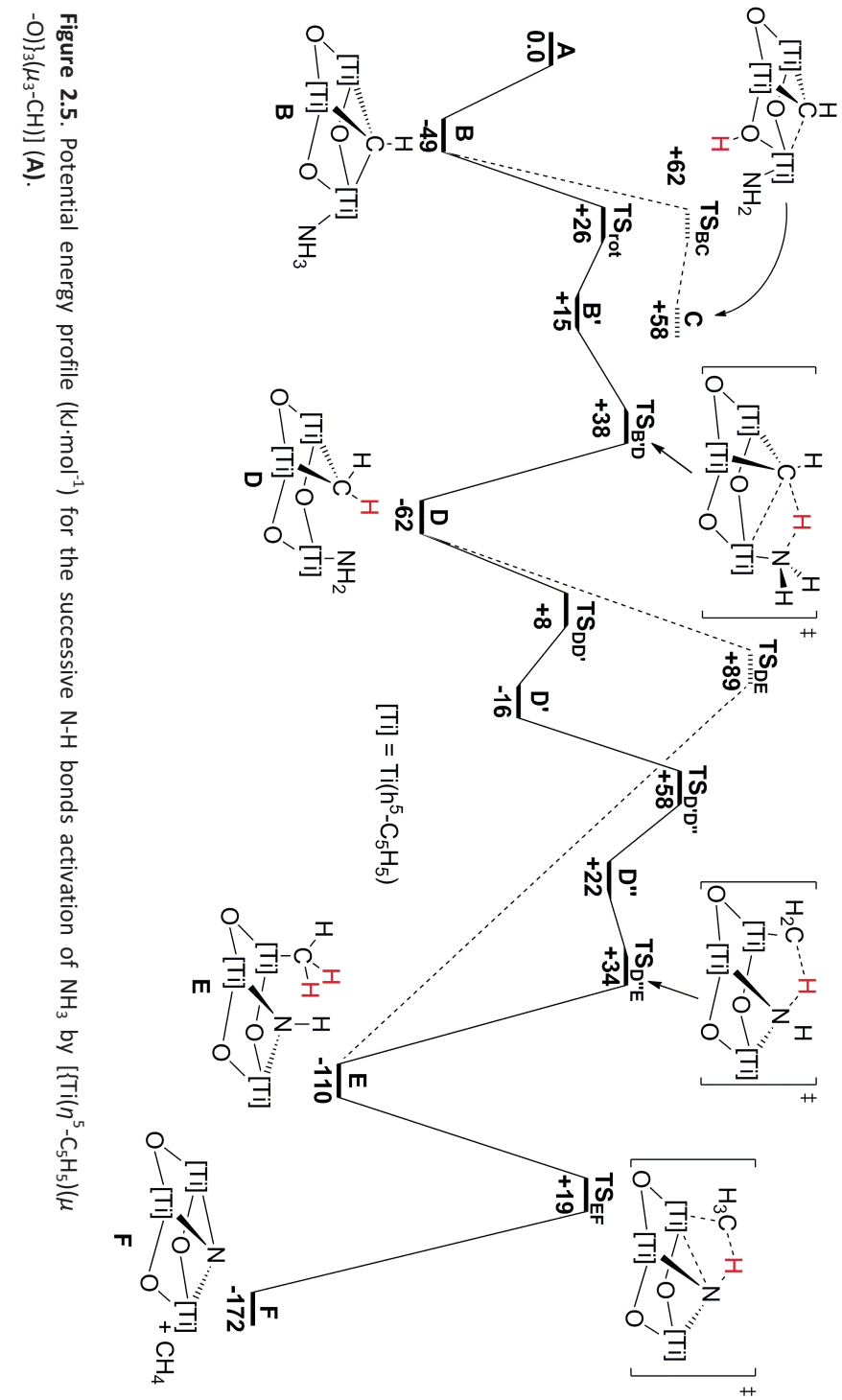
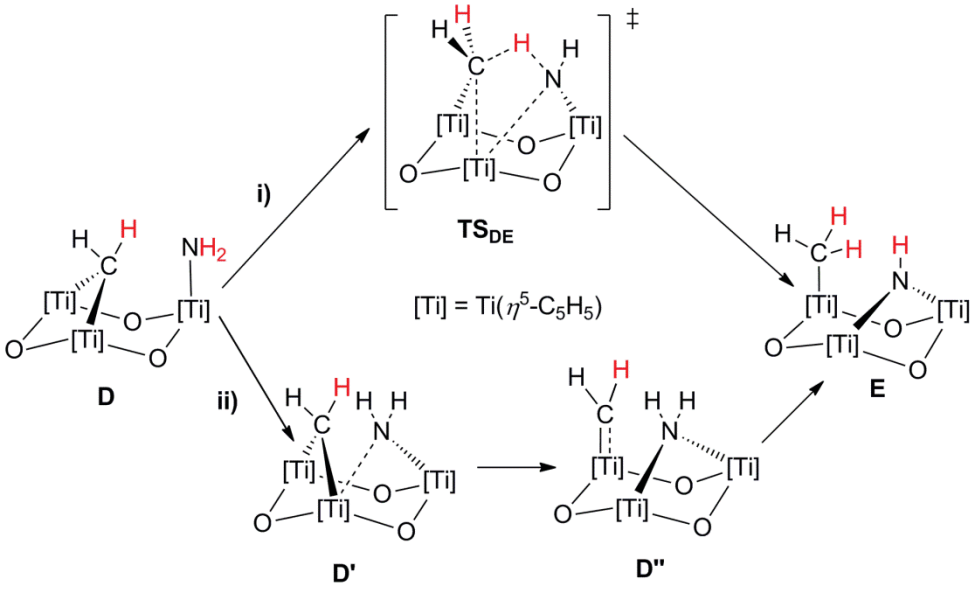


Figure 2.5. Potential energy profile (kJ·mol^{-1}) for the successive N-H bonds activation of NH_3 by [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-CH)] **(A)**.

Scheme 2.3.



In another possible mechanism for the second N-H activation, the ligands could rearrange previously to the hydron migration, following a stepwise mechanism to reduce the overall energy cost (Scheme 2.3ii). We located two new intermediates with different rearrangements of the alkylidene and amido ligands: the μ-alkylidene μ-amido-titanium complex **D'** and the alkylidene μ-amido-titanium complex **D''** (for molecular structures, see Annex). The relative energies of **D'** and **D''** with respect to the intermediate **D** are +46 and +84 kJ·mol^{-1}. In the first place, the terminal amido in **D** rearranges into a two-titanium bridging amido ligand in **D'**, with the energy cost to reach the **TS$_{DD'}$** being a modest 70 kJ·mol^{-1}. In **D'**, the former Ti-N bond distance lengthens to 2.05 Å, while the new Ti-N bond at the four-legs piano stool Ti centre is 2.31 Å. The formation of the new Ti-N bond increases the saturation of the Ti centre and in turns weakens the Ti-C$_{alkylidene}$ bond to be broken in the next step (lengthening of about 0.08 Å). Secondly, the bridging alkylidene in **D'** rearranges to a terminal alkylidene ligand, with the energy cost to reach the **TS$_{D'D''}$** again being a modest 74 kJ·mol^{-1}. In **D''**, the bond distance between the titanium and the alkylidene carbon is 1.87 Å, significantly shorter than in **D** (2.11 Å). The alkylidene carbon atom shows a strictly planar environment, suggesting that the group acts as a four-electron donor similar to the NH$_2$ moiety in **D** discussed above. In accordance, the HOMO molecular orbital corresponds mainly to a p-type orbital of the alkylidene carbon atom perpendicular to the CH$_2$ plane. At the HOMO, we also observe some stabilizing π-type overlapping with the titanium d orbitals (see Figure 2.6). Its energy is quite high (-4.6 eV), even higher than for **B'** (-5.5 eV), suggesting an even higher basicity of the carbon atom. Next, once **D''** is formed, an amido hydrogen can easily migrate as hydron to the alkylidene carbon via rotation of the terminal alkylidene ligand. The energy cost to reach **TS$_{D''E}$** is very low 12 kJ·mol^{-1}, in agreement with the predicted high basicity of the terminal alkylidene carbon. In **TS$_{D''E}$**, the calculated Ti-H distance is significantly larger (2.38 Å) than that found in the hydron migration

from oxo to alkylidyne group in related complexes (1.88 Å),[12a] suggesting that the Ti centre plays a minor role.

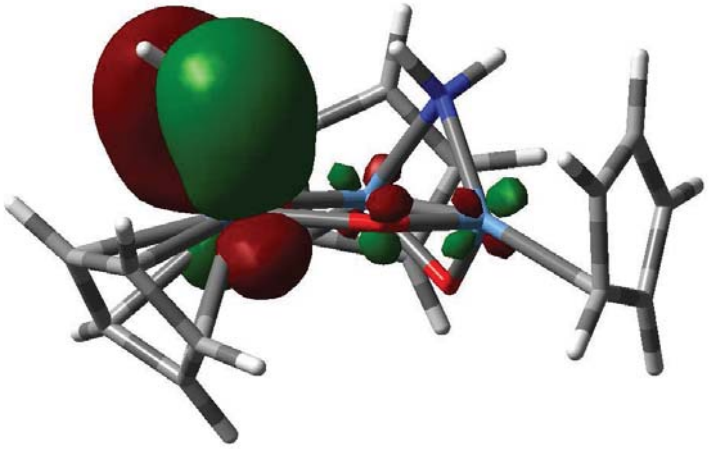


Figure 2.6. Representation of the p_π-d_π-type orbital HOMO for the alkylidene group in intermediate **D''**.

As observed in previous reactivity studies on these trinuclear titanium molecular systems,[12,32] the available empty low-energy titanium d orbitals could play an important role in this reaction via stabilization of the ligands such as the terminal amido, the bridging amido and the terminal alkylidene ligands. Nevertheless, in the case of the terminal alkylidene complex **D''**, the very low energy barrier calculated for the hydron migration shows that the **D''** species is not a very stable minimum and suggests that the existence or nonexistence of the terminal alkylidene complex as a reaction intermediate in the overall mechanism will depend on the particular nature of the system. A slight destabilization of this intermediate would cause the hydron transfer to pass from a three-step to a two-steps process. In summary, the overall energy barrier for the stepwise mechanism of the second N-H activation is determined by the energy difference between the lowest energy intermediate **D** and the highest energy point $\mathbf{TS_{D'D''}}$, amounting 120 kJ·mol^{-1}. The ligands of complex **D** rearrange intramolecularly (**D** → **D'** → **D''**), climbing the energetic ladder (-62 → -16 → +22 kJ·mol^{-1}) to reach a ligand disposition favorable for the hydron migration. This new estimation of the energy barrier is in better agreement with experimental observations than that for the direct hydron transfer ($\mathbf{TS_{DE}}$).

Finally, in the last N-H bond activation, the hydrogen atom at the imido group of **E** directly migrates to the alkyl group, releasing methane and forming the μ_3-nitrido titanium complex **F**. The process is exothermic (-62 kJ·mol^{-1}) and the energy barrier ($\mathbf{TS_{EF}}$) is computed to be 129 kJ·mol^{-1}, similar to that computed for the previous step (120 kJ·mol^{-1}). This last step can be regarded as the inverse process of the methane C-H σ-bond activation by metal nitride complexes.[33] Interestingly, there are several reported examples of arene or alkane C-H σ-bond

activation[34-35] via addition of the C-H bond across metal-nitrogen multiple bonds by early transition metal complexes, which include the titanium-imido complex [(^{t}Bu$_3$SiO)$_2$Ti(≡NSi^{t}Bu$_3$)].[36] Moreover, the reactivity of this latter complex has been theoretically investigated by the groups of Cundari[37] and Sakaki.[9] Our studied system, containing a nitrogen atom bridging two titanium atoms instead of Ti=N multiple bonds, shows a transition state geometry similar to those previously calculated for mononuclear Ti-imido complexes. In **TS$_{EF}$**, the breaking Ti-C bond distance lengths to 2.43 Å and the migrating H atom is at the middle position between C and N atoms, 1.39 and 1.28 Å for C-H and N-H distances, respectively. For the corresponding transition state in the mononuclear Ti-imido complex, the Ti-C, C-H and N-H distances are qualitatively similar 2.27, 1.44 and 1.39 Å, respectively.[9] These similarities suggest that the hydron transfer has the same nature in both complexes. Thus, as in previous studies,[9,37] we also propose that the bond cleavage occurs through a heterolytic manner.

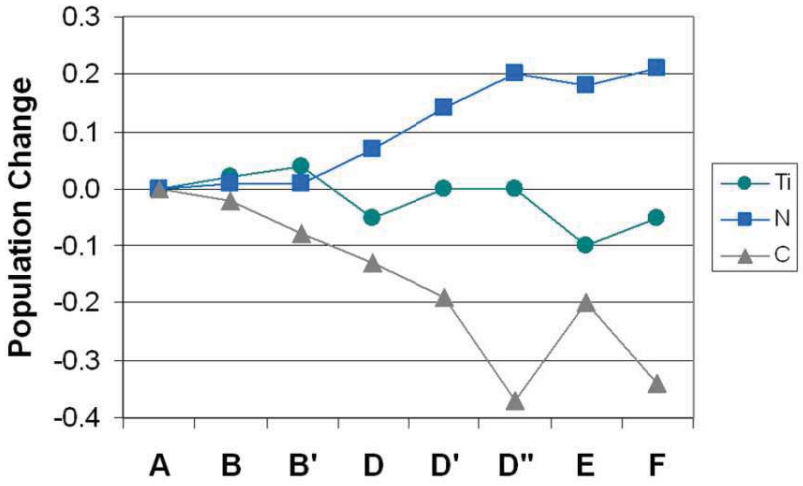


Figure 2.7. Population changes along the reaction coordinate for the atomic centres involved in the successive hydron transfer (Ti, average value of the three centres). The positive value represents the increase in Mulliken electron population, and the negative values the decrease.

To further clarify the nature of the successive N-H bonds activation of ammonia, we have analyzed the electron population changes along the reaction coordinate. Figure 2.7 shows the variation of atomic Mulliken electron population at the C and N atoms involved in hydron migration, as well as of the Ti centres (average values). As the reaction proceeds, the electron population of the C atom considerably decreases whereas the N atomic population increases. In both cases, it varies steadily, except in the case of the intermediate **D''**, where the formation of an unstable terminal Ti-alkylidene group causes an abrupt variation of electron population at the alkylidene carbon atom. On the other hand, the average population of the three Ti centres remains relatively unchanged. In the overall reaction, it slightly decreases, but in the ammonia intermediates, **B** and **B'**, it moderately increases, which is attributed to the charge transfer from the ammonia to the Ti centre. These results further support that the N-H

activation by the polynuclear Ti complex can be understood in terms of heterolytic N-H activation, in which Ti-alkylidyne, -alkylidene, -alkyl groups act as the basic centres promoting the bond cleavage.

2.5. Conclusions

In this Chapter, we have theoretical analyzed the N-H bonds activation of ammonia by the titanium μ_3-alkylidyne complexes [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-CR)] [R = H (**1**), Me (**2**)] to give the compound μ_3-nitride [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-N)] (**3**). DFT calculations have shown that upon successive N-H activation of ammonia, and the resulting species amide-μ-alkylidene, μ-imide alkyl and μ_3-nitride with methane elimination, the energy of the system goes downhill (-62, -110 and -172 kJ.mol^{-1}), providing the thermodynamic driving force for the reaction.

The mechanism for this sequential N-H bonds activation can be described as follows (see solid line in Figure 2.5.): i) first, the ammonia coordinates to a Ti centre *trans* to the alkylidyne group, and then isomerizes to adopt the *cis* arrangement, allowing the direct hydron transfer to the μ_3-alkylidyne group yielding a μ-alkylidene amido titanium complex, ii) next, the alkylidene and amido ligands rearrange to a favorable disposition for the hydron migration from the amido group to the alkylidene group, and iii) finally, in the μ-imido complex the hydron directly migrates to the alkyl group releasing methane and forming a μ_3-nitrido oxo titanium complex. All N-H bond activations occur via direct hydron transfer to carbon atom without direct participation of the titanium metal centres. The computed energy barriers of the successive steps are high enough to provide kinetic stability to the different intermediates, allowing their characterization.

The N-H bond activation of ammonia by terminal alkylidyne-titanium complexes had been predicted theoretically,[9] but never observed. Here we report a very closely related example, which supports previous predictions and promotes further investigations in titanium-alkylidyne species as potential catalysts in processes involving N-H activation.

In addition, the analysis of the molecular orbital structures of oxonitrido complex [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-N)], isoelectronic oxoligand [{Ti(η^5-C_5H_5)(μ-O)}$_3$(μ_3-CH)] and imido-nitrido derivative [{Ti(η^5-C_5H_5)(μ-NH)}$_3$(μ_3-N)] indicates that the apical nitrogen of oxonitrido could play an active role in the coordination behaviour and reactivity of the ligand. Experimental corroboration by reaction with hexacarbonyl complexes, allows us to suggest that the functionalization of the apical nitrogen atom (μ_3-N) in this compound provides great possibilities in future investigations to study the functionalization of small molecules or transformations that involves C-N bonds formation on the molecular metal oxide surface model Ti_3O_3.

2.6. References and notes

1 N. A. Lange, J. G. Speight, *Lange's Handbook of Chemistry*, 16th ed. McGraw-Hill, New York, **2005**.

2 a) M. Shultz, D. Milstein, *J. Chem. Soc. Chem. Commun.* **1993**, 318. b) W. Yao, O. Eisenstein, R. H. Crabtree, *Inorg. Chim. Acta* **1997**, *254*, 105.

3 a) Y. Peng, J.-D. Guo, B. D. Ellis, Z. Zhu, J. C. Fettinger, E. Rivard, P. P. Power, *J. Am. Chem. Soc.* **2009**, *131*, 16272. b) A. Jana, I. Objartel, H. W. Roesky, D. Stalke, *Inorg. Chem.* **2009**, *48*, 798. c) A. Jana, C. Shulzke, H. W. Roesky, *J. Am. Chem. Soc.* **2009**, *131*, 4600. d) P. A. Chase, D. W. Stephan, *Angew. Chem. Int. Ed.* **2008**, *47*, 7433.

4 a) M. L. H. Green, C. R. Lucas, *J. Chem. Soc. Dalton Trans.* **1972**, 1000. b) J. N. Armor, *Inorg. Chem.* **1978**, *17*, 203. c) J. E. Bercaw, D. L. Davies, P. T. Wolczanski, *Organometallics* **1986**, *5*, 443.

5 a) N. Russo, E. Sicilia, *J. Am. Chem. Soc.* **2001**, *123*, 2588. b) M. D. C. Michelini, E. Sicilia, N. Russo, *J. Phys. Chem. A* **2002**, *106*, 8937. c) E. Sicilia, N. Russo, *J. Am. Chem. Soc.* **2002**, *124*, 1471. d) M. D. C. Michelini, N. Russo, E. Sicilia, *Inorg. Chem.* **2004**, *43*, 4944. e) S. Chiodo, O. Kondakova, A. Irigoras, M. D. C. Michelini, N. Russo, E. Sicilia, *J. Phys. Chem. A* **2004**, *108*, 1069.

6 a) M. R. A. Blomberg, E. M. Siegbahn, M. Svensson, *Inorg. Chem.* **1993**, *32*, 4218. b) M. Hendrickx, M. Ceulemans, K. Gong, L. Vanquickenborne, *J. Phys. Chem. A* **1997**, *101*, 8540. c) Y. Nakao, T. Taketsugu, K. Hirao, *J. Chem. Phys.* **1999**, *110*, 10863. d) M. H. Chen, H. Lu, J. Dong, L. Miao, M. F. Zhou, *J. Phys. Chem. A* **2002**, *106*, 11456.

7 D. G. Musaev, K. Morokuma, *J. Am. Chem. Soc.* **1995**, *117*, 799.

8 T. Matsubara, *Organometallics* **2001**, *20*, 1462.

9 N. Ochi, Y. Nakao, H. Sato, S. Sakaki, *J. Am. Chem. Soc.* **2007**, *129*, 8615.

10 E. Khaskin, M. A. Iron, L. J. W. Shimon, J. Zhang, D. Milstein, *J. Am. Chem. Soc.* **2010**, *132*, 8542.

11 a) O. A. Kholdeeva, G. M. Maksimov, R. I. Maksimovskaya, M. P. Vanina, T. A. Trubitsina, D. Y. Naumov, B. A. Kolesov, N. S. Antonova, J. J. Carbó, J. M. Poblet, *Inorg. Chem.* **2006**, *45*, 7224. b) N. S. Antonova, J. J. Carbó, U. Kortz, O. A. Kholdeeva, J. M. Poblet, *J. Am. Chem. Soc.* **2010**, *132*, 7488.

12 a) J. J. Carbó, O. González-del Moral, A. Martín, M. Mena, J. M. Poblet, C. Santamaría, *Chem. Eur. J.* **2008**, *14*, 7930. b) J. J. Carbó, O. González-del Moral, A. Martín, M. Mena, J. M. Poblet, C. Santamaría, *Eur. J. Inorg. Chem.* **2009**, 643.

13 a) R. Andrés, M. V. Galakhov, A. Martín, M. Mena, C. Santamaría, *Organometallics* **1994**, *13*, 2159. b) R. Andrés, M. V. Galakhov, A. Martín, M. Mena, C. Santamaría, *J. Chem. Soc. Chem. Commun.* **1995**, 551.

14 L. Scoles, R. Minhas, R. Duchateau, J. Jubb, S. Gambarotta, *Organometallics* **1994**, *13*, 4978.

15 R. Andrés, M. V. Galakhov, M. P. Gómez-Sal, A. Martín, M. Mena, M. C. Morales-Varela, C. Santamaría, *Chem. Eur. J.* **2002**, *8*, 805.

16 a) F. N. Tebbe, G. W. Parschall, G. S. Reddy, *J. Am. Chem. Soc*. **1978**, 3611. b) B. J. J. Van de Heisteeg, G. Schat, O. S. Akkerman, F. Bickelhaupt, *Organometallics* **1985**, *4*, 1141. c) J. W. Park, L. M. Henling, W. P. Schaefer, R. H. Grubbs, *J. Am. Chem. Soc.* **1986**, *108*, 6402. d) F. Ozawa, J. W. Park, P. B. Mackenzie, W. P. Schaefer, L. M. Henling, R. H. Grubbs, *J. Am. Chem. Soc.* **1989**, *111*, 1319. e) P. B. Mackenzie, R. J. Coots, R. H. Grubbs, *Organometallics* **1989**, *8*, 8. f) J. W. Park, L. M. Henling, W. P. Schaefer, R. H. Grubbs, *Organometallics* **1991**, *10*, 171.

17 The structural relative dispositions of the different titanium substituents in **4** and **6** were confirmed by NOE homonuclear differential experiments.

18 H. W. Roesky, Y. Bai, M. Noltenmeyer, *Angew. Chem. Int. Ed. Engl.* **1989**, *28*, 754.

19 R. Serrano, J. C. Flores, P. Royo, M. Mena, M. A. Pellinghelli, A. Tiripicchio, *Organometallics* **1989**, *8*, 1404.

20 L. M. Babcock, V. W. Day, W. G. Klemperer, *J. Chem. Soc. Chem. Commun.* **1987**, 858.

21 A. Abarca, M. Galakhov, P. Gómez-Sal, A. Martín, M. Mena, J. M. Poblet, C. Santamaría, J. P. Sarasa, *Angew. Chem. Int. Ed.* **2000**, *39*, 534.

22 T. Carofiglio, C. Floriani, A. Sgamellotti, M. Rosi, A. Chiesi-Villa, C. Rizzoli, *J. Chem. Soc. Dalton Trans.* **1992**, 1081.

23 S. I. Troyanov, V. Varga, K. Mach, *J. Organomet. Chem.* **1991**, *402*, 201.

24 S. G. Blanco, M. P. Gómez-Sal, S. M. Carreras, M. Mena, P. Royo, R. Serrano, *J. Chem. Soc. Chem. Commun.* **1986**, 1572.

25 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, Jr. J. A. Montgomery, T. Vreven, K. N. Kudin, J. R. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprkch, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, C. González, J. A. Pople, *Gaussian 03*, revision B.03; Gaussian, Inc.: Pittsburgh, PA, **2004**.

26 a) R. G. Parr, W. Yang, *Density Functional Theory of Atoms and Molecules*. Oxford University Press: U. K. Oxford, **1989**. b) T. Ziegler, *Chem. Rev.* **1991**, *91*, 651.

27 a) C. Lee, C. Yang, R. G. Parr, *Phys. Rev.* B **1988**, *37*, 785. b) A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648. c) P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.

28 P. J. Hay, W. R. Wadt, *J. Chem. Phys.* **1985**, *82*, 270.

29 a) W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257. b) P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213. c) M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. Defrees, J. A. Pople, *J. Chem. Phys.* **1982**, *77*, 3654.

30 P. Gómez-Sal, A. Martín, M. Mena, M. C. Morales-Varela, C. Santamaría, *Chem. Commun.* **1999**, 1839.

31 J. J. Carbó, N. Martínez-Espada, M. Mena, M. E. G. Mosquera, J. M. Poblet, C. Yélamos, *Chem. Eur. J.* **2009**, *15*, 11619.

32 J. J. Carbó, A. Martín, M. Mena, A. Pérez-Redondo, J. M. Poblet, C. Yélamos, *Angew. Chem. Int. Ed.* **2007**, *46*, 11619.

33 M. Schlangen, J. Neugebauer, M. Reiher, D. Schröder, J. Pitarch López, M. Haryono, F. W. Heinemann, A. Grohmann, H. Schwarz, *J. Am. Chem. Soc.* **2008**, *130*, 4285.

34 a) A. E. Shilov, G. B. Shul'pin, *Chem. Rev.* **1997**, *97*, 2879. b) J. A. Labinger, J. E. Bercaw, *Nature* **2002**, *417*, 507.

35 C. C. Cummins, S. M. Baxter, P. T. Wolczanski, *J. Am. Chem. Soc.* **1988**, *110*, 8731.

36 a) P. T. Wolczanski, J. L. Bennett, *J. Am. Chem. Soc.* **1994**, *116*, 2179. b) P. T. Wolczanski, J. L. Bennett, *J. Am. Chem. Soc.* **1997**, *119*, 10696.

37 T. R. Cundari, T. R. Klinchkman, P. T. Wolczanski, *J. Am. Chem. Soc.* **2002**, *124*, 1481.

Chapter 3

Theoretical Studies on Asymmetric Hydroformylation by Rh-(*R,S*)-Binaphos Catalyst. Origin of Coordination Preferences and Stereoinduction

UNIVERSITAT ROVIRA I VIRGILI
MODELING OF HOMOGENEOUS CATALYSIS: FROM DFT TO QSPR APPROACHES
Sonia Aguado Ullate
DL:T. 279-2012

Chapter 3

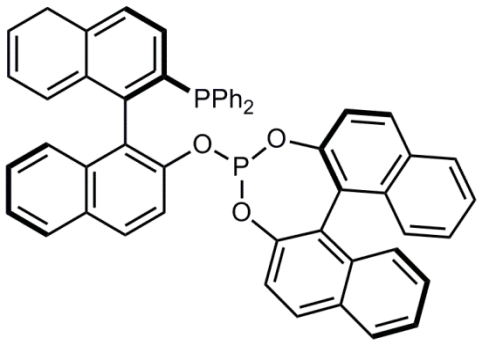
Theoretical Studies on Asymmetric Hydroformylation by Rh-(*R,S*)-Binaphos Catalyst. Origin of Coordination Preferences and Stereoinduction

We have undertaken theoretically investigated of the asymmetric hydroformylation of styrene by [HRh{(*R*,*S*)-binaphos}$(CO)_2$] catalyst, focusing on the origin of the ligand coordination preferences and stereoinduction. We evaluated the different factors governing the preference of the binaphos ligand to coordinate with phosphane moiety at the equatorial site and the phosphite at the apical site. The donor-acceptor interactions, obtained using a modified version of *Energy Decomposition Analysis* (EDA) based on orbital deletion, favour the phosphite moiety at the equatorial site. However, the electronic distortion and the steric effects inverse this tendency. Calculations also suggest that the coordination preference was transferred to the selectivity-determining transition state. We propose a sterochemical model based on quantitative quadrant-diagram representation obtained from a new molecular descriptor, the *distance-weighted volume* (V_W) which easily computed from ground-state structures. Repulsive interactions between the substrate and the apical phosphite were responsible for the enantiodifferentiation. The axial chirality of the phosphite discriminated one of the competitive equatorial-apical paths, whereas the axial chirality of the backbone discriminated one of the two enantiomers. Transition-state calculations revealed that the placement of phosphane at the apical site would lower enantioselectivity, explaining the poor performance of other phosphane-phosphite ligands. Finally, comparison with previous studies allowed the definition of several prerequisites for diphosphane ligands for high stereoselectivity: i) specific equatorial-apical coordination bringing chirality to the apical site; ii) combination of two stereogenic centres; and iii) rigid structures.

3.1. Introduction

As described in the *Chapter 1*, asymmetric hydroformylation (AHF) offers a potentially useful synthetic route to enantiomerically pure aldehydes, which can be used as precursors of biologically active compounds and fine chemicals.[1] A real breakthrough in this area was the publication of Rh/(*R,S*)-binaphos catalysts by Takaya and co-workers.[2] Binaphos is an unsymmetric bidentate phosphane phosphite ligand (Scheme 3.1), which has been successfully employed in the hydroformylation of a large variety of substrates to achieve high activity and stereoselectivity.[3] Currently, the broad application of the reaction is still a goal because of the intrinsic difficulties of simultaneously controlling enantio- and regioselectivity, as well as obtaining highly active catalysts. Thus, the understanding of the outstanding binaphos example would provide useful criteria for the design of novel catalytic systems effective in AHF.

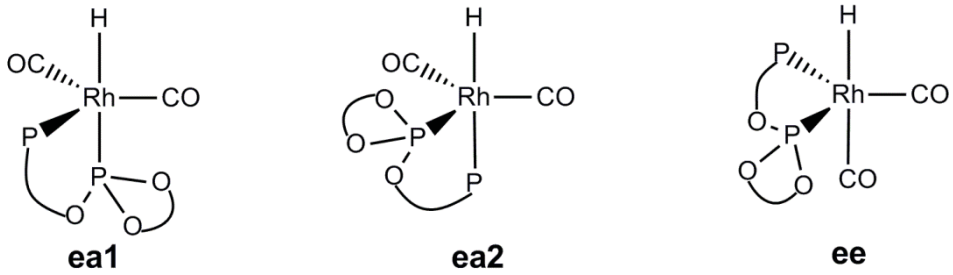
Scheme 3.1.



Frequently, the catalyst performance has been related to the coordination mode of the bidentate ligand to the transition metal.[4-5] Certainly, this can make a significant difference in the three-dimensional structure of the active complex, having a profound influence on the enantioinduction mechanism. It is known that the resting state in rhodium-catalyzed hydroformylation is a pentacoordinated complex [HRh(P-P)$(CO)_2$] with a trigonal-bipyramidal geometry. The two phosphorus donors of the bidentate ligand can coordinate either in an equatorial-equatorial (ee) or in an equatorial-apical (ea) mode. For unsymmetric bidentate ligands, such as binaphos, it is possible to distinguish between two ea isomers with either an apical phosphite (**ea1**) or an equatorial phosphite (**ea2**) ligand. Scheme 3.2 depicts the possible coordination modes for phosphane-phosphite ligands in trigonal-bipyramidal complexes. Studies on equatorial-apical coordinating unsymmetric diphosphane ligands have shown that there is a strong preference for the coordination of the least basic moiety in the equatorial plane of the complex,[6] in accordance with simple orbital preferences[7] and DFT calculations.[8] In phosphane-phosphite (P-OP) ligands, the phosphane is a good σ donor and the phosphite is a poorer σ donor and a better π acceptor.[9] The preferred ea coordination with the phosphite moiety in the equatorial position has been proposed by several authors.[4,10] Based on this coordination, Pizzano and co-workers have rationalized the observed enantioselectivity in

phosphane-phosphite ligands through stabilizing π-π stacking interactions between the vinyl arene substrates and the aryl substituents of the apical phosphane moiety.[4] Although the ligands do not possess a P-stereogenic phosphane, the diarylphosphino groups display a propeller-like chiral arrangement, which may be an important stereoregulating factor. From computationally studies, Carbó, Bo and co-workers have identified analogous π-π stabilizing interactions as key factors for enantioinduction in AHF of styrene by aminophosphane phosphinite (AMPP) ligands; the interaction is between the styrene and the aminophosphane moiety coordinated in apical mode.[11] However, they have also shown that when the aminophosphane is non P-stereogenic, the interactions at pro-*S* and pro-*R* transition states become similar,[11] leading to low enantioselectivities.

Scheme 3.2.



One clear exception to this rule is the binaphos ligand system. Takaya, Nozaki and co-workers have characterized a single isomer, in which the phosphite moiety occupies the apical position.[5a] From this coordination mode, Hermann and co-workers have developed a semiquantitative theoretical model that rationalizes the origin of stereodifferentiation.[12] For the structurally and electronically similar phosphane-phosphoramidite (yanphos) ligand, Zhang and co-workers have explained the enantioselectivity assuming the same coordination pattern.[13] Contrary to Pizzano's proposal for binaphos and yanphos ligands, the theoretical models suggest that ligand-substrate interactions governing enantioselectivity are repulsive in nature. In accordance with this, the same sense of enantioface was observed for different heteroatom substituted alkenes.[5a] Very recently, Jäkel and co-workers have re-examined the ligand coordination in the [HRh{(*R*,*S*)-binaphos}$(CO)_2$] complex.[14] The authors showed that, although the major chelating is an apical phosphite (**ea1**), the conformer with the phosphite in the equatorial position does also exist in equilibrium (**ea2**). The qualitative discussion of bonding and steric effects provided a hypothesis to understand the origin of coordination preference. Whereas bonding considerations place the better σ donor in the apical site and the better π acceptor in the equatorial site, the steric effects prefer to place the sterically larger phosphane moiety in the less crowded equatorial position. These arguments explain the experimental findings. However, a deeper understanding of the factors governing the coordination preferences of binaphos is important, as it is an outstanding example of an unsymmetric phosphane-phosphite ligand. Moreover, the origin of enantioselectivity was

attributed, in part, to the presence of only a single isomer, and because of these new findings we considered a more complicated explanation.

Despite the rapid evolution of computational methods and their application to asymmetric catalysis, there are only a few studies devoted to asymmetric hydroformylation. The most representative theoretical contributions to asymmetric hydroformylation have been recapitulated in *Chapter 1*. They include the first theoretical description of Rh-binaphos catalysts behaviour by Herrmann and co-workers,[12] the first quantitative theoretical assessment of enantioselectivity for AMPP ligands by our group of investigation,[11] and semiquantitative MM calculations in yanphos ligand system by Zhang and co-workers.[13] All these theoretical studies use transition state-based approaches (TS-based approach, for more information see *Chapter 1*), assuming that the alkene insertion into the Rh-H bond is the enantioselectivity-determining step. Also, purely qualitative models have been proposed for structures mimicking the alkene insertion transition state (TS).[5a,15] Moreover, some authors have reported attempts to find relationships between the enantioselectivity in AHF and the ground-state properties of the catalyst. For a series of biaryl-bridged bisphosphite ligands, Klosin, Whiteker and co-workers found a linear relationship between the enantioselectivity and the dihedral angle of the bridging aryl moieties, which, in turns, correlated with the computed bite angles in the intermediates.[16] However, the same authors showed that, for a family of bisphospholane ligands, the same type of correlation was unclear.[17] Hence, more sophisticated computational approaches have been tried to define the molecular descriptors from ground-state structures and to set quantitative relationships between them and enantioselectivity. These are emerging approaches in asymmetric organometallic catalysis,[18] but have not yet been explored in hydroformylation.

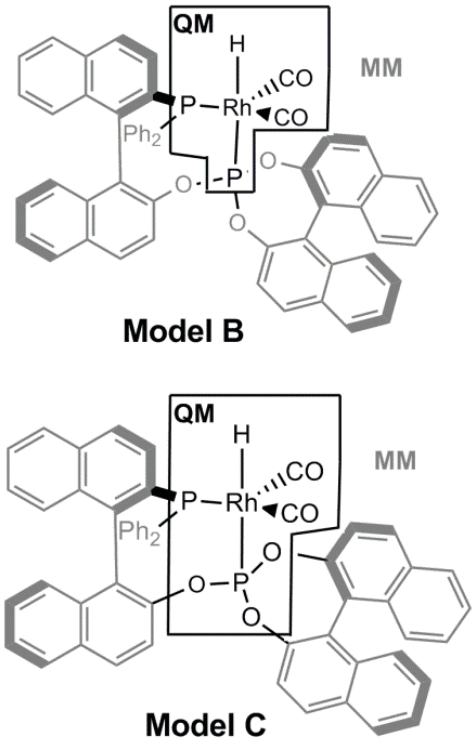
In this Chapter, and following the computational studies of our research group on Rh-catalyzed hydroformylation of alkenes,[8,11,19] we focus on Rh/(*R,S*)-binaphos system as a model of unsymmetric diphosphanes. First, we investigated the ligand coordination. Using a modified version of *Energy Decomposition Analysis* (EDA), we gained insight into the σ donor, π acceptor and electronic distortion properties of the two moieties. We used QM/MM methods as a tool to separate the electronic and steric effects by comparing different model systems. Our goal was to provide a quantitative theoretical characterization of the factors governing coordination preferences. Next, we re-examined the origin of enantioselectivity. We proposed a quantitative stereochemical model by defining a new molecular descriptor, the *distance-weighted volume* (V_W), easily obtained from ground-state structures. Finally, we analyzed the key transition states aiming to identify the intrinsic nature of the ligand-substrate interactions responsible for stereoinduction.

3.2. Computational details

The analysis of ligand coordination was carried out using the Amsterdam density functional program (ADFv2007).[20] For DFT calculations, the electronic configurations were described by a triple-ζ plus polarization Slater-type basis set, as included in ADF package. The 1s-3d electrons for Rh, the 1s electrons for C and O, and the 1s-2p electrons for P were treated as frozen

cores. We applied scalar relativistic corrections to them using the zero-order regular approximation (ZORA) with the core potentials generated using the DIRAC program.[20] Equilibrium structures and associated energies were obtained at the BP86 level, applying the X_α model with Becke's corrections[21] to describe exchange and the VWN parametrization[22] with Perdew's corrections[23] for correlation. For the hybrid QM/MM calculations, we applied the IMOMM method[24] as implemented in the ADF package.[25] The two different molecular QM/MM partitions (models **B** and **C**) are represented in Scheme 3.3. Both models included styrene in the QM part. The QM level was the same as mentioned above. SYBYL force field[26] was used to describe the atoms included in the MM part. The van der Waals parameters for the Rh atom were taken from the UFF force field,[27] and torsional contributions involving dihedral angles with the metal atom in terminal position were set to zero.

Scheme 3.3.



To search for possible conformational isomers, we defined a protocol based on restricted molecular dynamics simulations as implemented in ADF.[20] Figure 3.1 shows the workflow of our conformational search protocol. It consists of i) an initial quantum mechanics / molecular mechanics (QM/MM) optimization at the BP86:SYBYL level; ii) 100 ps of restricted MD simulation (time step = 0.5 fs) at 1000 K, fixing the rhodium coordination sphere and sampling a geometry each ps; iii) 100 partial geometry optimizations fixing the Rh coordination sphere by the inexpensive MM method; and iv) the refinement of the ten lowest energy structures by

full geometry optimization at the reliable QM/MM level. To ensure the reliability of the potential energy surface sampling, we repeated the restricted MD simulation at a higher temperature (2000 K).

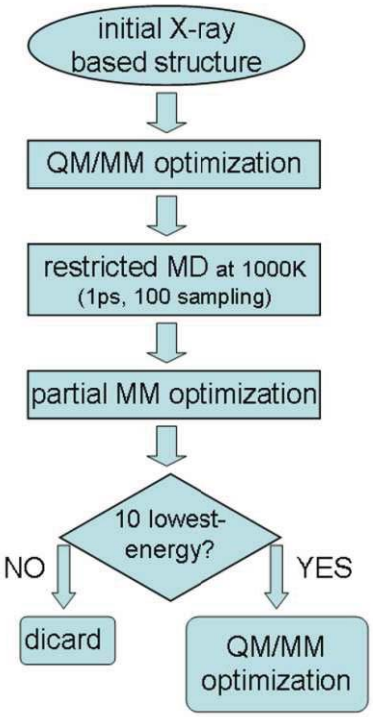


Figure 3.1. Workflow of the conformational search protocol.

To analyze the nature of Rh-phosphorus bonds, we employed the *Energy Decomposition Analysis* (EDA) method.[28] Within the EDA method, the interaction energy can be divided into three terms; i) the classical electrostatic interaction $\Delta E_{elastat}$ between the unperturbed charges of the fragments; ii) the Pauli term ΔE_{Pauli} which accounts for the repulsion between electrons of the same spin; and iii) the orbital term ΔE_{orb} that accounts for charge transfer and polarization effects.[29] The sum of electrostatic and Pauli terms is often called the steric term (ΔE_{ster}).

For C_1-symmetric complexes, we proposed a modified version of the EDA procedure based on orbital deletion, which allowed for the separating of the σ and π interactions in a physically meaningful manner.[30] To account for σ donation, referred as ΔE_{σ}(L→Rh), we repeated the EDA analysis while consecutively deleting all the virtual molecular orbitals except for the metal orbitals ($dz^2$and 5*s*-type) with the appropriate shape for σ donation. Repeating the same procedure but keeping the appropriate virtual orbitals of the ligand fragment (p_{π} orbitals) provided data for the π back-donation, referred as ΔE_{π}(Rh→L). This gave the energetic

contribution to the bonding from the interaction between the occupied orbitals of one fragment with the selected unoccupied orbitals of the other fragment.

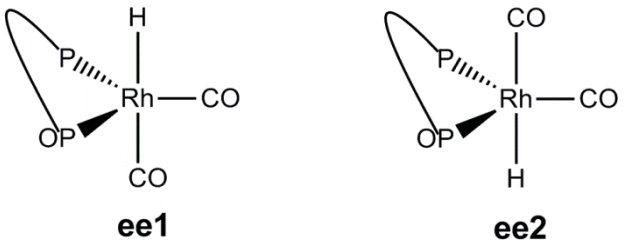
The analysis of enantioselectivity based on the transition states was performed with the Gaussian 09 package[31] because of the availability of very efficient transition state search algorithms and newly developed functional such as M06[32] and B97D[33] (see below). The QM/MM calculations used the ONIOM method[34] with the same molecular partition as model **C**. For the QM region we used the BP86 functional. For Rh and P atoms, the LANL2DZ pseudopotential was used,[35] supplemented by a d shell in the case of P.[36] The 6-31G(d, p) basis set[37] was used for the rest of the atoms. The MM calculations used the UFF force field.[27] Moreover, transition states were located at the B97D:UFF level, as well as, at the full DFT level with M06 and B97D functionals including solvent corrections with a continuum SMD model[38] as implemented in Gaussian 09.[31] All of transition states were characterized by a single imaginary frequency and the normal mode which corresponds to the expected reaction path.

3.3. Results and discussion

3.3.1. Conformational search and relative stabilities of the isomers

As stated in the *introduction* section, for pentacoordinated hydrido carbonyl complex [$HRh(CO)_2$\{(*R*,*S*)-binaphos\}], the diphosphorus ligand can coordinate as equatorial-equatorial (ee) or equatorial-apical, either with an apical phosphite (**ea1**) or with an apical phosphane (**ea2**, see Scheme 3.2). It is possible to generate two additional isomers for ee coordination by exchanging the relative positions of the hydride and carbonyl apical ligands with respect to the equatorial plane of the C_1-symmetric complex (**ee1** and **ee2**, Scheme 3.4). Thus, for each geometrical isomer **ee1**, **ee2**, **ea1**, and **ea2**, we generated initial structures using the ligand conformation in the X-ray structure [$Pt(CH_3)_2$\{(*R*,*S*)-binaphos\}].[39]

Scheme 3.4.



At the end of the conformational search protocol described in *computational details* section, only one conformer was found for each geometric isomer. This is due to the ligand rigidity induced by the binaphthyl fragments. In principle, the phenyl groups of the phosphane moiety could freely rotate yielding different isomers. However, a look at the MD simulation revealed that one of these phenyls was stuck by a π-π stacking interaction with a naphthyl group from

the phosphite in the backbone. The other phenyl from the phosphane flips around in a perpendicular arrangement with respect to the first one (see geometries in Figure 3.2). The same face-to-face arrangement between phenyl and naphthyl groups can be observed in the X-ray structure of [Pt(CH_3)$_2$\{(*R*,*S*)-binaphos\}].[39] The phenyl and naphthyl groups have a distance between ring centroids of 3.9 Å, the same as for the optimal centroid-centroid distance in the benzene dimer.[40] The calculated distances between ring centroids at the BP86:SYBYL level are very similar for all of the isomers (3.3-3.4 Å) and are somewhat shorter than experimental values. We have shown that MM calculations qualitatively reproduce π-π stacking interactions, but tend to overestimate their interaction energies and to give shorter distances.[11] This interaction together with the rigidity of binaphthyl fragments prevents conformational flexibility. Interestingly, in previous study of AMPP-type ligands, they proposed rigid backbones as a prerequisite to improve the performance of the catalyst.[11] Thus, the rigidity of binaphos could be one of the reasons for the high enantioselectivity.

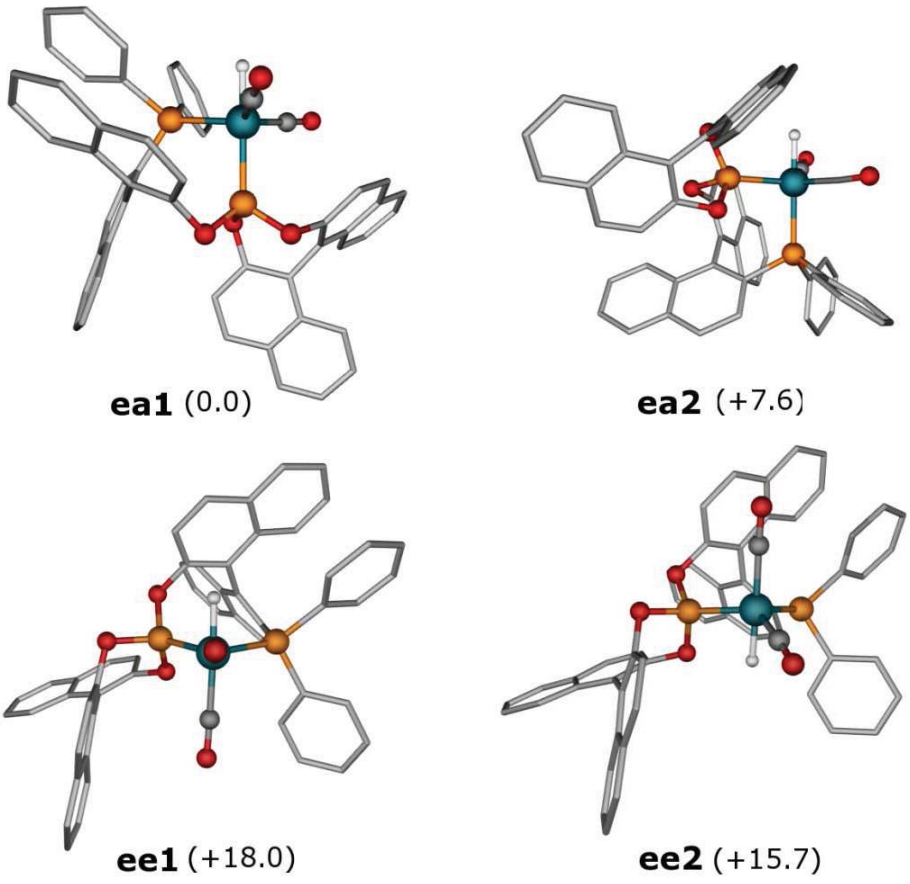


Figure 3.2. Calculated 3D molecular structures for the different isomers of the [HRh(CO)$_2$\{(*R*,*S*)-binaphos\}] complex. Relative energies in kJ.mol^{-1}.

In agreement with the experimental findings,[5a,14] the isomer with the equatorial-apical coordination with the phosphite in the apical site (**ea1**) was found to be the structure with the global energy minimum, 7.6, 15.7 and 18.0 kJ.mol^{-1} lower in energy than the **ea2**, **ee2** and **ee1** isomers, respectively. Consistent with recent observations, the **ea2** structure was the next most stable isomer, lying close in energy to **ea1**. Jäkel and co-workers have showed that there is a mixture of two equatorial-apical chelates for [HRh(CO)$_2${(*R*,*S*)-binaphos)}].[14] Therefore, we will focus on the analysis of the differences between the coordination and enantioinduction between the **ea1** and **ea2** isomers.

3.3.2. Analysis of coordination preferences

It is possible to identify and quantify the different effects governing the coordination preferences of the ligands by successively improving the model system.[41] We started by analyzing the bonding situation of the simplest model system [HRh(CO)$_2$(PH$_3$)(P(OH)$_3$)] (**A**) at the QM level. Using PH$_3$ as a model for the phosphane moiety and P(OH)$_3$ for phosphite, we put aside the steric effects of the bulky substituents but kept the electronic dissymmetry. Then, we performed QM/MM calculations on model **B**, in which the QM part was [HRh(CO)$_2$(PH$_3$)$_2$], and the rest of the binaphos ligand was in the MM part of the calculation. Within this partition, there was no charge transfer from the MM part, thereby resulting in a ligand with symmetric electronic properties and which accounted for the steric properties.[42] An additional model (**C**) included [HRh(CO)$_2$(PH$_3$)(P(OH)$_3$)] in the QM part. This model accounts for the synergism of the electronic effects in **A** and the steric effects in **B**. Finally, we compared them with the full QM calculations on the whole complex (**ea1** and **ea2**). Figure 3.3 shows a pictorial representation of the main effects analyzed with this procedure and Figure 3.4 summarizes the results.

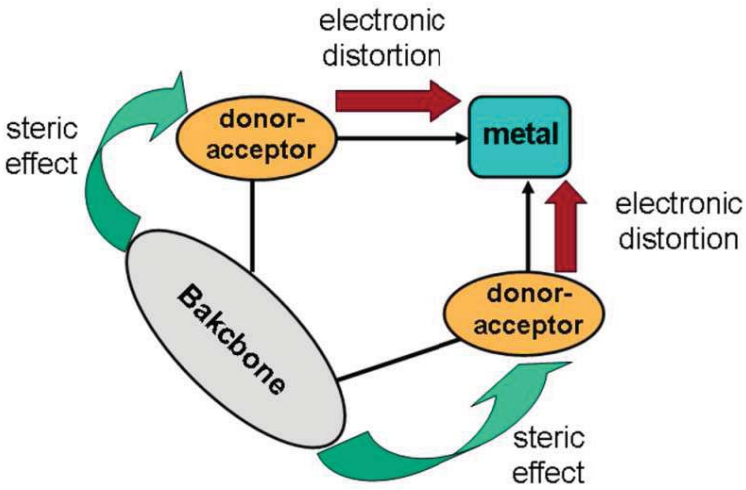


Figure 3.3. Pictorial representation of a bidentate ligand highlighting the key interaction with the metal centre.

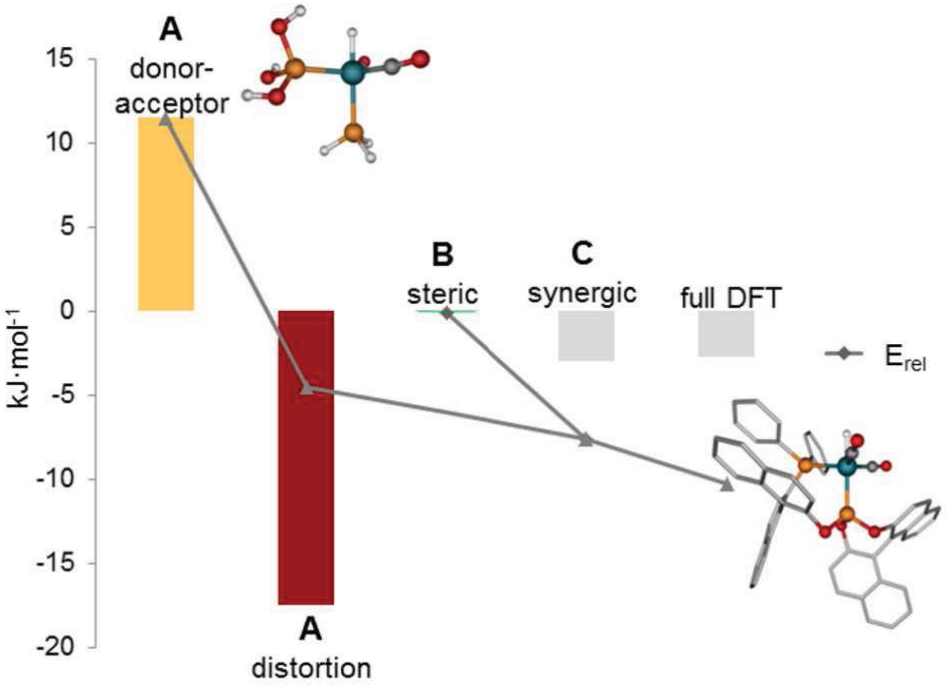


Figure 3.4. Contribution of the different effects (bars) on the coordination preferences of the binaphos ligand and the evolution of the relative energies (lines) of the two geometrical isomers **ea1** and **ea2**.

The calculated relative energies of **ea1** and **ea2** isomers for model **A** were 0.0 and +4.5 kJ.mol^{-1}, respectively, indicating an electronic preference to place the phosphite in the apical position. Even in the absence of steric effects, this appears to be in contradiction with previous qualitative[7] and quantitative[8] computational findings which predicting that the most basic phosphite moiety would be in equatorial position (**ea2**). Hofmann and Rossi, in their qualitative molecular orbital treatment of trigonal-bipyramidal [ML_5] complexes, concluded that, for d^8 transition metals, the apical bond should be stronger when considering only the σ bonding effects and that π acceptor interaction would favour the equatorial site.[7] Our group of investigation has quantitatively showed that for the electronically distinct phosphanes PH_3 and PF_3 the best σ donating (PH_3) ligand prefers coordination in the apical position for the [$HRh(CO)_2(PH_3)(PF_3)$(alkene)] complex.[8] To fully analyze the electronic coordination preferences, we divided the binding energy (ΔE_{bind}) of the model ligands into interaction (ΔE_{int}) and distortion (ΔE_{dist}) terms.[43] The ΔE_{dist} is defined as the energy cost to bring the fragments from their free geometries to those they adopt in the complexes. The ΔE_{int} is further divided using the *Energy Decomposition Analysis* (EDA) method.[28] Table 3.1 collects the main results of the EDA bonding analysis. First, we observed that the interaction energies (ΔE_{int}) follow the opposite trend to the relative energies. Although **ea2** is less stable, the sum of the interaction energies of both ligands is larger for **ea2** (-250 kJ.mol^{-1}) than for **ea1** (-240 kJ.mol^{-1}). The

difference in ΔE_{int} on going from **ea1** to **ea2** (-10 kJ.mol^{-1}) compares well with the difference between the orbital energies ΔE_{orb} (-13 kJ.mol^{-1}), as expected for bonds with mostly dative character.

Table 3.1. EDA partition energies (kJ.mol^{-1}) and metal-ligand bond lengths (Å) for $[HRh(CO)_2(PH_3)(P(OH)_3)]$ (**Aea1** and **Aea2**).

	Aea1			**Aea2**		
	$[PH_3]_{eq}$	$[P(OH_3]_{ap}$	sum	$[PH_3]_{ap}$	$[P(OH)_3]_{eq}$	sum
d(Rh-P)	2.407	2.349	-	2.413	2.334	-
ΔE_{int}	-89	-151	-240	-126	-124	-250
ΔE_{Pauli}	+465	+484	+949	+391	+582	+973
$\Delta E_{elestat}$	-386	-424	-810	-338	-493	-831
ΔE_{ster}	+79	+60	+139	+53	+89	+142
ΔE_{orb}	-168	-211	-379	-179	-213	-392
$\Delta E_{\sigma}(PR_3 \rightarrow Rh)$	-53	-68	-121	-71	-55	-126
$\Delta E_{\pi}(Rh \rightarrow PR_3)$	-30	-33	-63	-27	-43	-70

To evaluate of donor-acceptor interactions of the phosphane and phosphite ligands, we used a modified version of the EDA procedure based on orbital deletion.[30] This procedure allowed for a straightforward separation of σ and π interactions by evaluating the stabilization from the mixing of the occupied and unoccupied orbitals of the ligand and metal fragments. The decomposition of the orbital energy ΔE_{orb} in terms of σ donation ($\Delta E_{\sigma}(PR_3 \rightarrow Rh)$) and π back-donation ($\Delta E_{\pi}(Rh \rightarrow PR_3)$) is collected in Table 3.1. For both PH_3 and $P(OH)_3$, the estimated σ donation energy is larger for the apical site (-71 and -68 kJ.mol^{-1} *vs*. -53 and -55 kJ.mol^{-1}), and the π back-donation energy is larger for the equatorial site (-30 and -43 kJ.mol^{-1} *vs*. -27 and -33 kJ.mol^{-1}). Moreover, in agreement with general knowledge, the π interaction energy for the phosphite ligand was larger than for phosphane by 6 and 13 kJ.mol^{-1} for the apical and equatorial sites, respectively. Both ligands had larger interactions in the apical site; however, this preference was less pronounced in the phosphite case because the enhanced π acceptor interaction in the equatorial site almost overcomes the σ apical preference. Overall, the most favourable donor-acceptor situation places the phosphane in the apical site (**ea2**), with this situation favoured by 11.5 kJ.mol^{-1} (first bar in Figure 3.4). These results are in agreement with previous arguments by Hofmann and Rossi[7] and provide quantitative support for them.

Table 3.2 shows the values for the binding energies and the decomposition of the ligands in interaction and distortion terms. In agreement with the higher stability of the isomer **Aea1**, the sum of the binding energies for **Aea1** (-140 kJ.mol^{-1}) is larger than for **Aea2** (-132 kJ.mol^{-1}). This trend is mainly retrieved in the distortion term of the metal fragment induced by the coordination of the $P(OH)_3$ ligand in the equatorial position. The distortion term for $[P(OH)_3]_{eq}$

in **Aea2** is destabilized by 15 kJ.mol^{-1} more than for $[PH_3]_{eq}$ in **Aea1**.[44] Thus, the preference for phosphite coordination at the equatorial site induced by donor-acceptor interactions is balanced by the electronic distortion, which places the phosphite at the apical site (second bar in Figure 3.4).

Table 3.2. Energy decomposition (kJ.mol^{-1}) of the binding between P ligands and the Rh fragments in the **Aea1** and **Aea2** compounds.

	Aea1			**Aea2**		
	$[PH_3]_{eq}$	$[P(OH_3]_{ap}$	sum	$[PH_3]_{ap}$	$[P(OH)_3]_{eq}$	sum
ΔE_{bind}	-58	-82	-140	-54	-78	-132
ΔE_{int}	-89	-151	-240	-126	-124	-250
ΔE_{dist}(L)	+2	+4	+6	+4	+2	+6
ΔE_{dist}(TM)	+29	+66	+95	+68	+44	+112

One may think that the inversion of the trend could be attributed to the different steric hindrance of model ligands PH_3 and $P(OH)_3$. Starting from $[ML_4]$ square planar metal fragment, the coordination of the equatorial P ligand swings the carbonyl ligands back with a larger decrease in the $C_{carbonyl}$-Rh-$C_{carbonyl}$ angle for $[P(OH)_3]_{eq}$ than for $[PH_3]_{eq}$ (from 180° to 123° and to 129°, respectively). However, the different distortions can be also explained by electronic arguments based on frontier molecular orbitals (FMO). In fact, it was reported that metal fragment can be distorted to make filled orbitals that back-donate more accessible for the ligand.[45-46] Figure 3.5 depicts the frontier molecular orbitals of an idealized trigonal-bipyramid d^8 $[ML_5]$ complex and the qualitative effect of moving one equatorial ligand. As the equatorial P ligand approaches the metal, the antibonding interaction along the *x* axis increases in the $dx^2 - y^2$ orbital. To compensate, the other two equatorial ligands bend away releasing antibonding character along the *y* axis. The effect of back bending causes a significant energy increase in the d_{xy} orbital, which mixes with p_y orbitals and favours back-donation from the metal fragment in the equatorial plane. The other frontier orbitals do not suffer significant variations. Thus, equatorial ligands with short distances to the metal and π acceptor properties should promote ligand bending and, consequently, show larger distortions of the metal fragment. Accordantly, the $P(OH)_3$ ligand is the better π acceptor and shows shorter Rh-P distances in **Aea2** (2.33 Å) than PH_3 in **Aea1** (2.41 Å). The QM/MM calculations on model **B** considered the whole ligand in the MM part and symmetric PH_3 moieties in the QM part. The calculated $C_{carbonyl}$-Rh-$C_{carbonyl}$ angles are very similar for the **Bea2** (128°) and **Bea1** (129°) compounds. This indicates that the distortion in models complex **Aea2** are not a consequence of steric effects defined as destabilizing nonbonding type interactions but of the electronic distortion.[47] The **Bea1** isomer is slightly favoured by 1 kJ.mol^{-1} with respect to **Bea2** (third bar in Figure 3.4). Thus, the steric effects of the naphthyl and phenyl groups of the binaphos ligand also favour the placement of the phosphite moiety at the apical position. The new hybrid model **C** expands the QM region to include the oxygen atoms of phosphite moiety, accounting

for the sterics of the whole ligand and the electronics of the dissymmetric phosphorus moieties simultaneously. Thus, the computed energy difference (-7.6 kJ.mol^{-1}) is due to the electronic and steric effects (-4.6 kJ.mol^{-1}) and an additional synergic term (-3.0 kJ.mol^{-1}, fourth bar in Figure 3.4). Finally, we performed full DFT calculations which included the electronic effects of the naphthyl and phenyl groups of the ligand. This latter effect seems to somewhat favour the phosphite moiety in the apical position by about -2.7 kJ.mol^{-1} (fifth bar in Figure 3.4).

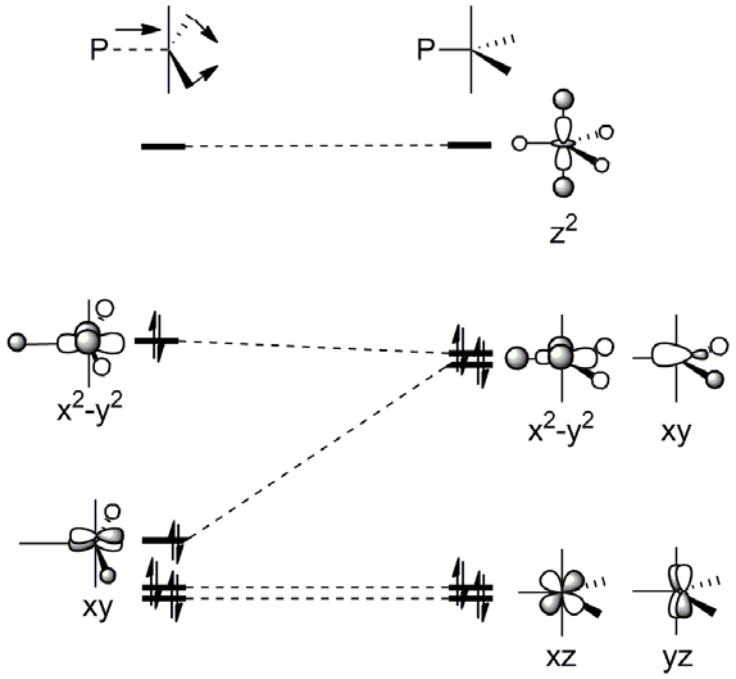


Figure 3.5. Qualitative frontier molecular orbital diagram for the elongation of an equatorial ligand in a d^8 [ML$_5$] complex.

In summary, electronic donor-acceptor interactions favour placing the phosphite moiety in the equatorial position, but one should also account for the electronic distortion of the metal fragment. The electronic distortion inverses the trend because the metal fragment is less distorted when phosphite is at the apical position. The steric effects of the bulky naphthyl and phenyl groups of binaphos also favour the placement of phosphite at the apical site. This latter effect is less pronounced, in contrast with previous proposals which attribute the coordination preferences to the steric effects of the bulky substituents as the main factor.[14] Overall, we predicted equatorial phosphane and apical phosphite (**ea1**) as the most stable isomer, in agreement with the observed major isomer.[14] Nevertheless, the small energy differences suggest that small tuning of ligand properties may have inverted the coordination preferences, which explains why other phosphane-phosphite ligands show coordination of the phosphite at the equatorial site (**ea2**).[4,10]

3.3.3. Coordination preferences at selectivity-determining step

Deuterioformylation with the Rh-binaphos catalyst indicates that alkene insertion is mostly irreversible,[48] and consequently, the alkene insertion into the Rh-H bond should be the enantioselectivity determining step. Carbó, Bo and co-workers proved this for rhodium catalysts modified with dissymmetric aminophosphane phophinite ligands.[11] In fact, all of the previous attempts to rationalize asymmetric hydroformylation are based on this step.[5a,12-13,15] Thus, we first analyzed the styrene insertion in model **A** in order to assess whether the electronic preference for apical phosphite in the resting state increases or decreases at the transition state. The computed **ea1** path is lower in energy than **ea2** by 7.5 kJ.mol^{-1},[49] increasing the electronic preference found in the resting state (4.5 kJ.mol^{-1}). The alkene insertion proceeded through rotation of the alkene moiety out of the equatorial plane of the trigonal-bipyramidal reactant, followed by the transfer of the hydride moiety. The rotational barrier of the alkene is largely determined by the amount of back-donation from the metal to the alkene moiety. On the one hand, when the π acceptor $P(OH)_3$ ligand is in the equatorial position (**ea2**), the competition for back-donation with the alkene is expected to be more effective. This would lead to a more facile rotation of the alkene moiety and, therefore, a low energy transition state. On the other hand, analogous to the hydrido carbonyl complex, the $[P(OH)_3]_{eq}$ induces a larger structural distortion as reflected in a larger P_{eq}-Rh-$C_{carbonyl}$ angle (125° *vs.* 122°). At this point, it seems that the proportion of the reaction that occurs through a channel with the phosphite moiety in the equatorial position (**ea2**), if it does, could be even lower than in the resting state. However, we cannot definitively discard a situation in which the **ea2** channel participates somewhat, and, therefore, in the next sections we will consider both paths.

3.3.4. Origin of enantioselectivity from ground-state properties

To avoid the complexity of transition state determinations, some attempts to rationalize enantioselectivity were used for structures with pseudo-five coordination environments and mimicking the geometry of the TS for alkene insertion.[5a,13,15] On the basis of this model and kinetic data, Landis and co-workers proposed a two-dimensional quadrant steric map for the determination of enantioselectivity for the Rh-(*S,S,S*)-bisdiazaphos catalyst.[15] In the quadrant diagram, the plane contains the Rh-H axis and is parallel to the alkene plane with the equatorial ligands lying behind the plane (Figure 3.6).[50] The model assumes an equatorial-apical coordination of the diphosphorus ligand and that the quadrant delimited by the two phosphorus atoms is blocked by steric effects. Here, we extended the model to account for two possible nonequivalent coordination sides of the alkene (paths **I** and **II**, in Figure 3.6).[5a,12] According to this model and for the styrene substrate, path **I** would favour the formation of an *S* isomer, but path **II** would favour the formation of an *R* isomer. Therefore, the mutual cancellation should be compensated by suitable backbone interactions favouring/disfavouring one of the two paths. In general, the quandrant-based stereochemical models proposed in the literature provide only qualitative information.

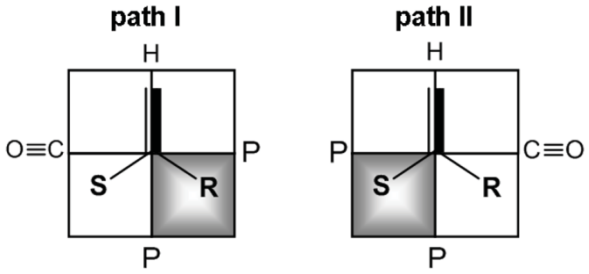


Figure 3.6. Initially proposed quadrant representation of the possible hydroformylation paths for diphosphorus ligands coordinated in the equatorial-axial mode.

Only very recently, Cavallo and co-workers have proposed a quantification of the different space occupation of the quadrants using the *percentage of the buried volume* (%V_{Bur}) as a molecular descriptor.[51] Inspired by this work, we defined a new molecular descriptor, the *distance-weighted volume* (V_W), which can easily be computed from the ground-state structures. The V_W gives a measure of the steric bulkiness of the ligand and its impact on the metal centre. We computed V_W using the [HRh(CO)$_2${(*R*,*S*)-binaphos)}] complexes aligned according to the chiral hypothesis in the quadrant representation. The Rh centre was placed at the origin with either carbonyl group defining path **I** or path **II** along the *z* axis and with the apical phosphorus atom in the *yz* plane (Figure 3.7). Then, we divided the *xy* plane along the *x* and *y* axes in the four regions (quadrants), for which we quantified the steric bulk. The quantification of the steric bulk considered three parameters: i) number of atoms in the front side of the catalyst ($z > 0$); ii) the size of the atom (r^3, r = van der Waals radius); and iii) the distance (d) from the atom to the metal centre. The factor r^3 was divided by d for each atom and the sum was extended for all of the atoms in the quadrant with $z > 0$. We expected that V_W would be an appropriate molecular descriptor when repulsive ligand-substrate interactions govern the enantioselectivity and that the larger the V_W value, the more disfavoured the path would be. For binaphos, the same sense of enantioface was observed for different heteroatom substituted alkenes, suggesting that the destabilizing nature of the interactions governs enantioselectivity.[5a] The V_W in the bottom two quadrants is more relevant for enantioselectivity because they indicate the orientation of the alkene substituent giving rise to the branched product. Table 3.3 shows the V_W values. For both **ea1** and **ea2**, the lowest value of V_W corresponded to the pro-*R* quadrant of path **II**. For path **I**, contrary to path **II**, the V_W value in the pro-*S* quadrant was lower than the pro-*R* quadrant but was still higher than in quadrant **II-*R***. Thus, we observed a relationship between the V_W and experimental selectivity.

These values allowed us to define a new sterochemical diagram (Figure 3.8). There are one sterically small quadrant (uncoloured), two quadrants that are sterically large (dark grey) and one that lies between the two extremes (light grey). This diagram rationalizes that pro-*R* transition state (bottom-right quadrant) being more favoured than the corresponding pro-*S* transition state (bottom-left quadrant) for path **II** and that the path **II** is, in general, favoured

over path **I** (Figure 3.8). Focussing on the favoured **ea1** coordination, we looked at the oriented geometries (Figure 3.7a and b). In path **II**, we observed an open space (Figure 3.7b, bottom-right), which gives a sterically nondemanding path for the pro-*R* transition state. The other side (Figure 3.8, bottom-left, **II-S**) is more sterically hindered (V_W = 9.3) due to the presence of the naphthyl fragment of the binaphthyl backbone bound to the apical phosphite (Figure 3.7b). In path **I**, most of the contribution to V_W comes from the naphthyl fragments of the apical phosphite binaphthyl, with a small additional contribution from the phenyl group of equatorial P substituent (Figure 3.7a). These results indicate that, in general, the steric influence of the apical P substituents governs the enantioselectivity. Moreover, they suggest that the axial chirality of the backbone discriminates between *R* and *S* isomers in path **II**, whereas the axial chirality of the phosphite moiety leads to enantiomeric discrimination of path **II** over **I**. This is consistent with the experimental findings showing that the absolute configuration of the backbone determines the sense of asymmetric induction; and that ligands with achiral diphenylphosphite moieties resulted in much lower enantioselectivity.[5a] The latter observation can be understood by a lower discrimination of path **II**. This would make path **I** and **II** competitive with each other giving a product with the opposite absolute configuration and resulting in an overall low enantioselectivity.

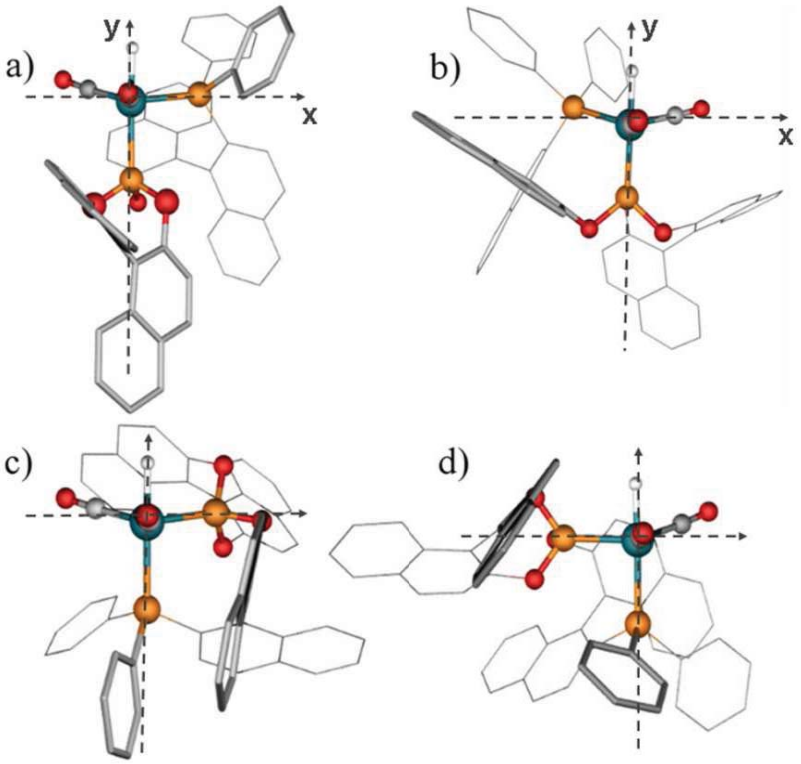


Figure 3.7. Molecular structure of the [HRh(CO)$_2${(*R*,*S*)-binaphos}] complex oriented by using the chiral hypothesis of the stereochemical model. The **ea1** coordination paths a) **I** and b) **II**; and the **ea2** coordination paths c) **I** and d) **II**.

Table 3.3. Values of V_W (Å^2) for the half of the bottom quadrants calculated from the **Cea1** and **Cea2** structures.

coordination		**I-*S***	**I-*R***	**II-*S***	**II-*R***
ea1	V_W	8.3	8.4	9.3	2.0
ea2	V_W	5.2	8.2	9.0	2.0

Turning our attention to the paths for **ea2**, the oriented geometries showed a pattern similar to path for **ea1** described above. Path **I** was sterically more hindered than path **II**, in this case, because the contribution of the naphthyl fragment of the binaphthyl backbone (Figure 3.7c). In path **II**, the pro-*S* side was more crowded due to one phenyl substituent of the apical phosphane (Figure 3.7d). This, along with the restricted phenyl rotation described above, is consistent with Pizzano's observations[4] which suggest that interaction between the substrate and the nonstereogenic apical phosphane moiety could explain enantioinduction. Finally, we could also conclude that the V_W molecular descriptor permits us to classify the steric properties of the Rh/binaphos catalyst and to predict the sense of enantioselectivity. An additional power of this modeling is the rapid identification of the preferred reaction paths from ground-state geometries, which can be obtained either computationally or experimentally.

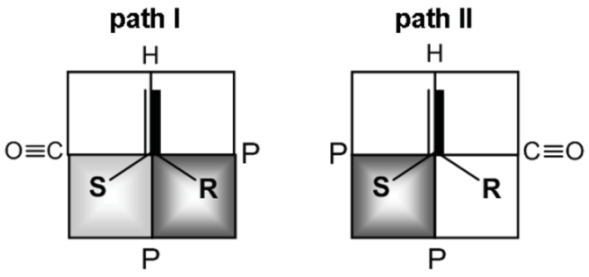


Figure 3.8. Newly proposed quadrant representation of the possible hydroformylation paths for binaphos ligands coordinated in the equatorial-apical mode.

3.3.5. Origin of enantioselectivity from transition state-based approaches

To more deeply analyze the factors governing enantioselectivity, and, considering that enantioselection might not arise purely from steric interactions, one might compute all of the possible transition state for styrene insertion. However, the small energy differences in the diastereomeric transition states (8 - 12 kJ.mol^{-1}) leading to enantiomerically enriched products are not always easily reproduced or rationalized. In a previous contribution to AHF based on transition state determination, our group of investigation showed that QM/MM methods overestimated the ligand-substrate interactions when stabilizing arene-arene interactions are involved.[11] Thus, when repulsive interactions control the selectivity, the sterics could be masked by the overestimated π-π contacts. In fact, considering only the lowest energy

rotation side of the styrene (Table 3.4), calculations of the all of the possible paths for **ea1** and **ea2** coordination in model **C** apparently fail in predicting the *R* isomer as the most favourable product (Table 3.5). Alternatively, we could perform full DFT calculations. However, most of the popular functionals, such as BP86, B3LYP or PBEh, are unable to describe noncovalent π-π interactions in their attractive regime.[52] Recently, Truhlar and co-workers recommended the M06 class of functionals for the study of noncovalent interactions.[32] Also, explicit dispersion corrections have been empirically added to conventional functionals (DFT-D) such as in B97D.[33] We tested these new functionals but they also seemed to fail to predict the right enantiomer (Table 3.6).

Table 3.4. Relative energies (in kJ.mol^{-1}) of the transition states for styrene insertion into the Rh-H bond at the BP86:UFF level using model **C**. From the alkene complex two rotations to reach the TS are considered: i) through the phosphane/phosphate side (labeled with P superscripts), or; ii) through the carbonyl side (labeled with CO superscript).

coordination	I-S^{CO}	I-S^{P}	I-R^{CO}	I-R^{P}	II-S^{CO}	II-S^{P}	II-R^{CO}	II-R^{P}
ea1	-15.3	-[a]	-11.3	-2.9	+8.9	+29.3	-[a]	0.0
ea2	+17.1[b]	+8.5	+23.1[b]	+15.6	-9.2	+53.9	+71.2	2.2

[(a)] The dihedral angle H-Rh-C=C is close to 0°, indicating that both rotations yield the same transition state. [(b)] Estimated value fixing the reaction centre form by the Rh, the hydride ligand, and the double bond of styrene.

Despite the methodological drawbacks of the QM/MM based TS approaches, we were able to obtain some interesting insight from a more profound analysis of the TS. Thus, in addition to relative energies, we divided the ligand-substrate contacts into interaction and steric deformation contributions (Table 3.5). We defined the interaction energy (IE) as the energy difference between the styrene and the binaphos ligand in the TS geometry and the two fragments at infinite distance in the geometries they adopt at the saddle point. The total steric deformation energy (ΔE_{Sdef}) is the sum of ligand and substrate contributions. For the ligand, we computed the energy difference between its geometry in the TS for styrene insertion and ethene insertion and for the substrate between its geometry in the TS of a real model (**C**) and a simplified model (**A**). In other words, we computed the energy difference between a sterically demanding situation and a nondemanding one. All of the IE values were stabilizing in nature and larger than the corresponding ΔE_{Sdef}. The largest values correspond to coloured quadrants (Figure 3.8), that is, to the regions with the most ligand-substrate interaction. Within the QM/MM calculations, these interactions were treated at the MM level, which gives overestimated energies and shorter distance between the arene rings.[11,53] Thus, short arene-arene contacts in the repulsive regime can become unrealistically attractive in MM calculations. On the other hand, the steric deformation energies followed the same trend as the steric diagrams derived from V_W; the ΔE_{Sdef} values are lower for the pro-*S* TS in path **I** and the pro-*R* TS in path **II**. Interestedly, we observed a correlation between the values of the V_W parameter and ΔE_{Sdef} energies, further proving the reliability of V_W as a molecular descriptor (Figure 3.9). In general, the overestimated interaction energies seem to outfit steric deformation; the pro-*S* TS of path **I-ea1** had the lowest energy structure (Table 3.5).

Table 3.5. Evaluation of the relative energies ($E_{rel.}$) of the transition states, ligand-substrate interaction energies (IE) and steric deformation energies (ΔE_{Sdef}) in kJ.mol^{-1}.[(a)]

coordination		**I-*S***	**I-*R***	**II-*S***	**II-*R***
ea1	$E_{rel.}$	-15.3	-11.3	+8.9	0.0
	IE	-30.7	-41.5	-27.1	-16.9
	ΔE_{Sdef}	+16.4	+18.9	+26.3	-2.4
ea2	$E_{rel.}$	+8.5	+15.6	-9.2	-2.2
	IE	-19.3	-30.0	-39.5	-20.0
	ΔE_{Sdef}	-8.4	+18.0	+12.8	+7.2

[(a)] Only the lowest energy rotation side of the styrene was considered (see Table 3.4).

Table 3.6. Relative energies (in kJ.mol^{-1}) of the transition states for styrene insertion into the Rh-H bond through **ea1** path. Full DFT calculations with M06 and B97D functionals using the lowest energy isomers of the QM/MM calculations.

functional	**I-*S***	**I-*R***	**II-*S***	**II-*R***
M06	-14.7	-0.2	+12.9	0.0
B97[(a)]	-7.2	-2.6	+26.8	0.0
B97:UFF[(a)]	-17.0	-8.3	+15.0	0.0
B97/SMD[(a)]	-0.6	+10.5	+30.5	0.0

[(a)] QM/MM calculations at the B97D:UFF level, full DFT calculations with B97D functional, and B97D calculation adding solvent (benzene) corrections via continuum SMD model.

In this context, it is difficult to judge whether the stabilizing or repulsive interactions prevail, but an additional consideration can rule out path **I**. In the analysis of the rate-determining step in the Rh/Xantphos-catalyzed hydroformylation, they showed that the energy of the TS for alkene coordination is relatively close to the TS for alkene insertion.[19b] Gleich and Hutter combined static and dynamic QM calculations to study alkene coordination, demonstrating the difficulties in finding the transition state for alkene coordination, which apparently only appears with sufficient steric repulsion between the incoming substrate and the metal fragment.[54] Thus, a reaction through the sterically crowded path **I** could not be prevented during alkene insertion but during alkene coordination. Unfortunately, it is unclear whether high kinetic barrier exists for the formation of one or more isomers of the styrene complex before alkene insertion. However, it is reasonable to think that this mechanism would help to discriminate in favour of path **II**.

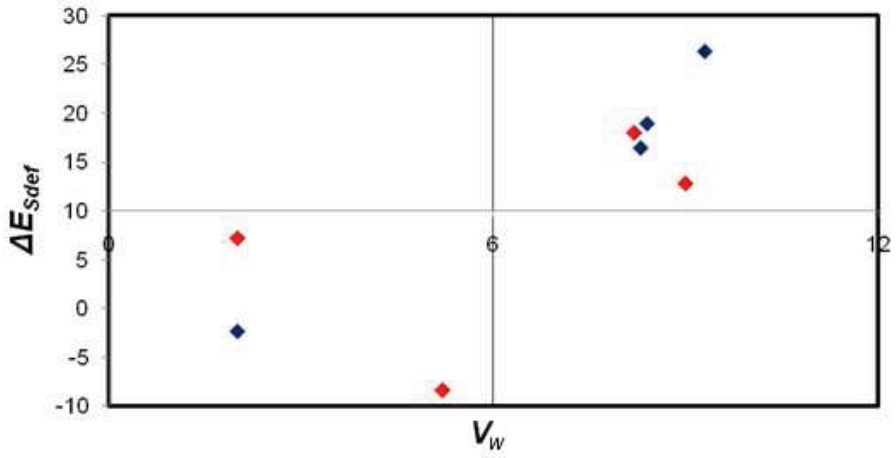


Figure 3.9. Comparison between V_W and ΔE_{Sdef} values. There is a fair correlation between the descriptor and steric distortion energy: large values of V_W correspond to high steric distortion energies. Blue dots correspond to **ea1** isomer and red dots to **ea2** isomer.

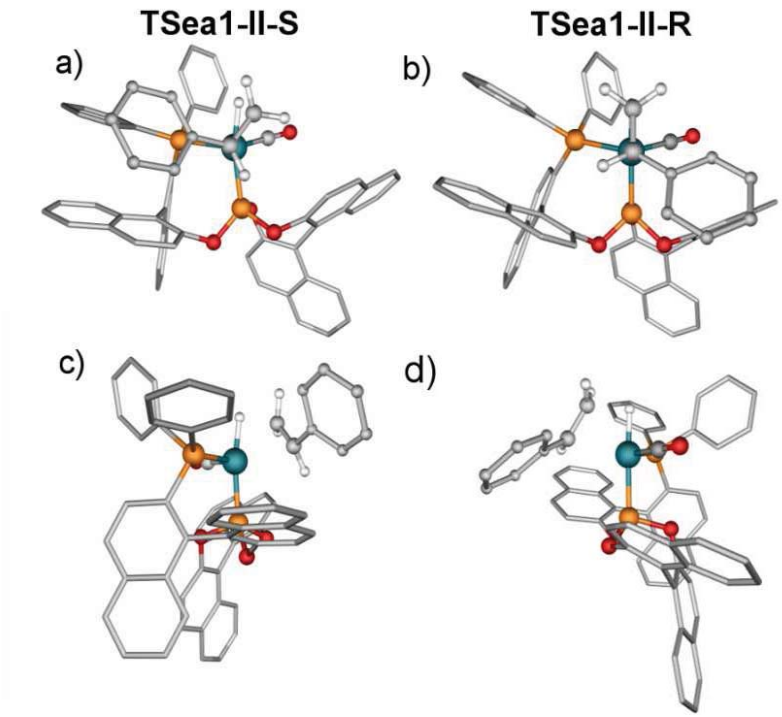


Figure 3.10. Molecular structures of pro-*S* and pro-*R* TS for styrene insertion: **TSea1-II-*S*** [a) front view, c) side view] and **TSea1-II-*R*** [b) front view, d) side view]. The hydrogen atoms of the phenyl and naphthyl groups are omitted for clarity.

Moreover, contrary to previous attempts to rationalize enantioselectivity exclusively through determination of transition states for alkene insertion, these results indicate that the situation is more complex and that other steps of the process need to be considered. In the experimentally consistent path **II**, the values of the ΔE_{Sdef} energies demostrate that the pro-*R* TS is less sterically demanding than the pro-*S* TS one for both **ea1** and **ea2** (Table 3.5). Interestedly, the difference in ΔE_{Sdef} energies between the pro-*S* and pro–*R* TS is significantly larger for **ea1** (29 kJ.mol^{-1}) than for **ea2** (6 kJ.mol^{-1}). In the latter, the substrate interacts with the less rigid apical diphenylphosphino moiety, allowing ligand reorganization with a low energy cost and leading to lower enantiodiscrimination. Although during our MD simulation we did not observe the complete rotation of the phenyl groups, they showed a relatively flexible behaviour. This would explain why phosphane-phosphite ligands analogous to binaphos with preferential apical phosphane coordination (**ea2**) yielded low enanatioselectivities.[4]

Finally, Figure 3.10 shows the geometries of the TS for the key path **II-ea1**. The most favoured pro-*R* transition state (**TSea1-II-R**) presents the substrate substituent in a rather open part of the space. On the other hand, in the competitive pro-*S* transition state (**TSea1-II-S**), the substrate substituent points toward one of the naphthyl fragments of the binaphthyl backbone. As a consequence, the structure of the TS is distorted toward square-pyramidal geometry, resulting in higher energy.[8]

3.4. Conclusions

In this Chapter, we performed a fundamental analysis of the coordination preferences of the phosphane-phosphite (*R*,*S*)-binaphos ligand in the [HRh(P-P)(CO)$_2$] complex, which acts as the catalyst for asymmetric hydroformylation. Using a range of computational tools, we were able to identify and quantify the different effects governing the observed preference for the phosphane moiety at the equatorial site and the phosphite at the apical site. As expected, in unsymmetric ligands, the evaluation of the electronic donor-acceptor interactions favoured placing the least basic phosphite moiety at the equatorial position. However, the trend was inversed as a consequence of the nonbonding steric effects induced by the bulky naphthyl and phenyl groups of binaphos and, to a large extends, by the larger electronic distortion induced by the better π acceptor phosphite moiety at the equatorial site. This balance explains the exceptional behaviour of this ligand. The ligand coordination has important implications in the mechanism of enantioinduction. In the stereoselectivity determining transition state, the key ligand-substrate interactions occur between the styrene and the apical ligand moiety. All of the evidences indicated that the coordination preference in the resting state was transferred to the transition states. Thus, the axial chirality of the binaphthyl group of the apical phosphite discriminates between two competing equatorial-apical paths (**I** and **II**) which favour the formation of different enantiomers. This avoids mutual cancellation of enantioselectivity and explains why the ligands with an achiral diphenylphosphite moiety resulted in much lower enantioselectivity. Within the preferred path (**II**), the second axially chiral centre in the ligand backbone discriminates between the *R* and *S* isomers, in agreement with the observed change in the sense of asymmetric induction upon change of absolute configuration of the backbone.

In all of these mechanisms, the ligand-substrate interactions are destabilizing in nature. The placement of phosphane moiety at the apical position would also induce some stereoselectivity through propeller-like chirality. However, the larger flexibility of the diphenylphosphino group would lead to lower enantiodiscrimination, explaining why other phosphane-phosphite ligands with preferential apical coordination of the nonstereogenic phosphane yield low enantioselectivities.

We built a sterochemical model based on quantitative quadrant-diagrams obtained from a newly defined molecular descriptor, the *distance-weighted volume* (V_W), which can be easily computed from ground state geometries. Giving a defined coordination, the newly developed model was able to predict that the alkene approach through region **I** is sterically hindered and that the preferred region **II** preferentially yielded the *R* enantiomer. This constitutes an alternative to more complex TS-based approaches, which could be less meaningful due to the difficulty of affordable methods to deal with nonbonding dispersion-type interactions. Currently, we are exploring this and other descriptors to set quantitative relationships between them and the enantioselectivity in AHF for a wide set of ligands. The alignment-dependent CoMFA-like approaches cannot consider because it is needs two alignments, the path **I** and **II** to explain enantioselectivity.

Finally, with these and previous results in hand, we believe that it is possible propose several requirements for the design of stereoselective Rh-diphosphane catalysts: i) equatorial-apical specific coordination with chirality at the apical site in order to allow diastereoisomeric ligand-substrate interactions; ii) combination of two stereogenic centres, each of which discriminate the competing equatorial-apical paths or the enantiomers; and iii) rigid structures for a more effective chiral induction.

3.5. References and notes

1 a) I. Ojima, *Catalytic Asymmetric Synthesis*, Wiley-VCH, New York, **2000**. b) E. N. Jacobsen, A. Pfaltz, H. Yamamoto, *Comprehensive Asymmetric Catalysis*, Springer-Verlag, New York, **1999**.

2 N. Sakai, S. Mano, K. Nozaki, H. Takaya, *J. Am. Chem. Soc*. **1993**, *115*, 7033.

3 N. Nozaki, *Chem. Rec.* **2005**, *5*, 376.

4 a) M. Rubio, A. Suárez, E. Álvarez, C. Bianchini, W. Oberhauser, M. Peruzzini, A. Pizzano, *Organometallics* **2007**, *26*, 6428. b) I. Arribas, S. Vargas, M. Rubio, A. Suárez, C. Domene, E. Alvarez, A. Pizzano, *Organometallics* **2010**, *29*, 5791.

5 a) K. Nozaki, N. Sakai, T. Nanno, T. Higashijima, S. Mano, T. Horiuchi, H. Takaya, *J. Am. Chem. Soc.* **1997**, *119*, 4413. b) M. Diéguez, O. Pàmies, A. Ruiz, S. Castillón, C. Claver, *Chem. Eur. J.* **2001**, *7*, 3086.

6 C. P. Casey, E. L. Paulsen, E. W. Beuttenmueller, B. R. Proft, B. A. Matter, D. R. Powell, *J. Am. Chem. Soc.* **1999**, *121*, 63.

7 A. R. Rossi, R. Hoffmann, *Inorg. Chem*. **1975**, *14*, 365.

8 E. Zuidema, E. Daura-Oller, J. J. Carbó, C. Bo, P. W. N. M. Leeuwen, *Organometallics* **2007**, *26*, 2234.

9 A. Suárez, M. A. Méndez-Rojas, A. Pizzano, *Organometallics* **2002**, *21*, 4611.

10 a) S. Deerenberg, P. C. J. Kamer, P. W. N. M. van Leeuwen, *Organometallics* **2000**, *19*, 2065. b) O. Pàmies, G. Net, A. Ruiz, C. Claver, *Tetrahedron: Asymmetry* **2001**, *12*, 3441. c) T. Robert, Z. Abiri, J. Wassenaar, A. J. Sandee, S. Romanski, J.-M. Neudörfl, H.-G. Schmalz, J. N. H. Reek, *Organometallics*, **2010**, *29*, 478. d) S. H. Chikkali, R. Bellini, G. Berthon-Gelloz, J. I. van der Vlugt, B. de Bruin, J. N. H. Reek, *Chem. Commun.* **2010**, *46*, 1244.

11 J. J. Carbó, A. Lledós, D. Vogt, C. Bo, *Chem. Eur. J.* **2006**, *12*, 1457.

12 D. Gleich, R. Schmid, W. A. Herrmann, *Organometallics* **1998**, *17*, 2141.

13 X. Zhang, B. Cao, Y. Yan, S. Yu, B. Ji, X. Zhang, *Chem. Eur. J.* **2010**, *16*, 871.

14 D. A. Castillo-Molina, C. P. Casey, I. Müller, K. Nozaki, C. Jäkel, *Organometallics* **2010**, *29*, 3362.

15 A. L. Watkins, C. R. Landis, *J. Am. Chem. Soc.* **2010**, *132*, 10306.

16 C. J. Cobley, R. D. J. Froese, J. Klosin, C. Quin, G. T. Whiteker, *Organometallics* **2007**, *26*, 2986.

17 A. T. Axtell, J. Klosin, G. T. Whiteker, *Organometallics* **2009**, *28*, 2993.

18 a) M. C. Kozlowski, S. L. Dixon, M. Panda, G. Lauri, *J. Am. Chem. Soc.* **2003**, *125*, 6614. b) S. Sciabola, A. Alex, P. D. Higginson, J. C. Mitchell, M. J. Snowden, I. Morao, *J. Org. Chem*. **2005**, *70*, 9025. c) J. L. Melville, K. R. J. Lovelock, C. Wilson, B. Allbutt, E. K. Burke, B. Lygo, J. D. Hirst, *J. Chem. Inf. Model.* **2005**, *45*, 971. d) M. Urbano-Cuadrado, J. J. Carbó, A. G. Maldonado, C. Bo, *J. Chem. Inf. Model.* **2007**, *47*, 2228. e) N. Fey *Dalton Trans.* **2010**, *39*, 296.

19 a) J. J. Carbó, F. Maseras, C. Bo, P. W. N. M. van Leeuwen, *J. Am. Chem. Soc.* **2001**, *123*, 7630. b) E. Zuidema, L. Escorihuela, T. Eichelsheim, J. J. Carbó, C. Bo, P. C. J. Kamer, P. W. N. M. van Leeuwen, *Chem. Eur. J.* **2008**, *14*, 1843.

20 a) ADF 2007.01. Department of Theoretical Chemistry. Vrije Universiteit. Amsterdam. b) E. J. Baerends, D. E. Ellis, P. Ros, *Chem. Phys*. **1973**, *2*, 41. c) L. Versluis, T. Ziegler, *J. Chem. Phys*. **1988**, *88*, 322. d) G. Te Velde, E. J. Baerends, *J. Comput. Phys*. **1992**, *99*, 84. e) C. Fonseca Guerra, J. G. Snijders, G. Te Velde, E. J. Baerends, *Theor. Chem. Acc*. **1998**, *99*, 391. f) G. Te Velde, F. M. Bickelhaupt, S. J. A. van Gisbergen, C. F. Guerra, E. J. Baerends, J. G. Snijders, T. Ziegler, *J. Comput. Chem*. **2001**, *22*, 931.

21 a) A. D. Becke, *J. Chem. Phys*. **1986**, *84*, 4524. b) A. D. Becke, *Phys. Rev*. *A* **1988**, *38*, 3098.

22 S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys*. **1980**, *58*, 1200.

23 a) J. P. Perdew, *Phys. Rev*. *B* **1986**, *33*, 8822. b) J. P. Perdew, *Phys. Rev. B* **1986**, *34*, 7406.

24 F. Maseras, K. Morokuma, *J. Comput. Chem*. **1995**, *16*, 1170.

25 T. K. Woo, L. Cavallo, T. Ziegler, *Theor. Chem. Acc.* **1998**, *100*, 307.

26 a) M. Clark, R. D. III Cramer, N. van Opdenbosch, *J. Comput. Chem*. **1989**, *10*, 982. b) U. C. Sing, P. A. Kollman, *J. Comput. Chem.* **1986**, *7*, 718.

27 A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. III Goddard, W. M. Skiff, *J. Am. Chem. Soc.* **1992**, *114*, 10024.

28 a) K. Morokuma, *J. Chem. Phys*. **1971**, *55*, 1236. b) T. Ziegler, A. Rauk, *Theor. Chem. Acc.* **1977**, *46*, 1. c) T. Ziegler, A. Rauk, *Inorg. Chem*. **1979**, *18*, 1755. d) T. Ziegler, A. Rauk, *Inorg. Chem*. **1979**, *18*, 1558.

29 F. M. Bickelhaupt, N. M. M. Nibbering, E. M. van Wezenbeek, E. J. Baerends, *J. Phys. Chem.* **1992**, *96*, 4864.

30 a) N. S. Antonova, J. J. Carbó, J. M. Poblet, *Organometallics* **2009**, *28*, 4283. b) N. S. Antonova, J. J. Carbó, J. M. Poblet, *Dalton Trans.* **2011**, *40*, 2975.

31 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Jr. Montgomery, T. Vreven, K. N. Kudin, J. R. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, C. González, J. A. Pople, *Gaussian 09*, revision A.02; Gaussian, Inc.: Pittsburgh, PA, **2009**.

32 Y. Zhao, D. G. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 157.

33 S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787.

34 S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, M. J. Frisch, *J. Mol. Struct. (Theochem)* **1999**, *461-462*, 1.

35 P. J. Hay, W. R. Wadt, *J. Chem. Phys.* **1985**, *82*, 270.

36 A. Höllwarth, M. Böhme, S. Dapprich, A. W. Ehlers, A. Gobbi, V. Jonas, K. F. Köhler, R. Stegmann, A. Veldkamp, G. Frenking, *Chem. Phys. Lett.* **1993,** *208*, 237.

37 a) M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. Defrees, J. A. Pople, *J. Chem. Phys.* **1982**, *77*, 3654. b) W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257. c) P. C. Hariharan, J. A. Pople, *Theor. Chem. Acc.* **1973,** *28*, 213.

38 A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6378.

39 K. Nozaki, N. Sato, Y. Tonomura, M. Yasutomi, H. Takaya, T. Hiyama, T. Matsubara, N. Koga, *J. Am. Chem. Soc.* **1997**, *119*, 12779.

40 a) S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, K. Tanabe, *J. Am. Chem. Soc.* **2002**, *124*, 104. b) M. O. Sinnokrot, E. F. Valeev, C. D. Sherill, *J. Am. Chem. Soc.* **2002**, *124*, 10887. c) M. O. Sinnokrot, C. D. Sherill, *J. Phys. Chem. A* **2004**, *108*, 10200.

41 E. Bustelo, J. J. Carbó, A. Lledós, K. Mereiter, M. C. Puerta, P. Valerga, *J. Am. Chem. Soc*. **2003**, *125*, 3311.

42 a) C. Bo, F. Maseras, *Dalton Trans.* **2008**, 2911. b) F. Maseras, *Chem. Commum.* **2000**, 1821.

43 L. R. Falvello, J. C. Gines, J. J. Carbó, A. Lledós, R. Navarro, T. Soler, E. P. Urriolabeitia, *Inorg. Chem*. **2006**, *45*, 6803.

44 Also the destabilizing term by the axial ligand favours the **Aea1** over the **Aea2** compound with modest calculated value of 2 $kJ.mol^{-1}$.

45 D. J. Wink, *Organometallics* **1991**, *10*, 442.

46 a) M. Elian, R. Hoffmann, *Inorg. Chem.* **1975**, *14*, 1058. b) T. A. Albright, R. Hoffmann, Y. Tse, T. D'Ottavio, *J. Am. Chem. Soc.* **1979**, *101*, 3812.

47 Note that there is not a precise and universal definition of what a steric effect is and that the definition within this QM/MM methodology is as arbitrary as any other. However, the QM/MM scheme provides a straightforward separation of the steric effects from the electronic ones, allowing them to be quantified.[42]

48 T. Horiuchi, E. Shirakawa, K. Nozaki, H. Takaya, *Organometallics* **1997**, *16*, 2981.

49 We calculated the TS for all of the possible paths of each ligand coordination. As expected for nonchiral ligands, all of them are isoenergetic within the same ligand coordination with energy differences < 1 $kJ.mol^{-1}$.

50 Here, we did not use the stereochemical model to rationalize the regioselectivity, instead considering only the branched products; therefore, the top quadrants are not discussed.

51 A. Poater, L. Cavallo, *Dalton. Trans.* **2009**, 8885.

52 a) Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2007**, *3*, 289. b) S. Tsuzuki, H. P. Lüthi, *J. Chem. Phys*. **2001**, *114*, 3949. c) E. J. Meijer, M. Sprik, *J. Chem. Phys.* **1996**, *105*, 8684.

53 C. D. Sherrill, B. G. Sumpter, M. O. Sinnokrot, M. S. Marshall, E. G. Hohenstein, R. C. Walker, I. R. Gould, *J. Comput. Chem.* **2009**, *30*, 2187.

54 D. Gleich, J. Hutter, *Chem. Eur. J.* **2004**, *10*, 2435.

Chapter 4

QSPR models for Predicting the Enantioselectivity and the Activity in Asymmetric Hydroformylation of Styrene Catalyzed by Rh-diphosphane

Chapter 4

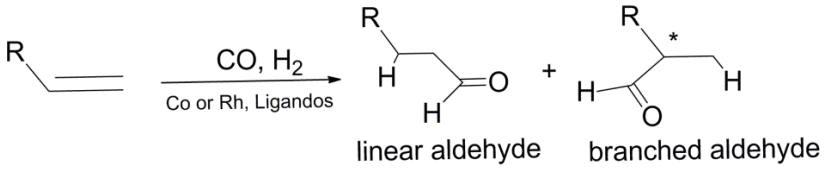
QSPR models for Predicting the Enantioselectivity and the Activity in Asymmetric Hydroformylation of Styrene Catalyzed by Rh-diphosphane

This Chapter develops Quantitative Structure-Selectivity and -Activity Relationships for the asymmetric hydroformylation of styrene by Rh-diphosphane. The QSPR models were generated using a 2D-QSPR and 3D-QSPR methodologies. Firstly, we used 3D steric- and electrostatic-type interaction fields derived from DFT calculation, and generated alignment-independent descriptors using *GRid INdependent Descriptor* (GRIND) methodology. The obtained QSPR models showed statistical significance and predictive ability. The most predictive model for enantioselectivity was obtained using steric-type 3D fields (MSF) based on the local curvature electron density isosurface that accounts for catalyst shape (r^2 = 0.92, q^2 = 0.68). We obtained the most predictive model for activity using a combination of shape- and electrostatic-based 3D fields (r^2 = 0.99, q^2 = 0.74). The use of chemically meaningful descriptors provides insight into the factors governing catalytic activity and selectivity. The worst predicted ligand kelliphite showed the lowest preference for equatorial-apical coordination and low selectivity, which suggests that its intrinsic enantiotopic differentiation capacity can be lost through the occurrence of bis-equatorial paths. The selective catalysts Rh-chiraphite, -binapine, -diazaphospholane, and -yanphos showed the same pattern as Rh-binaphos, for which the origin of stereoinduction is known: the chirality at the apical site discriminates one alkene coordination path and one enantiomer. Ligands with electron withdrawing groups at phosphorus atoms such as chiraphite, kelliphite, binaphos, and diazaphospholane reduce ligand basicity and promote catalytic activity effectively. However, more complex relationships underlie the origin of activity, and the shape of the catalyst also needs to be considered such as for Ph-BPE ligand. Comparison with previous studies suggests that reduction of the steric hindrance at the reaction centre which favours alkene coordination and insertion would also promote catalytic activity. Secondly, we used several atomic properties and generated alignment-independent descriptors using *Topological MAximum Cross Correction* (TMACC) approach. The obtained 2D-QSPR model showed statistical significance and predictive ability for activity when all atomic properties were considered. This confirms the hypothesis that several properties are necessary to model the activity; and that including properties related with basicity and steric effects is needed. A 2D-QSPR model for enantioselectivity was not possible because the lack of realistic description of catalyst shape and/or the lack of specifically designed descriptors.

4.1. Introduction

As described in previous chapters, asymmetric hydroformylation (AHF) offers a potentially very useful synthetic route to enantiomerically pure aldehydes.[1] Scheme 4.1 shows that the production of branched aldehydes can generate a chiral aldehyde enantioselectively. In asymmetric synthesis, hydrogenation is probably the most used, if not the only widely used, homogeneous catalytic process. Thus, it would be very valuable if we could provide criteria for the design of new active and selective chiral catalysts for AHF.

Scheme 4.1.



The previous studies on AHF revealed that TS-based modeling has some limitations. Carbó, Bo and co-workers showed that QM/MM methods overestimate ligand-substrate interactions when stabilizing arene-arene type interactions are involved, whereas full DFT calculations are unable to describe these noncovalent interactions.[2] Moreover for Rh-binaphos system, the discrimination of one of the paths does not occur at the selectivity-determining step but in a previous step of catalytic cycle, forcing to handle multiple kinetic scenarios (*Chapter 3*).[3] Alternatively, more sophisticated computational approaches defined molecular descriptors from ground-state structures, to look for quantitative relationships between descriptors and enantioselectivity, in an analogous way to that Quantitative Structure-Activity Relationship (QSAR) or -Property Relationship (QSPR) approaches used in drug design.[4] One of the most attractive aspects of these methods is that they generate simple equations, which allow predicting catalyst activity and selectivity a priori. Although QSAR techniques are not of common use in homogenous catalysis, there are some outstanding examples that correlate catalyst activity and selectivity.[5]

In asymmetric catalysis, the situation is even more challenging because the energy differences responsible of enantiomeric enrichment can be as low as 1-3 kcal.mol^{-1}. To identify Quantitative Structure-Selectivity Relationships, in principle, accounting for 3D features of the chiral catalyst is needed in order to reflect its enantiotopic differentiation capacity. It is possible to use geometrical descriptors based on steric size,[6] molecular topology,[7] continuous chirality quantification methods[8] that indirectly account for 3-dimensionality or 3D *molecular interaction field* (MIF) descriptors. Geometrical descriptors include easily accessible parameters from classical physical organic chemistry such as the *Taft-Charton steric* parameters,[6] but these are only defined for some specific functional groups, so cannot be applied to more complex ligand architectures.

The MIF-based approaches became the standard 3D-QSAR methods used in drug design, being CoMFA (*Comparative Molecular Field Analysis*) approach the most popular.[9] The generation of CoMFA descriptors involves three stages: i) generation of the molecular structures; ii) alignment of structures to obtain reproducible predictive spaces; and iii) computation of the MIFs at each point of a 3D grid surrounding the molecule. Initially, those methods used MM-based interactions, but semiempirical[10] and QM[11] interaction energies have been also used. CoMFA-like approaches enabled obtaining fairly good results in some asymmetric catalytic processes as well.[10-12] For Rh-diphosphane AHF catalysis, we have shown that two different orientations of the catalyst are required to explain enantioinduction (*Chapter 3*);[3] and consequently, it is not possible to apply an alignment-dependent approach straightforwardly because several alignment hypothesis should be combined simultaneously. Pastor and co-workers have introduced the *GRid-Independent descriptors* (GRIND), for which the alignment hypothesis is not required.[13] GRIND methods combined with empirical molecular interaction fields have showed their effectiveness in several asymmetric organometallic catalytic processes.[14] Similarly, Bo and co-workers developed a CoMFA-like alignment-independent (GRIND) protocol based on 3D molecular descriptors, which are derived from DFT calculations directly, named *Q-QSSR*.[15] A limitation of alignment-independent methodologies is the fact that molecular descriptors are not chiral, and thus unable to predict the absolute configuration of the product. Nevertheless, and we will show below, they are well suited to assess enantiomeric excess quantitatively.

Our recent research efforts have been aimed at rationalizing of the enantioselectivity outcome in Rh-catalyzed hydroformylation for binaphos ligand (*Chapter 3*)[3] Here, we introduce models for correlating catalyst structure with enantioselectivity and activity quantitatively for a set of 21 Rh-diphosphane catalysts. The dataset includes a variety of ligands that have different substituents, backbones and phosphane types. We combine the 3D alignment-independent protocol[15] and the 2D-type TMACC descriptors[16] together with previous mechanistic knowledge.[2-3,17] The use of chemically meaningful descriptors provides, in addition to the predictive models, chemical information to orient the design of ligands.

4.2. Computational details and methods

4.2.1. Ligand dataset and structure generation

The ligand dataset in this study is a library of 21 Rh-diphosphorus catalysts, which were mainly obtained from Ref. 18 (ligands **1**-**19**), and thus we made sure that reactions were conducted in the same laboratory and under comparable conditions (solvent, temperature, catalyst loading, etc.). Additionally, we add two ligands (**20** and **21**) from Ref. 19. The experimental conditions differ only slightly, and the catalysts can then be used to check the chemical meaning of the QSPR model and the descriptors by comparison with previously obtained stereo-chemical models based on transition state determinations.[2] Moreover, the selection showed in Figure 4.1 encompasses a wide enantioselectivity and activity ranges for the asymmetric hydroformylation of styrene by Rh-diphosphane complexes. The response variable for

selectivity is percentage of enantiomeric excess (*ee*) whereas for activity is percentage of conversion (*conv*). Figure 4.1 shows these values in parentheses.

The molecular structures were obtained by means of quantum mechanics/molecular mechanics (QM/MM) calculations[20] using the ADF2006.01 package.[21] The QM part was $[HRhCO(PH_3)_2]$ for diphosphane and diazaphospholane ligands, $[HRhCO(P(OH)_3)_2]$ for diphosphite ligands, whereas the geometry of dissymmetric ligands (**5**, **20** and **21**) was taken from our previous study (*Chapter 3*),[3] and previous studies of our research group,[2] respectively.

For the QM part, we selected DFT method with BP86 functional[22] and TZP basis set.[23] SYBYL force field[24] was used to describe the atoms included in the MM part.[25] To search for conformational isomers, we defined a protocol based on restricted Molecular Dynamics (MD) simulations as implemented in ADF software.[21] This protocol had been used successfully in our previous study of Rh-binaphos catalyst (*Chapter 3*).[3] It consists of i) initial QM/MM optimization at BP86:SYBYL level; ii) 100 ps of restricted MD simulation fixing the rhodium coordination sphere at 1000 K, sampling a geometry each ps; iii) 100 partial geometry optimizations fixing the Rh coordination sphere to select the 10 lowest energy geometries; and iv) final full geometry optimization of the selected structures at BP86:SYBYL level. The lowest energy isomer was then used to compute the DFT-based molecular interaction fields.

4.2.2. 3D-QSPR approach. Q-QSSR methodology

We selected the 3D-QSPR approach developed by Bo and co-workers, *Q-QSSR*, which uses alignment-independent 3D descriptors derived from DFT calculations.[15] Once the 3D structures are obtained, we can divide the working protocol in four stages: i) generation of the DFT molecular interaction fields; ii) filtering of the grid points; iii) computation of the molecular descriptors using GRIND methodology;[13] and iv) multivariate analysis.

We firstly generated the molecular surface from the DFT-derived electronic density distribution by performing single point calculations at BP86/TZP level on the whole structures obtained with the protocol defined above. The molecular surface was defined by an electronic density isosurface at a value of 0.002 au in a rectangular grid of points, which expands 5 Å beyond molecular boundaries and spaces 0.2 Å. Then, at each of the points of the isodensity surface we computed two different fields: the *Molecular Shape Field* (MSF) and the *Molecular Electrostatic Potential* (MEP). The MSF is a geometrical parameter based on the local curvature of isodensity surface and tries to account for steric effects of the catalyst.[13b] The MEP tries to assess the electronic properties of the catalyst.

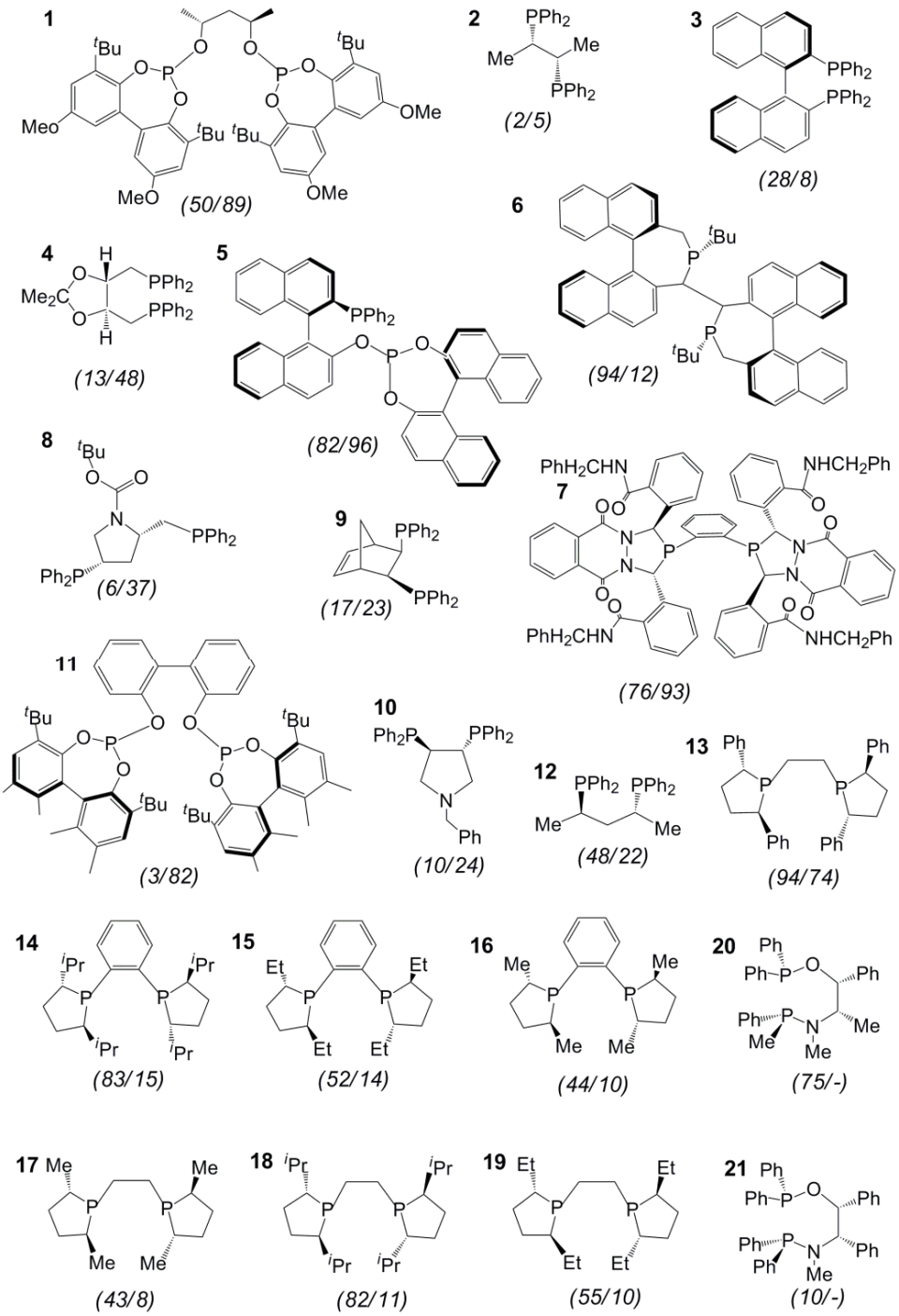


Figure 4.1. Ligand dataset and catalytic outcome for asymmetric hydroformylation of styrene by rhodium complexes (%*ee* / %*conv*).

The number of surface data points (from ~5000 to ~20000 depending on ligand size) needs to be filtered to reduce the dataset size and to select the most chemically relevant regions. We used two types of criteria for node selection: the shape MSF- and electrostatic MEP-based filtering. As in previous application of *Q-QSSR* approach,[15] we selected the most convex areas of the isosurface (*CONV*) as being the most descriptive of the molecular shape by keeping one node out of 15. In addition, we also explored filtering the points based on the MEP field for selecting the most basic areas of the catalyst (*BAS*), which contains the nodes that have the most negative values. In this case, a distance-based correction was introduced in the algorithms in order to avoid the selection of too few representative regions. The correction procedure is defined as follows: i) random selection of nodes; ii) evaluation of a quality value (from 0 to 1), which depends on the MEP values and the distance between each pair of nodes; iii) iteratively exchange of nodes showing the worst individual quality values until the global quality value behaves asymptotically; and iv) choosing the node selection with the best quality values. The details of this procedure will be published elsewhere.[15] Figure 4.2 shows the result of shape and electrostatic filtering for ligand **1,** where differences are clearly seen. In line with our general understanding, we can observe that the MSF filtering selects the bulky groups of the ligand while the MEP filtering selects the basic areas of the ligand such as oxygen and aromatic groups.

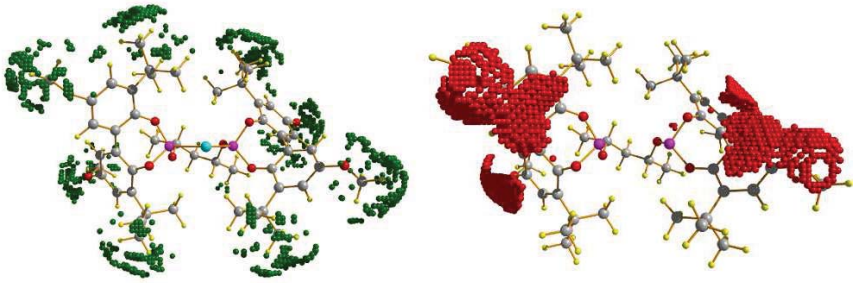


Figure 4.2. Example convexity MSF-based filtering (*CONV*, green) and basicity MEP-based filtering (*BAS*, red) criteria for ligand **1**. The *CONV* filtering selects lignad's regions associated with bulky groups such as ^{t}Bu and OMe, whereas the *BAS* filtering selects those associated with basic moieties such as the Ph and the oxygens.

Grid alignment independence was achieved using the GRIND methodology proposed by Pastor.[13] Within this methodology, we compute the product of the molecular field (MSF and MEP) for each pair of filtered grid points and then select the highest values of these products for each distance range. Thus, one builds the *correlograms* (see Figure 4.3), which are the true molecular descriptors that will enter multivariate analysis. We tested *correlograms* with distance intervals of 0.25, 0.5 and 1.0 Å and with distance range between 3 and 10 Å, 3 and 15 Å, and 3 and 20 Å. An optimal interval of 0.25 Å was found, and products of MSF and MEP values were computed between 3 and 20 Å. To sum up, with two molecular fields and two filtering criteria, we can build 4 different *correlograms*, that is, 4 different descriptors: MSF-

MSF_{CONV}, MEP-MEP_{CONV}, MSF-MSF_{BAS} and MEP-MEP_{BAS}. Additionally, we compared the results with a similar commercially-available alignment-independent approach (GRIND), which uses empirical interaction fields instead, as implemented in the ALMOND software.[26] The alignment-independent descriptors were built from the empirical *TIP* interactions generated with *C3* probes (an atom of carbon with sp^3 hybridization) which evaluate the shape of the catalyst in a similar way as MSF does. The *TIP* was computed using a grid-spacing 0.3 Å and the grid extending 1.4 Å beyond molecule boundaries.

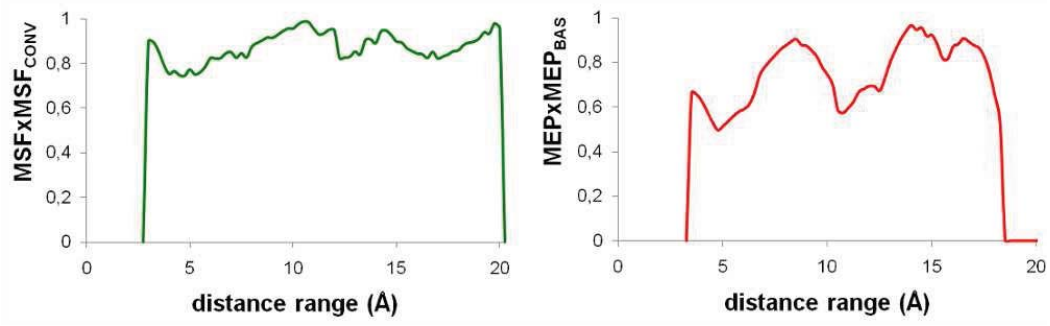


Figure 4.3. Example of MSF-MSF_{CONV} and MEP-MEP_{BAS} *correlograms* (the descriptors) for ligand **1**. The highest values of the products between fields values for pairs of filtered points (normalized) *vs.* distances between the nodes (in Å).

We used the *Partial Least Squares Regression* (PLSR)[27] as the multivariate regression technique. External and full-cross validations were considered for model building and evaluation. Different statistical parameters were employed for evaluating the predictive ability of models during the fitting and test stages, namely: the Pearson correlation coefficient (r^2), the determination coefficient (q^2), the sample standard error, and the slope and intercept of the fitted/predicted *vs.* observed values.[28]

4.2.3. 2D-QSPR approach. TMACC descriptors

We selected the 2D-QSPR approach developed by Hirst and co-workers, the *Topological MAximum Cross Correction*[16] (TMACC) methods. TMACC descriptors are based on concepts derived from the alignment-independent methodologies (see *Chapter 1*). These descriptors use the topological bond distances and physicochemical properties of the molecule, and only the maximum value calculated as the product of pair combinations of physicochemical properties for each distance is used to generate the TMACC descriptors.

We used 4 different types of atomic properties; *Gasteiger partial charge*, to represent electrostatics;[29] *Crippen-Wildman molar refractivity* (MR), to represent steric effect;[30] *Crippen-Wildman partition coefficients* (*logP*) to represent lipophilicity;[29] and *logS* parameters introduced by Xu and co-workers to represent solubility and solvation phenomena.[31] Property types which produce positive and negative values were considered as two separate properties.

This was the case for all property types except molar refractivity, for which all atomic values property were positive. To account for the different scales used by each atomic parameter, each contribution was rescaled by the largest absolute value, resulting in all values being confined within the range of +1 and -1.

Table 4.1. The activity bands used for TMACC descriptor interpretation.

activity contribution	colour
very positive	blue
positive	yellow
neutral	green
negative	orange
very negative	red
no contribution	black

TMACC descriptors were generated using the topological data of each molecule. All nonpolar hydrogen atoms were removed and their atomic value added to the heavy atom to which they were bonded. Polar hydrogen atoms were considered explicitly. All calculations were performed using Nottingham Cheminformatics Workbench (NCW), a package which provides the function of generating the TMACC descriptors and the TMACC interpretation.[32] The popular machine-learning workbench Weka was included to provide PLS modeling, and was validated using 10-fold cross-validation. The results of the PLS analysis were used to determine the atomic contribution of each atom. The interpretation was visualized by a colour scheme depicting activity contribution by atom. Table 4.1 shows the colour-coding of each atom according to its activity band.

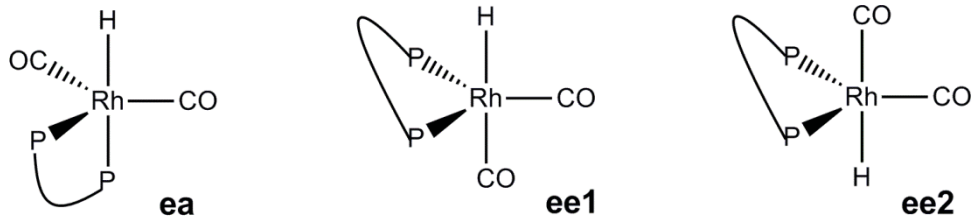
4.3. Results and discussion

4.3.1. Analysis of ligand dataset and molecular structures

All previous attempts to rationalize asymmetric hydroformylation were based on the transition state for alkene insertion into the Rh-H bond.[2-3,33] It was showed that alkene insertion is also the key transition state that determines the catalytic activity for usual kinetics of Rh-diphosphorus catalysts.[17b] This TS has a pseudofive coordination environment. To mimic the ligand arrangement in TS geometry, we optimized ground-state geometries of a pentacoordinated hydrido-bicarbonyl complex [$HRh(CO)_2$(P-P)], in which the diphosphorus ligand can coordinate in equatorial-apical (ea) and bis-equatorial (ee) modes. Assuming an apical preference of hydride ligand, there is one isomer for ea coordination (**ea**), and two for ee coordination when the ligand is C_1-symmetric (**ee1** and **ee2**) (Scheme 4.2). Additionally, for dissymmetric diphosphorus ligands (**5**, **20** and **21**), it is possible to distinguish two ea isomers depending on which ligand moiety is coordinated at the apical site. Coordination preferences

for ligand **5** have already been studied theoretically by us and showed that the preferred moiety in the apical site is the phosphite (*Chapter 3*).[3] Similarly, the ligands **20** and **21** have been also studied theoretically by our research group,[2] and showed that the aminophosphane moiety in the apical is preferred. Note, however, that for the alignment-independent methodologies the coordination mode of the ligand becomes less crucial because its shape is described by a two-dimensional graph, the *correlogram*.

Scheme 4.2.



For each geometrical isomer **ea**, **ee1** and **ee2**, suitable initial structures were generated, when possible, from analogous X-ray pentacoordinated structures. Then, for the three isomers, we applied the conformational search protocol described in the previous section to locate the species with the lowest energy. Except for the (*S,S*)-kelliphite ligand (**11**), our results indicate that all the analyzed ligands coordinate preferably in equatorial-apical manner. Ligand **11** showed the largest P-Rh-P bite angle (126° for ee coordination) along the ligand dataset, the rest of bite angles ranging from 86° for ligand **17** to 113° for **1**. The energy difference between ea and ee isomers for **17** was 14 kJ.mol^{-1}, decreasing to only 6 kJ.mol^{-1} for **1**. Interestingly, all catalysts exhibiting high enantioselectivity showed preferred ea coordination. This reinforces our previous suggestion that ea coordination brings a higher degree of chirality to the apical site thus favoring diasteroisomeric ligand-substrate interactions (*Chapter 3*).[3]

The optimized structures of the coordinated ligands were used to calculate the molecular fields. The inclusion of the whole complex with the metal and auxiliary ligands hindered the effect of the ligands, and gave poor statistical parameters. This effect was previously observed by Higgison, Morao and co-workers in analogous 3D-QSPR studies on asymmetric organometallic catalyisis.[14] Thus, the molecular interaction fields (MSF and MEP) were calculated on the isolated ligands at the geometry that they have in the pentacoordinated complex.

4.3.2. 3D-QSPR model for enantioselectivity. The Molecular Shape Field

Initially, we tried to build a QSPR model for predicting *ee* values in the 21-catalyst dataset (**set 1**) using the shape field with convexity node selection (MSF-MSF$_{CONV}$) as descriptor. After full cross-validation of the 21-catalyst dataset, the Pearson correlation coefficient (r^2) of the fitting stage is 0.83, whereas the predictive ability (q^2) calculated using the Leave-One-Out (LOO) cross-validation method is 0.51. Both values were obtained for the optimal number of

PLS factors (4 components). In drug design, a model is considered predictive when the q^2 is higher than 0.5 (halfway between perfect prediction 1.0 and no model at all 0.0). Thus, the value of 0.51 for q indicates that the model is in the limit of what one can consider as predictive.

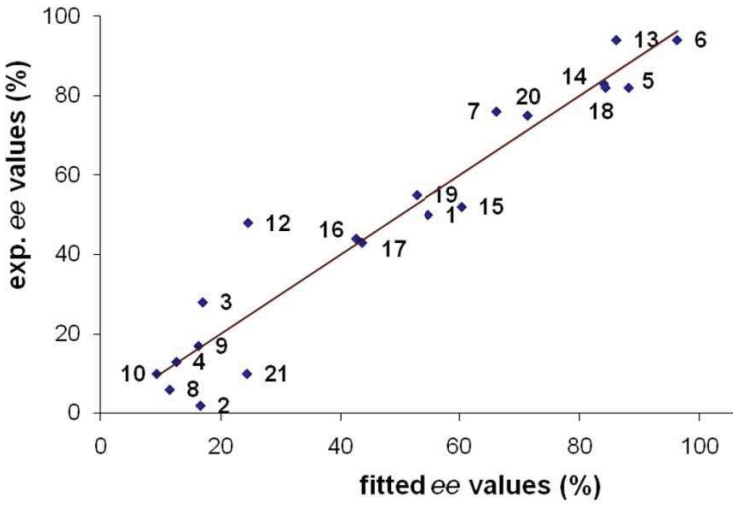


Figure 4.4. Experimental *vs* fitting %*ee* values for the dataset of 20 catalysts (**set 2**), using MSF-MSF$_{CONV}$ *correlograms* as descriptors for the 3D-QSPR modeling.

There may be outliers that influence the quality of the QSPR model. Indeed, ligand **11** has the largest residuals, the difference between experimental and predicted *ee* value for fitting and validation stages being 34% and 55%. As mentioned above, the most stable isomer was ea coordination for all the ligands but for **11**, which showed a preference for ee coordination. This means that ligand **11** was not only an outlier associated with the analysis of the response space, but also with the analysis of the chemical space. One of the assumptions of QSPR modeling is that the interaction between the substrate and the catalyst occurs in the same way. In previous studies, we showed that the interactions governing enantioselectivity occur mainly between the substrate and the apical ligand (*Chapter 3*).[3] Thus, for ee coordination, the substrate will meet an achiral CO group in apical, and yield a low enantioselectivity. Although the QSPR model reflects the potential enantioinduction of the ligand **11**, the preferred ee coordination does not allow transmitting the chiral information efficiently. When we set aside ligand **11**, the statistical parameters improved significantly, and this confirmed its nature as an outlier. The new QSPR model for the dataset of 20 ligands (**set 2**) had q^2 = 0.68 (in prediction) and r^2= 0.92 (in fitting) with 4 PLS factors. Figure 4.4 shows the experimental plotted against the fitted *ee* values. The coefficients of QSPR equation are provided in the Figure 4.5. Moreover, the standard error in these enantiomeric excess predictions is 17% (0.2 kcal.mol^{-1}), which is within the order of the discrepancy with experiments that is encountered for catalysts **20** (23%, 0.3 kcal.mol^{-1}) using QM/MM calculations of transition states.[2]

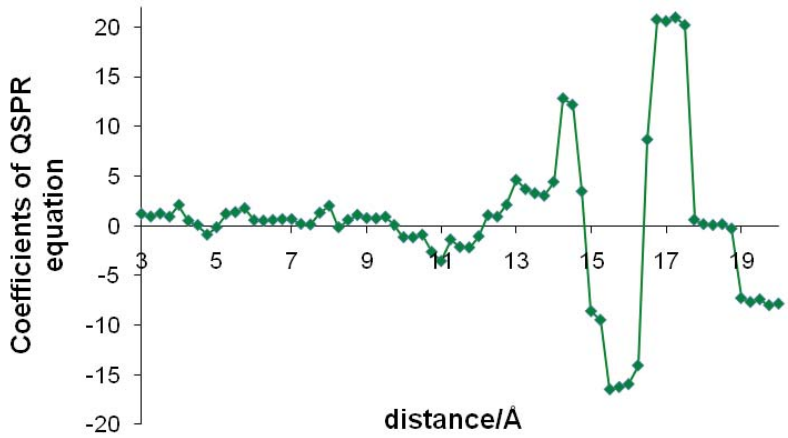


Figure 4.5. The coefficients of QSPR equation for the enantioselectivity model using MSFxMSF$_{CONV}$ descriptor.

Table 4.2. Statistical parameters for LOO cross-validation for enantioselectivity in **set 2** using different types of descriptors and filtering criteria.[(a)]

Filtering criteria	Descriptor	q^2	r^2
shape-based	MSF-MSF$_{CONV}$	0.68	0.92
	MEP-MEP$_{CONV}$	0.60	0.95
basicity-based	MSF-MSF$_{BAS}$	0.12	0.46
	MEP-MEP$_{BAS}$	0.41	0.85

[(a)] Slope = 0.97 and 0.95, and intercept = 4.2 and 2.4% for prediction and fitting, respectively; fitting error = 6.8% and predicting error = 19.6%.

We also tried the full set of available descriptors within *Q-QSSR* methodology. It is clear from Table 4.2 that the best results are obtained for shape-based descriptors MSF-MSF$_{CONV}$ and MEP-MEP$_{CONV}$ (q^2 > 0.6). The models generated using basicity-based descriptors MEP-MEP$_{BAS}$ and the MSF-MSF$_{BAS}$ cannot be accepted as predictive (q^2 < 0.5). These results are consistent with the idea that enantioselectivity is mainly determined by molecular shape recognition. It is also possible to combine the interaction fields in the representation of molecular structures by simply concatenating the *correlograms*. In the previous application of *Q-QSSR* approach, it was observed that MEP-MEP$_{CONV}$ descriptor contains a second level of steric information, which derives from node selection based on convexity.[15] However, combination of MSF-MSF$_{CONV}$ and MEP-MEP$_{CONV}$ did not improve the model in this case (q^2 = 0.68 and r^2 = 0.93); and therefore, we used the simpler MSF-MSF$_{CONV}$ model for further analysis. To check the robustness of this methodology further in correlating with enantioselectivity, we generated a new mathematical

model using ALMOND software,[26] in which GRIND methodology[13] is also implemented. The descriptors were built from the empirical *TIP* interaction fields, which evaluate the shape of the catalyst in a similar way as MSF does. The QSPR model showed a good predictive capability (q^2 = 0.59 and r^2 = 0.96 for the optimal number of PLS, 5 components).

We also carried out an external validation of the MSF-MSF$_{CONV}$ model by dividing the dataset into 4 series of training and test subsets of 15 and 5 catalysts, respectively. Then the 5 external catalysts were predicted with the model trained by the remaining catalysts. To ensure a balanced test subset on the %*ee* rank, we defined a procedure for selecting the subset randomly but covering a wide range of enantioselectivities: i) all catalysts were sorted in ascending %*ee* value order; ii) for each ith run, the first test object was the one ranked at position ith of the sorted data; and iii) then 5 iterations were carried out for selecting those ranked at ith + jth * 5, where jth starts at 1 and it is incremented by one in each iteration while ith + jth * 5 is lower than the size of the set. Table 4.3 shows the test subset composition for each run, the experimental and predicted %*ee* values in parentheses, and r^2 and q^2 for each test set. We observed good predictions for runs 1, 2, and 4, but not for run 3, in which the ligands **1** and **6** show the large discrepancies. This could very well be explained by the mediocre predicting ability (q^2 = 0.52) of the resulting training set,[34] which has only one ligand with an *ee* over 90%. Consequently, we expect a poor modeling of systems with high enantioselectivities such as **6** (*ee* = 94%). Ligand **1**, (2*R*,4*R*)-chiraphite, is predicted to have a high *ee* (90%) and, in fact, the ligand has showed *ee* values of up to 90% at low temperatures.[35] Nevertheless, in our dataset, the reaction was conducted at 80° C yielding low-medium enantioselectivity (*ee* = 50%).[18] This indicates that the ligand is highly selective, as predicted by the model, but very sensitive to experimental conditions. Interestedly, for **1** we computed the lowest preference for ea coordination among **set 2** (see above), which suggests that a nonselective ee path may be also operative at high temperatures reducing the overall enantioinduction. In summary, despite the observed discrepancies, the external validation achieved a 75% of success supporting further the predictive ability of the model.

Table 4.3. Experimental *vs.* predicted enantiomeric excess (*ee* in %) for different test subsets of 5 catalytic systems employing 3D-QSPR models generated from the corresponding training sets with MSF-MSF$_{CONV}$ descriptor.

run	**ligand** (exp,pred)	($\boldsymbol{r^2}$ / $\boldsymbol{q^2}$)[(a)]
1	**2**(2,23); **4**(13,14); **16**(44,43); **19**(55,51); **18**(82,82)	(0.96 / 0.89)
2	**8**(6,8); **9**(17,16); **12**(48,11); **2**0(75,54); **14**(83,81)	(0.88 / 0.62)
3	**10**(10,11**); 3**(28,7); **1**(50,91); **7**(76,90); **6**(94,43)	(0.58 / 0.05)
4	**21**(10,35); **17**(43,46); **15**(52,71); **5**(82,76); **13**(94,79)	(0.92 / 0.71)

[(a)] The r^2 and q^2 values correspond to the test set in prediction.

In an external blind test, we tried to carry out *a priori* prediction of the performance of a ligand from an independent laboratory. We tested whether it had been possible to anticipate

the excellent enantioselectivities (99% *ee*) of the recently tested phosphane-phosphoramidite ligand (yanphos).[36] To obtain the molecular structure, we assumed an ea coordination with the phosphoramidite moiety at the apical position, as proposed by Zhang and co-workers,[33c] and we optimized it at the QM/MM level. Then, we computed the MSF-MSF$_{CONV}$ descriptor using the same parameters as those used for building the QSPR model. In excellent agreement with experiment, we calculated an *ee* of 90%, and what is more important, the ligand was correctly classified as highly enantioselective.

4.3.3. 3D-QSPR model for activity. The Molecular Electrostatic Potential

We generated a new dataset with 19 ligands (**set 3**), using the percentage of conversion as the dependent variable. We removed the ligands **20** and **21** because the different origin of the data does not allow comparing conversion values. Table 4.4 collects the main statistical parameters obtained for the different descriptors. All defined descriptors yield models with poor predictive capacity (q^2 around 0.5). This is an indication that a property such as the activity requires to be modeled by more sophisticate descriptors, including both shape and basicity. When we combined shape- and electrostatic-based 3D fields via concatenation of *correlograms*, acceptable models were obtained. The concatenation of MEP-MEP$_{CONV}$ and MEP-MEP$_{BAS}$ descriptors achieved q^2 values up to 0.74. Figure 4.6 shows the experimentally measured plotted against the fitted *conv* values. The coefficients of QSPR equation are provided in the Figure 4.7.

Table 4.4. Statistical parameters for LOO cross-validation for activity in **set 3** using different types of descriptors and filtering criteria.

Filtering criteria	Descriptor	q^2	r^2
shape-based	MSF-MSF$_{CONV}$	0.52	0.71
	MEP-MEP$_{CONV}$	0.56	0.88
basicity-based	MSF-MSF$_{BAS}$	0.44	0.63
	MEP-MEP$_{BAS}$	0.47	0.56
combination[(a)]	MEP-MEP$_{CONV}$ & MEP-MEP$_{BAS}$	0.74	0.99

[(a)] Slope = 0.75 and 0.99, intercept = 0.49 and 15.8% for prediction and fitting, respectively, fitting error = 3.8% and predicting error = 16.6%.

Again, we carried out 4 runs to evaluate the external predictive ability of the model. Following the procedure described above, we divided the dataset into a training subset of 14 ligands and a test subset of 5, with the exception of run 4 consisting of 4 test ligands. Table 4.5 shows the test subset composition for each run, the experimental and predicted percentage of conversion, and the r^2 and q^2 for each test subset in prediction. We observed excellent predictions for three out of four external validations (run 1, 2 and 4). In run 3, the predicted activity for ligands **13** and **5**, both of which showed high conversion values (> 70%), is rather poor. Looking at Figure 4.6, we can observe that the available data are not uniformly

distributed, and most of the systems show low conversion. This could explain the discrepancies of high-selective catalysts such as **13** and **5**, especially when two of them are set aside of the training set such as in run 3.

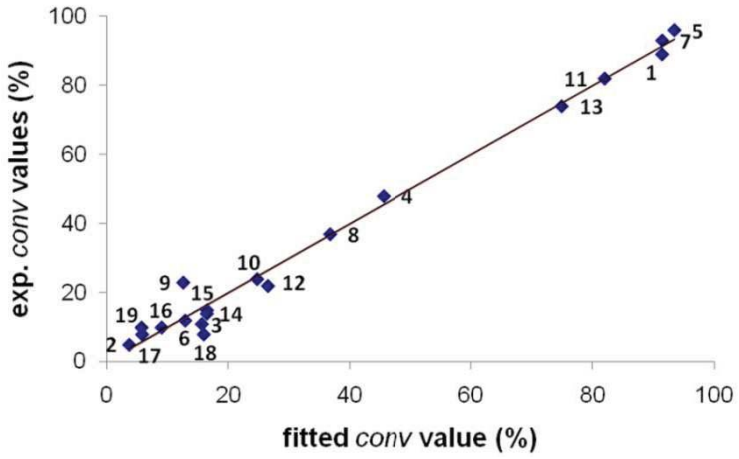


Figure 4.6. Experimental *vs* fitting *%conv* values for the dataset of 19 catalysts (**set 3**), using concatenated MEP-MEP$_{CONV}$ and MEP-MEP$_{CONV}$ *correlograms* as descriptors.

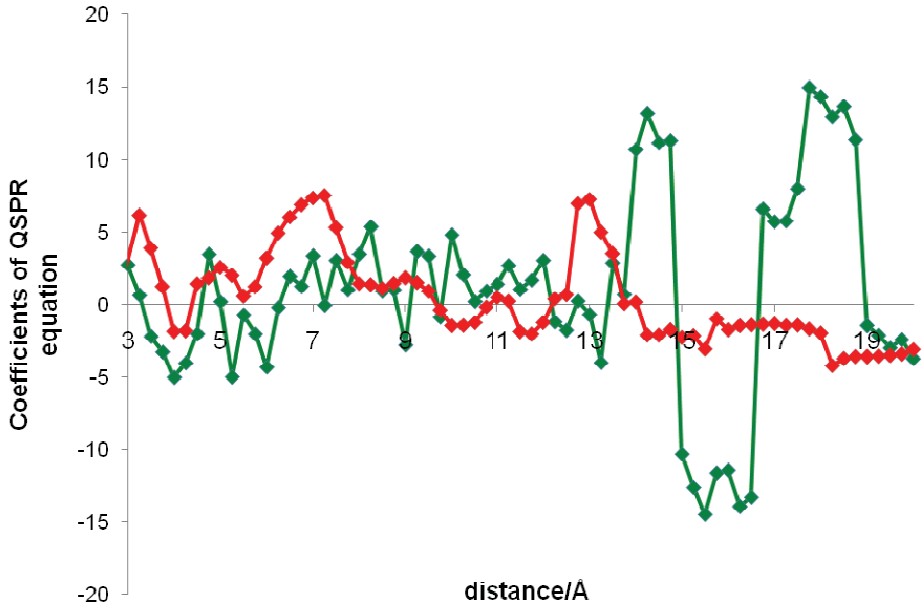


Figure 4.7. The coefficients of QSPR equation for the activity model. The green and red lines refer to MEP-MEP$_{CONV}$ and MEP-MEP$_{BAS}$ variable-types, respectively.

Table 4.5. Experimental *vs.* predicted activity (*conversion* in %) for different test sets of 5 catalytic systems employing 3D-QSPR models generated from the corresponding training subsets with concatenated MEP-MEP$_{CONV}$ and MEP-MEP$_{BAS}$ descriptors.

run	**ligand** (exp,pred)	(r^2 / q^2)[(a)]
1	**2**(5,6); **19**(10,5); **14**(15,14); **8**(37,34); **1**(89,83)	(1.00 / 0.98)
2	**3**(8,37); **18**(11,18); **12**(22,27); **4**(48,43); **7**(93,123)	(0.93 / 0.62)
3	**17**(8,0); **6**(12,30); **9**(23,9); **13**(74,29); **5**(96,61)	(0.82 / 0.41)
4[(b)]	**16**(10,12); **15**(14,23); **10**(24,23); **11**(82,76)	(0.99 / 0.96)

[(a)] The r^2 and q^2 values correspond to the test set in prediction.
[(b)] 4 test ligands as a consequence of the employed algorithm for set splits (see text).

4.3.4. Analysis of the enantioselectivity 3D-QSPR model and correlation with the sterochemical models

The main goal of QSPR approaches is to build mathematical models with predictive ability; however, obtaining additional chemical information to orient ligand design would be very valuable. To evaluate the exploratory ability of the QSPR models for understanding the mode of action of enantioinduction, we graphed the values of the product of the two-node molecular interaction field (*MSF-MSF*) by the equation coefficients (*Ci*) at each distance range (Figure 4.8).[37] This *coefficient-weighted correlogram* captures the 3D shape of catalyst structure and its impact in enantioselectivity, and transform it into a quantitative 2D graph, as appearance of an experimental spectrum. The peaks indicate the most positive and the most negative contributions to enantioselectivity, and indentify those independent variables (two-node interactions) that are directly correlated with the dependent variable (%*ee*). From the largest coefficient-weighted nodal interactions, it is possible to retrieve 3D information and to identify the regions of the ligands that contribute the most to enantioselectivity (Figure 4.9).

Figure 4.8 shows the *coefficient-weighted correlogram* of the enantioselectivity MSF-MSF$_{CONV}$ model for representative examples exhibiting either high enantioselectivities or with known sterochemical models. In the graph, we could identify different patterns depending on the number of peaks and the *correlogram* extent, which is related to the size of the ligand. For ligands **5** and **6**, there were three important peaks along the profile: two positives at 14.5 and 17.0 Å (**a** and **c**), and one negative at 16.0 Å (**b**). In **5**, backtracking of the variables associated to peaks **a** and **b** in 3D space highlighted very similar structural features. The two associated vectors were almost superimposed and connected the same ligand fragments (see Figure 4.9i). This suggests that the contributions of peaks **a** and **b** are mutually cancelled. Consequently, the contribution of peak **c**, with the largest positive coefficient-weighted nodal interaction, is the main responsible of the high computed enantioselectivity. In the 3D space, the **c** peak highlighted two patches: one at a phenyl substituent of phosphane moiety and the other at a naphthyl group of the phosphite moiety (highlighted in Figure 4.9i). Since the phenyl group does not belong to any stereogenic centre, it is reasonable to assign the key positive features

to the steric hindrance of the naphthyl group. In our previous study on AHF by Rh-binaphos (**5**) system (*Chapter 3*),[3] we found that the repulsive steric interactions between the substrate and the naphthyl groups of the apical phosphite are responsible for enantioinduction. This corroborates the interpretation of QSPR model.

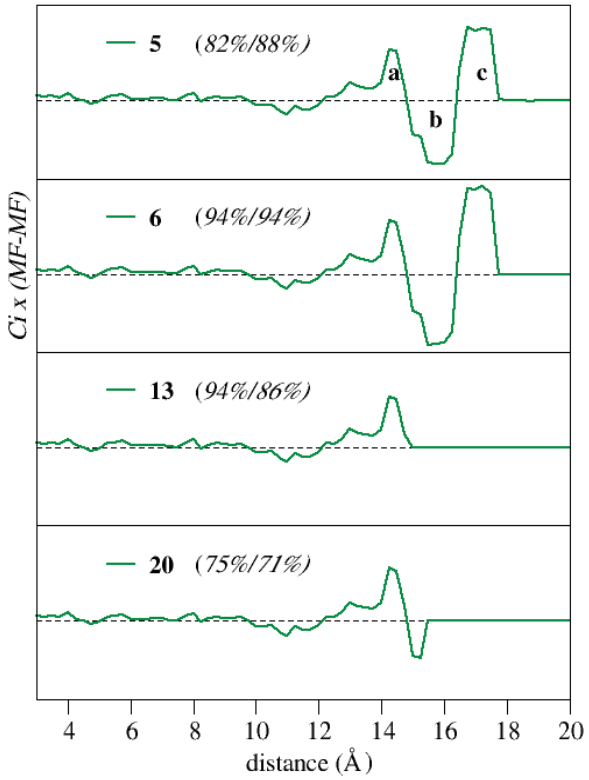


Figure 4.8. *Coefficient-weighted correlograms* of enantioselectivity model for ligands **5**, **6**, **13** and **20**. Product of the MSF-MSF$_{CONV}$ nodal interaction by the PLS coefficients of the enantioselectivity model at each distance range as a measure of the contribution of different catalyst regions to enantioinduction. In parenthesis values of the enantiomeric excess (experimental / LOO predicted).

The *coefficient-weighted correlogram* of ligand **6** (binaphine ligand) had a similar pattern to that of **5** (Figure 4.8). Focusing on peak **c**, its contribution arose from two naphthyl groups not involved in the backbone. One of the nodes of peak **c** is very close to other node highlighted for its negative contribution in peak **b** (Figure 4.9ii). Thus, the analysis of the QSPR model for **6** also identified a naphthyl group as the key fragment. Interestingly, in both ligands **5** and **6**, the node describing the protrusion of the key naphthyl group was at 9 Å from the metal centre. These facts suggest that binaphine (**6**) and binaphos (**5**) ligands possess some common 3D features responsible for the high enantioselectivity. Thus, we tested whether the

stereochemical model proposed for **5** (see *Chapter 3*)[3] is also valid for **6**. The quadrant-diagram representation of Figure 4.10 can be described as follows: i) for ea coordination of the ligand, the plane of the quadrant contains the Rh-H bond and is parallel to alkene plane; ii) the quadrant delimited by the two phosphorus is blocked by steric effects; and iii) one of two possible paths for substrate approach (**I** or **II**) is sterically hindered in order to avoid mutual cancellation of enantioselectivities. To quantify the steric bulk of each quadrant and its impact on the metal centre, we defined a geometrical descriptor called *distance-weighted volume* (V_W, *Chapter 3*).[3] The V_W values for (*S*)-binaphine (Figure 4.10) agreed with the stereochemical model. There are one sterically small quadrant (uncoloured), two sterically large (dark grey), and one intermediate (light grey). This diagram explains the observed preference for the product with *S* configuration;[18] path **I** is favored over path **II**, and in path **I**, the pro-*S* transition state is more favored than the pro-*R* for path **I**.

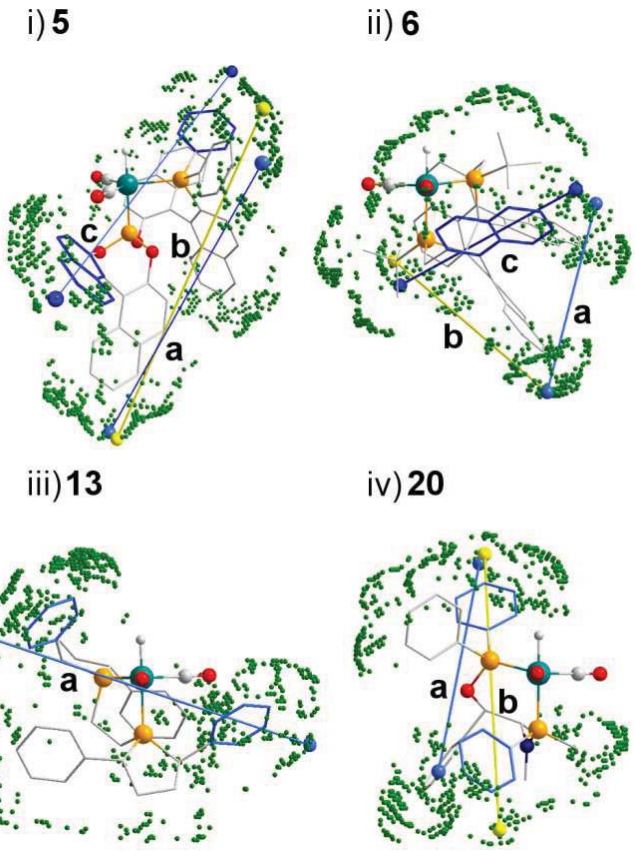


Figure 4.9. Backtracking of the most important MSF-MSF variables of the enantioselectivity model: i) catalyst **5**; ii) catalyst **6**; iii) catalyst **13**; and iv) catalyst **20**. Positive node interactions in dark blue, and negative in light yellow. The ligand substituents involved in key node interactions are highlighted in dark blue.

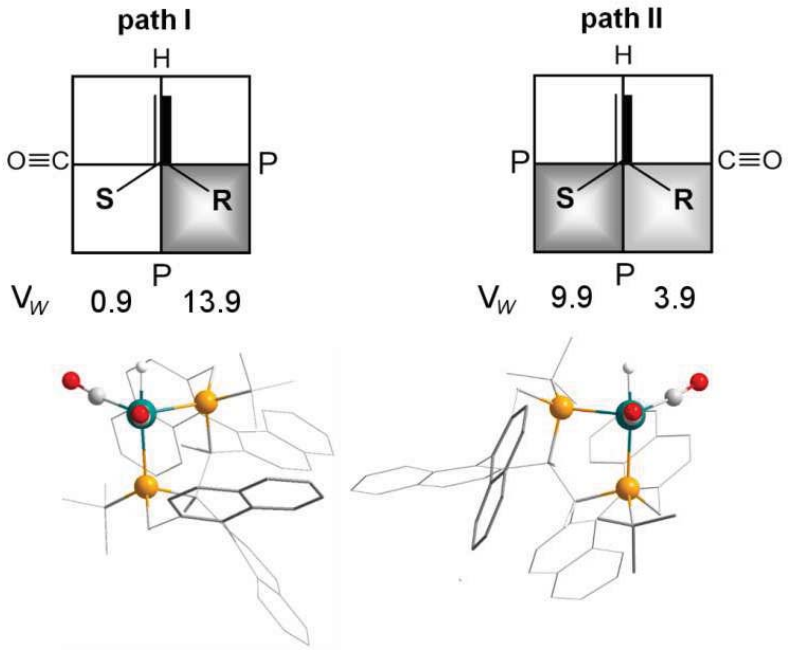


Figure 4.10. Quadrant-diagram representation of the possible hydroformylation paths for binaphine ligand, values of V_W in Å^2, and molecular structure of [HRh(CO)$_2${(*S*)-binaphine}] complex oriented following the chiral hypothesis of the stereochemical model.

The analysis of the external yanphos system showed the same features as binaphos (see Figures 4.11 and 4.12): i) the vectors associated with the positive **a** and the negative **b** peaks connect the same groups, cancelling each other; ii) the highest positive peak **c** highlighted naphthyl groups in the apical phosphoramidite moiety; and iii) the key node associated with a naphthyl fragment was located at ~9 Å of the metal centre. Other selective ligands such as (*2R*,*4R*)-chiraphite (**1**) and diazaphospholane (**7**) showed a similar pattern (Figure 4.12). For **1** the **c** peak highlighted the ^{t}Bu groups, while for **7**, it highlighted the chiral amine substituents. To sum up, the common 3D features of these ligands indicate that their enantioselectivity can be explained with the same stereochemical model as that proposed for Rh-binaphos (*Chapter 3*).[3]

Ligands **20** and **13** (P-sterogenic AMPP and (*R*,*R*)-Ph-BPE) showed only one peak with positive contribution at 14.5 Å (peak **a**, Figure 4.8). Backtracking the variables associated to peak **a** in 3D structure of **20** highlights two phenyls, one of the phosphinite moiety and the other one of the chiral aminophosphane (Figure 4.9iv). The latter selectivity-increasing group had been identified as the key group in inducing enantioselectivity by previous QM/MM calculations on transition states.[2] For ligand **13**, the model also points out two Ph groups at distinct phosphorus moieties as the key selectivity-determining molecular fragments. Analogously to what was found for ligands **5** and **6**, for ligands **13** and **20** the key nodes associated to the phenyl groups lay at the same distance from the metal centre (~7 Å). It seems that the model has recognized an optimal placement of different functional groups for the interaction with the substrate, and that the effective placement is achieved by coordinating a phophane moiety in the apical site, which is in agreement with previous hypotheses.[2-3,17,33,38] Thus, structural modifications around those specific areas would have a pronounced effect on enantioselectivity by tuning the interactions with the substrate.

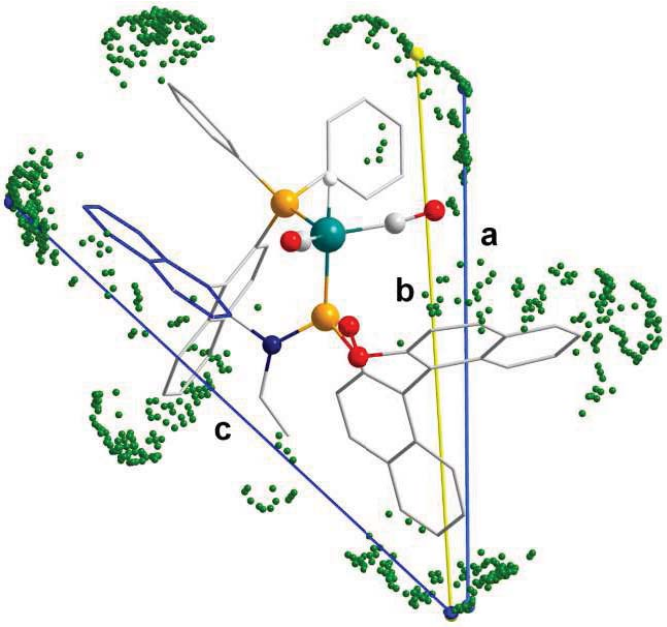


Figure 4.11. Backtracking of the most important MSF-MSF$_{CONV}$ variables of the enantioselectivity model of Rh-yanphos system. Positive node interactions in dark blue, and negative in yellow. The ligand substituent involved in key node interaction is highlighted in dark blue.

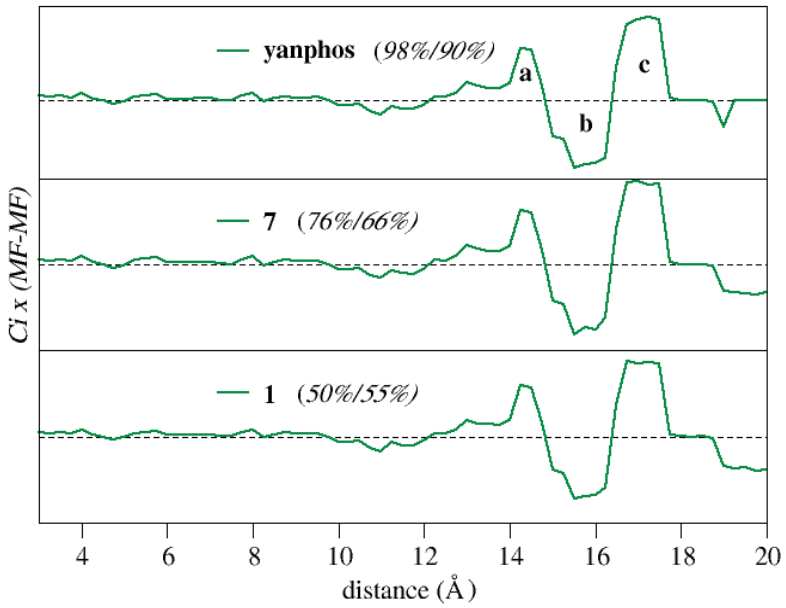


Figure 4.12. *Coefficient-weighted correlograms* of enantioselectivity model for ligands yanphos, **7** and **1**. Product of the MSF-MSF$_{CONV}$ nodal interaction by the PLS coefficients at each distance range. In parentheses values of the enantiomeric excess in % (experimental / LOO predicted).

4.3.5. Analysis of the activity 3D-QSPR model and correlation with ligand basicity

As stated in the introduction, the correlation of catalytic activity with the ligand basicity has been observed for specific families of ligands.[39] Thus, additionally, we tried to establish a correlation between the activity and the ligand basicity for the **set 3**. We used the HOMO energy of free ligands, formally corresponding to phosphorus lone pairs, as a measure of their donation ability; the lower in energy the diphosphane is the less basic it is. We were unable to find a correlation for the whole set. Nevertheless, if we only consider the diphospholane ligand family (**14**-**19**), or even the full family of phosphacycle moieties (**1**, **5**-**7**, **11**, **14**-**19**), fair correlations are observed (Figure 4.13). Ligands showing high activities (*conv* > 70%) have low-lying HOMOs(< -5.0 eV), whereas those showing low activities (*conv* < 20%) have high-lying HOMOs (> -5.0 eV). The use of HOMO energy as descriptor can be proved via correlation with the *Tolman Electronic Parameters* (χ)[40] of the phosphacyclic substituents. The HOMO energies for **13**, with the most electron-withdrawing Ph substituent (χ = 4.3), lays at -5.0 eV, whereas ligands **14**-**19** with electron-donating Me, Et and ^{i}Pr groups (χ = 2.6, 1.8 and 1.0), the energies lay at ~-4.7 eV.

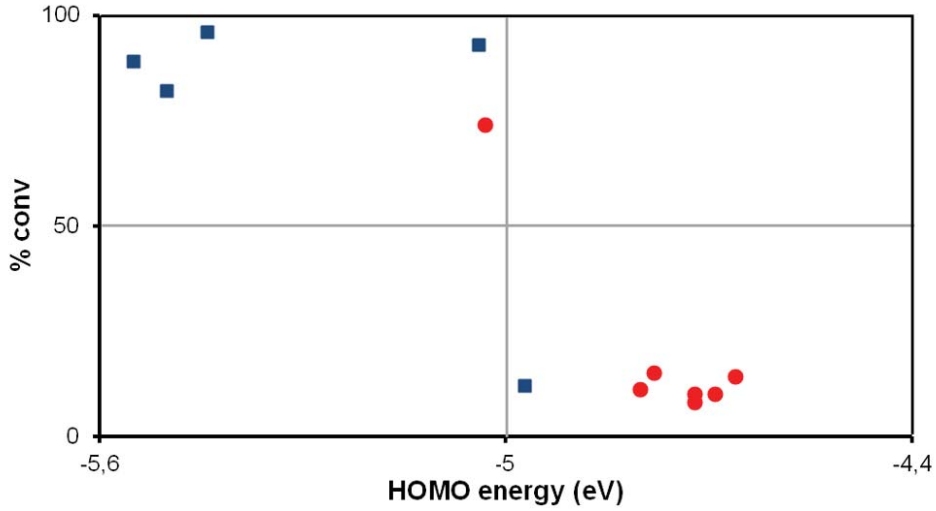


Figure 4.13. Correlation between the activity (percentage of conversion) and the HOMO energy (eV) of free diphosphanes. Circles correspond to diphospholane ligands and squares correspond to rest of ligands with phosphacycle moieties.

Although some correlation between activity and ligand basicity exists for specific ligand families, the 3D-QSPR modeling required shape-based descriptors in order to have predictive capacity for the whole **set 3** (Table 4.4). Figure 4.14 shows the coefficient-weighted *correlogram* for the activity model (MEP-MEP$_{CONV}$ & MEP-MEP$_{BAS}$) for selected examples. On the left-hand side of the graph, we plotted the shape-based MEP-MEP$_{CONV}$ descriptor, which resembles that found for enantioselectivity model (Figure 4.8). Depending on the ligand bulkiness, we observed three peaks at 14.5, 16.0 and 17.0 Å (**a**, **b** and **c**), or one peak at 14.5 Å. On the right-hand side of the graph, the basicity-based MEP-MEP$_{BAS}$ descriptor shows several positive peaks, the largest being at 7.0 Å (**d**), which vanishes for catalysts with low activities. Regarding the basicity-based descriptors, Figure 4.15 shows the superimposed *coefficient-weighted correlograms* of selected ligands, amplified around the region of peak **d**. The height of peak d (**5** > **13** ≈ **6** > **18**) can be correlated with the vales of HOMO energy; the lowest for **5** (-5.4 eV), somewhat higher for **6** and **13** (-4.97 and -5.03 eV), and the highest for **18** (-4.7 eV). To explain the observed differences between ligands **6** and **13**, we should look at the shape-based MEP-MEP$_{CONV}$ descriptors. When going from **6** to **13**, we observed that the negative contribution of peak **b** disappeared (Figure 4.14). Thus, although binaphine (**6**) and (*R*,*R*)-Ph-BPE (**13**) ligand have similar basicity, the shape of the (*R*,*R*)-Ph-BPE has a positive effect on activity. Considering the mechanistic knowledge,[17,40a-b] the positive effect could be very well related to a reduction of the steric hindrance on the reaction centre favouring the alkene coordination and hydride migration steps.

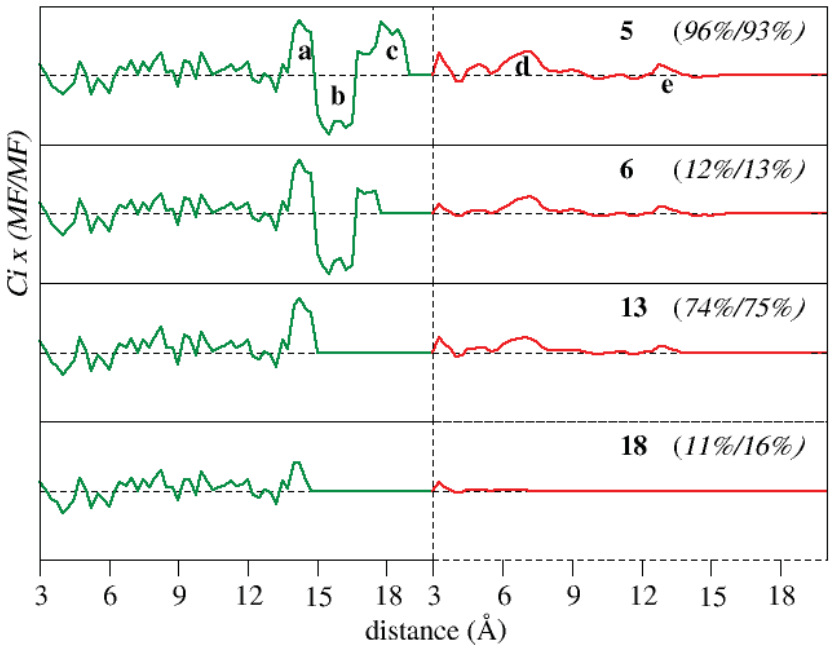


Figure 4.14. *Coefficient-weighted correlograms* of activity model for ligands **5**, **6**, **13** and **18**. Product of the MEP-MEP$_{CONV}$ (left) and MEP-MEP$_{BAS}$ (right) nodal interaction by the PLS coefficients at each distance range. In parentheses values of the conversion percentage (experimental / LOO predicted).

Figure 4.16 represents the backtracking of the variables associated with peak **d** for catalysts **5** and **13**. In both cases, one of the nodes is located at the phosphorus lone pair, and the other node highlights an electron-withdrawing substituent; the oxygen atoms in binaphos and the phenyl group in the (*R*,*R*)-Ph-BPE ligand. For the highly active catalysts chiraphite (**1**), diazaphospholane (**7**) and kelliphite (**11**), the vectors associated with peak **d** also connect a node located at the phosphorus lone pair and another node located near the heteroatom (O and N) substituents. This analysis further supports the correlation between high activity and low basicity; and highlights the importance of introducing electron-withdrawing groups to achieve more active catalysts.

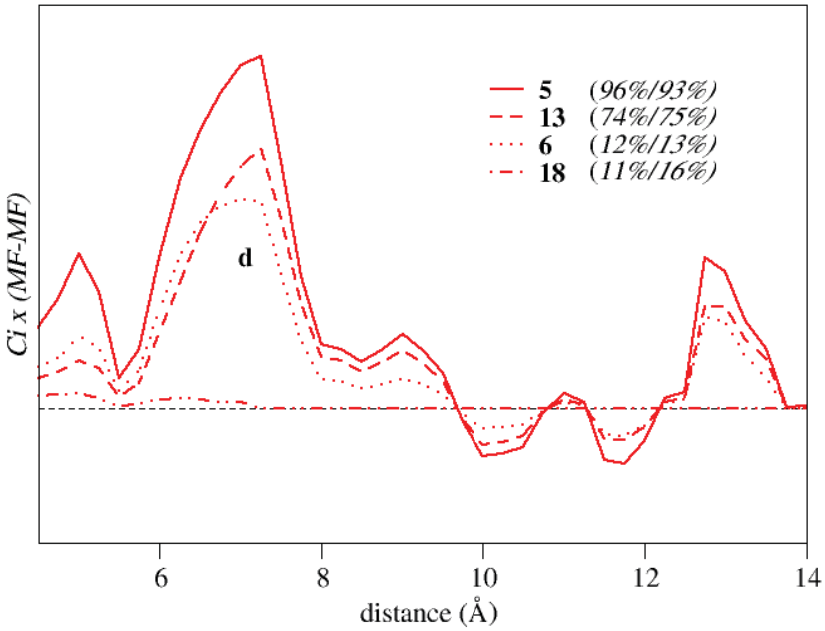


Figure 4.15. *Coefficient-weighted correlograms* of activity model amplified around peak **d** for ligands **5**, **6**, **13** and **18**. Product MEP-MEP$_{BAS}$ interaction by the PLS coefficients. In parentheses values of the conversion percentage (experimental / LOO predicted).

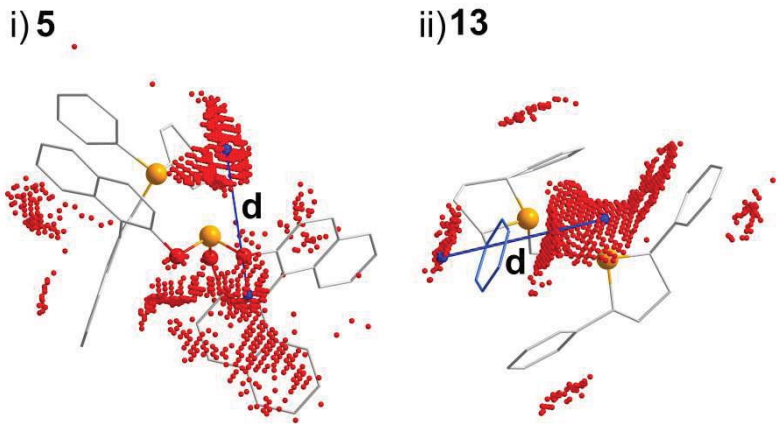


Figure 4.16. Backtracking of the most important MEP-MEP$_{BAS}$ variables of the activity model: i) ligand **5**; and ii) ligand **13**.

4.3.6. 2D-QSPR model for activity. The TMACC descriptors

During a 3-months stay in the theoretical group of Prof. Hirst (School of Chemistry, University of Nottingham, England), we tried to build a 2D-QSPR model for predicting the activity in the 19-dataset (**set 3**) using the atomic properties of the ligands and TMACC descriptors. For individual properties, the values of q^2 are similar and within the limits of prediction ability ($0.55 < q^2 < 0.65$, see Table 4.6). Nevertheless, the best model is obtained by combining of all properties, being the q^2 value around 0.71. The results agree with those obtained with 3D-QSPR model, suggesting that the activity required to be modelled by more sophisticate descriptors.

Table 4.6. Statistical parameters for 10-fold cross-validation for activity in **set 3** using different types of atomic properties.

Property	q^2	r^2
electrostatics	0.59	0.64
steric effect	0.64	0.69
lipophilicity	0.58	0.78
solubility	0.59	0.89
all properties	0.71	0.76

As mentioned in the *computational details and methods* section, the TMACC approach provides a method for interpreting the results, in which the activity contribution to each atom is represented by a colour (see Table 4.1). Figure 4.17 shows this representation for binaphos (**5**), diazaphospholane (**7**), kelliphite (**11**) and chiraphite (**1**). In agreement with results of 3D-QSPR models, the very positive activity contributions (dark blue) were located at heteroatoms such as oxygens and nitrogens. That is the oxygens of phosphite moieties in binaphos, kelliphite and chiraphite ligands; and oxygen and nitrogen atoms of the ligand skeleton in diazaphospholane system.

Unfortunately, this methodology does not allow the separation of the contribution for each property type. Therefore, in order to try to analyze further the origin of the activity, we compare the QSPR models obtained for each individual property. As an example, we analyzed the interpretation of QSPR models of the individual properties for the ligand **5** (Figure 4.18). The heteroatoms had very positive activity contributions only in the electrostatic model (Figure 4.18i), suggesting that the electrostatic is the property that contributes the most and that the other ones only complement the QSPR model. The same pattern was found for all ligands that contain heteroatoms (ligands **1**, **11** and **7**).

Figure 4.19 shows the visualization of QSPR models obtained for all-properties, single-electrostatic and single-steric in the case of low-enatioselective binapine (**6**) system. In this case, the naphthyl atoms closer to the reaction centre are the ones with the most negative

contribution to the activity for the all-properties model (Figure 4.19i). However, this is observed neither in the only-electrostatic model (Figure 4.19ii) nor in the only-steric model (Figure 4.19iii).This is a like that proposed in previous interpretation of 3D-QSPR models, in which the low catalytic activity of **6** was attributed to its shape.

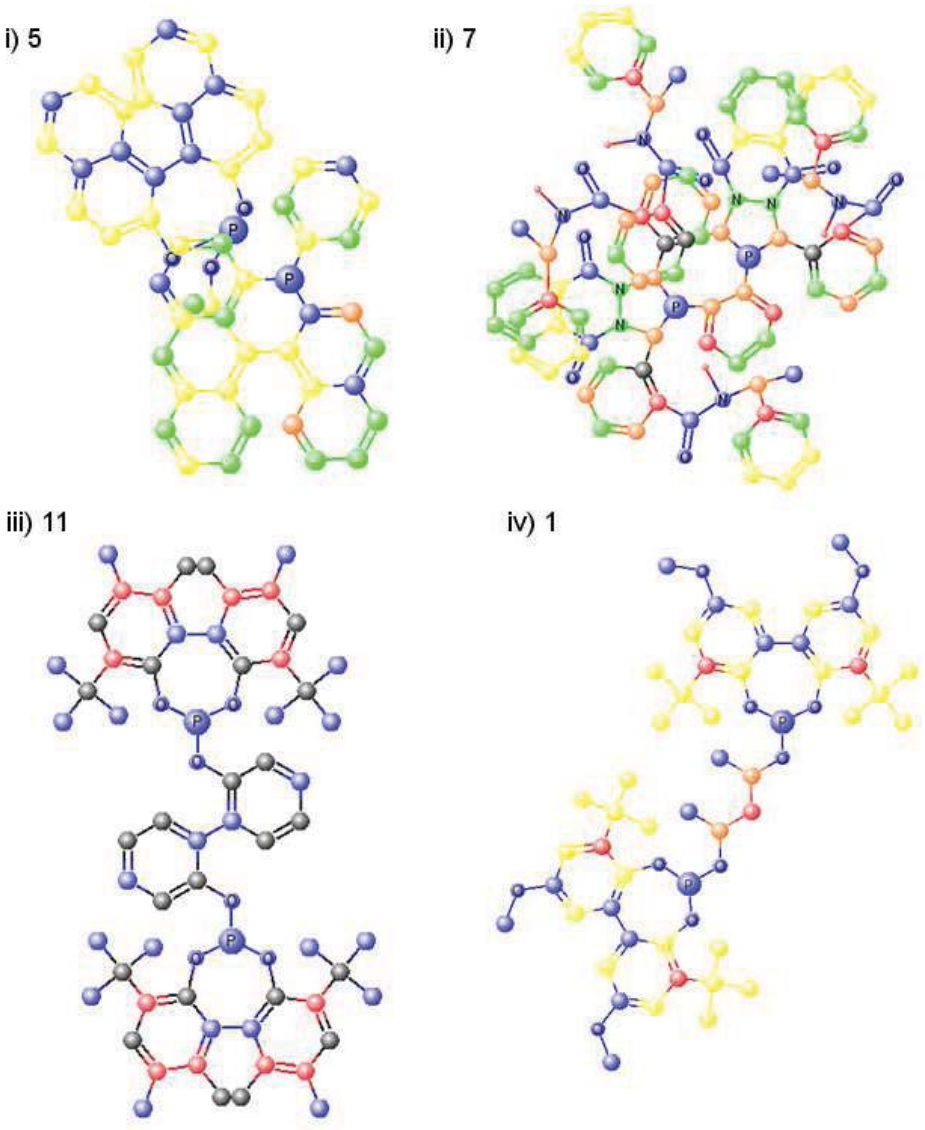


Figure 4.17. Representation of the TMACC interpretation of all-properties model: i) binaphos (**5**); ii) diazaphospholane (**7**); iii) kelliphite (**11**); and iv) chiraphite (**1**). TMACC activity colour scheme: red for very negative contribution to activity, orange for negative, green for neutral, yellow for positive and blue for very positive.

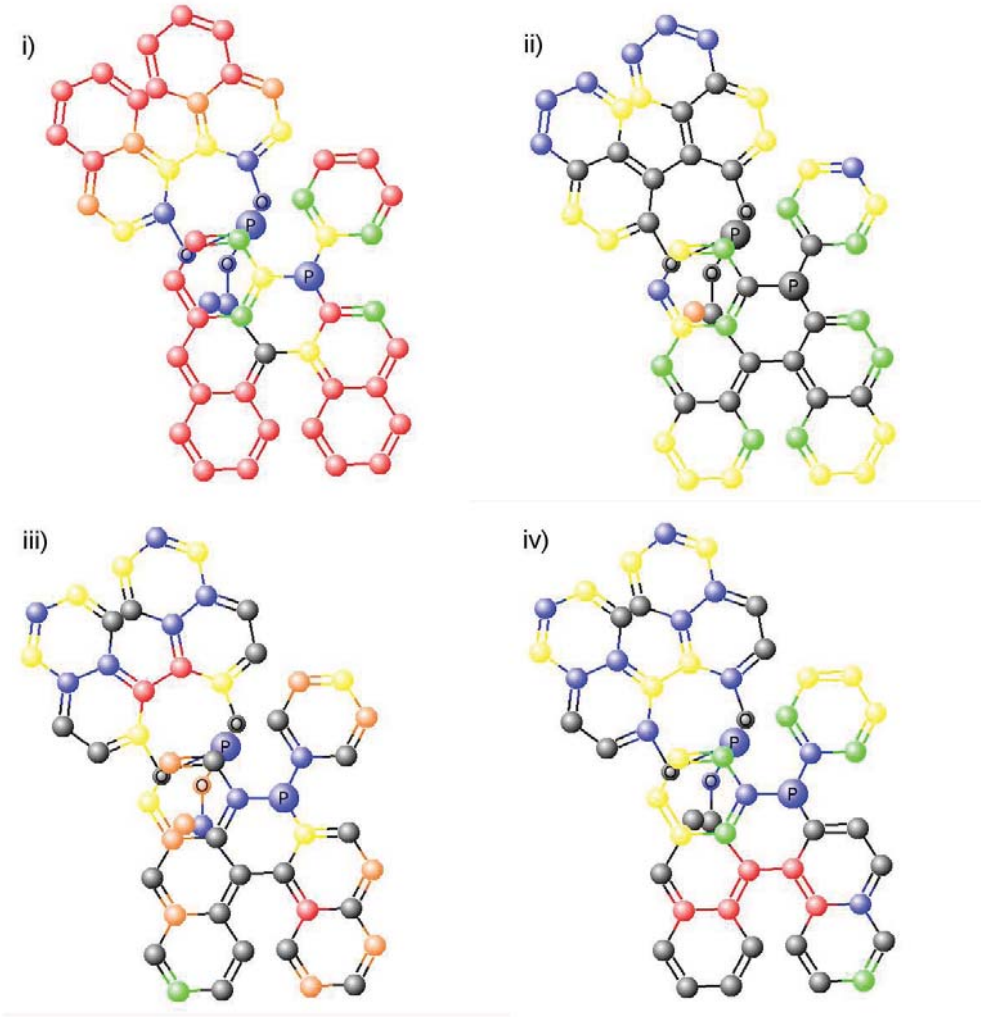


Figure 4.18. Representation of the TMACC interpretation of single-property models for binaphos (**5**): i) electrostatic; ii) steric effects; iii) lipophilicity; and iv) solubility. TMACC activity colour scheme: red for very negative contribution to activity, orange for negative, green for neutral, yellow for positive and blue for very positive.

We also tried to build TMACC models for enantioselectivity in the 21-dataset (**set 1**) and 20-dataset (**set 2**). However, and despite of the many efforts made, we were not able to find a 2D-QSSR model with predictive ability for enantioselectivity. In principle, this is consistent with our initial hypothesis that assumes that accounting for the shape of the catalyst is mandatory to model enantioinduction. Nonetheless, as mentioned in the *introduction* section, it is possible to use geometrical descriptors based on molecular topology that indirectly account for 3-dimensionality descriptors. In principle, the *Crippen-Wildman molar refractivity* (MR) describes the steric properties of the atoms and the structure; but they were proposed and

fitted for biological systems and for correlating activity. Thus probably, the descriptors are not transferable to organometallic complexes or to model the interactions that determinate the catalytic enantioselectivity. As commented in the *Chapter 1*, the 2D descriptors are quicker to be calculated, and in principle easier to be interpreted than 3D. However, 3D descriptors offer the most realistic way of representing a chemical system; and in the case of asymmetric organometallic catalysis, this seems to be crucial for obtaining predictive models.

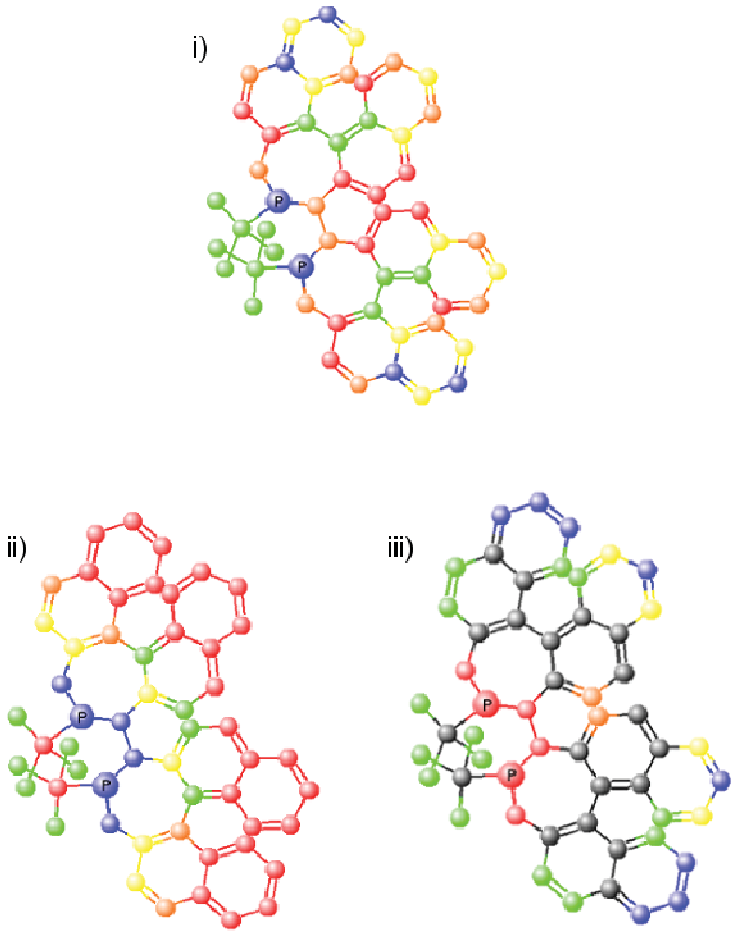


Figure 4.19. Representation of the TMACC interpretation for binaphine (**6**): i) all-properties; ii) single-electrostatic; and iii) single-steric property. TMACC activity colour scheme: red for very negative contribution to activity, orange for negative, green for neutral, yellow for positive and blue for very positive.

4.4. Conclusions

Two 3D-QSPR models were generated to predict enantioselectivity and activity for asymmetric hydroformylation of styrene catalyzed by rhodium-diphosphane catalysts, using a varied dataset of 21 ligands. The best predictive model for enantioselectivity was obtained using shape 3D fields (MSF) based on the local curvature of the electron density isosurface (r^2 = 0.92, q^2 = 0.68). Interestingly, the only ligand showing preference for bis-equatorial coordination (kelliphite, **11**) was identified as an outlier, and the worst predicted ligand in external validation was that with the lowest preference for equatorial-axial coordination, (2*R*,4*R*)-chiraphite (**1**). Although, the model predicts that both ligands are highly selective, low values of *ee* were observed. This suggests that the intrinsic enantiotopic differentiation capacity of kelliphite and chiraphite can be lost through the occurrence of non-selective bis-equatorial paths; and consequently, the equatorial-apical coordination is a prerequisite in Rh-diphosphane AHF. The most predictive model for activity was obtained using a combination of shape- and electrostatic-based 3D fields (r^2 = 0.99, q^2 = 0.74) for a dataset of 19 catalysts. The use of chemically meaningful descriptors provides insight into the factors governing catalytic activity and enantioselectivity. The models for enantioselectivity require shape-based descriptors to account for the steric effects induced by the ligand. In the case of the activity, it was possible to find a correlation with the ligand basicity, measured from the HOMO energies of free ligands, for structurally-related ligands. However, when a larger and more varied ligand set was considered, we required descriptors that accounted for both basicity and steric properties.

The QSPR analysis provides structural insight consistent with, and complementary to, the previously proposed stereochemical models for asymmetric hydroformylation. Thus, the stereochemical model previously proposed for Rh-binaphos catalyst can be also applied to explain the high enantioinduction observed for Rh-chiraphite, -binaphine, -diazaphospholane, and -yanphos systems. For binaphine and yanphos ligands, we assigned the main contribution to the selectivity to the naphthyl groups, identifying regions where steric hindrance would favor selectivity. For chiraphite and diazaphospholane ligands, the analysis of the QSPR equation highlighted the ^{t}Bu and chiral amine groups as the key enantioinducting moieties. Moreover, the effective placement of these groups to interact with the substrate is achieved through the coordination of phosphane moiety in the apical site of the complex.

The ligands that most effectively promote catalytic activity are those that carry electron-withdrawing groups in the vicinity of the phosphorus atoms via reduction of ligand basicity, as in chiraphite, binaphos, diazaphospholane and kelliphite ligands. Nevertheless, we found that more complex relationships account for the origin of activity. In the design of new active catalysts, the shape of the catalysts and the steric effects exerted on the reaction centre are also important as in the (*R*,*R*)-Ph-BPE ligand. Based on previous mechanistic knowledge, it is reasonable to propose that reduction of the steric hindrance on the reaction centre will favor the alkene coordination and hydride migration steps. Finally, this computational methodology can be used both to gain greater understanding of previously studied examples, to make

predictions of the enantioselectivity and activity of new cases, and to derive guidelines for new ligand design.

Additionally, we have tested the 2D-QSPR-type TMACC approach to model the activity and selectivity. The modeling used several types of atomic properties; electrostatic, steric, lipophilicity-related and solubility-related. For enantioselectivity, no models with predictive ability were obtained. This indicates that either 3D descriptors describing the shape of the catalyst or specifically designed 2D descriptors are crucial for predicting the enantioselectivity homogenous catalysis. For activity, a 2D-QSPR model with good statistical parameters was found when all atomic properties were considered that corroborates the hypothesis that several properties are necessary to predict the activity. Nevertheless, the precise interpretation of these 2D-QSPR models is complicated because the information obtained cannot be easily extrapolated to mechanistic interpretation.

4.5. References and notes

1 C. Claver, P. W. N. M. van Leeuwen, *Rhodium Catalyzed Hydroformylation*, Klumer Academic Publishers, Dordrecht, **2000**.

2 J. J. Carbó, A. Lledós, D. Vogt, C. Bo, *Chem. Eur. J.* **2006**, *12*, 1457.

3 S. Aguado-Ullate, S. Saureu, L. Guasch, J. J. Carbó, *Chem. Eur. J.* **2012**, *18*, 995.

4 a) A. G. Maldonado, J. A. Hageman, S. Mastroianni, G. Rothenberg, *Adv. Synth. Catal.* **2009**, *351*, 387. b) N. Fey, J. N. Harvey, *Coord. Chem. Rev.* **2009**, *253*, 704. c) N. Fey, *Dalton Trans.* **2010**, *39*, 296.

5 a) V. L. Cruz, J. Ramos, A. Muñoz-Escalona, A. Lafuente, B. Peña, J. Martínez-Salazar, *Polymer* **2004**, *45*, 2061. b) V. L. Cruz, J. Ramos, S. Martínez, A. Muñoz-Escalona, J. Martínez-Salazar, *Organometallics*, **2005**, *24*, 5095. c) V. L. Cruz, J. Martínez, J. Martínez-Salazar, J. Ramos, M. L. Reyes, A. Toro-Labbe, S. Gutierrez-Oliva, *Polymer* **2007**, *48*, 7672. d) G. Occhipinti, H. R. Bjørsvik, V. R. Jensen, *J. Am. Chem. Soc.* **2006**, *128*, 6952. e) G. Rothenberg, E. Burello, *Adv. Synth. Catal.* **2003**, *345*, 1334. f) G. Rothenberg, E. Burello, *Adv. Synth. Catal.* **2004**, *346*, 1844. g) E. Burello, P. Marion, J. C. Galland, A. Chamard, G. Rothenberg, *Adv. Synth. Catal.* **2005**, *347*, 803. h) J. A. Hageman, J. A. Westerhuis, H. W. Frühauf, G. Rothenberg, *Adv. Synth. Catal.* **2006**, *348*, 361. i) Z. Strassberger, M. Mooijman, E. Ruijter, A. H. Alberts, A. G. Maldonado, R. V. A. Orru, G. Rothenberg, *Adv. Synth. Catal.* **2010**, *352*, 2201.

6 a) J. J. Miller, M. S. Sigman, *Angew. Chem. Int. Ed.* **2008**, *47*, 771. b) M. S. Sigman, J. J. Miller, *J. Org. Chem.* **2009**, *74*, 7633. c) K. C. Harper, M. S. Sigman, *Proc. Natl. Acad. Sci. U.S.A*. **2011**, *108*, 2179. d) K. C. Harper, M. S. Sigman, *Science* **2011**, *333*, 1875.

7 C. Jiang, Y. Li, Q. Tian, T. You, *J. Chem. Inf. Comp. Sci.* **2003**, *43*, 1876.

8 a) K. B. Lipkowitz, S. Schefzick, D. Avnir, *J. Am. Chem. Soc.* **2001**, *123*, 6710. b) S. Alvarez, S. Schefzick, K. B. Lipkowitz, D. Avnir, *Chem. Eur. J.* **2003**, *9*, 5832.

9 R. D. III. Cramer, D. E. Patterson, J. D. Jeffrey, *J. Am. Chem. Soc.* **1988**, *110*, 5959.

10 S. Dixon, K. M. Jr. Merz, G. Lauri, J. C. Ianni, *J. Comp. Chem*. **2004**, *26*, 23.

11 P.-W. Phuan, J. C. Ianni, M. C. Kozlowski, *J. Am. Chem. Soc.* **2004**, *126*, 15473.

12 a) K. B. Lipkowitz, M. Pradhan, *J. Org. Chem.* **2003**, *68*, 4648. b) M. C. Kozlowski, S. Dixon, M. Panda, G. Lauri, *J. Am. Chem. Soc.* **2003**, *125*, 6614. c) M. Hoogenraad, G. M. Klaus, N. Elders, S. M. Hooijschuur, B. McKay, A. A. Smith, E. W. P. Damen, *Tetrahedron: Asymmetry* **2004**, *15*, 519. d) J. C. Ianni, V. Annamalai, P.-W. Phuan, M. C. Kozlowski, *Angew. Chem. Int. Ed.* **2006**, *45*, 5502. e) M. C. Kozlowski, J. C. Ianni, *J. Mol. Cat. A* **2010**, *324*, 141. f) S. E. Denmark, N. D. Gould, L. M. Wolf, *J. Org. Chem.* **2011**, *76*, 4337. g) J. L. Melville, K. R. J. Lovelock, C. Wilson, B. Albutt, E. K. Burke, B. Lygo, J. D. Hirst, *J. Chem. Inf. Model.* **2005**, *45*, 971.

13 a) M. Pastor, G. Cruciani, I. McLay, S. Pickett, S. Clementi, *J. Med. Chem*. **2000**, *43*, 3233. b) F. Fontaine, M. Pastor, F. Sanz, *J. Med. Chem*. **2004**, *47*, 2805. c) F. Fontaine, M. Pastor, I. Zamora, F. Sanz, *J. Med. Chem.* **2005**, *48*, 2687.

14 S. Sciabola, A. Alex, P. D. Higginson, J. C. Mitchell, M. J. Snowden, I. Morao, *J. Org. Chem.* **2005**, *70*, 9025.

15 M. Urbano-Cuadrado, J. J. Carbó, A. G. Maldonado, C. Bo, *J. Chem. Inf. Model.* **2007**, *47*, 2228.

16 a) J. L. Melville, J. D. Hirst, *J. Chem. Inf. Model*. **2007**, *47*, 626. b) B.M. Spowage, C.L. Bruce, J.D. Hirst, *J. Cheminf*. **2009**, *1*, 22.

17 a) E. Zuidema, E. Daura-Oller, J. J. Carbó, C. Bo, P. W. N. M. van Leeuwen, *Organometallics* **2007**, *26*, 2234. b) E. Zuidema, L. Escorihuela, T. Eichelsheim, J. J. Carbó, C. Bo, P. C. J. Kamer, P. W. N. M. van Leeuwen, *Chem. Eur. J.* **2008**, *14*, 1843. c) J. Carbó, F. Maseras, C. Bo, P. W. N. M. van Leeuwen, *J. Am. Chem. Soc.* **2001**, *123*, 7630.

18 A. T. Axtell, J. Klosin, K. A. Abboud, *Organometallics* **2006,** *25,* 5003.

19 R. Ewalds, E. B. Eggeling, A. C. Hewat, P. C. J. Kamer, P. W. N. M. van Leeuwen, D. Vogt, *Chem. Eur. J.* **2000**, *6*, 1496.

20 a) F. Maseras, K. Morokuma, *J. Comput. Chem*. **1995**, *16*, 1170. b) T. K. Woo, L. Cavallo, T. Ziegler, *Theor. Chem. Acta* **1998**, *100*, 307.

21 a) ADF 2006.01. Department of Theoretical Chemistry. Vrije Universiteit. Amsterdam. b) E. J. Baerends, D. E. Ellis, P. Ros, *Chem. Phys*. **1973**, *2*, 41. c) L. Versluis, T. Ziegler, *J. Chem. Phys*. **1988**, *88*, 322. d) G. Te Velde, E. J. Baerends, *J. Comput. Phys*. **1992**, *99*, 84. e) C. Fonseca Guerra, J. G. Snijders, G. Te Velde, E. J. Baerends, *Theor. Chem. Acc*. **1998**, *99*, 391. f) G. Te Velde, F. M. Bickelhaupt, S. J. A. van Gisbergen, C. Fonseca Guerra, E. J. Baerends, J. G. Snijders, T. Ziegler, *J. Comput. Chem*. **2001**, *22*, 931.

22 a) A. D. Becke, *J. Chem. Phys*. **1986**, *84*, 4524. b) A. D. Becke, *Phys. Rev*. *A* **1988**, *38*, 3098. c) S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys*. **1980**, *58*, 1200. d) J. P. Perdew, *Phys. Rev*. *B* **1986**, *33*, 8822. e) J. P. Perdew, *Phys. Rev. B* **1986**, *34*, 7406.

23 Slater-type basis set, as included in ADF package. The 1s-3d electrons for Rh, the 1s electrons for C, N and O, and the 1s-2p electrons for P were treated as frozen cores. We applied scalar relativistic corrections to them *via* the zeroth-order regular approximation (ZORA) with the core potentials generated using the DIRAC program.[21]

24 a) M. Clark, R. D. III Cramer, N. van Opdenbosch, *J. Comput. Chem*. **1989**, *10*, 982. b) U. C. Sing, P. A. Kollman, *J. Comput. Chem.* **1986**, *7*, 718.

25 A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. III Goddard, W. M. Skiff, *J. Am. Chem. Soc.* **1992**, *114*, 10024.

26 ALMOND 3.3.0 from Molecular Discovery, London, UK.

27 a) S. Wold, M. Sjostrom, L. Eriksson, *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109. b) P. Geladi, B. Kowlaski, *Anal. Chim. Acta* **1985**, *35*, 1.

28 D. M. Hawkins, S. C. Basak, D. Mills *J. Chem. Inf. Comput. Sci.* **2003**, *23*, 579.

29 J. Gasteiger, M. Marsili, *Tetrahedron* **1980**, *36*, 3219.

30 S. A. Wildman, G. M. Crippen, *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 868.

31 T. J. Hou, K. Xia, W. Zhang, X. J. Xu, *J. Chem. Inf. Comput. Sci*. **2004**, *44*, 266.

32 http://comp.chem.nottingham.ac.uk/ download/tmacc

33 a) D. Gleich, R. Schmid, W. A. Herrmann, *Organometallics* **1998,** *17,* 2141. b) D. Gleich, W. A. Herrmann, *Organometallics* **1999,** *18,* 4354. c). X. Zhang, B. Cao, Y. Yan, S. Yu, B. Ji, X. Zhang, *Chem. Eur. J.* **2010**, *16*, 871.

34 For the training sets in the runs 1, 2, 3 and 4, the q^2 in prediction for the LOO cross-validations are 0.62, 0.65, 0.52 and 0.63, respectively.

35 J. E. Babin G. T. Whiteker, *US Pat.*, 911518, **199**2 (WO 9303839).

36 Y. Yan, X. Zhang, *J. Am. Chem. Soc*. **2006**, *128*, 7198.

37 Related procedures have been used to retrieve information about sterochemical induction from QSPR models of asymmetric catalysis using Grid-Independent methodologies (see ref. 14 and 15).

38 A. L. Watkins, C. R Landis, *J. Am. Chem. Soc.* **2010**, *132*, 10306.

39 a) P. W. N. M. van Leeuwen, C. F. Roobeek, *J. Organomet. Chem.* **1983**, *258*, 343. b) W. R. Moser, C. J. Papile, D. A. Brannon, R. A. Duwell, S. J. Weininger, *J. Mol. Catal.* **1987**, *41*, 271. c) M. F. Haddow, A. J. Middleton, A. G. Orpen, P. G. Pringle, R. Papp, *J. Chem. Soc. Dalton Trans*. **2009**, 202.

40 C. A. Tolman, *Chem. Rev.* **1977**, *77*, 313.

Chapter 5

QSSR Models for Predicting the Enantioselectivity in Asymmetric Cyclopropanation of Alkenes using Quantitative Quadrant-Diagram Representation

Chapter 5

QSSR Models for Predicting the Enantioselectivity in Asymmetric Cyclopropanation of Alkenes using Quantitative Quadrant-Diagram Representation

In this Chapter, we present a new methodology to predict enantioselectivity based on quantitative quadrant-diagram representation of the catalysts and quantitative structure-selectivity relationship (QSSR) modeling. To account for quadrant occupation, we used two types of molecular steric descriptors: the *Taft-Charton steric* parameter (*v*(*Charton*)) and the *distance-weighted volume* (V_W). The former is easily accessible from tabulated values, but cannot be applied to complex ligand architectures, while the latter can be calculated for any type of ligand from its ground-state structure. By assigning the value of the steric descriptors to each of the positions of the quadrant-diagram, we generated the independent variables to build the multidimensional QSSR models. The methodology was applied to predict the enantioselectivity in the cyclopropanation of styrene catalyzed by copper complexes, a problem that has proven difficult to solve with quantum mechanics. The dataset comprised 30 chiral ligands belonging to four different oxazoline-based ligand families: bis(oxazoline) (Box), azabis(oxazoline) (AzaBox), quinolinyl(oxazoline) (Quinox) and pyridyl(oxazoline) (Pyox). In the *first-order* approximation, we generated QSSR models with good predictive ability (r^2 = 0.89 and q^2 = 0.88). The derived stereochemical model indicated that installing very large groups at two diagonal quadrants and leading free the other two might be enough to obtain an enantioselective catalyst. Fitting the data to higher-order polynomial, which included crossed terms between the descriptors of the quadrants, resulted in an improvement of the predicting ability of the QSSR model (r^2 = 0.96 and q^2 = 0.93). This suggests that bulky substituents operate sinergically, and that substituent interdependencies might occur between the substituents at different quadrants. Also, we tested the *Q-QSSR* approach used in *Chapter 4* that involves alignment-independent 3D descriptors derived from DFT calculations. This approach failed in obtaining predictive models because during the GRIND alignment independence procedure some information about the shape of the catalyst is lost. Consequently, the descriptors cannot capture the subtle differences between catalysts that are responsible of variations in the selectivity outcome.

5.1. Introduction

To computationally assess the key factors governing stereoselectivity, it is common practise to locate and analyze the transition states (TS) on the potential energy surface (PES). However, this task can become very difficult when the PES is rather flat in the area surrounding the TS. An example of such situation is the addition of alkenes to copper-carbene complexes, which is the key step in asymmetric cyclopropanation of alkenes with diazo esters by copper catalysts. In fact, some authors have found that there is no enthalpic barrier on the corresponding PES; and that the activation barrier can only be detected on the estimated Gibbs free-energy surface (GFES), which is extremely difficult to calculate for complex systems.[1-2] Recently, García and co-workers have shown that the existence or absence of the transition state is very dependent on the substrate type and on the theory level employed.[3]

As demonstrated in previous chapter, an alternative to TS-based approaches are the Quantitative Structure-Property Relationship (QSPR) approaches analogous to those used in drug design.[4] Within this context, one particularly valuable aspect of these methods is that enantioselectivity is assessed from ground-state structures through definition of molecular descriptors. In *Chapter 1*, we have described the different approaches and their applications in homogeneous catalysis. During last years, QSSR approaches based on 3D descriptors are emerging as an alternative in the theoretical study of asymmetric catalysis.[5-6] To account for 3D features of catalysis, it is also possible to use geometrical descriptors based on steric size of the ligand that indirectly account for 3-dimensionality. They include easily accessible parameters form classical physical organic chemistry such as the steric parameters originally developed by Taft[7] and modified by Charton.[8] Their advantage is that a priori it is not necessary to know the catalyst geometry; but as disadvantage, they are only defined for some specific functional groups and cannot be applied to more complex ligand architectures. Using these parameters, Sigman and co-workers have found classical linear free energy relationships (LFER) to predict enantioselectivity for several Nozaki-Hiyama-Kishi carbonyl alkylations.[9] Then, in more sophisticated approach, they were able to generate multidimensional relationships of free energy to *Taft-Charton steric* and electronic parameters, including crossed terms that account for interplays between substituents.[10]

Even before these quantitative approaches, quadrant-diagrams were used to qualitatively rationalize the steric properties of the catalysts. The diagrams divide the space around the metal centre in four regions (quadrants), which are differently occupied by the bulkiness of the ligand generating a chiral shape that is responsible of stereoinduction. Cavallo and co-workers did the first attempts to quantify the steric hindrance induced by the ligand within these diagrams using the *percent of buried volume* ($\%V_{Bur}$).[11] The $\%V_{Bur}$ is the amount of volume in the first coordination sphere of a metal occupied by a given ligand. The same authors developed this concept to account for the different distribution of the ligand around the metal centre through the construction of the topographic steric maps.[12] They represented the interaction surface of the ligand that was exposed towards the incoming reactants. Inspired by Cavallo's work, we have also contributed to the quantification of quadrant-diagram representations, proposing a new molecular descriptor the *distance-weighted volume* (V_W),

which gives a measure of the ligand bulkiness and its impact on metal centre (*Chapter 3*).[13] This has been successfully used to qualitatively rationalize the outcome in asymmetric hydroformylation catalyzed by Rh-binaphos complexes (*Chapter 3*),[13] for which the TS-modeling also showed some limitations.[13-14] In a step forward, here, we explore the use quantitative quadrant-diagram representations to build 3D-QSSR models using the steric descriptors *Taft-Charton steric* (ν(*Charton*)) and *distance weighted volume* (V_W) in order to account for quadrant occupation. Although the quadrant-diagram representation of catalysts has been extensively used to rationalize catalytic selectivity, to date, it still remains to be proved that these representations are quantitatively correlated with enantioselectivity.

As mentioned in *Chapter 1*, the asymmetric cyclopropanation of styrene with diazoesters catalyzed by copper complexes with chiral ligands allow obtaining cyclopropanes enantioselectively.[15] The reaction generates two diasteroisomers which have two stereogenic centres, and consequently, great efforts have been made to develop efficient diastero- and enantioselective catalysts. The widespread family of oxazoline-based ligands showed the best results,[16] Figure 5.1 displays several types of skeletons that form the family of oxazoline-based ligands studied in this work; bis(oxazoline) (Box), azabis(oxazoline) (AzaBox), quinolinyl(oxazoline) (Quinox), and pyridyl(oxazoline) (Pyox).

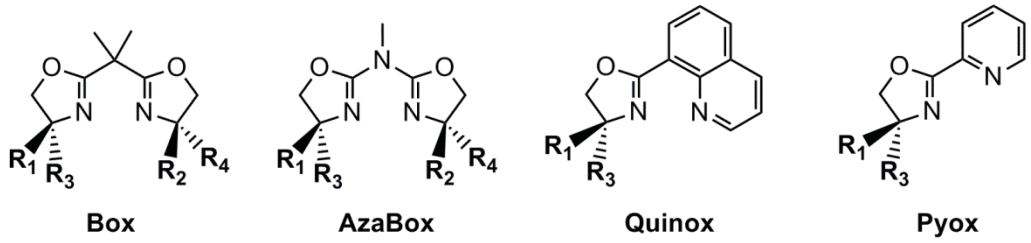


Figure 5.1. Oxazoline-based ligand families used in the asymmetric cyclopropanation of alkenes.

There are several experimental[17,18] and theoretical[1-3,19-20] studies on the mechanism of asymmetric cyclopropanation of alkenes with diazo compounds catalyzed by copper complexes. Thus, the catalytic cycle and the origin of enantioselectivity are now reasonably well established; the reaction proceeds via a Cu(I)-carbene complex and their role as reaction intermediates in cyclopropanation reactions has been unequivocally established.[21] The most favourable mechanism for the addition of alkenes copper-carbene complexes to give cyclopropanes is the *concerted pathway* (Figure 5.2), the direct addition of alkene to copper-carbene species being the step that controls the stereoselectivity.[1,19a]

The work in this Chapter has been performed in collaboration with Prof. J. I. García (Instituto de Síntesis Química y Cátalisis Homogénea, CSIC-Universidad de Zaragoza, Spain). We sake to test whether is possible to set quantitative multidimensional relationships of enantioselectivity and ground-state properties of the catalysts, leading to successful predictions of catalyst performance in cyclopropanation of styrene with diazo esters by

oxazoline-based copper complexes. To do that, we employ two different 3D-QSSR approaches. Firstly, the DFT-based alignment-independent approach used in previous chapter; and secondly, a new alignment-dependent approach based on quantitative quadrant-diagram representation of the catalysts. In addition, the use of chemically meaningful descriptors and models provides information to rationalize the stereoselectivity and to orient the ligand design.

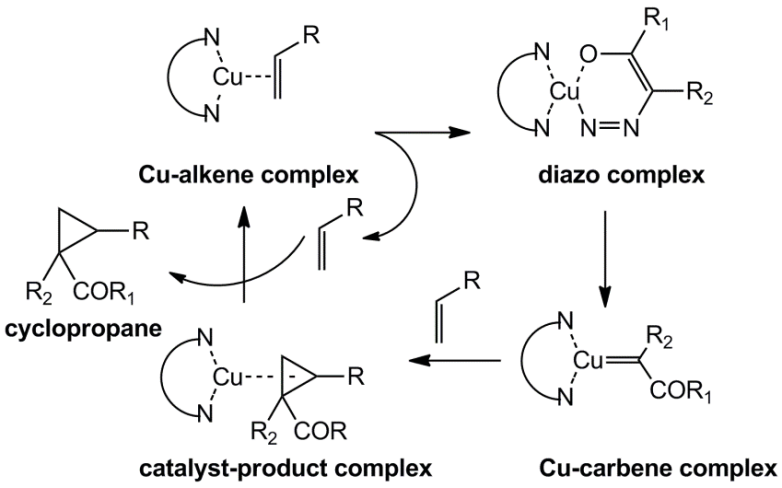


Figure 5.2. Concerted mechanism for the asymmetric cyclopropanation reaction of the alkenes with diazo compounds by copper catalysts.

5.2. Experimental background

The group of J. I. García has studied the asymmetric cyclopropanation by the copper catalyst. Figure 5.3 schematically shows the structures of the chiral ligands and the enantioselectivity outcome of previous catalytic experiments,[22] and two corresponding to complementary experiments performed in this work (ligands **14** and **23**). In all cases, the benchmark reaction was the asymmetric cyclopropanation of styrene with ethyl diazoacetate catalyzed by [(ligand)Cu(OTf)$_2$] complexes in dichloromethane, using the same 1:1 molar ratio of the reagents to straightforwardly compare the results. The *trans*/*cis* diastereoselectivity values fall between 60:40 and 73:27 ranges, the enantioselectivity being qualitatively similar in both *trans*- and *cis*-cyclopropanes.

The enantioselectivity observed for the new ligand **14**, with stereogenic centres at 4- and 5-positions of oxazoline ring, is very similar to that of ligand **13** (stereogenic centres at 4-position) and much higher than that of ligand **3** (stereogenic centres at 5-position). This result clearly indicates that only the substituents at the α-position with respect to chelating nitrogens (4-position) are close enough to the reaction centre as to influence reactivity. The ligand **23** led to an enantioselectivity very similar to other ligands bearing ^{t}Bu groups at C_2-symmetric positions.

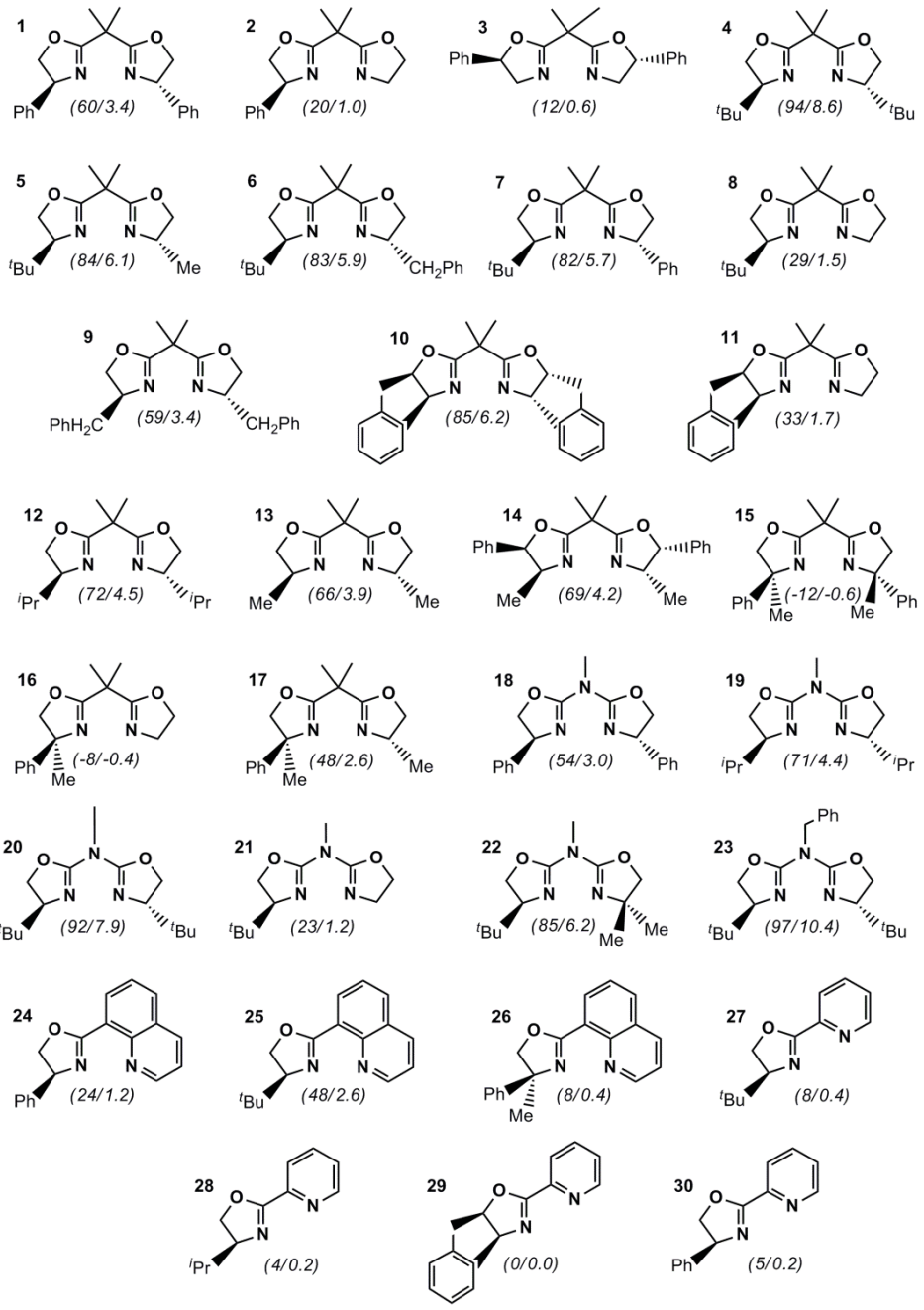


Figure 5.3. Ligand dataset and catalytic outcome for asymmetric cyclopropanation of styrene by the corresponding copper complexes (enantiomeric excess (*ee*) in % / *ee* expressed as energy differences, $\Delta\Delta G^{\ddagger}$, in kJ.mol^{-1} for the *trans* isomers). Negative values indicate that the opposite enantiomer is obtained as the major product.

5.3. Computational details and methods

To obtain the metal-ligand structures, we optimized the trigonal copper complex $[(L)CuCH_2]^+$ by means of DFT method[23] (B3LYP functional)[24] using the Gaussian03 series of programs.[25] All geometry optimizations were full using the 6-31G(d) basis set.[26] We used the simplest model carbene CH_2 in order to speed up the calculations and reduce the number of possible carbene isomers. For the conformational-rich substituents ^{i}Pr and CH_2Ph, we systematically searched for all the possible ligand isomers. Firstly, we selected the 3D-QSPR approach developed by Bo and co-workers, *Q-QSSR*, which uses alignment-independent three-dimensional descriptors derived from density functional theory (DFT) calculations.[6b] A detailed description of the methodology the *Q-QSSR* methodology is provided in *Chapter 4*.

Secondly, we used two different types of steric molecular descriptors to build the QSSR models based on quantitative quadrant-diagram representations of the catalysts: *Taft-Charton steric* parameter, ν(*Charton*),[8] and *distance-weighted volume*, V_W (*Chapter 3*).[13] The V_W gives a measure of the steric bulkiness of the ligand and its impact on the metal centre. Three different parameters were considered: i) the number of atoms in a given a region of the catalyst defined by the alignment hypothesis; ii) the distance of these atoms to the reactive centre, the copper atom, (d^l); and iii) the size of the atoms measured with the van der Waals radius (r^k). Depending on the balance between the size of the atom and its distance to metal centre (the exponents in equation 5.1), it is possible to define different variants of the descriptor, $V_{W,k,l}$. In equation 5.1, the *k* and *l* exponents took values of 0, 1, 2, and 3.

$$V_{W,k,l} = \sum_{i=1}^{N} \frac{r_i^k}{d_i^l} \tag{5.1}$$

For discussion in this Chapter, we use the two variants of $V_{W,k,l}$ ($V_{W,0,3}$ and $V_{W,1,3}$), that given the best results in the QSSR modeling; and a summary of the results for other types of V_W are given in the Annex. We used the *Partial Least Squares Regression* (PLSR)[27] as multivariate regression technique. External and full cross validation were considered for model building and evaluation. Different statistical parameters were employed for evaluating the predictive ability of models during the fitting and test stages, namely: the Pearson correlation coefficient (r^2), the determination coefficient (q^2), the standard error, and the slope and intercept of the fitted/predicted *vs.* observed values.[28]

5.4. Results and discussion

5.4.1. DFT-based alignment-independent QSSR models

Firstly, we tried to generate QSSR models using *Q-QSSR* methodology, which was successfully used to model the enantioselectivity in the asymmetric alkylation of aldehyde catalyzed by zinc-aminoalcohols[6b] and in the asymmetric hydroformylation of styrene by Rh-diphosphane (*Chapter 4*). We tried the full set of descriptors within the methodology. They included two

types of molecular fields, the *Molecular Shape Field* (MSF) and the *Molecular Electrostatic Potential* (MEP), and two types of filtering criteria, the MSF-based selecting convex areas (*CONV*) and the MEP-based selecting basic areas (*BAS*). This yields 4 different *correlograms* (descriptors): MSF-MSF$_{CONV}$, MSF-MSF$_{BAS}$, MEP-MEP$_{CONV}$ and MEP-MEP$_{BAS}$; as well as, numerous combination possibilities via concatenation of the descriptors (see *Chapter 4*). It is clear from the statistical parameters in Table 5.1 that none of the descriptors provided a mathematical model with predictive ability. The values of q^2 were all significantly lower than 0.5. Our tests also included different distance intervals (0.25, 0.5 and 1.0 Å) and distance ranges (between 3 and 10 Å, 3 and 15 Å, and 3 and 20 Å) for the *correlograms*; however, none resulting models improved sufficiently the values of statistical parameters to be considered an acceptable model. Table 5.1 shows the results obtained using an interval of 0.25 Å and distance range between 3 and 20 Å.

Table 5.1. Statistical parameters of LOO cross-validation for enantioselectivity using different descriptors and filtering criteria.

Filtering criteria	Descriptor	q^2	r^2
shape-based	MSF-MSF$_{CONV}$	0.10	0.39
	MEP-MEP$_{CONV}$	0.29	0.83
	MSF-MSF$_{CONV}$ & MEP-MEP$_{CONV}$	0.26	0.84
basicity-based	MSF-MSF$_{BAS}$	-0.01	0.52
	MEP-MEP$_{BAS}$	-0.27	0.08
combination	MEP-MEP$_{CONV}$ & MEP-MEP$_{BAS}$	0.08	0.89

To understand the failure of the alignment-independent GRIND descriptors, we can compare the *correlograms* of catalysts **5** and **8** (Figure 5.4). We observed that replacing one methyl group in **5** by hydrogen atom in **8** causes a dramatic drop in the selectivity, from 6.1 to 1.5 kJ.mol^{-1}. However, the MSF shape field was similar in both catalysts resulting in very similar *correlograms* (descriptors) as shown in Figure 5.4. We think that during alignment independence process, some information is lost. Thus, alignment-independent GRIND descriptors are not able to account for small structural differences and this can cause the failure of the methodology in some processes. For this reason, in next section, we generate new alignment-dependent descriptors, which derive from an alignment hypothesis based on a quadrant-diagram representation of the catalysts.

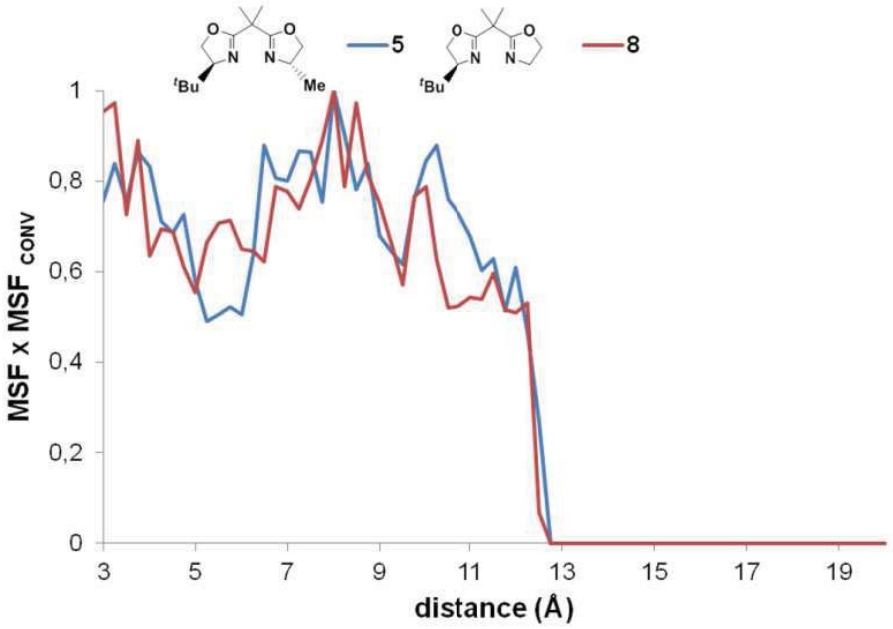


Figure 5.4. MSFxMSF$_{CONV}$ *correlograms* for the catalysts **5** ($\Delta\Delta G^{\ddagger}$ = 6.1 kJ.mol^{-1}) and **8** ($\Delta\Delta G^{\ddagger}$ = 1.9 kJ.mol^{-1}).

5.4.2. Steric molecular descriptors and quantitative quadrant-diagram representation

To qualitatively evaluate ligand effects on enantioselectivity, we used two types of molecular descriptors that account for steric effects: the *Taft-Charton steric* parameter, ν(*Charton*),[8] and the *distance-weighted volume*, $V_{W,k,l}$ (*Chapter 3*).[13] It was demonstrated that steric interactions involving the substituents of the oxazoline rings play a key role in the observed stereoselectivity.[19a] The *Taft-Charton steric* parameter is easily accessible from tabulated values and does not require additional calculations; however it is only defined for some specific functional groups and cannot be applied to more complex ligand architectures such as those in ligands **10**, **11** and **29**. Thus, we defined a subset (**set 1**) formed by ligands **1**-**9**, **12**-**28** and **30**. The variations of the ligand scaffold occur mainly in the substituents at the positions R_1, R_2, R_3 and R_4 (Figure 5.1 and 5.3). By assigning the value of ν(*Charton*) to each position, we generated four independent variable (ν(R_1), ν(R_2), ν(R_3) and ν(R_4)), which can be used to build polynomial models describing enantioselectivity. With regard to the spatial assignment of this model, we were implicitly assuming that the enantioselectivity can be rationalized by the well-known quadrant-based representation, in which the steric hindrance of each quadrant is evaluated by the *Taft-Charton steric* parameter associated with the corresponding substituent.

Following the hypothesis that enantioselectivity can be rationalized by quadrant-diagram representation, we also evaluated the *distance-weighted volume*, $V_{W,k,l}$ as a descriptor that quantifies the different space occupation of the quadrants. We computed $V_{W,k,l}$ using the ground-state geometries of $[(L)Cu(CH_2)]^+$ complexes aligned according to chiral hypothesis in

quadrant-diagram representation shown in Figure 5.5. The copper atom was placed at the origin, with the N atoms bound to copper in the *xz* plane at negative *z* values, and with the pseudo C_2-symmetry axis bisecting the N-Cu-N chelate angle aligned along the *z* axis. Then, we divided the *xy* plane along the *x* and *y* axis in four quadrants which correspond to the *-x/+y* (Q_1), *+x/+y* (Q_2), *-x/-y* (Q_3), and *+x/-y* (Q_4). This alignment hypothesis can be applied to the 4 different families of the oxazoline-ligand, for which it is possible to identify four common spatial regions, labelled as Q_1, Q_2, Q_3 and Q_4. For each quadrant, we considered the atoms in the front side of the catalyst ($z > 0$), along with different relations between the size of the atom (r^k, r = van der Waals radius), and the distance (d^l) from the atom to the metal centre (see *computational details* section). The $V_{W,k,l}$ can be computed for the 30 ligands (**set 2**), being each catalyst described by four descriptors: $V_{W,k,l}(Q_1)$, $V_{W,k,l}(Q_2)$, $V_{W,k,l}(Q_3)$ and $V_{W,k,l}(Q_4)$, the independent variables. For simplicity, hereinafter, we refer to the variable of both types of descriptors, $\nu(R_i)$ and $V_{W,k,l}(Q_i)$, as $\boldsymbol{Q_i}$. Also, we had to consider that the alignment hypothesis can lead to two different molecular structures, depending on how we arrange the substituents with regard to the *xz* plane. For example in ligand **2**, the alignment can place the Ph substituent at position Q_1 or at Q_4. Although both structures are chemically indistinguishable through C_2-rotation around *z* axis, the descriptor vector changes by permutation between two pairs of components (for example in ligand **2**, from [0.57,0,0,0] to [0,0,0,0.57] for ν(*Charton*)). In the case of C_2-symmetric ligands such as **1**, the descriptor vector remains unchanged. To account for this effect of symmetry, we duplicated the design matrix, defining for each of the catalysts two vectors that result from permuting two pairs of components $\boldsymbol{Q_1/Q_4}$ and $\boldsymbol{Q_2/Q_3}$. Consequently, the resulting models should be C_2-symmetric.

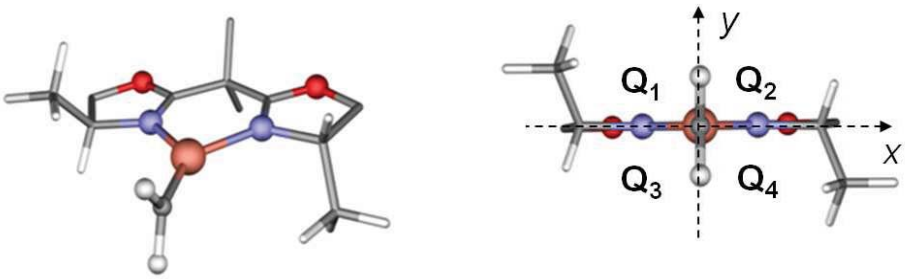


Figure 5.5. Molecular structure of $[(Me_2Box)Cu(CH_2)]^+$ complex (catalyst **13**). On the right the complex is oriented following chiral hypothesis.

Table 5.2 collects the values of the steric parameters for the different functional groups in order to perform an initial qualitative analysis of their predicting capabilities. The values of $V_{W,k,l}$ depend on the specific molecular structure of each $[(L)Cu(CH_2)]^+$ complex determined by the ligand scaffold; and therefore, Table 5.2 shows average values. Nevertheless, the values vary within a relatively narrow range (see Annex for complete list of parameters). A difficulty of this modeling is to deal with all of the possible conformational isomers. To handle with this problem, we systematically searched for all of the possible conformational isomers; and then, all of them were taken into account by Boltzmann-weighting the contribution of the minimized

conformations. The use of this type of strategy in QSSR approaches was reefed to as '3.5D QSSR models'.[5d]

Table 5.2. *Taft-Charton steric* (ν(*Charton*)) and averaged *distance-weighted volume* ($V_{W,k,l}$) parameters for the ligand substituents.

Substituent	ν(*Charton*)	$V_{W,0,3}$(x 10^{-1})	$V_{W,1,3}$(x 10^{-1})
H	0	0.08	0.09
Me	0.52	0.38	0.42
Ph	0.57[(a)]	0.68	0.94
CH_2Ph	0.70	0.80	1.03
^{i}Pr	0.76	0.72	0.80
^{t}Bu	1.24	1.07	1.23
Indonyl	-	0.60	0.73

[(a)] For Ph group the value 1.66 has been also defined (see text).

Both ν(*Charton*) and $V_{W,k,l}$ type descriptors gave the largest values for the ^{t}Bu group, which is present in the ligands yielding the largest enantioselectivity. When comparing ^{i}Pr and Ph groups, the ν(*Charton*) and $V_{W,0,3}$ seemed to do better than the $V_{W,1,3}$. The enantioinduction by ^{i}Pr-substituted ligands (72 and 71% *ee* for **12** and **19**) is larger than by the analogous Ph-substituted (60 and 54% *ee* for **1** and **18**), but according to $V_{W,1,3}$ descriptor, the size of ^{i}Pr group is smaller than the size of Ph (Table 5.2), whereas ν(*Charton*) and $V_{W,0,3}$ give larger values for ^{i}Pr than for Ph group. Likewise, comparing ^{i}Pr and CH_2Ph groups, it seems that the accuracy of the descriptors follows the order: ν(*Charton*) > $V_{W,0,3}$ > $V_{W,1,3}$. The enantioinduction by ^{i}Pr-substituted ligand **12** (72% *ee*) is larger than by the analogous CH_2Ph-substituted ligand **9** (59% *ee*), the evaluated size of the ^{i}Pr group being larger, similar or smaller for ν(*Charton*), $V_{W,0,3}$ or $V_{W,1,3}$.

We noted that Charton defined two different parameters for Ph group that should be applied depending on the situation (Table 5.2 and footnote). Here, the lower value better correlated with selectivity indicating a low steric induction, probably as a consequence of the flexibility of the Ph group, which reorganizes with the incoming substrate through rotation around the C-C_{ipso} bond. Contrary, calculations of $V_{W,k,l}$ on the quadrants occupied by Ph groups could not catch their flexibility because only one isomer was found, leading to values significantly larger than for Me group that induces comparable enantioinduction (compare for example, ligands **1** and **13** and values in Table 5.2). Nevertheless, the main power of $V_{W,k,l}$ is that allowed us describing all the catalysts.

5.4.3. Quadrant representation-based QSSR models: first-order approximation

Initially, we considered QSSR models fitted to first order polynomials with four variables (***Q₁***, ***Q₂***, ***Q₃*** and ***Q₄***) using enantioselectivity, expressed in terms of free energy ($\Delta\Delta G^{\ddagger}$) as the dependent variable (equation 5.2).

$$\Delta\Delta G^{\ddagger} = C_0 + \sum_{i=1}^{4} C_i Q_i \qquad (5.2)$$

The coefficients values (C_0, C_1, C_2, C_3 and C_4) were obtained using multivariate PLSR analysis (see above). Table 5.3 collects the main statistical parameters for the descriptors ν(*Charton*) and $V_{W,k,l}$. The Charton-based model had a good predictive power (q^2 = 0.78); and the discrepancy between q^2 and r^2 values (r^2 - q^2 < 0.2) were not too large, indicating that no major overfitting occurs. The QSSR models generated with $V_{W,0,3}$ and $V_{W,1,3}$ descriptors for the full set of 30 catalysts showed similar statistical parameters (q^2 > 0.7 and r^2 - q^2 < 0.2). Interestedly, these models follow the same order of accuracy as that defined in the qualitative analysis of descriptor values. Figure 5.6 shows the experimentally measured plotted against the fitted free energy values.

Table 5.3. Statistical parameters of LOO cross-validation for $\Delta\Delta G^{\ddagger}$ using different type of descriptors.[(a)]

Descriptor	q^2	r^2	Matrix dimension
ν(*Charton*)	0.78	0.81	54x4
$V_{W,0,3}$	0.76	0.79	60x4
$V_{W,1,3}$	0.70	0.73	60x4

[(a)] Slope (fitting/prediction), intercept (fitting/prediction), error (fitting/prediction): 0.81/0.80, 0.61/0.65 kJ.mol^{-1} 1.24/1.36 kJ.mol^{-1} for ν(*Charton*); 0.79/0.79, 0.68/0.64 kJ.mol^{-1}, 1.31/1.39 kJ.mol^{-1} for $V_{W,0,3}$; 0.73/0.73, 0.86/0.82 kJ.mol^{-1}, 1.48/1.57 kJ.mol^{-1} for $V_{W,1,3}$.

The final QSSR equation contains both positive and negative coefficients which, combined with the values of steric descriptors for a particular system, produces the corresponding predicted enantioselectivity. All the QSSR equations showed the same pattern in the coefficient signs, and that indicates that all of them describe the same stereochemical model (see Table 5.4). The values of the coefficients C_1 and C_4 were equal and positive, while those of the coefficients C_2 and C_3 were equal and negative. Thus, increasing the steric bulk in the quadrants Q_1 and Q_4 would favour one enantiomer ($\Delta\Delta G^{\ddagger}$ > 0), whereas the increase in Q_2 and Q_3 would favour the opposite enantiomer ($\Delta\Delta G^{\ddagger}$ < 0). Occupation of both types of quadrants would lead to mutual cancellation of enantioselectivity. Thus, the quadrant-based model is able to predict the absolute configuration and the quantitative enantiomeric excess induced by the catalysts.

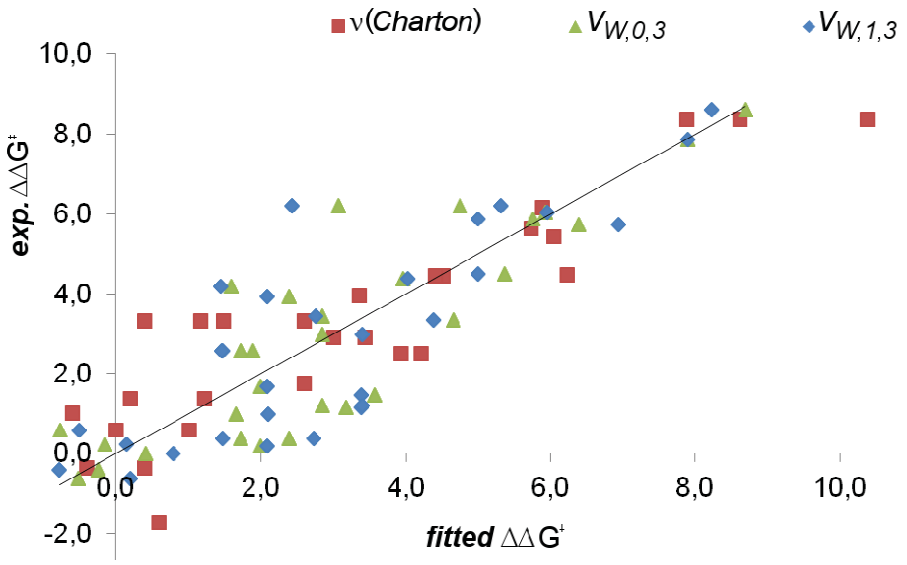


Figure 5.6. Experimental *vs.* fitting $\Delta\Delta G^{\ddagger}$ values in kcal.mol^{-1} using ν(*Charton*) and $V_{W,k,l}$ descriptors.

Table 5.4. Coefficients of the first-order equations for **set 1** and **set 2**.

Descriptor	C_0	C_1	C_2	C_3	C_4
ν(*Charton*)	-1.7	4.1	-1.8	-1.8	4.1
$V_{W,0,3}$	-0.7	39.4	-55.9	-55.9	39.4
$V_{W,1,3}$	-0.5	32.3	-63.9	-63.9	32.3

As expected, the coefficients associated with quadrants in diagonal arrangement were identical, that is, the coefficients were C_2-symmetric. However, when comparing the absolute values of C_1/C_4 coefficients with C_2/C_3, we observed certain asymmetry. We also noted that the independent terms (C_0) were not zero. This could be a consequence of the results deviation from ideal model (q^2 = 1), and/or of the lack of ligands with bulky groups at positions Q_2 and Q_3 giving $\Delta\Delta G^{\ddagger} < 0$. Despite of the limitations of the model, it was reasonable to propose a stereochemical model based on quadrant-diagram representation as depicted in Figure 5.7. Both diagrams are analogous but yield opposite enantiomers; the grey quadrant means that the space in this quadrant is occupied by the ligand. Thus, blocking alternative corner-sharing quadrants high enantioselectivities are expected. This is consistent with previously proposed models (Figure 5.8),[19] and roughly with experimental results. DFT calculations have shown that the key intermediate is a copper-carbene species, in which the carbene plane is perpendicular to the complex plane formed by the Cu atom and the two bound N atoms. Hence, the ester substituent of the carbene can place either above or bellow the plane. The enantioselectivity was rationalized from the relative energies of the transition

states for ethene addition. In the TS, the alkene approaches along Cu-N axis and parallel to it. The alkene slightly flips up or down of the catalyst plane ending in a pseudo-trans disposition respect to the ester group. This approach pushes back the ester group of the carbene as schematically illustrated in Figure 5.8 (and round-capped sticks in Figure 5.7). When the ester group meets a bulky group such the ^{t}Bu substituent, the resultant steric interaction causes an important deformation of the chelate complex, disfavouring the reaction channel. Thus, sterically blocking the Q_1 and Q_4 quadrants would lead to the preferred formation of (1*R*)-cyclopropanes, whereas blocking Q_2 and Q_3 quadrants would lead to the opposite 1*S* enantiomer. Both situations are illustrated in Figure 5.7. In the former, the attack of the alkene through either the up-right or the down-left quadrants would lead the ester group in the non-sterically demanding quadrant down-left and up-right, respectively.

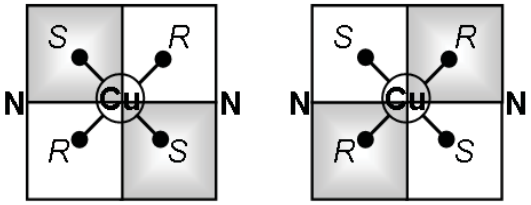


Figure 5.7. Initially proposed quadrant-diagram representation of the possible cyclopropanation paths for copper catalyst modified with oxazoline-based ligands considering a *first-order approximation*.

This initial analysis may suggest that the contribution of the quadrants to enantioselectivity is S_4-symmetric, the axis being perpendicular to the quadrant and passing through the canter of the diagram occupied by the copper atom. Thus, it is reasonable to try to set a simpler linear free energy relationship with the steric parameters that assumes an equal additive contribution of the steric effects in each quadrant. We generated new descriptors by summing the four values of each quadrant and keeping the sign found in the QSSR equation. The new descriptors ν_T(*Charton*) and $V_{TW,0,3}$ were defined in equations 5.3 and 5.4, respectively.

$$\nu_T(Charton) = \nu(Q_1) - \nu(Q_2) - \nu(Q_3) + \nu(Q_4) \tag{5.3}$$

$$V_{TW,0,3} = V_{W,0,3}(Q_1) - V_{W,0,3}(Q_2) - V_{W,0,3}(Q_3) + V_{W,0,3}(Q_4) \tag{5.4}$$

The linear regression between experimental $\Delta\Delta G^{\ddagger}$ and steric parameters gave a r^2 of 0.80 and 0.77 for ν_T(*Charton*) and $V_{TW,0,3}$, respectively (Figure 5.9). Considering only the Box ligand family, the correlation with ν_T(*Charton*) improved (r^2 = 0.87). The obtained correlations were statistically very close to those of multivariate analysis or even slightly better, and confirmed that a rough linear free energy relationship exists and that the proposed stereochemical model (Figure 5.7) is informative. These mathematical models can be expressed in an extended form showing the S_4 symmetry of the coefficients (equations 5.5 and 5.6).

$$\Delta\Delta G^{\ddagger}(\nu(Charton)) = -0.9 + 3.6\,\boldsymbol{Q_1} - 3.6\,\boldsymbol{Q_2} - 3.6\,\boldsymbol{Q_3} + 3.6\,\boldsymbol{Q_4} \quad (5.5)$$

$$\Delta\Delta G^{\ddagger}(V_{W,0,3}) = -1.1 + 40.6\,\boldsymbol{Q_1} - 40.6\,\boldsymbol{Q_2} - 40.6\,\boldsymbol{Q_3} + 40.6\,\boldsymbol{Q_4} \quad (5.6)$$

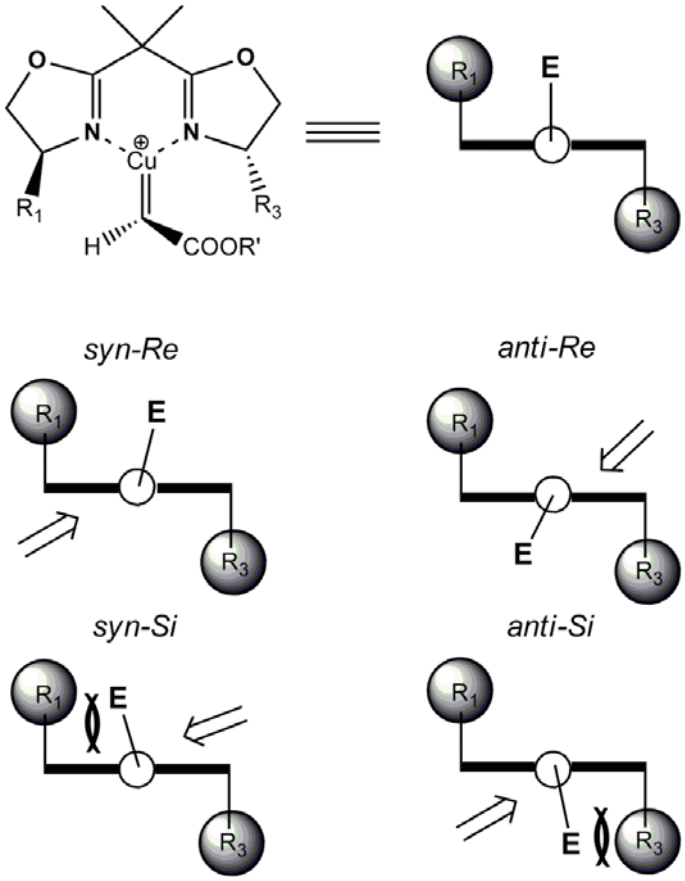


Figure 5.8. Schematic representation of the possible reaction channels and the possible steric interaction that explain enantioselectivity according to previously proposed models (ref. 19c).

Alternatively, to account for the symmetry of the stereochemical model, we can generate a new "virtual" datasets (**set 3** and **4**) by applying 4-fold "improper" rotation (S_4) to the quadrant positions. For odd operations (S_4^1 and S_4^3) the dependent variable ($\Delta\Delta G^{\ddagger}$) changes its sign, whereas for even operations (S_4^2 and S_4^4), it does not. This means that we quadruple the design matrix, defining for each of the catalysts four vectors. Moreover, the virtual datasets is wider and more representative because it covers the full positive and negative $\Delta\Delta G^{\ddagger}$ values (pro-*S* and –*R* catalysts). The statistical parameters of the virtual datasets models, q^2 = 0.89 and 0.88 for ν*(Charton*) and $V_{W,0,3}$, improved respect to the previous models. The equations 5.7

and 5.8 show the "virtual" models with symmetric coefficients and an independent term equal to zero. The new models are qualitatively and conceptually closer to the chemical behaviour of the catalytic system. For example, given an achiral catalyst with the four positions equally substituted, the equations would predict a $\Delta\Delta G^{\ddagger} = 0$, whereas previous equations would not.

$$\Delta\Delta G^{\ddagger}(\nu(Charton) = 0.0 + 3.0\,\boldsymbol{Q_1} - 3.0\,\boldsymbol{Q_2} - 3.0\,\boldsymbol{Q_3} + 3.0\,\boldsymbol{Q_4} \quad (5.7)$$

$$\Delta\Delta G^{\ddagger}(V_{W,0,3}) = 0.0 + 32.7\,\boldsymbol{Q_1} - 32.7\,\boldsymbol{Q_2} - 32.7\,\boldsymbol{Q_3} + 32.7\,\boldsymbol{Q_4} \quad (5.8)$$

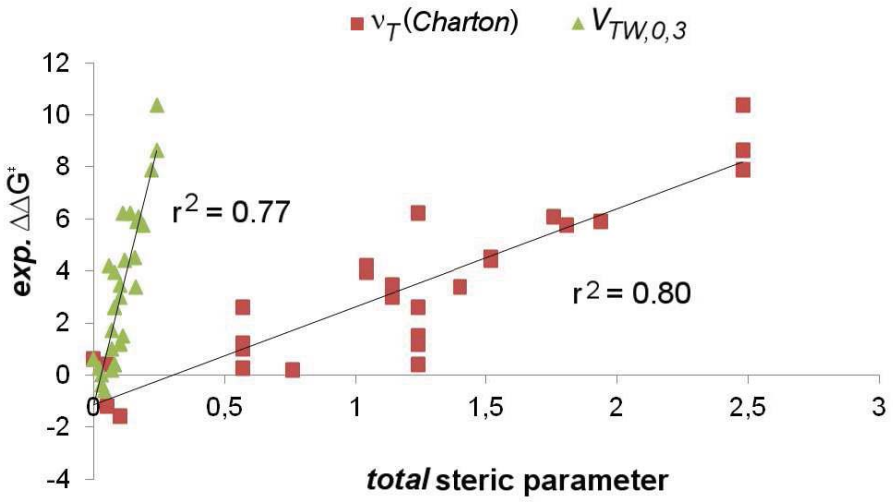


Figure 5.9. Experimental $\Delta\Delta G^{\ddagger}$ (in kJ.mol^{-1}) *vs. total* steric parameters ν_T(*Charton*) (squares) and $V_{TW,0,3}$ (dots); and linear regression coefficients (r^2).

To assess further how our qualitative and quantitative models account for the experimental results, we analyzed different prototypical cases of substitution pattern. Figure 5.10 collects all of the possible stereochemical situations based on quadrant-diagram representations. According to the derived models, the C_2- and pseudo-C_2-symmetric ligands bearing bulky substituents in position Q_1 and Q_4 should show high enantioinduction. As representative cases, the ligand **4** (*ee* = 94%) with two ^{t}Bu groups (Figure 5.10a) and **5** (*ee* = 84%) with one ^{t}Bu and one Me groups (Figure 5.10b) led to (*1R*)-cyclopropanes enantioselectively. In both cases, the two positive coefficients of QSSR equation added to increase the enantioselectivity, without counterbalance of substituents in positions Q_2 and Q_3. The catalysts with medium-size substituents in quadrants Q_1 and Q_4 (Figure 5.10e) such as the ligands **13** and **18** (*ee* = 66 and 54%) led to intermediate selectivity. They had the same mechanism of action as the high-selective ligand **4**, but the interactions of substituents with ester group seemed less destabilizing. Low enantioselectivities were observed for ligands bearing one or two quaternary stereogenic centres (Figure 5.10f and g) such as in ligands **16** and **15** (*ee* =-12% and -8 %). In ligand **16**, the two quadrants of one side (Q_1 and Q_3) were occupied, leading to

equally accessible paths for pro-*S* and pro–*R* products. In ligand **15**, all paths were similarly available or similarly hindered. The catalysts with only one substituent showed low enantioselectivities, even if the substituent was bulky such as the ^{t}Bu group (Figure 5.10h) in the ligand **8** (*ee* = 29 %). The observed *ee* was very close to that expected in a situation with three unique and equally accessible paths in the proportion 2 pro-*R* vs. 1 pro-*S* (33%).

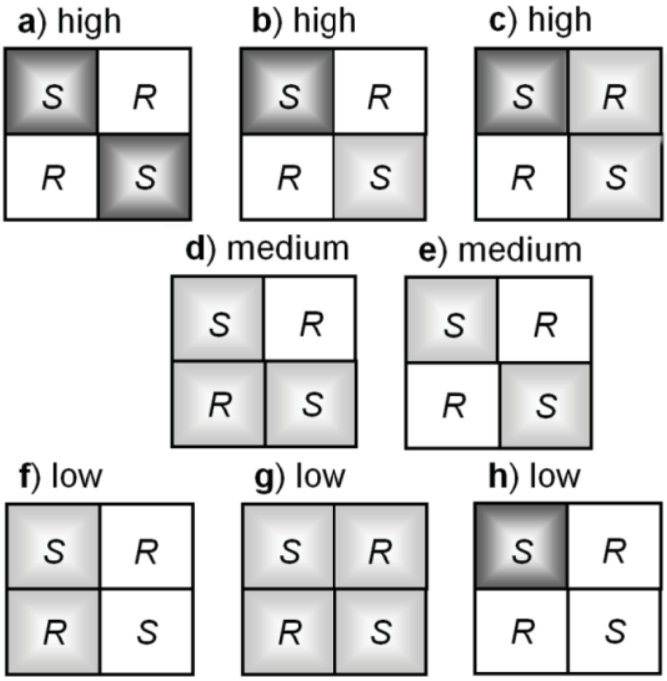


Figure 5.10. Quadrant-diagram representation of the possible cyclopropanation paths for different ligand substitution patterns.

Although the model was widely useful, we noted an outstanding exception. As reported in previous contributions, the steric effects of the substituents may not be strictly additive.[19a] The comparison of ligands **5** and **22** (Figure 5.10b and c) is an illustrative and puzzling example. Both catalysts showed very similar enantioselectivities (84 and 85%), although we would expect a decrease in *ee* on going from on going from ligand **5** to **22**, because the second Me group in Q_2 position would disfavour one of the *R* reaction channels. In fact for ligand **22**, all the models predicted $\Delta\Delta G^{\ddagger}$ values lower than experimental ones (ranging from 2.4 to 4.5 kJ.mol^{-1} for the fitting stage *vs.* 6.2 kJ.mol^{-1} experimentally). Previous calculation of the transition states for these species had showed that the presence of a second Me group does not appear to introduce new steric interactions.[27c] In the *syn-Re* path (Figure 5.8), the alkene flips down and the ester group in the carbene adopts a position that is intermediate between the two Me groups, avoiding the destabilizing steric interactions. On the other hand, in the *anti-Si* path (Figure 5.8), the alkene have also to flip down because the presence of the ^{t}Bu group in Q_1 position and leads to a disfavouring interaction between the ester group and the

Me group in position Q_4. This means that the steric interaction between the ester group and the bulky substituents on one side may be affected by the substituents on the other side. The different substituents may contribute synergically to the selectivity or other types of interplays may be present. Therefore, more complex mathematical models are worth to be explored in order to account for full catalyst features.

5.4.4. Quadrant representation-based QSSR models: higher-order approximation including crossed terms

To improve data description, we tried to fit the experimental values to polynomial functions including crossing and higher-order terms. Thus, we built polynomials of the form described in equation 5.9, where, a_i, b_i, c_i, d_{ij} and e_{ij} are the coefficient values to be found, $\boldsymbol{Q_i^n}$ the steric parameters raised to different powers, and $\boldsymbol{Q_i^n Q_j^m}$ the crossed terms evaluating the hypothesized interplay between substituent effects on selectivity. This approach is closely related to the three-dimensional correlations of steric parameters proposed by Sigman and co-workers to predict asymmetric catalyst performance.[10]

$$\Delta\Delta G^{\ddagger} = a_0 + \sum_{i=1}^{N} a_i Q_i + \sum_{i=1}^{N} b_i Q_i^2 + \sum_{i=1}^{N} c_i Q_i^3 + \sum_{i=1}^{N}\sum_{j>i}^{N} d_{ij} Q_i Q_{j+} \sum_{i=1}^{N}\sum_{j>i}^{N} e_{ij} Q_i Q_j^2 \qquad (5.9)$$

Table 5.5. Statistical parameters of LOO cross-validation for $\Delta\Delta G^{\ddagger}$ using different third-order polynomials and the non virtual datasets (**set 1** and **set 2**).[(a)]

	ν(*Charton*)		$V_{W,0,3}$		$V_{W,1,3}$	
Crossed terms	$\boldsymbol{q^2}$	$\boldsymbol{r^2}$	$\boldsymbol{q^2}$	$\boldsymbol{r^2}$	$\boldsymbol{q^2}$	$\boldsymbol{r^2}$
all	0.82	0.90	0.86	0.92	0.82	0.90
$\boldsymbol{Q_1^n Q_4^m, Q_2^n Q_3^m}$	0.88	0.94	0.76	0.79	0.70	0.74
$\boldsymbol{Q_1^n Q_2^m, Q_1^n Q_3^m}$ $\boldsymbol{Q_2^n Q_4^m, Q_3^n Q_4^m}$	0.74	0.83	0.76	0.79	0.70	0.73

[(a)] Slope (fitting/prediction), intercept (fitting/prediction), error (fitting/prediction): considering all terms :0.90/0.83, 0.32/0.49 kJ.mol^{-1}, 0.91/1.21 kJ.mol^{-1} for ν(*Charton*); 0.92/0.91, 0.26/0.18 kJ.mol^{-1}, 0.82/1.06 kJ.mol^{-1} for $V_{W,0,3}$; 0.90/0.86, 0.32/0.36 kJ.mol^{-1}, 0.90/1.22 kJ.mol^{-1} for $V_{W,1,3}$; considering $\boldsymbol{Q_1^n Q_4^m, Q_2^n Q_3^m}$: 0.94/0.89, 0.19/0.22 kJ.mol^{-1}, 0.70/1.01 kJ.mol^{-1} for ν(*Charton*); 0.79/0.79, 0.67/0.63 kJ.mol^{-1}, 1.31/1.39 kJ.mol^{-1} for $V_{W,0,3}$; 0.74/0.74, 0.85/0.81 kJ.mol^{-1}, 1.47/1.57 kJ.mol^{-1} for $V_{W,1,3}$; considering $\boldsymbol{Q_1^n Q_2^m, Q_1^n Q_3^m, Q_2^n Q_4^m, Q_3^n Q_4^m}$: 0.83/0.77, 0.56/0.70 kJ.mol^{-1}, 1.19/1.46 kJ.mol^{-1} for ν(*Charton*); 0.79/0.79, 0.68/0.64 kJ.mol^{-1}, 1.31/1.39 kJ.mol^{-1} for $V_{W,0,3}$; 0.73/0.73, 0.86/0.83 kJ.mol^{-1}, 1.48/1.58 kJ.mol^{-1} for $V_{W,1,3}$.

Table 5.5 and Figure 5.11 collect the results of the multivariable regressions for the different steric parameters and polynomials of the non virtual datasets (**set 1** and **set 2**). The coefficients of the models are given in Annex. When all the possible terms of the equation were considered, the new models improved significantly the correlation of the data, suggesting and enhanced predicting ability. For all the descriptors, we obtained $q^2 > 0.8$ and

r^2 > 0.9 (first row of Table 5.5), which are values typical of good precision models. In general, we observed an improvement of prediction for catalysts with a sole bulky substituent at position Q_1 or Q_4 (ligands **8**, **21**, **25** and **27**). For example for ligand **8**, the predicted $\Delta\Delta G^{\ddagger}$ in the fitting stage of Charton-based models diminished from +3.3 to +1.7 kJ.mol^{-1} on going from first- to third-order polynomial, becoming closer to the experimental value (+1.5 kJ.mol^{-1}). The improved prediction for ligands with a sole substituent indicates that the relation between the enantioselectivity and the steric size of the substituents is not strictly linear. The new polynomial also improved somewhat the prediction of ligand **22,** for which the interplay between substituents was hypothesized. When $V_{W,k,l}$ descriptors were used, the values vary from +3.1 and +2.4 kJ.mol^{-1} in first-order approximation to +4.9 and +4.6 kJ.mol^{-1} in the full third-order approximation, becoming closer to the experimental value (+6.2 kJ.mol^{-1}). Thus, although the description of ligand **22** features was not completely satisfactory, inclusion of crossed terms seemed to catch some interplay effects between substituents.

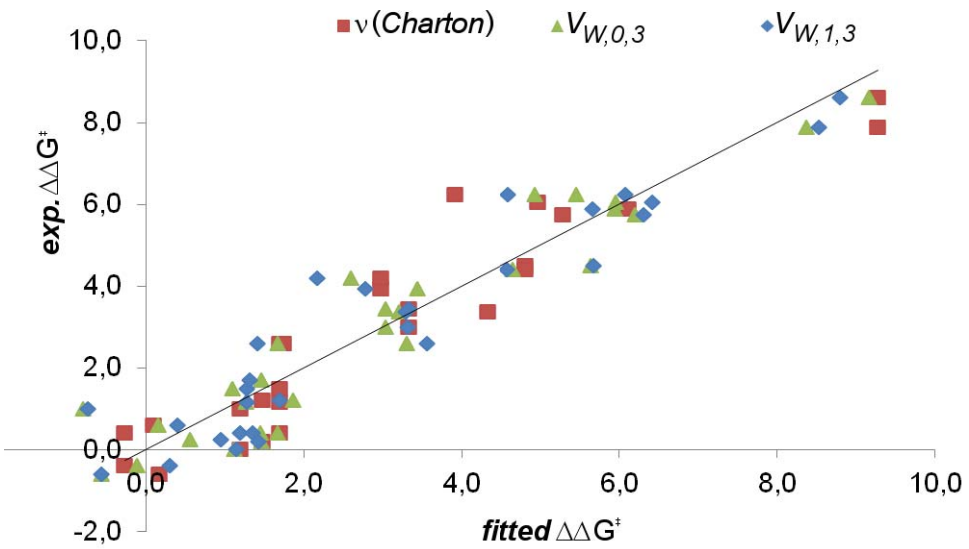


Figure 5.11. Experimental *vs.* fitting $\Delta\Delta G^{\ddagger}$ values in kcal.mol^{-1} for the third-order polynomial including all of the possible crossed terms for the non virtual datasets (**set1** and **set2**).

After full cross-validation of catalytic system, the dataset was divided into training (20 and 22 ligands for ν(*Charton*) and $V_{W,k,l}$ descriptors, respectively) and test subsets (7 and 8 ligands) to develop an externally validated QSSR model. We selected catalysts showing varied selectivities ranging from low to high values, and ensuring that the training models had good statistical parameters. A new QSSR model was generated with the training subset and predictions were made for the test subsets. Table 5.6 shows the results. There was a clear correlation between the experimental and the predicted values for all selectivity ranges. The three models classified correctly the catalysts as low-, medium- and high-selective ($\Delta\Delta G^{\ddagger}$ < 3, 3 < $\Delta\Delta G^{\ddagger}$ < 5, $\Delta\Delta G^{\ddagger}$ > 5 kJ.mol^{-1}, respectively), predicting roughly the same order as experiments.

To evaluate the specific interdependency between substituents, we can successively remove crossed terms and observe how the predicting ability varies. Removing the crossed terms between non-diagonal quadrants (second row of Table 5.5), the predictive ability for ν(*Charton*) remained practically invariant, whereas for $V_{W,k,l}$ decreased to values close to those of first-order polynomial model. Alternatively, removing the crossed terms between diagonal quadrants ($Q_1{}^nQ_4{}^m$ and $Q_2{}^nQ_3{}^m$), we observed a decrease of the correlation parameters for all the steric parameters (third row of Table 5.5), showing values very close to the first-order polynomial model. The results with $V_{W,k,l}$ parameters were more difficult to interpret because large groups had atoms in more than one quadrant. But, if we focus in the simpler ν(*Charton*) parameter, the comparison indicates that the most relevant interdependencies between substituents corresponds to groups placed in diagonal quadrants.

Table 5.6. Experimental *versus* predicted enantiomeric excess (expressed as $\Delta\Delta G^{\ddagger}$ in kJ.mol^{-1}) for a test subset (N = 7 or 8) employing QSSR models generated from the corresponding training set (N = 20 or 22) fitted to full third-order polynomial.[(a)]

Catalysts	experimental	ν(*Charton*)	$V_{W,0,3}$	$V_{W,1,3}$
4	8.6	9.2	9.6	9.1
5	6.1	-	6.0	6.3
6	5.9	6.0	6.1	5.8
7	5.7	5.2	6.7	7.1
13	3.9	2.9	3.4	2.8
18	3.0	3.2	3.0	3.6
21	1.2	1.6	1.5	1.4
28	0.2	1.4	1.6	1.6

[(a)] q^2/r^2, slope (fitting/prediction), intercept (fitting/prediction), error (fitting/prediction): 0.78/0.92, 0.92/0.81, 0.23/0.37 kJ.mol^{-1}, 0.79/1.35 kJ.mol^{-1} for ν(*Charton*); 0.79/0.90, 0.90/0.88, 0.23/0.22 kJ.mol^{-1}, 0.88/1.30 kJ.mol^{-1} for $V_{W,0,3}$; 0.77/0.86, 0.89/0.82, 0.32/0.41 kJ.mol^{-1}, 0.96/1.35 kJ.mol^{-1} for $V_{W,1,3}$.

The new derived models from third-order polynomial were mathematically more accurate but they were more difficult to interpret because the large number of variables, up to 30. Assuming that Q_1Q_4 and Q_2Q_3 crossed terms were the most relevant for selectivity modeling, we attempted to generate new models with a simpler polynomial equation but without losing predictive ability. In addition as done above, we built the new models from the virtual datasets in order to account for the symmetry of the system and to sane the lack of varied data. The equations 5.10 and 5.11 show the derived models for ν(*Charton*) and $V_{W,0,3}$ descriptors. The statistical parameters from LOO cross-validation (q^2/r^2 = 0.94/0.94 and 0.88/0.88, respectively) are very close to those of the third-order polynomial including all the terms (q^2/r^2 = 0.93/0.96 and 0.88/0.93, respectively). Looking at the coefficients of the different variables, we observed that the single terms showed the same sign as the first-order model,

positive coefficients for quadrants Q_1 and Q_4 and negative for Q_2 and Q_3. This is consistent with our previous proposed stereochemical model (Figure 5.7). Interestingly, the coefficients of the corresponding crossed terms keep the sign; $Q_1Q_4 > 0$ and $Q_2Q_3 < 0$. This indicates that the steric effect of the substituents in diagonal positions act synergically within the obtained models. Figure 5.12 represents the additive values of the multiplication of the QSSR coefficients (C_i) of v(*Charton*) model by each of the variables ($\boldsymbol{Q_i}$ and $\boldsymbol{Q_iQ_j}$) for each catalyst. The height of coloured bars in the graph provides an estimation of the relative contribution of each variable to the enantioselectivity in each catalyst.

$$\Delta\Delta G^{\ddagger}(v(Charton) = 0.0 + 2.2\,\boldsymbol{Q_1} - 2.2\,\boldsymbol{Q_2} - 2.2\,\boldsymbol{Q_3} + 2.1\,\boldsymbol{Q_4} + 2.1\,\boldsymbol{Q_1Q_4} - 2.1\,\boldsymbol{Q_2Q_3} \quad (5.10)$$

$$\Delta\Delta G^{\ddagger}(V_{W,0,3}) = 0.0 + 32.6\,\boldsymbol{Q_1} - 32.6\,\boldsymbol{Q_2} - 32.6\,\boldsymbol{Q_3} + 32.6\,\boldsymbol{Q_4} + 3.0\,\boldsymbol{Q_1Q_4} - 3.0\,\boldsymbol{Q_2Q_3} \quad (5.11)$$

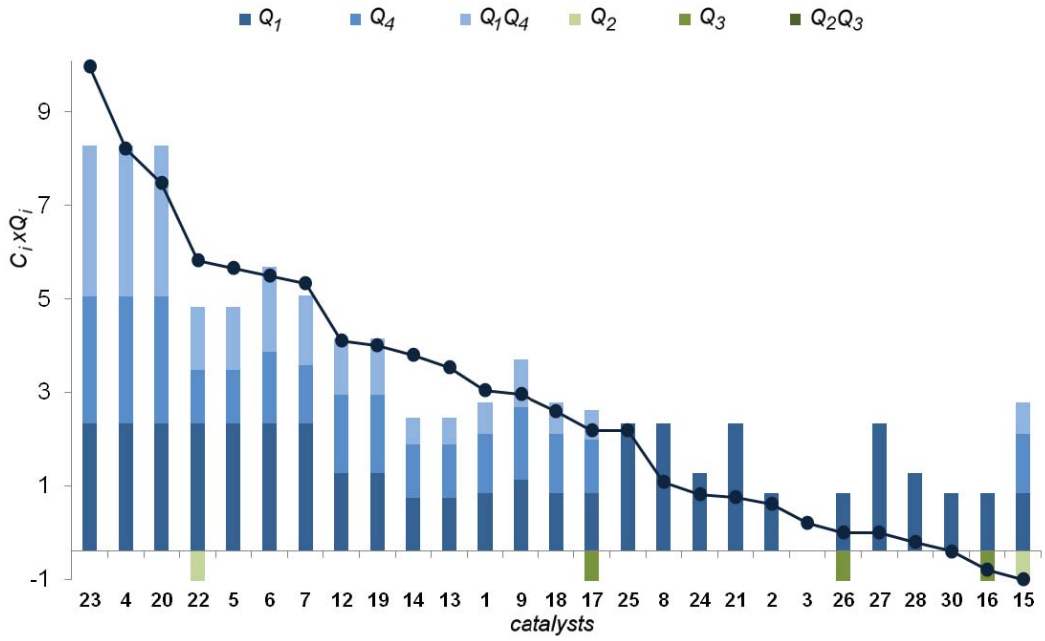


Figure 5.12. Product of the $\boldsymbol{Q_i}$ and $\boldsymbol{Q_iQ_j}$ quadrant variable by the PLS coefficients of the QSSR model following polynomial of equation 5.10 for each ligand as a measure of the contribution of different catalyst regions to enantioselectivity. The different contributions are separated by colours.

In summary, for Box and AzaBox ligand families, the inclusion of crossed terms between putatively independent variables complemented the information provided by the single terms,

leading to a better description the monosubstituted ligands as compared with their corresponding disubstituted examples. The improvement of the model is a consequence of catching non linear relationships and the interplay between the effects of the substituents; the synergistic relationship between diagonal positions (C_2-related) being the most important. Another important implication in catalyst design is that is necessary to install bulky groups in both R_1 and R_4 positions in order to obtain high enantioselectivities. Therefore, Quinox and Pyox ligand scaffolds will most likely no result in high enantioselectivities, because the need of a bulky substituent in one position and its C_2-*symmetric* position is not possible.

5.5. Conclusions

Two different quantitative structure selectivity relationships (QSSR) methodologies have been applied to the study of asymmetric cyclopropanation of styrene with diazo compounds by copper catalysts. Firstly, we have tried to generate a QSSR model with *Q-QSSR* approach, which uses alignment-independent three-dimensional descriptors derived from DFT calculations. Contrary to that found for asymmetric hydroformylation catalyzed by Rh-diphosphane (*Chapter 4*) we are unable to generate a model with predictive ability. We believe that this approach looses information about the shape of the catalyst during the GRIND alignment independence procedure; and consequently, it is not able to catch the subtle differences between catalysts that are responsible of variations in the selectivity outcome.

Secondly, the well-known quadrant-based diagram representation of the catalyst has been, for the first time, quantitatively correlated with the enantioselectivity using a new type of 3D-QSSR methodology. The derived models are able to predict the absolute configuration and the quantitative enantiomeric excess induced by the catalysts; thus proving that these quantitative relationships exit. The new type of 3D-QSSR approach consists of assigning a steric parameter to each of the spatial positions of the quadrant-diagram representation in order to form the matrix of the independent variable values for the multivariate analysis. To evaluate the steric hindrance, we have tested two types of steric parameters; the *Taft-Charton steric* and the *distance-weighted volume* ($V_{W,k,l}$). The *Taft-Charton steric* parameters are tabulated and do not require access to three-dimensional structures nor additional calculations. However, their use is restricted to a limited type of functional groups. The $V_{W,k,l}$ can be easily obtained from ground-state structures allowing quantifying any type of substituent. The new method predicts the enantiomeric excess showing good predictive abilities. The *Taft-Charton steric* parameters seem more robust than $V_{W,k,l}$ and give slightly better statistical values; however, they cannot be applied to the full set of synthesized ligands.

In the *first approximation*, we fit the experimental data to a first-order polynomial equation. This allows catching the main effects of the substituents and defining a simple stereochemical model to predict the outcome of new catalysts and to assist the design of *de novo* catalyst. The first-order approach predicts that installing very large groups at position R_1, R_4 or both, leading free positions R_2 and R_3, might be enough to obtain an enantioselective catalyst. To improve the model, we fitted the data to a third-order polynomial which relies on crossed terms between putatively independent variables. The new approximation improves appreciably the

predictive ability of the model. These results indicate that the relationship between enantioselectivity and steric size of substituents is not strictly linear, and that there is some interplay between substituent effects. By successively deleting terms from the full third-order polynomial model, we could identify synergistic effects between diagonal positions (C_2-related) as the most important interplay effects. Finally, we have provided a new methodology to quantitatively assess enantioselectivity from ground-state properties using the widely establish concept of representing a catalyst by a quadrant-diagram. The methodology is not limited to the use of steric descriptors, or to the model of enantioselectivity, as other types of descriptors (steric and electronic) and other catalytic variables (regioselectivity, chemselectivity and activity) could be also evaluated.

5.6. References and notes

1 T. Rasmussen, J. F. Jensen, N. Østergaard, D. Tanner, T. Ziegler, O.-O. Norrby, *Chem. Eur. J.* **2002**, *8*, 177.

2 G. Drudis-Solé, F. Maseras, A. Lledós, A. Vallribera, M. Moreno-Mañas, *Eur J. Inorg. Chem.* **2008**, 5614.

3 J. I. García, G. Jiménez-Osés, J. A. Mayoral, *Chem. Eur. J.* **2011**, *17*, 529.

4 a) A. G. Maldonado, J. A. Hageman, S. Mastroianni, G. Rothenberg *Adv. Synth. Catal.* **2009**, *351*, 387. b) N. Fey, J. N. Harvey, *Coord. Chem. Rev.* **2009**, *253*, 704. c) N. Fey, *Dalton Trans.* **2010**, *39*, 296.

5 a) K. B. Lipkowitz, M. Pradhan, *J. Org. Chem.* **2003**, *68*, 4648. b) M. C. Kozlowski, S. Dixon, M. Panda, G. Lauri, *J. Am. Chem. Soc.* **2003**, *125*, 6614. c) M. Hoogenraad, G. M. Klaus, N. Elders, S. M. Hooijschuur, B. McKay, A. A. Smith, E. W. P. Damen, *Tetrahedron: Asymmetry* **2004**, *15*, 519. d) J. L. Melville, K. R. J. Lovelock, C. Wilson, B. Allbutt, E. K. Burke, B. Lygo, J. D. Hirst, *J. Chem. Inf. Model.* **2005**, *45*, 971. e) S. Dixon, K. M. Jr. Merz, G. Lauri, J. C. Ianni, *J. Comp. Chem*. **2004**, *26*, 23. f) P.-W. Phuan, J. C. Ianni, M. C. Kozlowski, *J. Am. Chem. Soc.* **2004**, *126*, 15473. g) J. C. Ianni, V. Annamalai, P.-W. Phuan, M. Panda, M. C. Kozlowski, *Angew. Chem. Int. Ed.* **2006**, *45*, 5502. h) M. C. Kozlowski, J. C. Ianni, *J. Mol. Cat. A* **2010**, *324*, 141. i) S. E. Denmark; N. D. Gould, L. M. Wolf, *J. Org. Chem.* **2011**, *76*, 4337.

6 a) S. Sciabola, A. Alex, P. D. Higginson, J. C. Mitchell, M. J. Snowden, I. Morao, *J. Org. Chem.* **2005**, *70*, 9025. b) M. Urbano-Cuadrado, J. J. Carbó, A. G. Maldonado, C. Bo, *J. Chem. Inf. Model.* **2007**, *47*, 2228.

7 a) Jr. R. W. Taft, *J. Am. Chem. Soc*. **1952**, *74*, 2729. b) Jr. R. W. Taft, *J. Am. Chem. Soc*. **1953**, *75*, 4538. c) Jr. R. W. Taft, *Steric Eff. Org Chem.* **1956**, 556.

8 a) M. Charton, *J. Am Chem. Soc*. **1969**, *91*, 615. b) M. Charton, *J. Am Chem. Soc*, **1975**, *97*, 1552. c) M. Charton, *J. Am Chem. Soc*. **1975**, *97*, 3691. d) M. Charton, *J. Am Chem. Soc*. **1975**, *97*, 3694. e) M. Charton, *J. Am Chem. Soc*. **1975**, *97*, 6159. f) M. Charton, *J. Am Chem. Soc*, **1975**, *97*, 6472.

9 a) J. J. Miller, M. S. Sigman, *Angew. Chem., Int. Ed.* **2008**, *47*, 771. b) M. S. Sigman, J. J. Miller, *J. Org. Chem.* **2009**, *74*, 7633.

10 a) K. C. Harper, M. S. Sigman, *Proc. Natl. Acoad. Sci. U.S.A.* **2001**, *108*, 2179. b) K. C. Harper, M. S. Sigman, *Science* **2011**, *333*, 1875.

11 A. Poater, L. Cavallo, *Dalton Trans*. **2009**, 8878.

12 a) F. Ragone, A. Poater, L. Cavallo, *J. Am. Chem. Soc.* **2010**, *132*, 4249. b) A. Poater, F. Ragone, R. Mariz, R. Dorta, L. Cavallo, *Chem. Eur. J*. **2010**, *16*, 14348.

13 S. Aguado-Ullate, S. Saureu, L. Guasch, J. J. Carbó, *Chem. Eur. J.* **2012**, *18*, 995.

14 J. J. Carbó, A. Lledós, D. Vogt, C. Bo, *Chem. Eur. J.* **2006**, *12*, 1457.

15 P. W. N. M. van Leewen, *Homogeneous Catalysis*, Published by Kluwer Academic Publishers, **2004**.

16 a) A. K. Glosh, M. Packiarajan, J. Capiello, *Tetrahedron: Asymmetry* **1998**, *9*, 1. b) J. S. Johnson, D. A. Evans, *Acc. Chem. Res*. **2000**, *33*, 325. c) G. Desimoni, G. Faita, P. Quadrelli, *Chem. Rev*. **2003**, *103*, 3199. d) H. Lebel, J.-F. Marcoux, C. Molinaro, A. B. Charette, *Chem. Rev*. **2003**, *103*, 977.

17 H. Fristchi, U. Leuteneger, A. Ptaflz, *Helv. Chim. Acta* **1988**, *71*, 1553.

18 a) R. G. Salomon, J. K. Kochi, *J. Am. Chem. Soc.* **1973**, *95*, 3300. b) J. Green, E. Sinn, S. Woodward, R. Butcher, *Polyhedron* **1993**, *12*, 991. c) M. M. Díaz-Requejo, P. J. Pérez, M. Brookhart, J. L. Templeton, *Organometallics* **1997**, *16*, 4399. d) M. M. Díaz-Requejo, T. R. Belderrain, M. C. Nicasio, F. Prieto, P. J. Pérez, *Organometallics* **1999**, *18*, 2601. e) M. Bühl, F. Terstegen, F. Löffler, B. Meynhardt, S. Kierse, M. Müller, C. Näther, U. Lüning, *Eur. J. Org. Chem.* **2001**, 2151.

19 a) J. M. Fraile, J. I. García, V. Martínez-Merino, J. A. Mayoral, L. Salvatella, *J. Am. Chem. Soc.* **2001**, *123*, 7616. b) J. M. Fraile, J. I. García, M. J. Gil, V. Martínez-Merino, J. A. Mayoral, L. Salvatella, *Chem. Eur. J.* **2004**, *10*, 758. c) J. I. García, G. Jiménez-Osés, V. Martínez-Merino, J. A. Mayoral, E. Pires, I. Villalba, *Chem. Eur. J.* **2007**, *13*, 4064.

20 B. F. Straub, I. Gruber, F. Rominger, P. Hofmann, *J. Organomet. Chem.* **2003**, *684*, 124.

21 B. F. Straub, P. Hofmann, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 1288.

22 a) J. M. Fraile, J. I. García, A. Gissibl, J. A. Mayoral, E. Pires, O. Reiser, M. Roldán, I. Villalba, *Chem. Eur. J.* **2007**, *13*, 8830. b) J. I. García, B. López-Sánchez, J. A. Mayoral, E. Pires, I. Villalba, *J. Catal.* **2008**, *258*, 378. c) J. I. García, G. Jiménez-Osés, B. López-Sánchez, J. A. Mayoral, A. Vélez, *Dalton Trans.* **2010**, *39*, 2098.

23 a) R. G. Parr, W. Yang, *Density Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, UK, **1989**. b) T. Ziegler, *Chem. Rev.* **1991**, *91*, 651.

24 a) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785. b) A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648. c) P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.

25 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M.A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Jr. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Revision A.02; Gaussian, Inc.: Wallingford, CT, **2003**.

26 a) M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. Defrees, J. A. Pople, *J. Chem. Phys.* **1982,** 77, 3654. b) W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972,** *56*, 2257. c) P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973,** *28*, 213.

27 a) S. Wold, M. Sjostrom, L. Eriksson, *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109. b) P. Geladi, B. Kowlaski, *Anal. Chim. Acta* **1985**, *35*, 1.

28 D. M. Hawkins, S. C. Basak, D. Mills, *J. Chem. Inf. Comput. Sci.* **2003**, *23*, 579.

Chapter 6

Concluding remarks

Chapter 6

Concluding remarks

DFT and hybrid DFT/MM methods permit the characterization of the reaction mechanism, with a view to identifying the activity and selectivity-determining step and analyzing the factors governing the catalytic outcome. Although TS-based approaches using DFT and DFT/MM methods have demonstrated their accuracy, even of stereochemical predictions, we have identified some limitations. For asymmetric hydroformylation of styrene by Rh-diphosphanes, one limitation is that describing nonbonding interactions with DFT and MM methods is troublesome. Moreover, for Rh-binaphos system, we observed that several steps have to be considered simultaneously to account to enantioselectivity. For asymmetric cyclopropanation alkenes, we realized that the *Potential Energy Surface* (PES) can be very flat in the area surrounding the TS, making its location very difficult. Thus, as an alternative to TS-based approach, we have shown that the QSPR-type modeling can be used to predict and rationalize the outstanding asymmetric organometallic catalyzed processes without prior experimentation synthesis or time-consuming DFT calculations all reaction paths. For the studied systems, it was possible to derive a parameter from catalyst structure that correlates with the activity or enantioselectivity. However, it seems that each process requires its specific descriptors, and that the same QSPR methodologies cannot describe any property and any catalytic system. Thus, we believe that much effort needs to be done in order to find "universal" descriptors that can be transferred to different catalytic process with special emphasis in asymmetric catalysis. Besides the methodological knowledge, this Thesis provides insight into the chemistry of some outstanding catalytic processes. The specific conclusions are divide in blocks, each of which is related with the different studied chemical processes: ammonia activation, asymmetric hydroformylation of styrene by Rh-diphosphane, and asymmetric cyclopropanation of alkenes by Cu complexes.

With respect to the DFT study on ammonia activation by μ_3-alkylidyne fragments supported on a titanium molecular oxide

- The DFT studies on the N-H bonds activation of ammonia by the titanium μ_3-alkylidyne complexes [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-CR)] show that upon successive N-H activation the intermediate species amido-, imido- and nitrido-titanium go downhill in energy. The analysis of their electronic structure indicates a multiple bond character of the terminal amido and the bridging nature of imido ligand in the reaction intermediates.

- The mechanism can be divided in three stages: i) coordination of the ammonia to a titanium centre located *trans* with respect to the alkylidyne group, and then, the

isomerization to adopt the *cis* arrangement, allowing the direct hydron migration to the μ_3-alkylidyne group to yield the amido μ-alkylidene complexes; ii) hydron migration from the amido moiety to the alkylidene group, and finally; iii) hydron migration from the μ-imido complex to the alkyl group to afford the oxo μ_3-nitrido titanium complex with alkane elimination. The N-H bond activation of ammonia by terminal alkylidyne-titanium complexes had been predicted theoretically, but never reported until now.

With respect to the DFT and QSPR studies on asymmetric hydroformylation by Rh-diphosphanes

- The conformational search protocol based on molecular dynamics calculations can be used to explore the conformational space and account for the possible isomers in Rh-diphosphane complexes.

- We have performed a fundamental analysis of the coordination preferences of phosphine-phosphite (*R*,*S*)-binaphos ligand in [HRh(P-P)$(CO)_2$] complex, which acts as catalyst for asymmetric hydroformylation. The use of different molecular regions in DFT/MM methods and of a modified version of EDA procedure based orbital deletion can be used to identify and quantify the different effects governing the coordination preferences of binaphos ligand. The evaluation of donor-acceptor interactions favours placing the least basic phosphite moiety in the equatorial position. However, the trend is inversed as a consequence of the nonbonding steric effects induced by the bulky naphthyl and phenyl groups of binaphos, and in major extends, of the larger *electronic distortion* induced by the better π acceptor phosphite moiety in the equatorial site.

- For binaphos system, we propose a sterochemical model based on quantitative quadrant-diagram representation obtained from a new molecular descriptor, the *distance-weighted volume,* V_W, which is easily computed from ground-state structures. Repulsive interactions between the substrate and the apical phosphite are responsible for the enantiodifferentiation. The axial chirality of the phosphite discriminates one of the competitive equatorial-apical paths, whereas the axial chirality of the backbone discriminates one of the two enantiomers.

- Quantitative Structure-Selectivity and -Activity relationships for a varied dataset of 21 rhodium-diphosphane catalysts on asymmetric hydroformylation of styrene have been generated using a 2D-QSPR and 3D-QSPR methodologies, which include the TMACC approach and the CoMFA-like alignment-independent (GRIND) protocol based on 3D molecular descriptors derived from DFT calculations (*Q-QSSR*).

- The most predictive model for selectivity is obtained using shape 3D fields (MSF) based on the local curvature electron density isosurface (r^2 = 0.92, q^2 = 0.68). The most predictive model for activity is obtained using a combination of shape- and electrostatic-based 3D fields (MEP, r^2 = 0.99, q^2 = 0.74).

- The worst predicted ligands are the kelliphite and chiraphite, which shows the lowest preference for equatorial-apical coordination and low enantioselectivity, suggesting that their intrinsic enantiotopic differentiation capacity can be lost through the occurrence of bis-equatorial paths.

- The stereochemical model previously proposed for Rh-binaphos catalyst can be also applied to explain the high enantioinduction observed for Rh-chiraphite, -binaphine, -diazaphospholane, and -yanphos systems because we observed the same pattern in the analysis of QSPR model via *coefficient-weighted correlograms*.

- Ligands with electron withdrawing groups at phosphorus atoms such as chiraphite, kelliphite, binaphos, and diazophospholane reduce ligand basicity and effectively promote catalytic activity. However, more complex relationships underlie the origin of activity, and the shape of the catalyst needs also to be considered such as for Ph-BPE ligand. By comparison with previous studies, we propose that reduction of the steric hindrance at the reaction centre favouring alkene coordination and insertion would also promote catalytic activity.

- The 2D-QSPR-type TMACC model has predictive ability for activity when the all atomic properties are considered. The interpretation of TMACC results confirms the importance of the basicity and steric hindrance in the catalyst activity.

With respect to the QSSR studies on asymmetric cyclopropanation of alkenes

- The *Q-QSSR* approach fails in predicting the enantioselectivity for asymmetric cyclopropanation of alkenes by oxazoline-based Cu complexes. Probably, this approach looses information about the shape of the catalyst during the GRIND alignment independence procedure.

- The well-known quadrant-diagram representation of the catalysts has been, for the first time, correlated with enantioselectivity using a new type of 3D-QSSR methodology.

- Both descriptors the *Taft-Charton steric* parameter and the *distance-weighted volume* can quantify the steric hindrance of the quadrants in this representation.

- The enantioselectivity data fits to first-order polynomial equations that catches the main effects of the substituents and define a simple stereochemical model, which predicts that installing bulky groups at alternative corner-sharing quadrants would increase enantioselectivity. Fitting the data to a third-order polynomial relaying on crossed terms between quadrants, the correlation improves somewhat. This indicates that the relationship between enantioselectivity and steric size is not strictly linear and there is some interplay between substituents effects; the synergistic effects between diagonal substituents being the most important.

List of publications

S. Aguado-Ullate, J. J. Carbó, O. González-del Moral, A. Martín, M. Mena, J. M. Poblet, C. Santamaría.

"Ammonia Activation by μ_3-Alkylidyne Fragments Supported on a Titanium Molecular Oxide Model"

Inorg. Chem. **2011**, *50*, 6269-6279.

S. Aguado-Ullate, J. J. Carbó, O. González-del Moral, M. Gómez-Pantoja, A. Hernán-Gómez, A. Martín, M. Mena, J. M. Poblet, C. Santamaría.

"Discovering the chemical reactivity of the molecular oxonitride [{Ti(η^5-C_5Me_5)(μ-O)}$_3$(μ_3-N)]"

J. Organomet. Chem. **2011**, *696*, 4011-4017.

S. Aguado-Ullate, S. Saureu, L. Guasch, J. J. Carbó.

"Theoretical studies on asymmetric hydroformylation by Rh-(*R*,*S*)-BINAPHOS catalyst. Origin of coordination preferences and stereoinduction."

Chem. Eur. J. **2012**, *18*, 995-1005.

S. Aguado-Ullate, L. Guasch, M. Urbano-Cuadrado, C. Bo, J. J. Carbó.

"3D-QSPR models for Predicting the Enantioselectivity and the Activity in Asymmetric Hydroformylation of Styrene Catalyzed by Rh-diphosphane."

(*to be submitted*).

S. Aguado-Ullate, M. Urbano-Cuadrado, J. I. García, C. Bo, J. J. Carbó.

"Predicting the Enantioselectivity of Copper-catalyzed Cyclopropantion of alkenes using Quantitative Quadrant-Diagram Representation of the Catalysts."

(*submitted*).

S. Aguado-Ullate, V. González-González, C. Muller, J. J. Carbó.

"Theoretical studies on hydroformylation of alkenes by Rh-Phosphabenzene catalysts. The origin of enhanced activity of π acceptor ligands."

(*in preparation*).

Annex

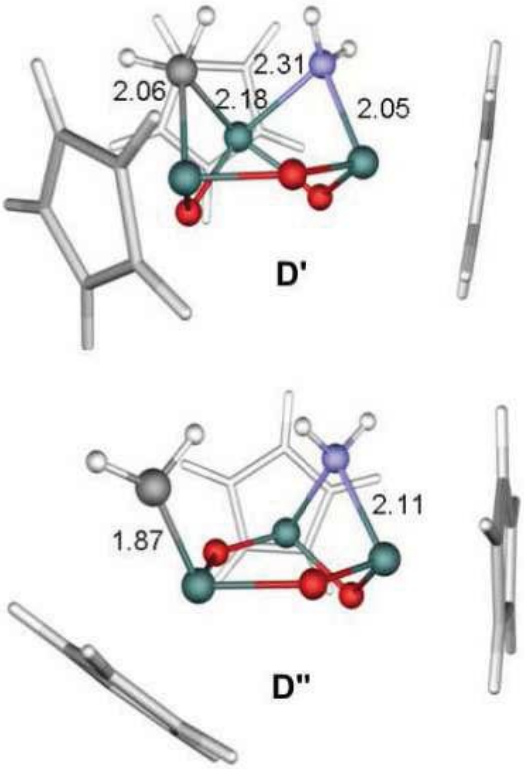


Figure A2-1. Computed molecular structures (Å) of the amido-alkylidene intermediates in the stepwise mechanism of the second N-H bond activation **D'** and **D''** (*Chapter2*).

Table A4-1. Relative energies (in kJ.mol^{-1}) of **ee1**, **ee2**, **ea** isomers for complexes [RhH(CO)$_2$(L)]. The structures derived from the conformational search protocol for the catalysts dataset used in *Chapter 4*.

Ligand	Isomers	Relative energy	Bite angle (P-Rh-P)	Ligand	Isomers	Relative energy	Bite angle (P-Rh-P)
1	**ee1**	16,8	113,3	**11**	**ee1**	16,4	98,5
	ee2	14,7	104		**ee2**	13,9	120,1
	ea	0,0	99,3		**ea**	0,0	121,8
2	**ee1**	24,1	86,9	**12**	**ee1**	13,9	94,3
	ee2	35,5	88,1		**ee2**	5,9	95,5
	ea	0,0	84,3		**ea**	0,0	92,5
3	**ee1**	19,3	88,8	**13**	**ee1**	10,1	87,2
	ee2	19,5	88,8		**ee2**	10,0	87,13
	ea	0,0	87,1		**ea**	0,0	85,5
4	**ee1**	16,6	102,1	**14**	**ee1**	22,7	87,8
	ee2	21,4	101,2		**ee2**	23,7	88,27
	ea	0,0	94,6		**ea**	0,0	86,1
5	**ee1**	0,0	94,4	**15**	**ee1**	18,1	86,6
	ee2	18,0	99,1		**ee2**	19,6	86,8
	ea	15,7	101,6		**ea**	0,0	85,6
	ea2	0,0	94,7	**16**	**ee1**	11,4	87,3
6	**ee1**	14,9	91,2		**ee2**	11,5	87,2
	ee2	14,0	89,0		**ea**	0,0	85,5
	ea	0,0	89,5	**17**	**ee1**	13,5	86,2
7	**ee1**	74,2	86,3		**ee2**	13,5	86,2
	ee2	67,1	86,9		**ea**	0,0	84,5
	ea	0,0	85,6	**18**	**ee1**	46,2	86,8
8	**ee1**	33,9	105,8		**ee2**	41,1	87,6
	ee2	27,1	106,5		**ea**	0,0	86,2
	ea	0,0	98,8	**19**	**ee1**	14,2	87,7
9	**ee1**	16,7	87,8		**ee2**	14,2	87,7
	ee2	16,1	87,7		**ea**	0,0	86,1
	ea	0,0	86,7				
10	**ee1**	24,0	90,2				
	ee2	27,0	90,3				
	ea	0,0	85,7				

Table A5-1.Values of $V_{W,0,3}$ parameters for the ligands in **set 2** defined in *Chapter 5*.

Ligand	Q_1	Q_2	Q_3	Q_4
1	0.06	0.01	0.01	0.06
2	0.09	0.00	0.02	0.00
3	0.00	0.00	0.00	0.00
4	0.12	0.00	0.00	0.12
5	0.11	0.00	0.00	0.06
6	0.11	0.01	0.00	0.07
7	0.10	0.02	0.00	0.11
8	0.11	0.00	0.00	0.00
9	0.11	0.01	0.02	0.07
10	0.07	0.00	0.00	0.07
11	0.07	0.00	0.00	0.00
12	0.08	0.00	0.01	0.09
13	0.04	0.00	0.00	0.04
14	0.03	0.00	0.00	0.03
15	0.06	0.04	0.04	0.06
16	0.07	0.00	0.04	0.00
17	0.09	0.00	0.04	0.03
18	0.06	0.01	0.01	0.06
19	0.05	0.00	0.00	0.07
20	0.11	0.00	0.00	0.11
21	0.10	0.00	0.00	0.00
22	0.11	0.03	0.00	0.03
23	0.12	0.00	0.00	0.12
24	0.09	0.00	0.02	0.03
25	0.11	0.03	0.00	0.00
26	0.09	0.00	0.04	0.03
27	0.08	0.00	0.00	0.00
28	0.07	0.00	0.00	0.00
29	0.03	0.00	0.00	0.00
30	0.03	0.00	0.01	0.00

Table A5-2. Values of $V_{W,1,3}$ parameters for the ligands in **set 2** defined in *Chapter 5*.

Ligand	Q_1	Q_2	Q_3	Q_4
1	0.09	0.02	0.02	0.09
2	0.12	0.00	0.02	0.00
3	0.00	0.00	0.00	0.00
4	0.14	0.00	0.00	0.13
5	0.12	0.00	0.00	0.08
6	0.12	0.01	0.00	0.07
7	0.12	0.02	0.00	0.15
8	0.12	0.00	0.00	0.00
9	0.15	0.01	0.03	0.08
10	0.09	0.00	0.00	0.09
11	0.08	0.00	0.00	0.00
12	0.09	0.00	0.01	0.10
13	0.04	0.00	0.00	0.04
14	0.03	0.00	0.00	0.03
15	0.09	0.04	0.04	0.09
16	0.09	0.00	0.05	0.00
17	0.13	0.00	0.05	0.03
18	0.08	0.01	0.01	0.08
19	0.06	0.00	0.00	0.08
20	0.13	0.00	0.00	0.13
21	0.12	0.00	0.00	0.00
22	0.12	0.03	0.00	0.03
23	0.14	0.00	0.00	0.13
24	0.13	0.00	0.02	0.03
25	0.12	0.03	0.00	0.00
26	0.13	0.00	0.05	0.03
27	0.10	0.00	0.00	0.00
28	0.08	0.00	0.00	0.00
29	0.04	0.00	0.00	0.00
30	0.04	0.00	0.01	0.00

Table A5-3. Statistical parameters of LOO cross-validation for $\Delta\Delta G^{\ddagger}$ models using different type of $V_{W,k,I}$ descriptors defined in *Chapter 5*.[(a)]

Descriptor	q^2	r^2	Matrix dimension
$V_{W,0,1}$	0.11	0.29	46x4
$V_{W,0,2}$	0.55	0.65	46x4
$V_{W,0,3}$	0.74	0.80	46x4
$V_{W,1,1}$	0.42	0.57	46x4
$V_{W,1,2}$	0.41	0.63	46x4
$V_{W,1,3}$	0.69	0.75	46x4

[(a)] Slope (fitting/prediction), intercept (fitting/prediction), error (fitting/prediction): 0.29/0.20, 2.89/3.35 kJ.mol^{-1}, 2.37/2.64 kJ.mol^{-1} for $V_{W,0,1}$; 0.66/0.63, 1.38/1.63 kJ.mol^{-1}, 1.67 /1.88 kJ.mol^{-1} for $V_{W,0,2}$; 0.80/0.78, 0.80/0.86 kJ.mol^{-1}, 1.27/1.43 kJ.mol^{-1} for $V_{W,0,3}$; 0.57/0.51, 1.60/2.03 kJ.mol^{-1}, 1.85/2.14 kJ.mol^{-1} for $V_{W,1,1}$; 0.61/0.52, 1.39/2.03 kJ.mol^{-1}, 1.72/2.15 kJ.mol^{-1} for $V_{W,1,2}$; 0.75/0.73, 0.98/1.08 kJ.mol^{-1}, 1.40/1.57 kJ.mol^{-1} for $V_{W,1,2}$.

Table A5-4. Equation coefficients of the Taft-Charton based models using different third-order polynomials and the non virtual datasets (**set 1** and **set 2**). The third column included crossed terms of diagonal quadrants, and the fourth column included crossed terms of the non diagonal quadrants.

Coefficients	all	$Q_1^n Q_4^m$, $Q_2^n Q_3^m$	$Q_1^n Q_2^m$, $Q_1^n Q_3^m$ $Q_2^n Q_4^m$, $Q_3^n Q_4^m$
C_0	0.0	0.0	0.0
Q_1	2.2	2.2	0.8
Q_2	-2.2	-2.2	-0.8
Q_3	-2.2	-2.2	-0.8
Q_4	2.2	2.2	0.8
Q_1^2	0.4	0.4	0.9
Q_2^2	-0.4	-0.4	-0.9
Q_3^2	-0.4	-0.4	-0.9
Q_4^2	0.4	0.4	0.9
Q_1Q_2	0.0	-	0.0
Q_1Q_3	0.0	-	0.0
Q_1Q_4	2.2	2.2	-
Q_2Q_3	-2.2	-2.2	-
Q_2Q_4	0.0	-	0.0
Q_3Q_4	0.0	-	0.0
$Q_1^2Q_2$	0.3	-	0.0
$Q_1^2Q_3$	0.0	-	0.0
$Q_1^2Q_4$	0.7	0.7	-
$Q_2^2Q_1$	-0.3	-	0.0
$Q_2^2Q_3$	-0.7	-0.7	-
$Q_2^2Q_4$	0.0	-	0.0
$Q_3^2Q_1$	0.0	-	0.0
$Q_3^2Q_2$	-0.7	-0.7	-
$Q_3^2Q_4$	-0.3	-	0.0
$Q_4^2Q_1$	0.7	0.7	-
$Q_4^2Q_2$	0.0	-	0.0
$Q_4^2Q_3$	0.3	-	0.0
Q_1^3	0.9	0.9	1.0
Q_2^3	-0.9	-0.9	-1.0
Q_3^3	-0.9	-0.9	-1.0
Q_4^3	0.9	0.9	1.0

Table A75-5 Equation coefficients of the $V_{W,0,3}$ based models using different third-order polynomials and the non virtual datasets (**set 1** and **set 2**). The third column included crossed terms of diagonal quadrants, and the fourth column included crossed terms of the non diagonal quadrants.

Coefficients	all	$Q_1^nQ_4^m$, $Q_2^nQ_3^m$	$Q_1^nQ_2^m$, $Q_1^nQ_3^m$ $Q_2^nQ_4^m$, $Q_3^nQ_4^m$
a_0	0.0	0.0	0.0
Q_1	33.1	33.0	32.3
Q_2	-33.1	-33.0	-32.3
Q_3	-33.1	-33.0	-32.3
Q_4	33.1	33.0	32.3
Q_1^2	-157.5	-156.6	3.3
Q_2^2	157.5	156.6	-3.3
Q_3^2	157.5	156.6	-3.3
Q_4^2	-157.5	-156.6	3.3
Q_1Q_2	0.0	-	0.0
Q_1Q_3	0.0	-	0.0
Q_1Q_4	360.1	359.4	-
Q_2Q_3	-360.1	-359.4	-
Q_2Q_4	0.0	-	0.0
Q_3Q_4	0.0	-	0.0
$Q_1^2Q_2$	2.778	-	0.0
$Q_1^2Q_3$	-17.9	-	0.0
$Q_1^2Q_4$	35.8	35.7	-
$Q_2^2Q_1$	-2.8	-	0.0
$Q_2^2Q_3$	-35.8	-35.7	-
$Q_2^2Q_4$	17.9	-	0.0
$Q_3^2Q_1$	17.9	-	0.0
$Q_3^2Q_2$	-35.8	-	0.0
$Q_3^2Q_4$	-2.8	-	0.0
$Q_4^2Q_1$	35.8	35.7	-
$Q_4^2Q_2$	-17.9	-	0.0
$Q_4^2Q_3$	2.8	-	0.0
Q_1^3	-8.1	-8.1	0.3
Q_2^3	8.1	8.1	-0.3
Q_3^3	8.1	8.1	-0.3
Q_4^3	-8.1	-8.1	0.3

Table A5-6. Equation coefficients of the $V_{W,1,3}$ based models using different third-order polynomials and the non virtual datasets (**set 1** and **set 2**). The third column included crossed terms of diagonal quadrants, and the fourth column included crossed terms of the non diagonal quadrants.

Coefficients	all	$Q_1^nQ_4^m$, $Q_2^nQ_3^m$	$Q_1^nQ_2^m$, $Q_1^nQ_3^m$ $Q_2^nQ_4^m$, $Q_3^nQ_4^m$
a_0	0.0	0.0	0.0
Q_1	26.1	26.1	26.3
Q_2	-26.1	-26.1	-26.3
Q_3	-26.1	-26.1	-26.3
Q_4	26.1	26.1	26.3
Q_1^2	3.1	3.1	3.1
Q_2^2	-3.1	-3.1	-3.1
Q_3^2	-3.1	-3.1	-3.1
Q_4^2	3.1	3.1	3.1
Q_1Q_2	0.0	-	0.0
Q_1Q_3	0.0	-	0.0
Q_1Q_4	2.8	2.8	-
Q_2Q_3	-2.8	-2.8	-
Q_2Q_4	0.0	-	0.0
Q_3Q_4	0.0	-	0.0
$Q_1^2Q_2$	0.0	-	0.0
$Q_1^2Q_3$	0.0	-	0.0
$Q_1^2Q_4$	0.3	0.3	-
$Q_2^2Q_1$	0.0	-	0.0
$Q_2^2Q_3$	-0.3	-0.3	-
$Q_2^2Q_4$	0.0	-	0.0
$Q_3^2Q_1$	0.0	-	0.0
$Q_3^2Q_2$	-0.3	-0.3	-
$Q_3^2Q_4$	0.0	-	0.0
$Q_4^2Q_1$	0.3	0.3	-
$Q_4^2Q_2$	0.0	-	0.0
$Q_4^2Q_3$	0.0	-	0.0
Q_1^3	0.4	0.4	0.4
Q_2^3	-0.4	-0.4	-0.4
Q_3^3	-0.4	-0.4	-0.4
Q_4^3	0.4	0.4	0.4