



Durand, D., & Fey, N. (2019). Computational Ligand Descriptors for Catalyst Design. *Chemical Reviews*, 119(11), 6561-6594.
<https://doi.org/10.1021/acs.chemrev.8b00588>

Peer reviewed version

License (if available):
Other

Link to published version (if available):
[10.1021/acs.chemrev.8b00588](https://doi.org/10.1021/acs.chemrev.8b00588)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via ACS Publications at <https://doi.org/10.1021/acs.chemrev.8b00588> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/pure/user-guides/explore-bristol-research/ebr-terms/>

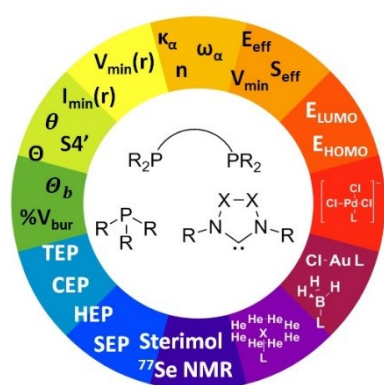
Title: Computational Ligand Descriptors for the Design of Catalyst involving Phosphines and Carbenes

Authors: Derek J. Durand and Natalie Fey*

*Corresponding author, email: Natalie.Fey@Bristol.ac.uk

School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

Suggested Table of Contents Graphic:



Abstract

Ligands, especially phosphines and carbenes, can play a key role in modifying and controlling homogeneous organometallic catalysts, and they often provide a convenient approach to fine-tuning the performance of known catalysts. The measurable outcomes of such catalyst modifications (yields, rates, selectivity) can be set into context by establishing their relationship to steric and electronic descriptors of ligand properties, and such models can guide the discovery, optimisation and design of catalysts.

In this review we present a survey of calculated ligand descriptors, with a particular focus on homogeneous organometallic catalysis. A range of different approaches to calculating steric and electronic parameters are set out and compared, and we have collected descriptors for a range of representative ligand sets, including 30 monodentate phosphorus(III) donor ligands, 23 bidentate P,P-donor ligands and 30 carbenes, with a view to providing a useful resource for analysis to practitioners. In addition, several case studies of applications of such descriptors, covering both maps and models, have been reviewed, illustrating how descriptor-led studies of catalysis can inform experiments and highlighting good practice for model comparison and evaluation.

Contents

Abstract.....	1
1. Introduction	3
2. Ligand Descriptors.....	7
2.1 Monodentate Phosphorus(III)-Donor Ligands	7
2.1.1 Individual Descriptors	7
2.1.1.1 Electronic Descriptors	8
2.1.1.2 Steric Descriptors	13
2.1.2 Descriptor databases	17
2.1.2.1 Molecular Electrostatic Potentials	17
2.1.2.2 Ligand Knowledge Bases	20
2.1.2.3 Reaction-Specific Descriptors	23
2.2 Bidentate P-donor Ligands.....	24
2.2.1 Individual Descriptors	24
2.2.2 Descriptor Databases	28
2.2.2.1 Ligand Knowledge Bases	28
2.2.2.2 Reaction-Specific Descriptors	32
2.3 Carbenes	32
2.3.1 Individual Descriptors	32
2.3.1.1 Electronic Descriptors	32
2.3.1.2 Steric Descriptors	38
2.3.2 Descriptor databases	39
2.3.2.1 Molecular Electrostatic Potential Data	39
2.3.2.2 Ligand Knowledge Base (LKB-C)	41
2.3.2.3 Electronic Structure Trends	44
2.4 Other Ligands	45
3. Case Studies	47
3.1 Mapping Chemical Space	47
3.2 Analysis of Catalyst Performance.....	50
3.2.1 Single Class of Ligand	51
3.2.2 Comparison of different ligand classes	56
4. Summary and Outlook	60
Acknowledgements.....	62
Supporting Information	62
Author Biographies	62
References	62

1. Introduction

The development of homogeneous (organometallic) catalysts has turned a significant corner in the last decade: not only does the field now make frequent use of computational mechanistic studies to confirm hypotheses about likely reaction pathways,¹⁻⁸ but researchers have also embraced data-led approaches, combining large-scale experimentation with suitable descriptors to fit statistical models, for the discovery, optimisation, and indeed design of catalysts for a broad range of reactions.⁹⁻¹⁵

This should not come as a surprise, as it builds on a long tradition of using stereoelectronic parameters, Tolman's perhaps most prominently among them,¹⁶ in this field.¹⁷ Homogeneous catalysis is not (yet) data-rich enough to be considered amenable to "Big Data" approaches,¹⁸⁻²¹ and machine-learning approaches, while used in this area,^{9, 22-25} are still very much in their infancy,²⁶ but this provides a convenient opportunity to survey and collate available descriptors.

The role of computational mechanistic studies in the design and optimisation of catalysts will be reviewed in detail elsewhere (Mu-Hyun Baik *et al.*, this issue), and here we have focussed on providing a survey of calculated descriptors (the term "parameters" tends to be used interchangeably in the present review^a) focussed on the ancillary ligands used to fine-tune the properties of transition metal centres known to provide active catalysts. In addition, we have favoured relatively large-scale studies of ligand effects on catalyst properties and performance, as well as those with explicit experimental applications that seek insights leading to reaction optimisation and catalyst design. Our main focus is on two "blockbuster" classes of ligands, the family of mono- and bidentate phosphines and related phosphorus(III)-donor ligands, as well as carbenes; a brief list of descriptor-based approaches to other ligands has been included in section 2.4. Many aspects of this field have been reviewed previously and each section highlights relevant reviews and seeks to avoid excessive overlap; in practice, that means studies considered here have been published in the last decade, with most falling within the last 5-6 years. Where we have compiled descriptor data, we have cast a wider net, both in terms of publication date and by including experimentally-determined descriptors where these are standard references in the field.

Despite the title of this review, the term "catalyst design" appears to mean many things to many people. Identified by Houk and Liu as one of the "holy grails" of computational organic chemistry and biochemistry,²⁸ perhaps the best-case scenario will see an entirely computational process leading up to a single experiment. This would most likely combine mechanistic study with data analysis and predictions from some form of regression model, fitted by humans or machine-learned, which leads to a reliable suggestion of a catalyst fitting a user-defined set of selection criteria (such as cost, activity, selectivity, availability or toxicity). While there have been considerable successes for computational prediction in (organometallic) homogeneous catalysis,¹⁻³ they have usually relied on extensive ground work by experimental and computational chemists, as well as benefitting substantially from a productive dialogue between experts.¹² Timescales for computational analysis and prediction will continue to be shortened by growing computational power, comprehensive descriptor databases and better approaches to the location and analysis of feasible reaction pathways,^{29, 30} albeit at computational cost, but it seems prudent to acknowledge that catalyst discovery, optimisation and

^a The organometallic literature tends towards using the term "parameter", whereas (molecular) "descriptors" are more commonly used in the context of chemical data analysis. As Livingstone puts it: "All parameters are thus descriptors, but not vice versa. (27) Livingstone, D. *A Practical Guide to Scientific Data Analysis*; John Wiley & Sons Ltd.: Chichester, West Sussex, PO19 8SQ, UK, 2009.)

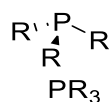
design are very much convergent fields of research, and here we have included these approaches as key application areas for ligand descriptors.

While we readily acknowledge that computational studies of catalysts can be affected by methodological issues around the level of theory, especially deciding which density functional, as well as basis set, solvation, dispersion and free energy effects are likely to give reliably good agreement with available experimental data,^{3, 31-34} ligand descriptors are designed to capture trends in ligand properties, such that method choices matter less,²⁰ as long as the approach is computationally consistent and robust.³⁵ Subtle variations in electronic effects are much more likely to require calculations at a higher level of theory and on complexes specific to the catalytic process of interest, than a broad and transferable comparison of catalyst properties. Throughout this review we will thus note the level of theory used, but not discuss this aspect in detail.

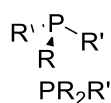
With a view to facilitating data analysis and the sharing of workflows and statistical models for catalyst optimisation, discovery and design, we have compiled published data, from both experimental and computational studies, for a representative subset of our “blockbuster” classes of ligands, *i.e.* phosphorus(III) donor and carbene ligands. These ligands are summarised in Table 1. They were selected to reflect experimental utilisation and commercial availability of ligands, (our) access to ligand descriptors, as well as applications to catalysis, and they are by no means complete; where additional ligands have been computationally characterised, we have mentioned this throughout the review. We also note that there are some inconsistencies in ligand naming across the published literature and suggest strongly that the molecular structures listed here are used as the main approach to determining what is meant, rather than relying overly on names/abbreviations.

Table 1: Representative ligand sets for which experimental and calculated descriptors have been compiled.

a) Monodentate P(III)-Donor Ligands

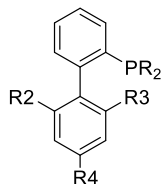


Ligand	R	Ligand	R
P1	H	P11	OMe
P2	Me	P12	OEt
P3	Et	P13	OPh
P4	iPr	P14	C ₆ F ₅
P5	nBu	P15	F
P6	tBu	P16	o-tol
P7	Cy	P17	p-tol
P8	Ph	P18	p-F-C ₆ H ₄
P9	Bn (CH ₂ Ph)	P19	p-Cl-C ₆ H ₄
P10	NMe ₂	P20	p-OMe-C ₆ H ₄



Ligand	R	R'
P21	Me	Ph

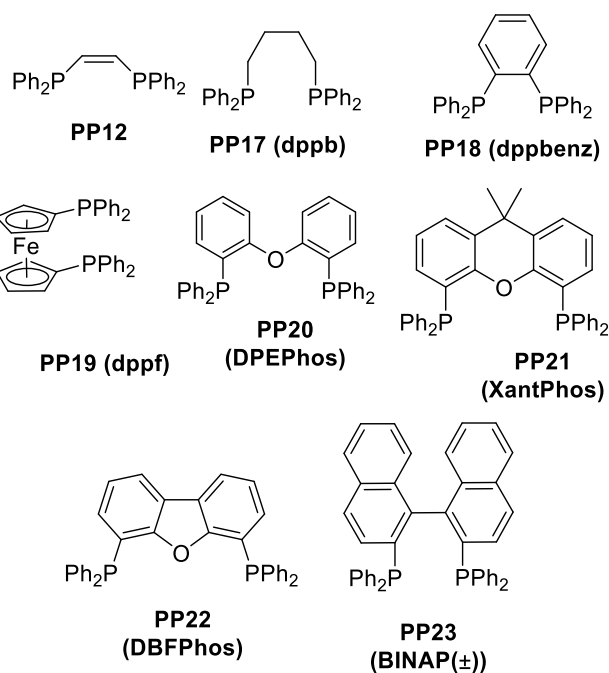
P22	Ph	Me
P23	Et	Ph
P24	Ph	Et
P25	Ph	OMe



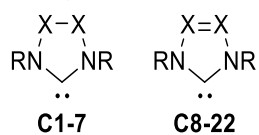
Ligand	Name	R	R2	R3	R4
P26	JohnPhos	tBu	H	H	H
P27	CyJohnPhos	Cy	H	H	H
P28	MePhos	Cy	CH ₃	H	H
P29	SPhos	Cy	OMe	OMe	H
P30	XPhos	Cy	iPr	iPr	iPr

b) Bidentate P,P-Donor Ligands

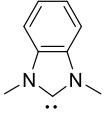
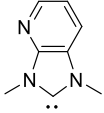
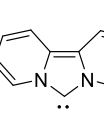
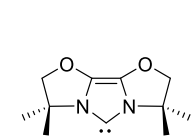
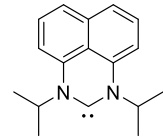
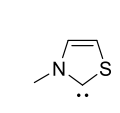
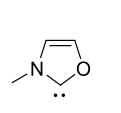
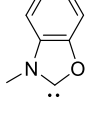
Ligand	R
PP1	Ph
PP2	Me
PP3	tBu
PP4	F
Ligand	R
PP5	H
PP6	Me
Ligand	R
PP7	Me
PP8	Ph
PP9	tBu
PP10	OMe
PP11	Cy
Ligand	R
PP13	Me
PP14	Ph
PP15	F
PP16	tBu



c) Carbenes



Ligand	Acronym	R	X
C1	-	H	CH ₂
C2	SIMe	Me	CH ₂
C3	S <i>i</i> Pr	<i>i</i> Pr	CH ₂
C4	SIPh	Ph	CH ₂
C5	S <i>i</i> Pr	2,6-(<i>i</i> Pr) ₂ -C ₆ H ₃	CH ₂
C6	SIMes	2,4,6-(Me) ₃ -C ₆ H ₂	CH ₂
C7	SIXy	2,6-(Me) ₂ -C ₆ H ₃	CH ₂
C8	-	H	CH
C9	IMe	Me	CH
C10	liPr	<i>i</i> Pr	CH
C11	l ^t Bu	<i>t</i> Bu	CH
C12	ICy	Cy	CH
C13	SIPh	Ph	CH
C14	IBn	CH ₂ Ph	CH
C15	lPr	2,6-(<i>i</i> Pr) ₂ -C ₆ H ₃	CH
C16	IMes	2,4,6-(Me) ₃ -C ₆ H ₂	CH
C17	IXy	2,6-(Me) ₂ -C ₆ H ₃	CH
C18	lAd	1-Ad	CH
C19	^{Me} IMe	Me	C(CH ₃)
C20	^F IMe	Me	C(F)
C21	^{Cl} IMe	Me	C(Cl)
C22	^{NO₂} IMe	Me	C(NO ₂)

		
C23	C24	C25
		
C26	C27	
		
C28	C29	C30
Ligand	Abbreviation used	
C23	BlmN(Me) ₂ (acronym: BMe)	
C24	Py(b)ImN(Me) ₂	
C25	DpyIm	
C26	lBioxMe ₄	
C27	PerN(iPr) ₂	
C28	ThNMe	
C29	OxNMe	
C30	BOxNMe	

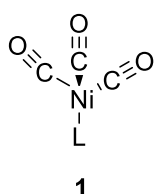
2. Ligand Descriptors

2.1 Monodentate Phosphorus(III)-Donor Ligands

Ligands with a phosphorus(III) donor atom enjoy persistent popularity for organometallic and coordination chemistry.³⁶⁻⁴² Key representatives of this class of ligands coordinate transition metals well, and they have been tested in a broad and varied range of catalytic cycles, with a particular focus on the late transition metal complexes active for cross-coupling,⁴²⁻⁴⁸ hydrogenation,^{49, 50} hydroformylation^{38, 51-53} *etc.* Beyond their experimental utility, these ligands have been characterised with perhaps the most extensive range of calculated descriptors of their steric and electronic properties, suggesting a fertile environment for data-led studies in this field. A range of reviews have touched on the importance of quantitative approaches to this area, both by us^{12, 17, 20, 54} and other groups,^{13, 14, 51, 55, 56} and here we have focussed on descriptor updates and modifications described or applied from around 2012 onwards, albeit mentioning older studies to set these into context.

2.1.1 Individual Descriptors

There is an obvious appeal in identifying simple, usually linear, relationships between one or just a few steric and electronic parameters and an experimental measurement: visualisation is usually straightforward, making it easy to identify and understand trends, as well as attempting prediction.^{36, 39} The use of individual descriptors of P-donor ligands received a considerable boost with the publication of Tolman's seminal review,¹⁶ which compiled data from infra-red (IR) spectroscopical



measurements on a tetrahedral nickel complex ([Ni(CO)₃L]) **1**, as well as describing and presenting a steric measure for these ligands. The former, IR-derived, descriptor is now commonly referred to as the Tolman Electronic Parameter (TEP), while the latter tends to be called the Tolman Cone Angle (TCA), and both continue to inform

ligand selection,¹⁷ as well as inspiring the development of alternative, calculated descriptors, as demonstrated below.

2.1.1.1 Electronic Descriptors

Tolman's keystone review¹⁶ focussed on the highest CO stretch, of A_1 symmetry, measured in the IR spectrum of $[\text{Ni}(\text{CO})_3\text{L}]$ complexes **1** and used that as a proxy for the strength of the metal-ligand interaction. Due to the cone-like shape of both PR_3 ligands and the $\text{Ni}(\text{CO})_3$ metal fragment, this was considered to be largely free of steric effects for any but the largest ligands, and to capture the net electron donation from the ligand to the metal centre, modifying the strength, force constant and so stretching frequency of the CO bonds.

With growing computational resources, it became possible to calculate this parameter on a relatively large scale,¹⁷ either with density functional theory (DFT, initially reported as the Calculated Electronic Parameter, CEP),⁵⁷ or with semi-empirical calculations on a different complex (*trans*- $\text{RhL}_2(\text{CO})\text{Cl}$), labelled as the Semi-empirical Electronic Parameter, SEP),⁵⁸ and such calculations opened up the possibility of considering novel, toxic and unstable compounds.⁵⁷ As set out in a number of earlier studies (see for example references 57-59) and reviews,^{17, 54, 55} carbonyl stretching frequencies and related measures of net electron-donation derived from a range of different metal carbonyl complexes tend to correlate highly and have thus been used almost interchangeably, and indeed across different ligand classes, as discussed by Gusev.⁵⁹ We have collated some representative examples for our ligand set in Table 2, all rounded to 4 significant figures in recognition of instrumental limitations and computational noise (note that data in the accompanying spreadsheet (ESI) are included as quoted in the relevant publications). Indeed, recent analyses of experimental³⁶ and calculated⁶⁰ data have used IR-derived descriptors (TEP and phosphine oxides) to capture ligand electronic effects.

Table 2: Infra-red stretching frequencies (cm^{-1}) and related descriptors, for monodentate P(III)-donor ligands (see Table 1a for ligand details), measured experimentally or calculated. P24 & P25 excluded due to lack of data. TEP values in bold are from reference 61. All data rounded to 4 significant figures.

Ligand No.	Phosphine Ligand	TEP ^a	CEP ^b	SEP ^c	Ni stretch ^d	Au stretch ^d	(LTEP) W _a ^e	(MLEP) W _a ^f	Ni(CO) ₃ L ^g		Ir(Cp)(CO)L ^g
									v(CO) Calc	v(CO) Exp	v(CO) Calc
P1	PH ₃	2083	2171	-	2031	2166	2130	2130	2082	2083	2060
P2	P(Me) ₃	2064	2152	2102	2011	2145	Multiple values	2106	2065	2064	2037
P3	P(Et) ₃	2062	-	2105	-	-	-	2123	20623	2062	2034
P4	P(iPr) ₃	2060	-	2096	-	-	-	-	2059	2059	2032
P5	P(nBu) ₃	2060	-	2106	-	-	-	-	-	-	-
P6	p(tBu) ₃	2056	-	2084	1999	2123	-	-	2055	2056	2028
P7	P(Cy) ₃	2056	-	2104	2000	2125	-	-	2057	2056	2032
P8	P(Ph) ₃	2069	-	2123	-	-	-	2103	2068	2069	2049
P9	P(Bn) ₃	2066	-	2120	-	-	-	-	-	-	-
P10	P(NMe ₂) ₃	2062	2151	-	-	-	-	2108	-	-	-
P11	P(OMe) ₃	2080	2171	2160	-	-	2119	2118	2080	2080	2059
P12	P(OEt) ₃	2076	-	2161	-	-	-	-	20798	2076	2055
P13	P(Oph) ₃	2085	-	2156	-	-	-	-	-	-	-
P14	P(C ₆ F ₅) ₃	2090	-	2127	-	-	-	2126	-	-	-
P15	PF ₃	2111	2201	2158	2058	2186	2159	2177	2111	2111	2115
P16	P(o-tol) ₃	-	-	2124	-	-	-	-	-	-	-
P17	P(p-tol) ₃	2067	-	2122	-	-	-	-	-	-	-
P18	P(p-F-C ₆ H ₄) ₃	2071	-	-	-	-	-	-	-	-	-
P19	P(p-Cl-C ₆ H ₄) ₃	2073	-	-	-	-	-	-	-	-	-
P20	P(p-OMe-C ₆ H ₄) ₃	2066	-	-	-	-	-	-	-	-	-
P21	P(Ph)(Me) ₂	2065	-	2108	-	-	-	-	2067	2065	2045
P22	P(Me)(Ph) ₂	2067	-	2121	-	-	-	-	2067	2067	2043
P23	P(Ph)(Et) ₂	2064	-	-	-	-	-	-	-	-	-
P26	JohnPhos	-	-	-	2001	2115	-	-	-	-	-
P27	CyJohnPhos	2056^h	-	-	-	-	-	-	-	-	-
P28	MePhos	205^h	-	-	-	-	-	-	-	-	-
P29	SPhos	2054^h	-	-	-	-	-	-	-	-	-
P30	XPhos	2053^h	-	-	-	-	-	-	-	-	-

^a References 16, 61; ^b reference 57; ^c reference 58; ^d reference 62; ^e reference 63; ^f reference 64; ^g reference 59; ^h reference 61.

The group of Cremer⁶³ reviewed such prior work, as well as related descriptors, in a 2014 study and queried the validity of using normal vibrational modes from experimental measurements or calculations, as these couple with the M-C stretching modes, making it more difficult to isolate the net electronic effects of different ligands on the CO bond strength, as bond length and strength cease to be directly related. Using DFT calculations (with a focus on M06/aug-cc-pVTZ data), they determined the contribution of this coupling in a range of metal carbonyls and introduced a local TEP (LTEP) derived from the nickel tricarbonyl complexes used most commonly, considering 42 representative ligands across a range of different chemistries, including 8 P-donor ligands and 2 simple, acyclic carbenes. This mode-decoupled descriptor is designed to capture the CO bond strength more accurately than parameters based on normal modes, facilitating the comparison of different ligand classes (an issue explored in greater detail below). They presented local CO stretching frequencies (LTEP ω_α) and force constants (LTEP κ_α) for the nickel carbonyl complex and processed these to give a general LTEP descriptor (LTEP n) capturing bond orders, suitable for the comparison of different ligand types. They provided detailed analyses of data and molecular orbitals to set their conclusions into context, as well as demonstrating how LTEP descriptors can also allow comparison of the electronic influence of different transition metal centres.

Following on from their study reported in 2014,⁶³ Cremer and co-workers later reported a more extensive analysis of ligand properties using 181 [Ni(CO)₃L] complexes **1**,⁶⁴ capturing a wider range of steric properties as well as additional ligand classes, again using local vibrational modes. Their ligand set included 20 P-donor and 15 carbene ligands. From their analysis of the relationship between Ni-L and CO bond strengths, they concluded that their LTEP ω_α descriptor is insufficient for this larger dataset, and proposed a new electronic parameter, MLEP (κ_α (Ni-Y)), calculated from local stretching force constants. While experimental data could be used to determine MLEP, they provided calculated data for all ligands considered based on M06/aug-cc-pVTZ geometries; they also noted that other DFT approaches would give similar results. In line with their earlier work, they explored further simplification by using relative bond strength orders, showing high linear correlations between the descriptors across all ligand classes considered. They indicated that these descriptors implicitly capture steric effects, arguing that while the C-Ni-L bending force constant gives an indication of steric hindrance, there is no need for this as overall ligand effects have been captured already. While both this⁶⁴ and a recent Perspective article from this group⁶⁵ hinted at planned extensions to other M-L interactions, we have not been able to locate further published data which might lead to a more general measure of ligand properties.

Also in 2014, the groups of Ciancaleoni and Belpassi⁶² reviewed whether the relationship between CO stretches derived from different transition metal complexes could always be described by a simple linear relationship, highlighting some important failures, especially where different ligand types were considered together. They were particularly interested in gold carbonyl complexes ($[(\text{CO})\text{AuL}]^{+/0}$) and used charge displacement analysis to assess and compare the net charge transfer (CT) with calculated electronic parameters (BLYP/TZ2P) for their gold, as well as the standard nickel carbonyl complexes **1**, considering a range of ligands, including 10 mixed alkyl- and aryl-phosphines and 4 N-heterocyclic carbenes (NHCs). This analysis shows that while the relationship between TEP and ligand net donor ability is strong for the nickel complex across different ligand classes, this is not the case for the gold carbonyl, leading them to suggest that the CO stretch is no longer a good measure of ligand donor ability in this complex. Interestingly, this is not revealed by a scatter plot of calculated CO frequencies, instead requiring charge decomposition analysis. Detailed orbital and charge analyses have been used to justify these observed differences in the interactions between ligands and different metal complexes. While this highlights the importance of considering the coordination environment when

assessing ligand properties, it also suggests that either a reaction-specific descriptor or more than a single electronic parameter may be needed for catalyst design.

In an effort to develop a simple set of steric and electronic parameters, independent of metal coordination for monodentate phosphines and phosphites, a Venezuelan group led by Coll⁶⁶ have analysed local ionisation energies ($I_{\min}(r)$) derived from DFT geometry optimisations (B3LYP/6-31++G**) of 43 P-donor ligands, alongside the global minimum of the electrostatic potential proposed by Suresh's group ($V_{\min}(r)$, see section 2.1.2.1). They observed a high linear correlation between the minimum of the local ionisation energy and the TEP, leading them to suggest that this captures the polarisability of ligands. They also derived a cone angle measure from these calculations (discussed in the next section, 2.1.1.2).

The interpretation of organometallic and coordination chemistry tends to rely on describing the bonding between ligands and metals in terms of bonding interactions with σ - and π -symmetry. A range of energy decomposition schemes, previously reviewed in reference 17, have been used to attempt to unravel and quantify the contributions made by different types of bond, with a particular focus on the extent of π -backbonding in complexes of phosphines. Recent work by Ardizzoia and Brenna⁶⁷ used such an approach, combining the Extended Transition State (ETS) method⁶⁸ with the Natural orbitals for chemical valence (NOCV) approach^{69, 70} to analyse the σ -donation and π -backdonation components of metal-phosphine interactions. From their application of this ETS-NOCV scheme to 41 $[\text{Ni}(\text{CO})_3(\text{PR}_3)]$ complexes **1**, they also identified a σ -backdonation term. They fitted a multivariate regression model based on three descriptors, E_{σ}^d , E_{π}^{bd} and E_{σ}^{bd} , to predict the TEP and so derived a single electronic parameter, T^{phos} . In addition, they explored substituent effects on T^{phos} , updating Tolman's work for predicting TEP data by adding substituent contributions to a computational approach. For a subset of ligands, they also derived a steric parameter, discussed below (section 2.1.1.2).

Table 3: Individual descriptor data for monodentate P(III)-donor ligands (see Table 1a for details of ligands). Ligands P16 and P27-30 excluded from this table due to a lack of published data.

Ligand No.	Phosphine Ligand	LTEPa		MLEPa				Coll et al. b		Ciancaleonic		Ardizzoia and Brennad			
		K_{α} (C≡O) (mdyn Å ⁻¹)	n (BSO) (C≡O)	K_{α} (Ni-Y) (mdyn Å ⁻¹)	ω_{α} (Ni-Y) (cm ⁻¹)	n (BSO) (Ni-Y)	K_{α} (C≡O) (mdyn Å ⁻¹)	$I_{\min}(r)$ vdW (eV)	$V_{\min}(r)$ vdW (kcal mol ⁻¹)	CT (Ni)	CT (Au)	E_{σ}^d (kcal mol ⁻¹)	E_{π}^{bd} (kcal mol ⁻¹)	E_{σ}^{bd} (kcal mol ⁻¹)	T^{phos} (cm ⁻¹)
P1	PH ₃	18.33	2.4	0.90	274.60	0.43	18.33	0.287	17.94	0.02	0.26	-	-	-	-
P2	P(Me) ₃	Multiple values	1.07	299.80	0.51	17.92		0.241	-26.72	0.09	0.33	-25.83	-13.04	-5.43	-13.56
P3	P(Et) ₃	-	-	0.99	288.90	0.48	18.03	0.228	-28.42	-	-	-28.03	-13.56	-5.76	-16.00
P4	P(iPr) ₃	-	-	-	-	-	-	0.222	-27.17	-	-	-	-	-	-
P5	P(nBu) ₃	-	-	-	-	-	-	0.226	-29.24	-	-	-25.79	-12.11	-5.01	-15.89
P6	p(tBu) ₃	-	-	-	-	-	-	0.220	-25.48	0.11	0.4	-24.49	-11.11	-3.93	-15.43
P7	P(Cy) ₃	-	-	-	-	-	-	0.227	-27.99	0.11	0.43	-	-	-	-
P8	P(Ph) ₃	-	-	0.99	288.40	0.48	17.87	0.250	-21.28	-	-	-22.73	-13.45	-5.61	-6.96
P9	P(Bn) ₃	-	-	-	-	-	-	-	-	-	-	-22.69	-13.75	-5.12	-7.18
P10	P(NMe ₂) ₃	-	-	0.96	284.10	0.46	17.94	0.241	-24.76	-	-	-25.23	-13.74	-4.57	-12.72
P11	P(OMe) ₃	18.14	2.383	1.13	308.40	0.54	18.12	0.281	-24.46(-27.82)	-	-	-23.35	-17.43	-6.17	0.05
P12	P(OEt) ₃	-	-	-	-	-	-	0.283	-28.86(-29.97)	-	-	-24.09	-16.9	-5.79	-2.88
P13	P(Oph) ₃	-	-	-	-	-	-	0.307	-14.81(-18.97)	-	-	-21.03	-20.02	-5.53	7.77
P14	P(C ₆ F ₅) ₃	-	-	0.74	250.00	0.36	18.26	0.313	-6.65	-	-	-16.42	-15.7	-5.91	8.93
P15	PF ₃	18.83	2.445	1.26	325.90	0.60	18.11	0.380	6.79(-8.56)	0.09	0.21	-18.99	-30.63	-8.19	35.08
P17	P(<i>p</i> -tol) ₃	-	-	-	-	-	-	0.244	-24.11	-	-	-23.29	-13.17	-5.58	-8.51
P18	P(<i>p</i> -F-Ph) ₃	-	-	-	-	-	-	0.264	-15.32	-	-	-21.95	-13.75	-5.48	-5.24
P19	P(<i>p</i> -Cl-Ph) ₃	-	-	-	-	-	-	0.265	-14.35	-	-	-21.79	-14.17	-5.66	-3.89
P20	P(<i>p</i> -OMe-Ph) ₃	-	-	-	-	-	-	0.240	-25.72	-	-	-	-	-	-
P21	P(Ph)(Me) ₂	-	-	-	-	-	-	0.246	-24.39	-	-	-24.76	-13.2	-5.12	-11.89
P22	P(Me)(Ph) ₂	-	-	-	-	-	-	0.249	-22.93	-	-	-23.73	-13.59	-5.57	-8.57
P23	P(Ph)(Et) ₂	-	-	-	-	-	-	0.240	-25.76	-	-	-24.03	-12.38	-5.21	-11.89
P24	P(Et)(Ph) ₂	-	-	-	-	-	-	-	-	-	-	-23.78	-13.35	-5.38	-8.40
P25	P(OMe)(Ph) ₂	-	-	-	-	-	-	-	-	-	-	-23.73	-16.06	-5.73	-3.85
P26	JohnPhos	-	-	-	-	-	-	-	-	0.07	0.35	-	-	-	-

^a Reference 64; ^b reference 66; ^c reference 62; ^d reference 67.

2.1.1.2 Steric Descriptors

While electronic descriptors are often assessed in terms of their relationship with the TEP or a suitable proxy (see above, section 2.1.1.1), several steric descriptors have been proposed since Tolman's work on cone angles,¹⁶ and these have been reviewed by a number of authors.^{17, 54, 71-73} With a view to exploring some of the applications of such steric measures for catalyst design, it may be helpful to provide a brief summary of how these descriptors can be calculated.

The Tolman cone angle was originally derived from space-filling models of different ligands in $[\text{Ni}(\text{CO})_3\text{L}]$ complexes **1**,¹⁶ with a Ni-P distance of 2.28 Å, rotating substituents to achieve the smallest possible cone angle. Problems with conformational freedom in this approach have been discussed extensively and are summarised in reference 17; we note that for those with large computational budgets automated approaches such as AARON can now be used to sample multiple conformers even across catalytic manifolds.⁷⁴ Cone angles can be derived from structural coordinates, including from DFT-optimised geometries, and a straightforward process for their calculation has been described by Smith and Coville.⁷⁵ In this, a half-cone angle θ_i is measured for each substituent according to: $\theta_i = \alpha + \frac{180}{\pi} \sin^{-1} \left(\frac{r_{vdW}}{d} \right)$, where d is the distance between the metal centre and an atom A on the surface of the cone, considered likely to give the largest half-cone angle, α is the P-M-A angle and r_{vdW} is the van der Waals radius of A. The half cone angles are then added up according to: $\theta = \frac{2}{3} \sum_{i=1}^3 \frac{\theta_i}{2}$. This approach has been implemented in the Solid-G programme developed by Guzei and Wendt,⁷⁶ which is currently free to download.⁷⁷

Sigman's group used this Solid-G implementation to generate cone angles for 38 P-donor ligands in their analysis of Suzuki reactions.³⁶ The mapping of the average local ionisation energy by Coll's group,⁶⁶ discussed in the context of electronic parameters above, has also been used to calculate cone angles, which were found to correlate highly with Tolman's data.¹⁶ Wu and Doyle considered cone angles calculated from molecular mechanics in their analysis of nickel catalysts.³⁹

Related measures include the exact cone angle,⁷⁸ which provides a mathematically exact approach to determining a cone angle, avoiding selection of a suitable atom A, as well as accommodating different metal coordination environments. The group developing this descriptor has published data for more than 150 metal complexes derived from B3LYP/6-31G*, using the LANL2DZ basis set on metal-optimised geometries,⁷⁸ as well as making a Mathematica package available for download.⁷⁹ More recently, Petitjean has proposed an approach to cone angle calculations,⁸⁰ which does not rely on atomic radii and has been implemented in a freeware programme.⁸¹

An alternative, albeit closely related, steric descriptor is the solid (cone) angle Ω , generally described as the area of the shadow cast by a ligand onto a sphere which has the transition metal atom at the centre. It was the focus of Guzei and Wendt's development of the Solid-G programme⁷⁶ and an exact solid angle calculation has been described by Allen and co-workers;⁸² both have been implemented in software available for download.^{77, 79} An analysis of ligand effects in gold(I) catalysis undertaken by the groups of Sigman and Toste³⁷ mentions that the solid cone angle were considered, but did not give high correlations with the experimental data in this case.

Orpen's $S4'$ descriptor⁸³ has its origins in crystallography, but was later used by Cundari's group as a computationally convenient measure of ligand steric bulk.⁵⁸ This descriptor captures the structural change in a metal complex due to differences in steric bulk and is calculated as $S4' = \sum \angle MPA - \sum \angle APA$. It has been incorporated in Bristol's Ligand Knowledge Bases (LKBs),³⁵ discussed below (section 2.1.2.2).

Even though the % Buried Volume descriptor, $\%V_{\text{bur}}$, was originally developed for N-heterocyclic carbenes (NHCs, see section 2.3.1.2),^{84, 85} which, due to the more two-dimensional nature of their steric demands, were not described well by cone angles, it has been applied to phosphines and carbenes⁷² and merits inclusion here. As with most steric parameters, atomic coordinates from both crystallography and calculations can be used, and the descriptor determines what percentage of a metal-centred sphere of defined radius is occupied by the ligand.^{72, 85} Variations in metal-donor distance and sphere radius affect this parameter,⁷² but it has been shown to correlate highly with cone angles.⁷² Cavallo and co-workers have implemented an application for the calculation of $\%V_{\text{bur}}$ online,⁸⁶ and this has been updated and extended to include topographic steric maps more recently.^{87, 88} The descriptor has been applied in the analysis of Buchwald-type biaryl P-donor ligands by several groups.^{39, 61, 89} In addition, it was used to assess the steric effects of several phosphine and carbene ligands on calculated barriers for the copper-catalyzed boracarboxylation of styrene (Scheme 1).⁶⁰

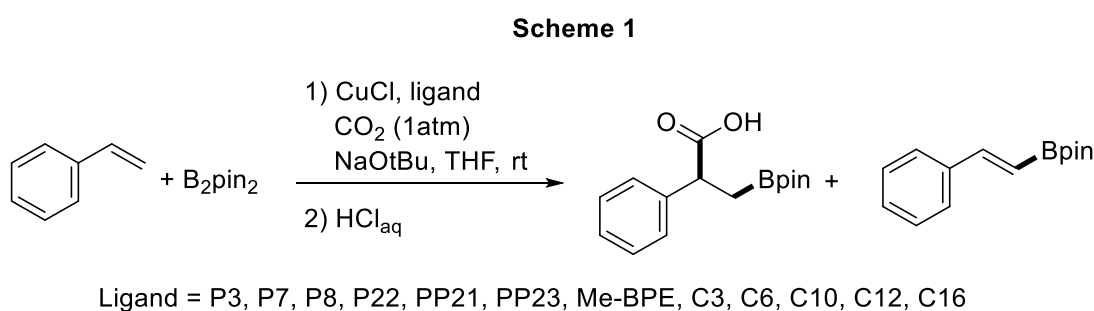


Table 4 lists a range of steric descriptors for our core set of ligands.

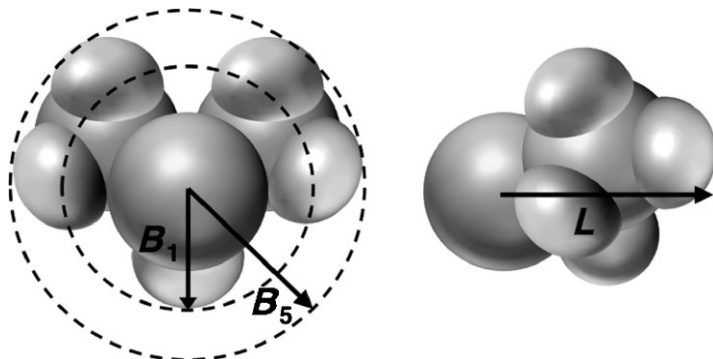


Figure 1: Illustration of key Sterimol descriptors used by Sigman; parameterisation of an isopropyl group. Reprinted with permission from reference 90. Copyright 2013, American Chemical Society.

The Sterimol descriptors do not exactly fit the present category of individual descriptors, as they are actually a family of five parameters, originally proposed by Verloop (see reference 91 and citation in reference 90 for details) for quantitative structure-activity relationships in drug design, but they are generally determined and used together. In applications to asymmetric catalysis, Sigman's group found three of these descriptors most useful for chiral ligands,⁹⁰ B_1 , B_5 , and L (Figure 1). These describe the minimum width orthogonal to the primary bond (B_1), the maximum width along the same axis (B_5), and the length of the substituent through the primary bond (L). A later study of gold catalysis uses the L/B_1 average as a single steric parameter for aryl and biaryl phosphine ligands.³⁷ These descriptors have been summarised in Table 5. A study of copper-catalysed asymmetric conjugate addition by the groups of Paton and Fletcher⁹² also used Sterimol descriptors in their analysis (see section 2.1.2.3).

Table 4: Individual steric descriptors for P(III)-donor ligands. Ligands P1, P10, P25 and P28 excluded due to lack of data.

Ligand No.	Phosphine Ligand	S4 (Cooney) (°) ^a	S4' (Orpen) (°) ^b	%V _{bur} ^c	Tolman Cone Angle (°) ^d	Solid Cone Angle Θ (°) ^e	Exact Cone Angle (°) (Pd) ^f		Exact Cone Angle (°) (Pt) ^f		Exact Cone Angle (°) (Ni) ^f	
							Min	Max	Min	Max	Min	Max
P2	P(Me) ₃	52	46.6	26.1	118	124	120	-	126	-	125	-
P3	P(Et) ₃	35	33.2	32.7	132	143	136	169	143	176	140	171
P4	P(iPr) ₃	35	26.1	37.6	160	163	169	177	177	185	173	181
P5	P(nBu) ₃	35	28.2	30.4	132	148	143	169	143	176	141	171
P6	p(tBu) ₃	17	2.6	42.4	182	182	188	-	196	-	192	-
P7	P(Cy) ₃	21	17.9	37.1	170	-	-	-	-	-	-	-
P8	P(Ph) ₃	35	27.6	34.5	145	129	170	-	178	-	175	-
P9	P(Bn) ₃	40	27.2	-	<u>165</u>	-	-	-	-	-	-	-
P11	P(OMe) ₃	63	-	-	107	-	-	-	-	-	-	-
P12	P(OEt) ₃	61	-	-	109	-	-	-	-	-	-	-
P13	P(OPh) ₃	76	-	35.4	128	-	-	-	-	-	-	-
P14	P(C ₆ F ₅) ₃	48	-	42.6	184	-	-	-	-	-	-	-
P15	PF ₃	64	81.0	-	-	-	-	-	-	-	-	-
P16	P(o-tol) ₃	39	14.6	46.7	194	142	176	208	184	219	180	211
P17	P(p-tol) ₃	35	34.3	33.0	145	135	171	-	178	-	175	-
P18	P(p-F-C ₆ H ₄) ₃	-	-	-	145	129	171	-	178	-	175	-
P19	P(p-Cl-C ₆ H ₄) ₃	-	-	-	145	129	170	-	178	-	175	-
P20	P(p-OMe-C ₆ H ₄) ₃	-	-	-	145	139	170	172	178	178	173	175
P21	P(Ph)(Me) ₂	47	39.2	-	122	126	149	-	156	-	153	-
P22	P(Me)(Ph) ₂	41	33.7	-	136	124	151	-	160	-	156	-
P23	P(Ph)(Et) ₂	-	-	-	136	137	153	173	160	180	157	177
P24	P(Et)(Ph) ₂	-	-	-	140	140	150	169	156	178	153	173
P26	JohnPhos	-	-	30.8	181	-	-	-	-	-	-	-
P27	CyJohnPhos	-	-	-	<u>193</u>	-	-	-	-	-	-	-
P29	SPhos	-	-	-	<u>208</u>	-	-	-	-	-	-	-
P30	XPhos	-	-	48.8	221	-	-	-	-	-	-	-

^a Reference 58; ^b reference 83; ^c references 36, 72; ^d references 71, 36, 78; ^e reference 82; ^f reference 78.

Table 5: Sterimol descriptors reported in reference 36. (Ligands P1-P2, P10-P15 and P28 omitted due to lack of data.)

Ligand No.	Phosphine Ligand	Mean B ₁	Mean B ₅	Mean L	Mean L/B ₁	Mean B ₁ *L
P3	P(Et) ₃	3.25	4.87	4.88	1.50	15.87
P4	P(iPr) ₃	3.61	4.90	5.44	1.51	19.61
P5	P(nBu) ₃	4.36	5.94	5.00	1.15	22.26
P6	P(tBu) ₃	4.11	4.90	5.42	1.32	22.26
P7	P(Cy) ₃	4.19	6.67	6.39	1.52	26.76
P8	P(Ph) ₃	4.09	6.28	5.86	1.43	23.98
P9	P(Bn) ₃	3.93	7.31	6.77	1.72	26.58
P16	P(<i>o</i> -tol) ₃	4.40	6.34	6.15	1.40	27.05
P17	P(<i>p</i> -tol) ₃	5.03	7.51	5.86	1.17	29.46
P18	P(<i>p</i> -F-C ₆ H ₄) ₃	4.26	6.97	5.85	1.37	24.96
P20	P(<i>p</i> -OMe-C ₆ H ₄) ₃	5.12	8.41	6.42	1.25	32.88
P21	P(Ph)(Me) ₂	3.04	6.34	6.09	2.01	18.49
P22	P(Me)(Ph) ₂	3.25	6.33	6.03	1.86	19.61
P23	P(Ph)(Et) ₂	3.25 ^a	6.31 ^a	6.04 ^a	1.86 ^a	19.60 ^a
P24	P(Et)(Ph) ₂	3.26	6.39	6.00	1.84	19.58
P26	JohnPhos	3.97	6.11	7.69	1.94	30.58
P27	CyJohnPhos	4.20	6.27	7.45	1.78	31.27
P29	SPhos	4.34	7.08	7.76	1.79	33.70
P30	XPhos	4.32	7.83	9.50	2.20	41.04

^aP23 L/B₁ and B₁*L are calculated from given values in the ESI of reference 36.

2.1.2 Descriptor databases

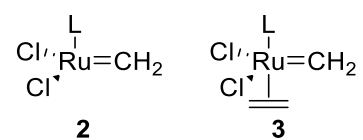
Individual steric and electronic parameters can work well for the analysis of experimental data where a single effect dominates, facilitating visualisation and the detection of correlations,^{13, 14, 36} but such circumstances are rare in organometallic catalysis and usually a wider range of descriptors need to be considered to capture both steric and electronic effects. Where groups with an application-led experimental focus are concerned,^{36, 39} these descriptor databases may well have been compiled from a number of different sources, and they can reflect a combination of experimental and calculated parameters. Examples of these approaches will be reviewed in greater detail in section 3.2 below.

Here, our focus will be on databases of calculated descriptors, usually capturing both steric and electronic effects, intended to be used together for multivariate data analysis. Such databases have usually been developed and curated by the same group, often over a significant length of time, and most are featured in earlier reviews of this field,^{12, 17, 20, 54, 93, 94} allowing us to just give a brief overview to set descriptor data into context, and then explore possible applications. The focus of such approaches varies, ranging from ligand-focussed databases intended to be transferable to a broad range of chemistries,^{35, 95-97} to data specific to an application of interest³⁹ but also taking in databases that apply a consistent computational approach to the analysis of different processes,⁹⁸⁻¹⁰⁰ in some cases facilitating the consideration of several ligand classes.¹⁰¹

2.1.2.1 Molecular Electrostatic Potentials

Suresh and co-workers have published a substantial body of work focussed on calculated molecular electrostatic potentials (MESPs) for P-donor ligands,^{98, 101-103} from which steric and electronic descriptors can be determined as summarised in reference 17. The electronic descriptor corresponds to the local minimum of the electrostatic potential, V_{\min} ,¹⁰² closely related to the local ionisation energies used by Coll and co-workers⁶⁶ (see section 2.1.1.1). Comparison of QM/MM and QM calculations allowed a separation into steric and electronic effects for phosphines, giving rise to the S_{eff} and E_{eff} descriptors respectively, presented in 2007 and summarised in Table 6.⁹⁸ Similar descriptors have been developed for carbenes, as discussed below (section 2.3.2.1).

The minimum electrostatic potential, V_{\min} ,¹⁰² was used by Wu and Doyle as an electronic descriptors for analysing nickel catalysts for Suzuki coupling;³⁹ they did not observe strong correlation with their experimental data, focussing instead on steric descriptors.



The MESP approach has been applied to first generation ruthenium metathesis catalysts,¹⁰⁴ generating steric and electronic descriptors for 21 phosphines in two different coordination environments relevant to the catalytic cycle (the active catalyst $[\text{Cl}_2(\text{PR}_3)\text{Ru}=\text{CH}_2]$ **2** with $\text{L} = \text{PR}_3$ generated by ligand dissociation from the pre-catalyst, and the ethene coordinated complex **3**), which have been determined at the BP86/def2-TSVPP level of theory. Differences between MESP values at the P nucleus for complexes compared to a PH_3 -coordinated reference have been used as a measure of the combined steric and electronic effects (V_{SE}), while frozen structures were used to derive steric descriptors (V_{S}), and from these two measures, electronic descriptors (V_{E}) could be determined.

Correlations with other steric and electronic descriptors were found to be high, helping to set these parameters into context. By relating these descriptors to calculated barriers, the authors were able to quantify the importance of steric and electronic effects on key barriers along the reaction pathway, allowing them to formulate some general ligand design criteria, highlighting that bulky and electron-rich phosphines would make the best spectator ligands in this case. A closely-related approach was then applied to second generation, carbene-based ruthenium complexes,¹⁰⁰ which will be discussed in section 2.2.2.1 below, with the relevant data collected in Table 19.

More recently, this approach has been applied to palladium-catalysed oxidative addition reactions, considering a range of ligands and substrates.¹⁰¹ In this study, MESP's have been calculated for uncoordinated ligands, including 20 phosphines, as well as for PdL₂ and PdL complexes, generating descriptors specific to ligands and palladium complexes, intended to relate to catalyst activation. These descriptors have then been related to calculated energy barriers for the addition of Ph-X to Pd(alkene) complexes across different ligand classes (Scheme 2), supporting catalyst tuning. Again, these descriptors have been summarised in Table 19 below, while ligand-only MESP data can be found in Table 6.

Scheme 2: Ligands considered in MESP study of PdL₂ and PdL complexes.¹⁰¹

R _{1,2,3} = H R _{1,2,3} = Me R _{1,2,3} = Et R _{1,2,3} = iPr R _{1,2,3} = tBu R _{1,2,3} = Cy R _{1,2,3} = SiMe ₃ R _{1,2,3} = Thiophene R _{1,2,3} = SMe R _{1,2,3} = Ph R _{1,2,3} = Ph-F R _{1,2,3} = Ph-Cl R _{1,2,3} = Ph-CF ₃ R _{1,2,3} = F R _{1,2} = H, R ₃ = CF ₃ R _{1,2} = Cl, R ₃ = Ph R _{1,2} = Cl, R ₃ = Me R ₁ = H, R ₂ = Me, R ₃ = Ph	X, Y = H, Me X, Y = H, F X, Y = Me, H X, Y = Me, COOMe X, Y = Me, CN X, Y = Me, NO ₂ X, Y = Me, Cl X, Y = Me, CF ₃ X, Y = Me, F X, Y = CF ₃ , H	R = H Me Et Ph NH ₂ NMe ₂ SiMe ₃ Br Cl F	R = H Me Et Ph CN SiMe ₃ Br Cl CF ₃ F

Table 6: MESP descriptors for P(III)-donor ligands.^{98, 101-103} Ligands P9, P14, P19-20 and P25-30 excluded due to lack of data.

Ligand No.	Phosphine Ligand	V _{min} (ONIOM) (kcal mol ⁻¹) ^a	MESP (steric) (kcal mol ⁻¹) ^b	θ MESP (°) ^b	V _{min} (real) (kcal mol ⁻¹) ^b	E _{eff} (kcal mol ⁻¹) ^a	S _{eff} (kcal mol ⁻¹) ^a
P1	PH ₃	-28.22	0.0	87	-28.22	0.00	0.00
P2	P(Me) ₃	-30.62	-2.4	108	-43.02	12.40	2.40
P3	P(Et) ₃	-33.35	-5.1	132	-43.51	10.54	4.79
P4	P(iPr) ₃	-37.34	-9.1	168	-44.47	6.71	9.24
P5	P(nBu) ₃	-	-	-	-43.71	-	-
P6	P(tBu) ₃	-	-	-	-	1.95	14.70
P7	P(Cy) ₃	-37.83	-9.6	172	-44.99	7.16	9.61
P8	P(Ph) ₃	-34.62	-6.4	144	-34.85	-0.13	5.98
P10	P(NMe ₂) ₃	-	-	-	-37.79	-	-
P11	P(OMe) ₃	-	-	-	-26.12	-	-
P12	P(OEt) ₃	-	-	-	-27.85	-	-
P13	P(Oph) ₃	-	-	-	-15.88	-	-
P15	PF ₃	-	-	-	-	-	-
P16	P(<i>o</i> -tol) ₃	-39.88	-11.7	190	-31.16	-	-
P17	P(<i>p</i> -tol) ₃	-	-	-	-	-	-
P18	P(<i>p</i> -F-C ₆ H ₄) ₃	-	-	-	-29.03	-	-
P21	P(Ph)(Me) ₂	-32.83	-4.6	128	-40.41	7.58	4.61
P22	P(Me)(Ph) ₂	-32.47	-4.3	125	-36.76	4.29	4.25
P23	P(Ph)(Et) ₂	-33.01	-4.8	129	-40.76	-	-
P24	P(Et)(Ph) ₂	-35.12	-6.9	148	-37.23	2.11	6.90

^a Reference 98; ^b reference 103.

2.1.2.2 Ligand Knowledge Bases

A consortium of authors based at the University of Bristol, including one of us (NF), have developed several so-called Ligand Knowledge Bases (LKBs), capturing ligand effects across a range of representative coordination environments with DFT calculations (BP86/6-31G* and LACV3P on metal centres). For monodentate P-donor ligands,^{35, 96} LKB-P descriptors have been harvested from calculations on the complexes shown in Figure 2.

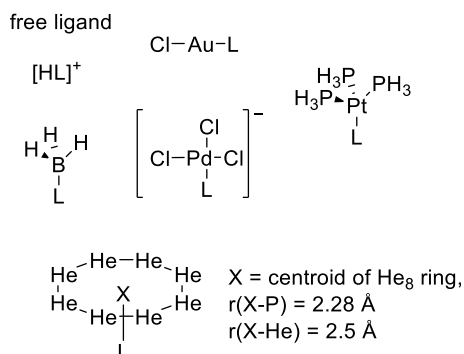


Figure 2: Complexes used in LKB-P.^{35, 96}

Table 7: Descriptors in LKB-P.^{35, 96}

Descriptor	Derivation (Unit)
<i>Free Ligand</i>	
E_{HOMO}	Energy of highest occupied molecular orbital (Hartree)
E_{LUMO}	Energy of lowest unoccupied molecular orbital (Hartree)
He_8_steric	Interaction energy between singlet L in ground state conformation and ring of 8 helium atoms; $E_{ster} = E_{tot}(\text{system}) - [E_{tot}(He_8) + E_{tot}(L)]$ (kcal mol ⁻¹)
<i>Protonated Ligand ([HL]⁺)</i>	
PA	Proton affinity (kcal mol ⁻¹)
<i>Borane Adduct (H₃B.L)</i>	
Q(B fragm.)	NBO charge on BH ₃ fragment
BE(B)	Bond energy for dissociation of P-ligand from BH ₃ fragment (kcal mol ⁻¹) ^a
P-B	P-B distance (Å)
$\Delta P-A(B)$	Change in average P-A bond length compared to free ligand (Å)
$\Delta A-P-A(B)$	Change in average A-P-A angle compared to free ligand (°)
<i>Gold Complexes ([AuClL])</i>	
Q(Au fragm.)	NBO charge on AuCl fragment
BE(Au)	Bond energy for dissociation of L from [AuCl] fragment (kcal mol ⁻¹) ^a
Au-Cl	$r(\text{Au-Cl})$ (Å)
P-Au	$r(\text{Au-P})$ (Å)
$\Delta P-A(\text{Au})$	Change in average P-A bond length in complex compared to free ligand (Å)
$\Delta A-P-A(\text{Au})$	Change in average A-P-A angle compared to free ligand (°)
<i>Palladium Complexes ([PdCl₃L]⁻)</i>	
Q(Pd fragm.)	NBO charge on [PdCl ₃] ⁻ fragment
BE(Pd)	Bond energy for dissociation of L from [PdCl ₃] ⁻ fragment (kcal mol ⁻¹) ^a
Pd-Cl <i>trans</i>	$r(\text{Pd-Cl})$, <i>trans</i> to ligand (Å)
P-Pd	$r(\text{Pd-C})$ (Å)
$\Delta P-A(\text{Pd})$	Change in average P-A bond length compared to free ligand (Å)
$\Delta A-P-A(\text{Pd})$	Change in average A-P-A angle compared to free ligand (°)

<i>Platinum complexes ([Pt(PH₃)₃L])</i>	
Q(Pt fragm.)	NBO charge on [(PH ₃) ₃ Pt] fragment
BE(Pt)	Bond energy for dissociation of P-ligand from [Pt(PH ₃) ₃] fragment (kcal mol ⁻¹) ^a
P-Pt	P-Pt distance (Å)
ΔP-A(Pt)	Change in average P-A bond length compared to free ligand (Å)
ΔA-P-A(Pt)	Change in average A-P-A angle compared to free ligand (°)
<(H ₃ P)Pt(PH ₃)	Average (H ₃ P)Pt(PH ₃) angle (°)
<i>Cumulative</i>	
S4' calc	(Σ <ZPA – Σ <APA), where Z=BH ₃ , [PdCl ₃] ⁻ , [Pt(PH ₃) ₃], [AuCl] (°)

^a BE = [E_{tot}(fragment)+E_{tot}(L)]-E_{tot}(complex)

The full set of 28 LKB-P descriptors is listed in Table 7 and includes some single-effect parameters, such as the He₈_steric, S₄' calc steric descriptors, and frontier molecular orbital energies and proton affinities (E_{HOMO} and PA are related to the phosphorus lone pair, while E_{LUMO} generally aligns with the likely acceptor orbital for M-L backbonding).¹⁰⁵ In addition, the database contains a range of measures of the ligand and fragment responses to BH₃ adduct formation and metal coordination by [AuCl], [PdCl₃]⁻ and [Pt(PH₃)₃]. These descriptors will be affected by both steric and electronic effects, but their interpretation can be facilitated by analysing their relationship with the single-effect parameters above. These were selected to be computationally and chemically robust, making them generally straightforward to calculate, as well as representative and transferable to different coordination environments.

To date, descriptors for 366 monodentate P-donor ligands have been published,^{35, 96, 97, 106} with further data held in-house. Correlations with other descriptors for this class of ligands have been explored,^{17, 35, 96} and the descriptors have been used individually (multivariate linear regression, MLR) and in derived variables (in partial-least squares and principal component regression, PLSR and PCR respectively) to fit more sophisticated models for the prediction of TEP,⁹⁵ as well as the interpretation and prediction of experimental and calculated data.^{35, 96} In addition, descriptors have been processed with Principal Component Analysis (PCA) to facilitate visualisation of the data; this will be discussed further in section 3.1 below. We also note that one of the LKB-P descriptors, the Au-Cl distance, has been used recently in an analysis of gold(I) catalysts by the groups of Toste and Sigman, discussed in section 3.2 below.³⁷

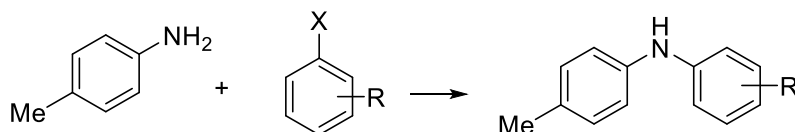
Table 8 summarises a subset of descriptors for monodentate P-donor ligands from LKB-P, with the full set included in the ESI.

Table 8: Subset of LKB-P descriptors (see ESI for full dataset for these ligands).^{35, 95, 96}

Ligand No.	Phosphine Ligand	E _{HOMO}	E _{LUMO}	PA	He ₈ _steric	S4' calc	BE(B)	P-B	BE(Au)	P-Au	Au-Cl	BE (Pd)	P-Pd	Pd-Cl <i>trans</i>	BE(Pt)	P-Pt
P1	PH ₃	-0.246	0.000	190.60	2.30	66.9	26.10	1.941	58.40	2.257	2.288	29.10	2.244	2.357	8.80	2.328
P2	P(Me) ₃	-0.190	0.033	233.00	3.00	39.4	39.20	1.926	72.42	2.272	2.309	38.50	2.268	2.378	12.90	2.331
P3	P(Et) ₃	-0.187	0.026	239.10	6.00	39.0	39.00	1.934	74.62	2.280	2.318	37.00	2.289	2.378	17.10	2.338
P4	P(iPr) ₃	-0.180	0.024	244.60	12.18	28.4	38.60	1.941	74.56	2.284	2.320	37.17	2.317	2.381	14.40	2.367
P5	P(nBu) ₃	-0.187	0.027	243.30	6.10	40.2	39.10	1.934	75.18	2.281	2.320	37.20	2.287	2.377	16.70	2.342
P6	p(tBu) ₃	-0.173	0.028	250.80	23.38	7.4	39.60	1.958	81.87	2.300	2.328	23.90	2.418	2.382	9.80	2.390
P7	P(Cy) ₃	-0.177	0.026	250.40	16.65	28.6	38.80	1.941	74.83	2.289	2.322	36.79	2.315	2.380	13.32	2.369
P8	P(Ph) ₃	-0.187	-0.051	241.30	8.00	32.6	36.50	1.941	70.61	2.278	2.310	31.90	2.304	2.368	16.90	2.332
P9	P(Bn) ₃	-0.196	-0.038	241.61	12.65	38.1	36.74	1.965	70.63	2.278	2.317	38.76	2.287	2.370	13.28	2.326
P10	P(NMe ₂) ₃	-0.161	0.020	245.30	12.13	34.4	39.30	1.927	73.65	2.291	2.344	34.35	2.298	2.379	22.47	2.339
P11	P(OMe) ₃	-0.207	0.000	223.60	4.24	40.2	34.60	1.889	69.15	2.247	2.301	37.98	2.254	2.375	20.94	2.285
P12	P(OEt) ₃	-0.213	0.004	229.70	3.00	45.9	36.20	1.898	71.45	2.250	2.304	34.40	2.247	2.368	21.60	2.290
P13	P(OPh)P	-0.201	-0.047	222.10	5.94	48.9	31.20	1.894	55.47	2.254	2.296	31.81	2.233	2.350	19.81	2.254
P14	P(C ₆ F ₅) ₃	-0.224	-0.091	219.00	11.60	35.7	26.30	1.946	59.34	2.267	2.299	32.70	2.259	2.343	11.90	2.321
P15	PF ₃	-0.282	-0.035	163.60	1.50	62.0	31.10	1.867	56.86	2.213	2.269	37.00	2.192	2.343	22.50	2.234
P16	P(o-tol) ₃	-0.182	-0.049	244.10	30.11	31.0	30.20	1.977	67.07	2.294	2.320	25.85	2.323	2.382	6.33	2.376
P17	P(p-tol) ₃	-0.180	-0.052	247.60	6.60	33.6	38.20	1.940	72.89	2.279	2.312	32.10	2.299	2.370	18.30	2.346
P18	P(p-F-C ₆ H ₄) ₃	-0.193	-0.056	236.50	8.30	32.7	36.20	1.941	69.28	2.275	2.311	35.30	2.301	2.365	16.80	2.336
P19	P(p-Cl-C ₆ H ₄) ₃	-0.201	-0.070	232.80	8.20	33.7	35.10	1.937	67.69	2.273	2.309	37.10	2.295	2.360	17.00	2.333
P20	P(p-OMe-C ₆ H ₄) ₃	-0.168	-0.037	251.60	8.30	31.7	37.90	1.942	72.39	2.278	2.319	30.70	2.305	2.374	16.20	2.345
P21	P(Ph)(Me) ₂	-0.188	-0.039	237.05	3.23	36.1	39.29	1.929	72.21	2.271	2.311	37.28	2.278	2.372	16.53	2.330
P22	P(Me)(Ph) ₂	-0.189	-0.048	239.20	4.83	34.6	37.62	1.933	70.77	2.276	2.313	35.33	2.286	2.369	16.02	2.333
P23	P(Ph)(Et) ₂	-0.193	-0.047	240.08	7.29	36.4	38.73	1.932	72.90	2.275	2.315	37.38	2.287	2.374	15.76	2.341
P24	P(Et)(Ph) ₂	-0.188	-0.048	240.76	5.41	34.3	37.54	1.935	71.36	2.277	2.315	33.94	2.304	2.371	16.11	2.339
P25	p(OMe)(Ph) ₂	-0.192	-0.059	239.45	5.15	45.0	39.39	1.918	73.01	2.273	2.314	37.27	2.274	2.367	20.45	2.317
P26	JohnPhos	-0.173	-0.052	253.20	45.88	19.0	31.54	1.961	70.73	2.298	2.329	22.54	2.366	2.380	-0.42	2.395
P27	CyJohnPhos	-0.178	-0.052	253.87	21.37	25.2	35.65	1.949	74.56	2.286	2.327	34.81	2.305	2.375	10.38	2.387
P28	MePhos	-0.180	-0.048	251.03	23.48	32.2	33.61	1.948	72.05	2.285	2.328	28.17	2.312	2.375	5.64	2.385
P29	SPhos	-0.170	-0.038	255.96	24.06	26.0	34.04	1.952	74.96	2.287	2.338	29.27	2.324	2.379	4.40	2.399
P30	XPhos	-0.181	-0.050	255.65	22.15	28.6	36.82	1.946	76.78	2.289	2.331	31.29	2.313	2.381	8.71	2.397

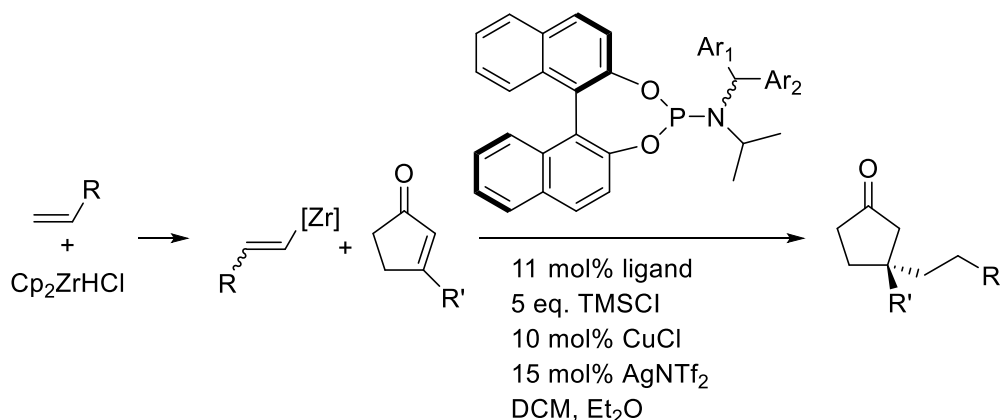
2.1.2.3 Reaction-Specific Descriptors

Scheme 3



The group of Doyle, in collaboration with Merck Research Laboratories, have utilised reaction-specific ligand descriptors and machine learning to predict the performance of Buchwald-Hartwig cross-coupling of aryl halides with 4-methylaniline (Scheme 3).⁹ They generated a set of 120 descriptors from DFT (B3LYP/6-31G*) calculations automated to run in the software package Spartan, capturing properties of the additives, aryl halides, bases and ligands used. For the four biaryl phosphine ligands considered, 64 ligand descriptors were harvested, consisting of electrostatic charges and NMR shifts for the 21 atoms (17 C, 4 H) shared across the ligands, as well as the frequency and intensity of ten shared vibrational modes, the dipole moment of the ligand molecule and the electrostatic charge on the phosphorus atom. These 120 descriptors, when used in conjunction with the random forest machine-learning algorithm, allowed for good prediction of reaction performance within their data set ($R^2=0.92$). They note, however, that this methodology is prone to predictive limitations when the training and test sets have significant structural dissimilarity. While the present contribution was already in review, this approach has been criticised in detail, see references 107-110.

Scheme 4



The (*S_a*,*R*,*R*) BINAP backbone was the most successful from four backbones screened. 32 Ar₁/Ar₂ structures were screened, with a 1-naphthylene group proving most successful.

The groups of Paton and Fletcher⁹² have recently reported the use of quantitative structure-selectivity relationships (QSSR) models to guide the development of ligands to achieve improved enantioselectivity in a copper-catalysed asymmetric conjugate addition reaction (Scheme 4). 15 chiral phosphoramidite ligands (see Scheme 4) were initially screened in this reaction, and the resulting selectivity was then examined using a database of 28 calculated descriptors (B97D/6-31G(d) incl. solvation), capturing the structures, energies, charges and spectroscopic properties of both the whole ligands, and of the aromatic substituents on these ligands. Regression models were then built, and their predictive performance evaluated, to select the most suitable subset of descriptors, which included the HOMO energy of the aryl rings and a subset of Sterimol parameters. This approach was supported by DFT calculations of the key selectivity-determining steps of the reaction. While

experimental screening with ligands selected from these predictions gave improved enantioselectivity, yields were initially low and required further experimental optimisation.

2.2 Bidentate P-donor Ligands

Chelating P,P-donor ligands share many of the desirable characteristics of their monodentate equivalents, as well as providing a potentially more well-defined coordination environment due to occupying two (usually *cis*) sites on a transition metal centre. Despite their synthetic popularity and utility,^{40, 51, 111, 112} their systematic characterisation through calculated ligand descriptors appears, to the best of our knowledge, less common-place, although earlier work by the group of Rothenberg is of note here.¹¹³⁻¹¹⁵

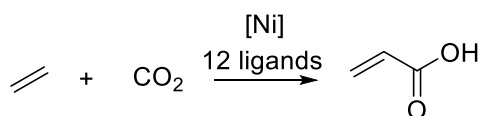
2.2.1 Individual Descriptors

This dearth may arise because one of the oldest and most commonly used descriptors,^{40, 111, 112, 116-119} the ligand bite angle ($\angle P-M-P$) is very easy to measure from atomic coordinates. This descriptor likely measures the net interaction with the metal centre, capturing a mixture of steric and electronic effects,¹¹⁷ which complicates its utilisation for catalyst design as only net effects are measured. In addition, unless so-called “natural” bite angles are determined, which is challenging as these require a molecular mechanics calculation with a modified force field,^{40, 116} the metal, its oxidation state and electronic configuration will affect the favoured geometry, altering the ligand field stabilisation energy and so imposing a structural “demand” on the ligand. This gives rise to multiple bite angles, extracted from different coordination environments, rather than a single, definitive protocol. Nevertheless, the bite angle has a status for bidentate ligands which is equivalent to Tolman’s descriptors for monodentates, in that it is frequently reported as the first characterisation of novel ligands.⁵⁴

Tolman’s 1977 review actually included cone angles for a range of bidentate ligands,¹⁶ but this descriptor has not been adopted widely. The group of Weigand have described an extension of the solid angle to bidentate ligands,¹²⁰ focussing on bidentate P,P-donor ligands. From this, using structural data mined from the Cambridge Structural Database¹²¹ for 282 square planar platinum complexes, they derived a “generalised equivalent cone angle”, θ_b . Data were processed in-house, and they also extracted bite angles for their complexes. While this work was based on crystal structure data, a similar approach could be employed using calculated geometries and this descriptor has been included in Table 9 below.

In a review of steric descriptors for ligands in organometallic chemistry, Clavier and Nolan have included buried volume, $\%V_{bur}$, data (see section 2.1.1.2) for a number of bidentate ligands,⁷² calculated for each half of the ligand using $[(AuCl)_2(PP)]$ complexes, as well as for the entire ligand in $[Pd(Cl_2)(PP)]$ complexes.⁸⁵ Again, they have derived these data from crystal structure geometries, but it would be feasible to generate a similar dataset from calculated coordinates.

Scheme 5



The steric effects of bidentate ligands for the nickel-catalysed coupling of carbon dioxide and ethene¹²² (Scheme 5) have been captured with the buried volume descriptor,^{85, 87} based on BP86-optimised

geometries. The authors noted that bite angles are not able to fully capture differences in steric hindrance. They combined % V_{bur} data with the Mulliken charge on the Ni atom as their electronic descriptor, with a view to identifying and utilising correlations between these parameters and calculated barriers along the catalytic cycle.

Buried volumes and solid cone angle descriptors (also summarised in section 2.1.1.2) were calculated for simplified surrogate ligands to represent a range of bidentate P,P-donor ligands by the groups of Sigman and Tan,¹²³ using the Solid-G programme of Guzei and Wendt.⁷⁶ Figure 3 illustrates the relationship between ligand and surrogate structures used in this case, with surrogates optimised at the M06-2X/def2TZVP level of theory; and the authors noted a linear correlation with the SambVca 2.0 buried volumes⁸⁷ in their references.¹²³ Electronic properties of the ligands were represented by the mean Pd-Cl distance in $[Cl_2Pd(PP)]$ complexes extracted from LKB-PP_{screen}¹²⁴ (discussed in section 2.2.2), as well as calculated NBO charges for simplified phosphine selenides and experimentally-determined ³¹P NMR chemical shifts (see also section 2.2.2 below).

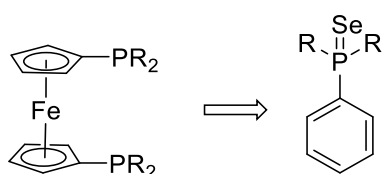


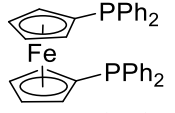
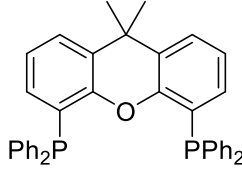
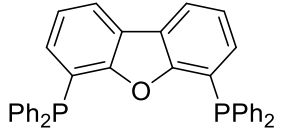
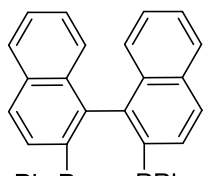
Figure 3: Comparison of ligand and surrogate structures used by Sigman and Tan.¹²³

A study by Lu and co-workers,⁶⁰ covering a range of different ligand classes and their effects on copper-catalysed boracarboxylation of styrene with CO_2 , includes 3 bidentate P,P donor ligands. However, only the charges on the α -carbon of styrene in the relevant copper complex have been used in this case, and the bulk of their analysis focusses on other ligands.

A range of individual descriptors have been collated in Table 9 for our core set of bidentate P,P-donor ligands.

Table 9: Individual descriptors for bidentate, P,P-donor ligands (units in ° unless otherwise stated). Ligands P2,P4, P10, P13 and P20 excluded due to lack of data.

Ligand No.	Bidentate Phosphine Ligand	Bite angle ^a	Error	Solid cone angle ^b	Exact solid cone angle ^b	P-Pt(Å) ^a	Av eq. cone angle ^a	Error	Av gen eq. cone angle (θ_b) ^a	Error	θ_b (P∩P) ^a	θ_b (SenSe) ^a
PP1		72.6	± 1.3	168	167.4	2.282	168.1	± 4.4	168.1	± 3.3	167.9	141.0
PP3		75.2	± 0.9	-	-	2.301	188.6	± 4.1	189.5	± 2.6	-	-
PP5		69.5	± 0.4	-	-	2.297	163.4	± 4.4	164.0	± 3.9	-	-
PP6		70.9	± 1.0	-	-	2.263	166.8	± 3.0	166.1	± 2.5	166.8	140.4
PP7		85.6	± 0.8	156	161.4	2.275	155.8	± 2.9	155.5	± 1.6	-	-
PP8		85.5	± 1.0	178	180.7	2.267	178.2	± 3.4	177.6	± 2.5	178.0	140.4
PP9		88.5	± 0.6	202	204.7	2.298	201.6	± 2.0	202.6	± 1.4	-	-
PP11		87.3	± 1.5	-	-	2.273	191.4	± 2.9	191.0	± 2.4	-	-
PP12		86.0	± 0.8	-	-	2.256	176.6	± 2.9	175.4	± 2.3	-	-
PP14		92.9	± 2.0	183	190.4	2.266	183.6	± 3.9	182.9	± 2.9	179.8	139.6
PP16		99.0	± 1.6	210	215.9	2.297	210.1	± 2.3	211.0	± 1.3	-	-
PP17		-	-	-	-	-	-	-	-	-	186.1	140.2
PP18		86.8	± 0.5	-	-	2.235	179.0	± 3.7	176.8	± 2.7	174.0	140.4

PP19		99.6	± 2.0	192	204.8	2.286	191.9	± 2.7	192.2	± 2.2	189.2	139.4
PP21		102.7	± 3.3	192	212.5	2.297	192.2	± 3.2	193.2	± 2.2	-	-
PP22		-	-	-	-	-	-	-	-	-	-	-
PP23		93.0	± 1.5	-	199.7	2.287	189.0	± 2.9	189.4	± 1.8	-	-

^a Reference 120; ^b reference 82.

2.2.2 Descriptor Databases

2.2.2.1 Ligand Knowledge Bases

Bristol's Ligand Knowledge Base approach (section 2.1.2.2) has been extended to bidentate ligands, and descriptors have been published as LKB-PP for 324 P,P-donor ligands, as well as 24 P,N-donor ligands, again calculated at the BP86/6-31G*, LACV3P on metals level of theory.¹²⁵⁻¹²⁷ In addition, the approach has been used to screen ligands assembled systematically from 25 backbones and 11 substituents, using less rigorous conformational search and optimisation protocols (LKB-PP_{screen}),¹²⁴ with a view to improving sampling of bidentate ligand space of the standard LKB-PP.

Some of the descriptors have been calculated from truncated ligand structures, designed to provide overlap with monodentate P-donor ligands (FMO energies and proton affinities), but while the database shares the LKB philosophy of robustness and transferability, again calculating the same ligand in different coordination environments with a standard DFT approach (BP86/6-31G* & LACV3P on M basis sets), the complexes reflect the chemistry of chelating ligands in this case, as shown in Figure 4. Descriptors have been listed in Table 10, with a subset of ligand data listed in Table 11 (see ESI for full set).

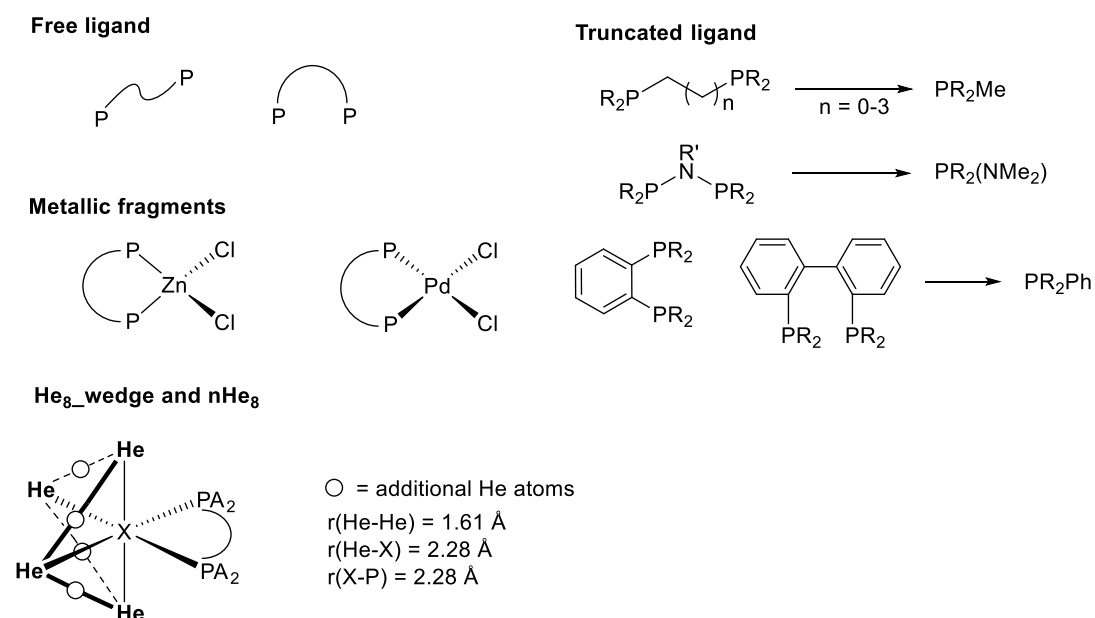


Figure 4: Complexes considered for LKB-PP and LKB-PP_{screen}.¹²⁴⁻¹²⁶

Table 10: Descriptors used in LKB-PP calculations (P1, P2 denote the two donor atoms).¹²⁴⁻¹²⁶

Descriptor	Derivation (Unit)
<i>Truncated Ligand</i> (L_{P1} , L_{P2})	
$E_{\text{HOMO}_{P1}}, E_{\text{HOMO}_{P2}}$	Energy of highest occupied molecular orbital (Hartree)
$E_{\text{LUMO}_{P1}}, E_{\text{LUMO}_{P2}}$	Energy of lowest unoccupied molecular orbital (Hartree)
PA_{P1}, PA_{P2}	Proton affinity, $PA = E(L_{P1}) - E([L_{P1}-H]^+)$ (kcal mol ⁻¹)
<i>Free Ligand</i>	
He ₈ _wedge	Interaction energy between ligand in chelating conformation and wedge of 8 He atoms, ^a $E_{\text{He8}_\text{wedge}} = E(\text{He}_8(\text{PP})) - E(\text{He}_8) - E((\text{PP}))$ (kcal mol ⁻¹)

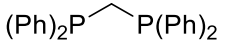
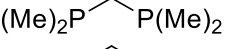
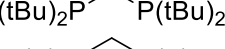
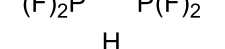
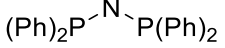
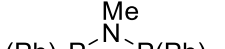
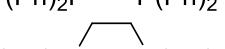
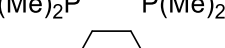
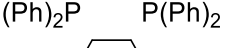



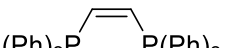
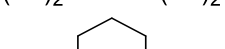
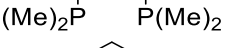
nHe ₈	Interaction energy between ligand in chelating conformation and wedge of 8 He atoms, ^b $E_{n_{\text{He}_8}} = E(\text{He}_8(\text{PP})) - E(\text{He}_8) - E(\text{PP})$ (kcal mol ⁻¹)
<i>Zinc complexes Zn(PP)Cl₂</i>	
BE(Zn)	Bond energy for dissociation of PP-ligand from metal fragment (kcal mol ⁻¹)
Zn–Cl	Average Zn–Cl distance (Å)
∠P1–Zn–P2	Ligand bite angle in complex (degrees)
ΔP1–R(Zn), ΔP2–R(Zn) ^c	Change in average P–R distances cf. PP (Å)
ΔR–P1–R(Zn), ΔR–P2–R(Zn) ^c	Change in average R–P–R angles cf. PP (degrees)
ΔZn–P1, ΔZn–P2	Change in Zn–P distances cf. ligand 1 (Å)
Q(Zn)	NBO charge on ZnCl ₂ fragment
<i>Palladium complexes Pd(PP)Cl₂</i>	
BE(Pd)	Bond energy for dissociation of PP-ligand from metal fragment (kcal mol ⁻¹)
Pd–Cl	Average Pd–Cl distance (Å)
∠P1–Pd–P2	Ligand bite angle in complex (degrees)
ΔP1–R(Pd), ΔP2–R(Pd) ^c	Change in average P–R distances cf. PP (Å)
ΔR–P1–R(Pd), ΔR–P2–R(Pd) ^c	Change in average R–P–R angles cf. PP (degrees)
ΔPd–P1, ΔPd–P2	Change in Pd–P distances cf. ligand 1 (Å)
Q(Pd)	NBO charge on PdCl ₂ fragment

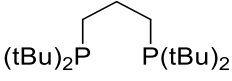
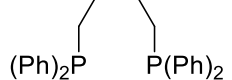
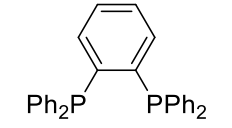
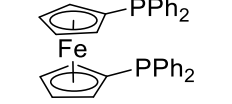
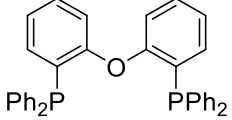
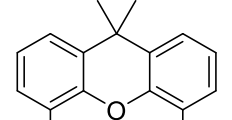
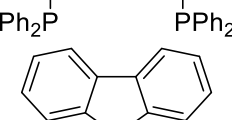
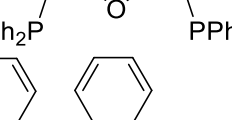
^a P atoms in fixed positions, fixed “P–X” distance = 2.28 Å; ^b Fixed “P–X” distances (2.28 Å), P atoms position free; ^c R = Substituents on P atoms.

These databases contain two bite angles, ∠P1–Pd–P2, determined from a [PdCl₂(PP)] complex, where the Pd(II) centre favours a square-planar coordination geometry, and ∠P1–Zn–P2, extracted from a [ZnCl₂(PP)] complex, which should allow ligands to adopt something closer to their “natural” bite angle. Electronic differences between ligands mean these descriptors are not purely steric, but they have been shown to correlate highly with crystallographic bite angles.¹²⁵ The LKB-PP also includes two purely steric parameters, derived from the repulsive interaction energies between ligands in their chelating orientation and a wedge of 8 helium atoms, positioned to capture interactions with *cis* ligands in an octahedral coordination environment.¹²⁶

Similar to the monodentate LKB-P, descriptor data have been used not just to generate maps of ligand space (section 3.1), but also to fit and predict different external datasets relevant to organometallic chemistry and catalysis. Both we¹²⁵ and others⁴⁰ have noted that further testing and applications crucially depend on the availability of large and varied experimental datasets.

Table 11: Subset of LKB-PP descriptors.^{125, 126} See ESI for full dataset.

Ligand No.	Bidentate Phosphine Ligand	PA_P1	PA_P2	He ₈ _wedge	nHe ₈	BE(Zn)	Mean Zn-Cl	<P1-Zn-P2	BE(Pd)	Mean Pd-Cl	<P1-Pd-P2
PP1		239.20	239.20	11.19	11.10	29.85	2.269	69.5	86.13	2.379	74.4
PP2		233.00	233.00	12.72	12.40	27.78	2.258	68.3	82.09	2.37	75.4
PP3		245.67	245.67	27.53	27.05	30.71	2.281	71.1	85.31	2.384	77.3
PP4		222.41	222.41	10.75	10.61	21.65	2.686	65.2	65.04	2.351	75.1
PP5		241.21	241.21	8.82	8.71	30.51	2.262	66.3	83.03	2.368	73.1
PP6		241.21	241.21	13.18	9.65	31.73	2.264	65.9	85.74	2.366	73.0
PP7		233.00	233.00	15.99	14.78	34.77	2.268	82.0	99.99	2.388	88.3
PP8		239.20	239.20	17.24	15.95	34.78	2.276	82.5	94.42	2.383	87.4
PP9		245.67	245.67	29.64	32.76	38.52	2.299	85.7	95.53	2.391	90.3
PP10		227.36	227.36	12.80	12.38	31.89	2.264	78.7	97.35	2.385	86.0
PP11		245.42	245.42	19.65	2.35	41.24	2.291	84.2	97.93	2.387	88.9
PP12		238.11	238.11	12.97	12.18	35.09	2.274	79.8	95.71	2.381	87.9
PP13		233.00	233.00	21.19	15.82	37.77	2.282	92.6	102.14	2.389	97.6
PP14		239.20	239.20	22.98	17.61	38.61	2.295	90.9	94.06	2.379	94.9
PP15		197.71	197.71	15.47	12.69	18.55	2.236	84.6	75.28	2.347	96.7

PP16		245.67	245.67	30.49	40.46	37.84	2.308	95.2	88.81	2.397	98.2
PP17		239.20	239.20	26.17	22.56	33.83	2.288	93.4	86.68	2.375	94.4
PP18		241.30	241.30	14.35	13.65	36.17	2.277	77.5	95.66	2.376	86.3
PP19		241.30	241.30	29.05	23.34	36.24	2.294	98.7	87.02	2.37	100.8
PP20		241.30	241.30	37.66	27.02	35.03	2.289	106.7	74.51	2.375	106.8
PP21		241.30	241.30	36.84	25.45	39.77	2.297	107.8	76.65	2.373	108.8
PP22		241.30	241.30	47.82	47.33	31.50	2.299	132.7	55.67	2.366	116.0
PP23		241.30	241.30	29.63	26.54	31.19	2.295	93.6	83.36	2.379	95.0

2.2.2.2 Reaction-Specific Descriptors

The groups of Frenking and Girolami¹²⁸ have reported a calculated (RI-BP86/def2-TZVP) version of a TEP-like descriptor derived from $[\text{Ni}(\text{CO})_2(\text{PP})]$ complexes, including 27 bidentate ligands, which, along with additional data harvested from these calculations, they then used to analyse ligand properties and to interpret the bonding in nickel dihydride complexes by energy decomposition analysis.

In a study of palladium-catalysed amine arylations using ligand descriptors to support the interpretation of experimental results, the groups of Sigman and Tan note that they were not able to find simple multivariate models using LKB-PP and LKB-PP_{screen} descriptors.¹²³ However, they used the average Pd-Cl distance in $[\text{Pd}(\text{Cl}_2)(\text{PP})]$, published in LKB-PP_{screen} for a range of ligands relevant to their study, as a measure of *trans* influence and σ -donor ability. As noted in section 2.1.2.1, these were combined with solid cone angles, NBO charges, as well as experimental NMR measurements.

2.3 Carbenes

Although carbene ligands have, at times, replaced P(III)-donor ligands in successful organometallic catalysis, not in the least in ruthenium-catalysed metathesis (see also section 3.2),^{11, 129, 130} in more general terms they are more likely to support *complementary* reactivity¹³¹⁻¹³⁹ than to displace other classes of ligands completely. As was the case for P(III)-donor ligands, a number of steric and electronic descriptors have been used in isolation (section 2.2.1), and several descriptor databases (section 2.2.2) have also been proposed, so this section will follow a similar structure. Many of these carbene studies have been reviewed previously, see, for example, references 72, 140-142, again allowing us to focus on more recent work and application examples. In addition, Munz has recently presented a review of the experimental and computational studies of carbene electronic structure, which draws on a range of different descriptors from experiment and calculations as well.¹⁴³

2.3.1 Individual Descriptors

2.3.1.1 Electronic Descriptors

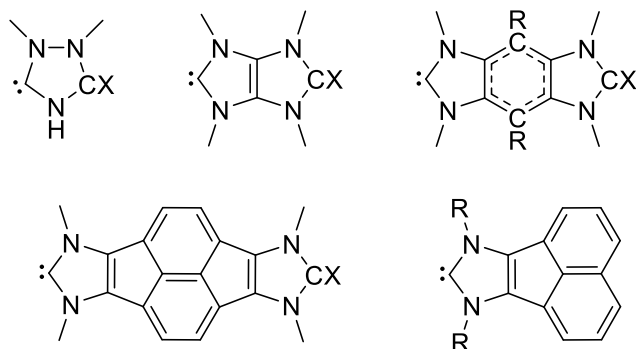
As described in section 2.1.1.1 above, the Tolman Electronic Parameter (TEP) and related descriptors have been used extensively for P(III)-donor ligands, making this a tempting parameter to consider for other ligand classes. However, experimentally it proved to be difficult to synthesise some of the $[\text{Ni}(\text{CO})_3\text{L}]$ complex reliably for large N-heterocyclic carbene ligands (NHCs).^{144, 145} Critical assessment of the data obtained also highlighted that the TEP stretching frequency showed a much smaller response to changes in the ligand structure than observed for P-donor ligands, as summarised in a recent review of experimentally-measured electronic parameters by Huynh.¹⁴⁶ That, and the criticisms of Cremer and co-workers about using normal mode vibrations discussed in section 2.1.1.1^{63, 65} notwithstanding, Gusev presented DFT-calculated TEP data for 76 NHC ligands (mPW1PW91/6-311+G(d,p) for all atoms apart from 6-311+G(2d) on Ni), which allowed a ranking of their net electron donation.¹⁴⁷

Gusev's TEP electronic descriptor has been used recently for the analysis of calculated reaction barriers,⁶⁰ although the authors focussed on %V_{bur} for the 6 NHC ligands considered, after noting that the correlation between such barriers and the TEP was poor, in contrast to the high correlations with electronic property descriptors identified for monophosphines.

Gusev⁵⁹ and others (see references 146 and 142 for overviews of experimental work) have also explored the utility of other transition metal complexes for producing data which could then be brought into line with TEP, based on determining a suitable linear relationship between different

complexes^{141, 142, 146} and so processing data to produce a TEP-equivalent value. The complexes used most commonly for these alternative measurements are [RhCl(CO)₂(NHC)] and [IrCl(CO)₂(NHC)], but other metal centres have also been considered.⁵⁹

Scheme 6



More recently, Gusev and Peris¹⁴⁸ have used a calculated TEP descriptor (mPW1PW91, range of basis sets of TZP quality and better), which, when compared to calculated CO stretches in a range of metal complexes, can be used to give a new descriptor, Δ TEP, as a measure of the electronic communication between two metal centres in dinuclear M-NHC-M' complexes, more commonly assessed by electrochemical measurements. They moved beyond consideration of the ligand by comparing the TEP from the standard nickel complex with calculated CO stretches in other metal fragments. This allowed them to identify the net electron donicity of metal fragments, and so to quantify the extent of electron communication for five ligands (Scheme 6), as well as highlighting that such communication is mostly due to differences in σ -bonding.

The groups of Cianceleoni and Belpassi⁶² reported calculated carbonyl stretching frequencies for the nickel and gold complexes of general form $[M(CO)_nL]^{\pm/0}$ for a diverse range of ligands, including 4 NHCs. They then extracted a charge transfer parameterⁱⁱ from these DFT calculations (BLYP/TZ2P) and used this to highlight that for the gold complex, the CO stretching frequency does not provide a good measure of the net charge transfer, as discussed above. They also included a brief comparison of phosphine and carbene ligands, noting that, according to their analysis, the phosphines donate more electron density to gold than carbenes, a result they describe as “unconventional”.

The dataset of 181 nickel complexes, presented by Cremer and co-authors,⁶⁴ to provide calculated MLEP, an alternative electronic parameter to the standard TEP derived from local vibrational modes, included 15 carbene ligands. The MLEP descriptors have been calculated at the M06/aug-cc-pVTZ of theory and further analysis of force constants revealed differences between Arduengo carbenes (NHCs) and other carbene ligands due to differences in their interaction with the metal centre, with the NHCs withdrawing charge density from the metal and showing stronger Ni-C bonding.

Table 12 includes experimental and calculated electronic parameters related to CO stretching frequencies for our core set of carbenes.

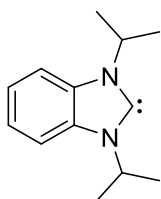
ⁱⁱ See section 2.3.2.3 for other applications of charge transfer analysis.

Table 12: Descriptors related to CO stretching frequencies for carbene ligands (see Table 1c for structures). C4 removed due to lack of data.

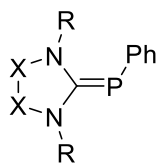
Ligand No.	Carbene Ligand	Acronym	Gusev NHC Paper ^a			Nolan Review ^b		HEP ^c
			TEP[Ni(CO) ₃ NHC]	Repulsiveness	ΔH(kcal/mol)	TEP[Ir(Cl)(CO) ₂ (NHC)]	v _{av} (CO)	
C1	slmN(H) ₂	-	2057.5	0.3	22.6	-	-	-
C2	slmN(Me) ₂	SIMe	2054.7	1.8	19.5	2052.0	2026.0	-
C3	slmN(iPr) ₂	SIiPr	2051.9	2.3	18.6	-	-	-
C5	slmN(Dipp) ₂	SIPr	2051.5	1.6	20.2	2051.1	2024.9	177.5
C6	slmN(Mes) ₂	SIMes	2051.2	1.9	19.5	2050.8	2024.6	177.6
C7	slmN(2,6-Me-Ph) ₂	SIXy	-	-	-	2051.2	2025.0	-
C8	lmN(H) ₂	-	2058.1	0.0	23.3	-	-	-
C9	lmN(Me) ₂	IMe	2054.1	1.1	21.1	2051.2	2025.0	-
C10	lmN(iPr) ₂	liPr	2051.5	1.6	20.2	2050.3	2024.0	180.6
C11	lmN(tBu) ₂	l ^t Bu	2050.6	7.0	8.2	2048.9	2022.3	177.6
C12	lmN(Cy) ₂	ICy	2049.7	1.6	20.2	2049.5	2023.0	181.2
C13	lmN(Ph) ₂	IPh	2053.6	2.3	18.6	-	-	-
C14	lmN(Bn) ₂	IBn	-	-	-	2050.3	2024.0	179
C15	lmN(Dipp) ₂	IPr	2050.5	1.8	19.7	2050.2	2023.9	177.5
C16	lmN(Mes) ₂	IMes	2050.5	1.5	20.6	2049.6	2023.1	177.2
C17	lmN(2,6-Me-Ph) ₂	IXy	-	-	-	2050.3	2024.0	-
C18	lmN(Ad) ₂	IAd	2045.8	8.0	6.2	2048.3	2021.6	-
C19	lm(Me) ₂ N(Me) ₂	^{Me} IMe	2051.7	1.4	20.7	-	-	-
C20	lm(F) ₂ N(Me) ₂	^F IMe	2059.1	1.0	20.9	-	-	-
C21	lm(Cl) ₂ N(Me) ₂	^{Cl} IMe	2059.0	1.3	20.4	-	-	-
C22	lm(NO ₂) ₂ N(Me) ₂	^{NO₂} IMe	2068.6	1.6	19.0	-	-	-
C23	BImN(Me) ₂	BMe	2057.0	1.6	19.8	-	-	-
C24	Py(b)ImN(Me) ₂	-	2058.6	1.6	19.8	-	-	-
C25	DPylm	-	2055.9	0.8	21.8	-	-	-
C26	IBioxMe ₄	-	2049.5	2.8	17.7	2050.3	2024.0	-
C27	PerN(iPr) ₂	-	2055.4	6.5	9.0	-	-	-
C28	ThNMe	-	2061.5	0.9	21.2	-	-	-
C29	OxNMe	-	2065.3	0.4	22.0	-	-	-
C30	BOxNMe	-	2066.6	0.6	21.3	-	-	-

^a Reference 147; ^b reference 142; ^c reference 146.

Scheme 7a

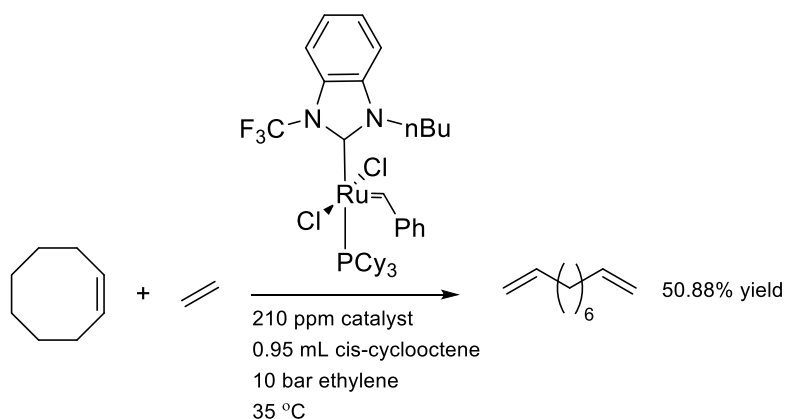


Scheme 7b

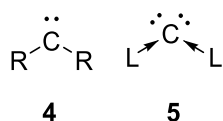


Experimentally, carbene ligand bonding to metal centres has also been assessed by a range of different descriptors derived from NMR studies (reviewed, for example, in references 142, 146). Different nuclei and complexes have been used, such as the ^{13}C chemical shift of the carbene carbon *trans* to the ligand (L) of interest in $[\text{PdBr}_2(\text{iPr}_2\text{-bimy})\text{L}]$ (Scheme 7a) proposed by Huynh (and called the Huynh Electronic Parameter, HEP),^{146, 149} the ^{77}Se chemical shifts for selenium adducts of carbenes,^{146, 150-153} and the ^{31}P chemical shifts for NHC-phosphinidene adducts (Scheme 7b).^{142, 146, 150, 151, 153}

Scheme 8



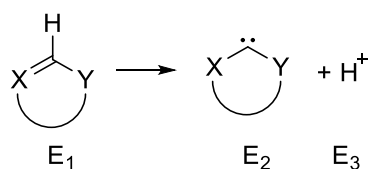
Such chemical shifts should be amenable to calculation as well. Indeed, the analysis of the selectivity of ethenolysis of cyclic alkenes catalysed by ruthenium-NHC complexes (Scheme 8), reported by a consortium of authors,¹⁵⁴ was supported by both experimental and calculated (PBE0/TZ2P) chemical shift tensor data for selenium NHC complexes, used as a measure of ligand electronic properties.



In an effort to distinguish between carbenes **4** and carbenes **5** (CL_2 compounds with carbon(0), formally with two lone pairs), calculated ^{13}C chemical shifts of protonated and parent carbenes (PBE1PBE/6-311++G*//PBE1PBE/6-31+G*) have been used to assess the electronic properties of 8 carbenes and 13 carbenes.¹⁵⁵ Some of the authors involved in this work have also used a wider range of calculated descriptors, including ^{31}P chemical shifts of carbene-phosphinidene adducts (Scheme 7b) and donor carbon ^{13}C chemical shifts for the parent carbene and their *cis*- $[\text{RhCl}(\text{CO})_2\text{L}]$ complexes, to evaluate the effects of structural modifications on NHC properties.^{156, 157} The initial experimental work by Ganter and co-workers^{151, 152} on carbene-selenium and carbene-phosphinidene adducts was followed up in a study by Vummaleti, Nolan, Cavallo and co-workers¹⁵³ for a larger set of carbene ligands, using both experimental and DFT-calculated (BP86/TZ2P) ^{31}P and ^{77}Se chemical shifts, together with a more extensive analysis of the experimental data using a wider range of calculated

descriptors. More recently, a further computational study of carbene-phosphinidenes (Scheme 7b), correlating their ^{31}P chemical shifts with a range of calculated descriptors to analyse the bonding observed, has been reported for 21 structurally varied carbenes,¹⁵⁸ confirming and reinforcing the computational analysis presented by Vummaleti et al.¹⁵³ These studies will be discussed in greater detail in section 2.3.2 on descriptor databases for carbenes, below. Table 13 collects NMR-derived descriptors from both experimental and calculated data.

Scheme 9



While not applied to catalysis so far, we note that Ramsden and Oziminski have proposed a calculated descriptor,^{159, 160} the Carbene Relative Energy of Formation (CREF, B3LYP/6-311++G(d,p)), which seeks to capture the energy required to deprotonate the heterocyclic precursor E_1 of a broad and varied range of carbenes E_2 (Scheme 9). They proposed that this energy of deprotonation, calculated from the zero-point energy-corrected potential energies of parent carbene and protonated form, will give an indication of the NHC σ -donor strength, excluding contributions from metal to ligand π -bonding and substituent steric effects.

Table 13: Electronic descriptors reported for carbenes (see Table 1c for ligand structures) and derived from NMR shifts. C1, C3-4, C7-8, C14, C17, C20-24, C26, C28-30 omitted due to lack of data.

Ligand No.	Carbene Ligand	Acronym	⁷⁷ Se NMR Shift (ppm) (Acetone-D ₆) ^a	⁷⁷ Se NMR Shift (ppm) (CDCl ₃)[DFT] ^b	⁷⁷ Se iso (ppm)(CDCl ₃) ^c	Calculated		¹³ C NMR Shift (D ₆ -DMSO) ^e	³¹ P NMR Shift Calculated (Experimental) ^f	³¹ P NMR Shift (ppm) (benzene-D ₆) [DFT] ^b
						¹³ C NMR Shift (ppm) ^d	¹³ C NMR Shift (Protonated) ^d			
C2	sImN(Me) ₂	SIMe	-	-	-	236.1	153.3	-	-	-
C5	sImN(Dipp) ₂	SIPr	181.0	190.0 [1347.2]	-	-	-	-	-12.0 (-10.2)	-10.2 [383.1]
C6	sImN(Mes) ₂	SIMes	116/113	109.7 [1462.6]	95.0	-	-	160.2	(-10.4)	-10.4 [387.3]
C9	ImN(Me) ₂	IMe	-	29.9 [1531.1]	-	214.9	133.8	-	-	-
C10	ImN(iPr) ₂	liPr	-3.0	-	-	-	-	-	(-61.2)	-61.2 [430.2]
C11	ImN(tBu) ₂	l ^t Bu	197.0	182.6 [1354.9]	-	-	-	-	-	-
C12	ImN(Cy) ₂	ICy	-4.0	-22.1 [1617.0]	-	-	-	-	-	-
C13	ImN(Ph) ₂	IPh	-	-	-	-	-	-	-	-
C15	ImN(Dipp) ₂	IPr	87.0	90.0 [1441.9]	-	-	-	139.4	-17.6 (-18.9)	-18.9 [389.6]
C16	ImN(Mes) ₂	IMes	35.0	26.7 [1535.7]	26.0	-	-	139.9	(-23.0)	-23.0 [402.1]
C18	ImN(Ad) ₂	IAd	-	196.9 [1360.6]	-	-	-	-	-	-
C19	Im(Me)2N(Me) ₂	^{Me} lMe	3.0	-	-	-	-	-	-62.3 (-53.5)	[424.0]
C21	Im(Cl) ₂ N(Me) ₂	^{Cl} lMe	-	-	-	-	-	-	-43.1	-
C25	DPyIm	-	-	-	-	200.3	111.4	-	-	-
C27	PerN(iPr) ₂	-	364.0	-	-	-	-	-	-	-

^a Reference 152; ^b reference 153; ^c reference 154; ^d reference 155; ^e reference 149; ^f references 146, 158.

2.3.1.2 Steric Descriptors

Cone angles and related measures can be determined for carbenes, but carbenes are less cone-shaped and often less symmetrical in terms of substituents, making such measure more problematic, especially for NHCs.¹⁶¹ As set out in the review by Clavier and Nolan,⁷² the “percent buried volume” (%V_{bur}) was proposed as a suitable alternative, capturing the percentage of the total volume of a sphere, centred on the metal and having a defined radius occupied by the ligand. Any source of cartesian coordinates can be used for such calculations, which are very fast, and the effect of altering the M-L distance and sphere radius have been explored extensively, with standard settings implemented in convenient online tools (SambVca and SambVca 2.0).⁸⁵⁻⁸⁸

The %V_{bur} descriptor has been reviewed extensively,^{54, 72, 141, 161} including demonstrations of its correlations with cone angle for P-donor ligands,^{17, 72} and, more recently, this has been supplemented by steric maps which provide a graphical representation of the steric profile of ligands¹⁶¹ and can be produced using SambVca 2.0.⁸⁷ Carbenes can undergo conformational change in response to the coordination environment and this has been explored in terms of %V_{bur} and steric maps in reference 162. Large-scale automated sampling of conformers, *e.g.* as facilitated by AARON, can help to capture such effects on catalysis,⁷⁴ emphasising the importance of such descriptors. Outside of databases, %V_{bur} appears to be used most commonly to assess the impact of structural changes to the NHC substituents and backbone. This descriptor has been applied to the analysis of carbene experimental data by a consortium of authors exploring the selectivity of ruthenium-NHC catalysts for metathesis,¹⁵⁴ along with chemical shift tensor data for selenium NHC complexes as discussed above (section 2.3.1.1).

Along with calculated TEP descriptors (section 2.3.1.1), Gusev’s study of 76 NHCs¹⁴⁷ includes a proposal for a steric parameter *r*. This “repulsiveness” parameter (Table 12) has been derived from an analysis of calculated reaction enthalpies for the decarbonylation of [Ni(CO)₃L], defined as $r = 10 \times (3.493 - d(\text{Ni-C}))$, with *d*(Ni-C) as the distance from Ni to the C atoms of NMe substituent on L in [Ni(CO)₂L] complexes; the data range from 0.0 to 8.0 and measure the repulsive interaction between the NHC and the carbonyl ligands.

Table 14 summarises individual steric descriptors for a range of carbene ligands.

Table 14: %V_{bur} data for carbenes (see Table 1c for ligand structures). (C7, C14, C17, C20, C22, C24-C30 omitted due to lack of data.)

Ligand No.	Carbene Ligand	Acronym	Nolan ^a		SambVca Paper ^b
			%V _{Bur} (2 Å)	%V _{Bur} (2.28 Å)	%V _{Bur}
C1	sImN(H) ₂	-	-	-	19
C2	sImN(Me) ₂	SIMe	-	-	25.4
C3	sImN(iPr) ₂	SliPr	28.2	24.2	-
C4	sImN(Ph) ₂	SIPh	-	-	31.6
C5	sImN(Dipp) ₂	SIPr	47.0	41.5	35.7
C6	sImN(Mes) ₂	SIMes	-	-	32.7
C8	ImN(H) ₂	-	-	-	18.8
C9	ImN(Me) ₂	IMe	26.3	22.6	24.9
C10	ImN(iPr) ₂	liPr	27.5	23.5	-
C11	ImN(tBu) ₂	l ^t Bu	39.6	35.1	35.5
C12	ImN(Cy) ₂	lCy	27.5	23.6	-

C13	ImN(Ph) ₂	IPh	-	-	30.5
C15	ImN(Dipp) ₂	IPr	45.4	39.0	33.6
C16	ImN(Mes) ₂	IMes	36.5	31.2	31.6
C18	ImN(Ad) ₂	IAd	39.8	35.3	36.1
C19	Im(Me) ₂ N(Me) ₂	^{Me} IMe	26.2	22.6	-
C21	Im(Cl) ₂ N(Me) ₂	^{Cl} IMe	26.3	22.7	-
C23	BImN(Me) ₂	BMe	-	-	25.1

^a Reference 72; ^b reference 85.

2.3.2 Descriptor databases

As noted above, one of the key features of descriptor databases is that these were generally built over relatively long timeframes and the descriptors have often been curated to fit in with the authors' design philosophy, as can be seen for the MESP and LKB approaches. As before, earlier work has been reviewed previously,^{12, 54} but the comprehensive studies by Jensen and co-workers are of particular note in this area.^{11, 163}

2.3.2.1 Molecular Electrostatic Potential Data

Suresh and co-workers have applied their analysis of the molecular electrostatic potential (MESP) to a range of carbenes^{99, 100} and used this approach to quantify the steric and electronic factors that affect carbenes in a range of settings, including alkene metathesis,¹⁰⁰ CO₂ fixation¹⁶⁴ and the oxidative addition step to palladium catalysts (see also section 3.2.2).¹⁰¹

Initial analysis⁹⁹ considered carbenes in isolation, determining the absolute minimum of the MESP at the carbene lone pair (V_{\min}), together with the MESP at the carbene nucleus (V_C). The relationship of both descriptors with TEP (Gusev¹⁴⁷ is cited, but it is not clear what the source of TEP data is) is described well by a linear equation and they noted that functional choice affects the magnitude of descriptor values, but not the trends observed, reporting B3LYP/6-311++G** data in the main body of their work. They also noted the great sensitivity of V_{\min} to the environment of the lone pair, suggesting a steric contribution, whereas V_C is considered less likely to be affected, leading them to propose that V_{\min} measures the electron-rich character of free NHC ligands, whereas V_C measures the metal-coordinating power of the ligand. Comparison with data for free P-donor ligands reported previously⁹⁸ led them to note that NHC ligands coordinate metal centres more strongly than phosphines, in line with other work.

In a later study,¹⁰⁰ MESP-related data, calculated with BP86/def2-TSVPP, have been used to investigate ligand effects on the coordination of substrates to a ruthenium metathesis catalyst, with a view to determining the steric and electronic effects on the binding energy of ethene and the activation energy for formation of the metallacyclobutane intermediate. This study followed the same process as their work on first generation catalysts,¹⁰⁴ reviewed in section 2.1.2.1, but their descriptors have been calculated in subtly different ways to take account of differences in coordination between the ligand classes: The combined stereoelectronic effect was derived from the sum of the MESP values at the NHC carbon nucleus and at the CH₂ carbon nucleus, and individual steric and electronic descriptors were again determined from calculations on the active form of the catalyst **2** and the ethene-bound complex **3** (L=NHC), with comparison between full substituents and H-substituted ligands used to determine individual steric and electronic contributions. The steric descriptor was found to correlate highly with % V_{bur} , and the analysis highlighted the different steric demands of different ethene coordination modes; the MESP calculations used the lowest energy binding mode, which was found to be different for ligands with larger substituents, necessitating separate analyses of the two groups of ligands. This led them to conclude that steric control dominates for bulky N substituents on the NHC

ligands, whereas a mixture of steric and electronic effects contribute to the reactivity of smaller ligands, in line with results reported by other authors.^{11, 165}

Most recently, Suresh's group¹⁰¹ have investigated the oxidative addition step of monoligated palladium catalysts with both general V_{\min} and V_D (where D is the donor atom) MESP data, and reaction-specific analysis of the Pd nucleus in PdL₂ and PdL complexes of these ligands (see sections 2.1.2.1 and 3.2.2). This study included 10 carbenes, along with phosphines, alkenes and alkynes. These descriptors have been related to calculated barriers for the oxidative addition of Ph-X, allowing comparison of different ligand classes and leading to a common design strategy, which will be discussed in greater detail in section 3.2.2 below; the general ligand descriptors have been collected in Table 15, with reaction-specific data collated in Table 19.

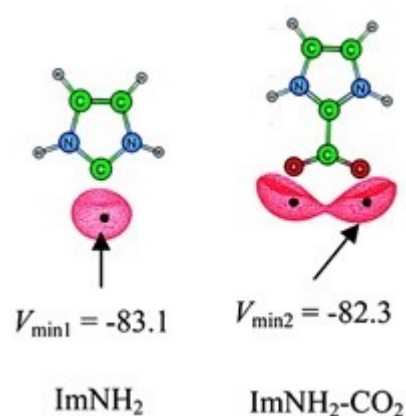


Figure 5: MESP minima at both the carbene lone pair and the oxygen lone pairs. Reprinted with permission from reference 164. Copyright 2012, American Chemical Society.

While not applied to organometallic catalysis, Ajitha and Suresh have also used the MESP approach to a model for organocatalytically-active NHC ligands,¹⁶⁴ namely capturing the formation of an NHC-CO₂ adduct. In this case, their analysis combined the MESP minimum at the lone pair region ($V_{\min1}$) with a second V_{\min} at the oxygen lone pair ($V_{\min2}$, Figure 5). The data have been processed further by calculating the change compared to an unsubstituted reference system where all sites have been occupied with Hs. These descriptors were then related to the energy released on adduct formation, allowing correlations to be established. This study presented data for 54 carbenes, including substantial variation of N and backbone substituents. V_{\min} data have been collected in Table 15.

Table 15: MESP-related descriptors for carbene ligands, see Table 19 for reaction-specific data used with this approach (free ligand V_{\min} , V_C ^{99, 101, 164})

Ligand No.	Carbene Ligand	Acronym	$V_{\min1}$ (kcal mol ⁻¹) ^a	$V_{\min2}$ (kcal mol ⁻¹) ^a	V_{\min} (kcal mol ⁻¹) ^b	V_C (kcal mol ⁻¹) ^b
C1	sImN(H) ₂	-	-	-	-80.87	-9266.10
C2	sImN(Me) ₂	SIMe	-	-	-81.38	-9271.35
C3	sImN(iPr) ₂	SliPr	-	-	-82.13	-9273.92
C8	ImN(H) ₂	-	-83.1	-82.3	-79.07	-9265.13
C9	ImN(Me) ₂	IMe	-87.2	-86.0	-80.43	-9273.06
C10	ImN(iPr) ₂	liPr	-87.5	-87.1	-81.68	-9275.36

C11	ImN(tBu) ₂	tBu	-	-	-78.98	-9278.44
C12	ImN(Cy) ₂	ICy	-86.8	-85.5	-82.79	-9276.91
C13	ImN(Ph) ₂	lPh	-75.5	-82.0	-	-
C16	ImN(Mes) ₂	lMes	-83.9	-84.8	-	-
C19	Im(Me) ₂ N(Me) ₂	MeIMe	-88.2	-86.5	-84.79	-9278.80
C20	Im(F) ₂ N(Me) ₂	FlMe	-73.7	-76.8	-69.46	-9258.89
C21	Im(Cl) ₂ N(Me) ₂	ClMe	-72.9	-77.4	-69.20	-9259.65
C22	Im(NO ₂) ₂ N(Me) ₂	NO ₂ Me	-54.7	-65.6	-50.75	-9235.71
C23	BimN(Me) ₂	BMe	-79.3	-80.8	-76.02	-9263.77
C24	Py(b)ImN(Me) ₂	-	-75.9	-80.6	-	-
C25	DPyIm	-	-77.9	-85.9	-74.54	-9265.40

^a Reference 164; ^b reference 99.

2.3.2.2 Ligand Knowledge Base (LKB-C)

The Ligand Knowledge Base approach has also been applied to carbenes and related C-donor ligands, giving LKB-C.¹⁶⁶ While the initial work included 100 carbenes, calculated with BP86/6-31G*, LACV3P on metals, substantial ligand data are currently held in-house and in preparation for publication.¹⁶⁷ The philosophy and approach in this database are very similar to work on P-donor ligands, but some of the complexes (Figure 6) and the descriptors derived from these calculations (Table 16) have changed to accommodate the differences in coordination behaviour and ligand shape, as alluded to in other studies.

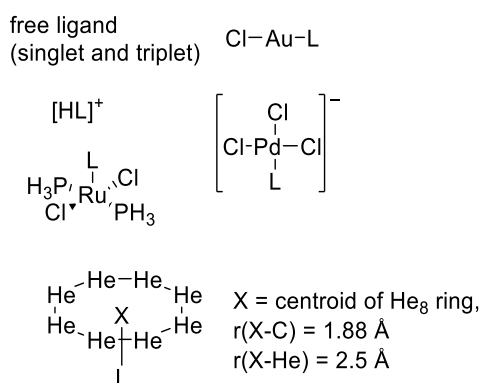


Figure 6: Complexes used in LKB-C.¹⁶⁶

In brief, the He₈_steric descriptor uses a shorter distance between the ligand and the ring of helium atoms, and a square-based pyramidal ruthenium(II) fragment has been added as the angular changes around the metal centre capture differences in NHC size and coordination behaviour. The energy difference between triplet and singlet electronic configurations of the free ligand has also been included, allowing an indirect assessment of ligand stability (discussed further in section 2.3.2.3). Table 16 summarises the descriptors used in LKB-C, while Table 17 presents a subset of calculated descriptors, with the full data summarised in the supporting information; processing of these descriptor to give a map of carbene chemical space will be discussed in section 3.1 below.

Table 16: Descriptors in LKB-C.¹⁶⁶

descriptor ^a	derivation (unit)
<i>Free Carbene Species (L, descriptors for singlet and triplet configurations)</i>	
E _{HOMO(s)}	energy of highest occupied molecular orbital (Hartree)
E _{LUMO(s)}	energy of lowest unoccupied molecular orbital (Hartree)
E _{t-s}	E _{t-s} = E(triplet) – E(singlet) (kcal mol ⁻¹)
He ₈ _steric	interaction energy between singlet L in ground state conformation and ring of 8 helium atoms; E _{ster} = E _{tot(system)} – [E _{tot(He₈)} +E _{tot(L)}] ^b (kcal mol ⁻¹)
<i>Protonated Ligand ([HL]⁺)</i>	
PA	proton affinity; calculated as the difference between the energy of the neutral and protonated singlet L (kcal mol ⁻¹)
<i>Gold Complexes ([AuClL])</i>	
Q(Au fragm.)	NBO charge on AuCl fragment
BE(Au)	bond energy for dissociation of L from [AuCl] fragment (kcal mol ⁻¹) ^c
Au-Cl	r(Au-Cl) (Å)
Au-C	r(Au-C) (Å)
Δ C-A (Au)	change in av. r(C-A) in complex compared with singlet L (Å)
Δ A-C-B (Au)	change in av. <(A-C-B) in complex compared with singlet L (°)
<i>Palladium Complexes ([PdCl₃L]⁻)</i>	
Q(Pd fragm.)	NBO charge on [PdCl ₃] ⁻ fragment
BE (Pd)	bond energy for dissociation of L from [PdCl ₃] ⁻ fragment (kcal mol ⁻¹) ^c
Pd-Cl <i>trans</i>	r(Pd-Cl), <i>trans</i> to ligand (Å)
Pd-C	r(Pd-C) (Å)
Δ C-A (Pd)	change in av. r(C-A) in complex compared with singlet L (Å)
Δ A-C-B (Pd)	change in av. <(A-C-B) in complex compared with singlet L (°)
<i>Ruthenium Complexes (trans-[RuCl₂(PH₃)₂L])</i>	
Q(Ru fragm.)	NBO charge on [RuCl ₂ (PH ₃) ₂] fragment
BE (Ru)	bond energy for dissociation of L from [RuCl ₂ (PH ₃) ₂] fragment (kcal mol ⁻¹) ^c
Ru-C	r(Ru-C) (Å)
Ru-Cl	av. r(Ru-Cl)
Ru-P	av. r(Ru-P)
Δ C-A (Ru)	change in av. r(C-A) in complex compared with singlet L (Å)
Δ A-C-B (Ru)	change in av. <(A-C-B) in complex compared with singlet L (°)
< Cl-Ru-Cl	< Cl-Ru-Cl (°)
< P-Ru-P	< P-Ru-P (°)

^a All calculations were performed on isolated molecules; ^b centroid-donor distance = 1.88 Å; ^c BE = [E_{tot(fragment)} + E_{tot(L)}] – E_{tot(complex)}

To establish some of the descriptors, correlations with other data were explored in this initial report, including comparing the He₈_steric descriptors with %V_{bur} data; this showed similar trends were captured by the two steric descriptors. However, detailed analysis suggested some differences in the responsiveness of data, with He₈_steric achieving slightly improved resolution, possibly due to capturing the steric demands of a metal coordination environment better. Correlations between electronic descriptors were also explored, illustrating that the Pd-Cl distance *trans* to the carbene ligand appears to capture ligand effects similar to TEP and CEP data.

Table 17: Subset of LKB-C descriptors. See ESI for full dataset and Table 1c for ligand structures.¹⁶⁶

Ligand No.	Carbene Ligand	Acronym	E _{t-s}	PA	He ₈ steric	BE (Au)	Au-Cl	BE (Pd)	Pd-Cl <i>trans</i>	BE (Ru)	Ru-C	< Cl-Ru-Cl	< P-Ru-P	Mean Ru-Cl	Mean Ru-P
C1	sImN(H) ₂	-	102.95	268.79	2.24	91.66	2.306	61.02	2.378	83.37	1.924	174.7	174.7	2.441	2.333
C2	sImN(Me) ₂	SIMe	73.17	274.63	17.94	91.20	2.309	50.51	2.383	69.24	1.960	155.2	173.3	2.431	2.338
C3	sImN(iPr) ₂	SiPr	85.69	280.28	25.67	92.46	2.314	53.56	2.385	67.77	1.972	156.4	170.1	2.434	2.337
C4	sImN(Ph) ₂	SIPh	50.38	281.85	12.50	94.03	2.310	55.90	2.376	69.93	1.938	157.5	162.2	2.422	2.346
C5	sImN(Dipp) ₂	SIPr	60.46	279.12	20.85	87.03	2.316	49.22	2.377	58.31	1.937	163.5	153.5	2.427	2.341
C6	sImN(Mes) ₂	SIMes	42.06	283.25	17.40	90.95	2.318	48.90	2.379	67.45	1.951	162.9	154.4	2.432	2.338
C7	sImN(2,6-Me-C ₆ H ₃) ₂	SIXy	69.52	279.70	18.66	89.11	2.314	47.81	2.379	65.78	1.950	162.3	154.1	2.431	2.340
C8	ImN(H) ₂	-	82.53	265.89	2.24	90.29	2.304	61.62	2.374	80.27	1.944	176.1	175.4	2.449	2.330
C9	ImN(Me) ₂	IMe	83.66	273.63	13.62	90.44	2.310	50.87	2.382	68.54	1.973	157.3	173.0	2.434	2.336
C10	ImN(iPr) ₂	IiPr	84.27	279.32	24.21	92.35	2.313	52.37	2.387	67.35	1.982	157.7	172.3	2.438	2.334
C11	ImN(tBu) ₂	I ^t Bu	81.75	282.09	67.59	88.12	2.318	44.51	2.385	29.09	2.007	146.1	160.6	2.440	2.349
C12	ImN(Cy) ₂	ICy	84.42	282.62	24.63	92.92	2.314	52.74	2.386	67.56	1.984	157.2	172.8	2.438	2.334
C13	ImN(Ph) ₂	IPh	68.34	273.20	19.42	82.35	2.312	44.93	2.374	60.04	1.953	160.9	162.4	2.428	2.341
C14	ImN(Bn) ₂	IBn	-	278.88	19.21	91.27	2.310	54.41	2.377	67.53	1.973	157.9	170.7	2.436	2.338
C15	ImN(Dipp) ₂	IPr	81.38	282.15	19.57	90.42	2.317	45.49	2.374	59.75	1.951	164.3	155.0	2.428	2.335
C16	ImN(Mes) ₂	IMes	47.61	283.01	15.70	90.45	2.315	48.54	2.379	66.13	1.954	164.7	155.7	2.434	2.335
C17	ImN(2,6-Me-C ₆ H ₃) ₂	IXy	75.11	281.02	15.42	90.23	2.314	49.11	2.376	65.99	1.954	164.5	155.3	2.434	2.336
C18	ImN(Ad) ₂	IAd	84.89	285.03	74.33	88.51	2.325	44.33	2.384	36.42	2.043	122.9	177.9	2.435	2.335
C19	Im(Me) ₂ N(Me) ₂	^{Me} I Me	86.57	281.30	12.89	94.12	2.312	53.36	2.384	72.02	1.980	157.9	172.9	2.438	2.334
C20	Im(F) ₂ N(Me) ₂	^F I Me	100.26	265.92	11.04	89.16	2.304	53.29	2.374	68.57	1.967	157.7	172.6	2.432	2.337
C21	Im(Cl) ₂ N(Me) ₂	^{Cl} I Me	-	264.63	12.85	87.28	2.304	52.83	2.371	66.93	1.967	157.3	171.0	2.432	2.339
C22	Im(NO ₂) ₂ N(Me) ₂	^{NO₂} I Me	-	243.67	16.86	79.27	2.294	54.56	2.358	61.38	1.942	157.4	168.2	2.425	2.350
C23	BImN(Me) ₂	BMe	77.28	273.03	16.22	90.13	2.306	51.94	2.375	68.91	1.957	156.3	174.2	2.426	2.340
C24	Py(b)ImN(Me) ₂	-	-	266.60	12.80	89.79	2.303	56.18	2.371	70.88	1.949	155.7	172.0	2.425	2.341
C25	DPyIm	-	46.46	275.43	17.01	89.49	2.306	35.46	2.381	69.30	1.969	156.7	177.5	2.428	2.339
C26	IBioxMe ₄	-	63.65	279.23	28.70	92.61	2.316	53.36	2.384	57.48	2.003	142.3	171.2	2.427	2.337
C27	PerN(iPr) ₂	-	56.65	278.49	44.38	83.37	2.311	52.25	2.373	44.66	2.042	133.8	167.9	2.464	2.349
C28	ThNMe	-	68.24	261.19	4.93	86.35	2.305	48.25	2.374	70.91	1.931	159.4	173.6	2.424	2.340
C29	OxNMe	-	81.90	257.53	3.05	86.72	2.299	47.79	2.371	71.92	1.910	162.3	171.7	2.421	2.339
C30	BOxNMe	-	73.51	260.53	3.15	86.84	2.295	50.04	2.366	74.31	1.901	161.9	171.2	2.418	2.342

2.3.2.3 Electronic Structure Trends

The effect of structural changes on carbene stability and electronic properties has been investigated by multiple groups. These studies tend to focus on establishing carbene reactivity and electronic structure, as well as decomposing bonding contributions, with a view to establishing the extent of σ - and π -bonding to metal-carbene interaction. Frenking's group have made a considerable contribution in this area,¹⁶⁸⁻¹⁷¹ using energy and charge partitioning approaches to assess the bonding in a wide range of transition metal-carbene complexes. By comparing calculated structural parameters with quantitative measures of backbonding,¹⁶⁹ their work has helped to lay the foundations for other analyses (see for example references 142, 143, 153, 158, 172, 173); we note that other approaches to the analysis of backbonding have also been reported.¹⁷⁴ When initially reported, such calculations tended to be focussed on comparing carbenes with each other, but some of these descriptors have found their way into catalytic applications more recently and so merit inclusion here.

Carbene reactivity has been related to singlet-triplet gaps in a number of early studies,^{175, 176} with later formalisation to predicting the likelihood of dimerization as discussed in detail by Cavallo and co-workers.¹⁷⁷ This computational study highlighted the balance between steric and electronic effects by using %V_{bur} and E_{S-T} in a multivariate model to predict the enthalpy of dimerization, which could allow the evaluation of new designs. This energy gap has been used as a descriptor by a number of groups interested in catalysis, including LKB-C¹⁶⁶ (section 2.3.2.2).

Phukan and collaborators,^{155-157, 178-184} have again used a range of descriptors to assess carbenes; while the exact selection varied with application and no single coherent database has been presented, capturing carbene properties across a wide range of structural modifications, most of these studies have included calculated singlet-triplet gaps as a measure of the thermodynamic stability of carbenes, along with HOMO-LUMO gaps as a measure of the kinetic stability (PBE1PBE/6-31G*, SDD on metals). Studies have included comparison of carbenes with silylenes, germlyenes and abnormal carbenes,^{178, 182} analysis of ring size and heterocycles on carbene properties,¹⁷⁹ the introduction of boron substituents to carbene (NHC and PHC) backbones,¹⁵⁶ the effect of additional rings and carbonyl substituents on normal and abnormal NHCs,¹⁵⁷ analysis of remote carbenes,^{181, 183} consideration of adduct formation for normal and abnormal NHCs,¹⁸⁰ and small molecule activation by cyclic (alkyl)(amino) carbenes (CAACs).¹⁸⁴ Overlap with our core set of carbenes (Table 1c) is quite poor as these studies have been focussed on novel/unusual structures, so data have not been compiled in this case.

These studies also reported a number of additional descriptor and energy calculations, allowing orbital analysis, predicting redox potentials, metal and fragment binding energies, nucleophilicity, electrophilicity, proton affinity, assessing aromaticity through Nucleus Independent Chemical Shifts (NICS¹⁸⁵) *etc.*, as well as a range of calculated NMR data to assess different ligand structures. For the latter, the ³¹P chemical shifts of carbene-phosphinidene adducts (Scheme 7b) have been calculated for a range of carbenes,^{157, 181, 182} relating to other experimental and computational studies as noted in section 2.2.1.1 above.

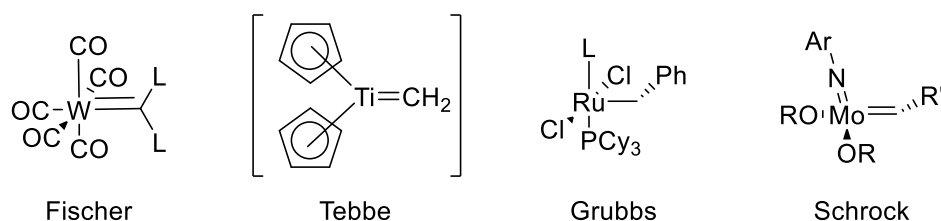
A combined experimental and computational study reported by a consortium of authors led by Nolan and Cavallo¹⁵³ included both ⁷⁷Se and ³¹P chemical shifts for 24 selenoureas and 11 carbene-phosphinidene (Table 13). The experimental NMR measurements were supplemented by an extensive computational analysis, not just of DFT-calculated NMR shielding (BP86/TZ2P), but also including charge analyses, orbital interactions and bond energy decomposition analysis. Correlations between experimental and calculated data, as well as further analysis of different calculated descriptors

allowed the authors to establish the reasons for the experimentally observed changes in chemical shift and to relate these to donor/acceptor properties of NHCs. In addition, relationships between the two datasets could be explored.

More recently, in a computational study of carbene-phosphinidenes,¹⁵⁸ calculated descriptors derived from orbitals, charges and energy decomposition analyses have been used to establish relative π -acceptor strength of a structurally-varied set of 21 carbenes. The DFT-calculated (BP86/def2-TZVPP//BP86/def2-SVP) ³¹P chemical shifts were found to correlate strongly with experimental data, and analysis of the correlation between these and a range of bonding parameters allowed the authors to establish the usefulness of these chemical shifts, in line with earlier work, including that by Vummaleti and co-authors.¹⁵³ They also noted that correlation between ³¹P chemical shifts and both singlet triplet and HOMO-LUMO energy gaps of free carbenes are low, suggesting that this NMR-derived descriptor captures a different aspect of carbene electronic structure which the authors identified as the relative π -acidity, ostensibly free of other effects.

The groups of Frison and Huynh^{173, 186} have also used frontier molecular orbital analyses, along with calculated proton affinities, HOMO-LUMO and singlet-triplet gaps to analyse trends in the electronic structures of carbenes and related divalent carbon donor compounds (using B3LYP/aug-cc-pVTZ calculations, with ECP basis sets on metals); as with the work of Phukan and collaborators, their main focus was on the effect of structural changes. In their second study,¹⁸⁶ they established the usefulness of their approach by relating ligand properties to a range of measures examining the metal-ligand bonding, presenting multiple high correlations between ligand descriptors and metal complex data for 14 NHCs.

Scheme 10



Different metathesis catalysts (Scheme 10) have been investigated using calculated descriptors derived from orbital analysis and NBO analysis.¹⁸⁷ In this case, the focus was on 11 different metal carbenes, taking advantage of the mechanistic convergence for Fisher, Rebbe, Grubbs and Schrock carbene complexes, for which property data were extracted from PBEPBE-D3(BJ)/cc-pVQZ, SDD on metals//PBE/DNP calculations. Principal component analysis (the approach is discussed in greater detail in section 3.1) of the MO and NBO data, treated both separately and as a single set, was then used to select the most important descriptors, *i.e.* those that loaded highly. More detailed analysis of those individual descriptors allowed the authors to develop a general activity trend for these metathesis catalysts, based on a key step in the reaction pathway, and to propose orbital-based descriptors as an indication of chemical activity for such catalysts.

2.4 Other Ligands

While we have deliberately focussed this review on ligand classes that have found very widespread use, it is worth noting that calculated descriptors, both individual and in databases, have been used

for the analysis of other ligands as well; some recent examples are summarised in Table 18 and a subset of these will be considered in the section on data analysis (section 3.2 below). As with P-donor ligands and carbenes, DFT calculations have been used most commonly to optimise structures, and descriptors often include the familiar steric descriptors, such as cone angles and %V_{bur}, as well as IR-derived data to determine electronic properties.

Table 18: Overview of calculated descriptors for other types of ligands described recently.

Ligands, Application	Descriptors	Reference
Salen and acacen' ligands (16) coordinated in oxo-Mn(salen) complexes, axial side occupied by donor ligands/halide	Geometric and structural data harvested from DFT calculations, analysed with PCA.	188
Dithiolate ligands in MN ₂ S ₂ complexes, evaluation for possible synthetic applications	Consideration of cone, solid and wedge angles, %V _{bur} , IR stretches.	189
N- and N,N-donor ligands (filtered to 115 molecules), Fe(II/III) redox couples/spin-crossover complexes using artificial neural networks.	Large database of simple descriptors (topology, size, elements, electronegativity), pruned for high correlation with redox potentials.	15, 23, 190, 191
Bidentate pyrrolide, indolide, aryloxide, and bis(thiolate) ligands (12), applied to analysis of titanium-catalysed hydroamination	%V _{bur} , natural ligand donor parameter (LDP) developed by Odom's group, ^{192, 193} ligand properties derived from simplified, monodentate ligands X on [NCr(NiPr ₂) ₂ X]	194
Cyclopentadienyl ligands (22) in Rh(III)-catalysed C-H activations	NMR, CO stretching, redox potential, charges, cone angles, Sterimol parameters	195
P,N-donor and Cp/Cp* ligands (11), coordinated to Ruthenium catalysts for alkene isomerisation, study of selectivity and activity	Initial calculation of 308 descriptors, reduced through further analysis to 6 key descriptors, analysis discussed in section 3.2 below.	196
Asymmetric bidentate ligands (19) with range of donor groups coupled via <i>ortho</i> -phenylene bridge, coordination to Rh(CO) ₂ fragments.	IR stretching frequencies L ₂ EP in isostructural Rh complexes	197
Pyridine-oxazoline ligands (36), analysis of Pd-catalysed redox-relay Heck reaction and Ru-catalysed Carroll rearrangement	Ligand descriptors (IR vibrational modes, NBO charges, Sterimol parameters), metal complex-derived descriptors (PdCl ₂ (LL), metal NBO charges, M-L bonding orbital energies, Sterimol parameters, %V _{bur} , structural parameters including bite angles, M-D distances etc.).	198
α-Amino acid ligands (37), used in Pd-catalysed C-H functionalisation reactions	Molecular descriptors harvested from calculations on ligand and Pd-coordinated ligand, including NBO charges, %V _{bur} , structural parameters.	199

3. Case Studies

As section 2 has demonstrated, there is no shortage of descriptors which seek to capture the steric and electronic properties of ligands, both in isolation, and when coordinated to transition metal complexes. Not surprisingly, descriptors seeking to capture the same effect are often correlated (indeed, correlation has frequently been used to socialise users to a new descriptor^{17, 35, 72, 125, 166}), and the differences between such parameters can be subtle, making it challenging to determine which subset of descriptors would be “best” for a problem in hand. Where curated descriptor databases have been presented, data for each ligand calculated in different coordination environments will give rise to descriptors that are correlated with each other,^{35, 95} further complicating the selection and comparison of regression models. Faced with such a lack of certainty, different groups of researchers have adopted different strategies and philosophies, and this section will set out some case studies to illustrate how calculated descriptors have contributed to the discovery, optimisation and design of catalysts, maintaining our focus on homogeneous organometallic systems. These case studies fall into two main categories, the mapping of chemical space (section 3.1) and the analysis of catalyst performance (section 3.2).

3.1 Mapping Chemical Space

Property descriptors can be used to illustrate how similar molecules are to each other, and indeed Tolman’s 1977 review¹⁶ included a scatter plot of cone angles and electronic parameters for all the monodentate P-donor ligands considered.¹⁷ Similarly, other steric and electronic descriptors have been used in this fashion (Cundari’s SEP and S4’;⁵⁸ Suresh’s E_{eff} and S_{eff},⁹⁸ both reviewed in reference 17), highlighting the extent of sampling of ligand space and allowing unusual ligand structures to be set into context. Larger descriptor databases can also be processed to produce such “maps” of chemical space, and this area has recently been reviewed,²⁰ allowing us here to pick out some highlights.

Arguably, the biggest impact of the ligand knowledge base approach has arisen from the processing of descriptors by principal component analysis (PCA), producing such maps of the relevant ligand space for LKB-P,^{35, 96, 97, 106} LKB-PP¹²⁴⁻¹²⁷ and LKB-C.^{166, 167} PCA is a statistical projection technique often used in image processing to identify the main variation in a dataset (see reference 20 and further references cited therein). Within each ligand set, LKB descriptors are highly correlated as they focus on the same ligand in different coordination environments,³⁵ hampering visualisation of the data in simple scatter plots, as well as interpretation in terms of familiar steric and electronic effects. PCA can be used to derive new descriptors (principal components, PCs), which are linear combinations of the original parameters optimised to capture most of the variation in the dataset in as few dimensions as possible, with the added advantage that PCs are orthogonal and so not correlated. Plotting the principal component scores for ligands for the first few PCs (PCs 1 and 2 usually capture around 60 % of the variation in the data set for LKBs) shows that ligands with similar properties have similar scores and so appear close together, while greater differences are shown by increased distances between data points. PCs can also be used as variables in multivariate regression analysis (principal component regression, PCR),⁹⁵ and a related approach for the derivation of latent variables forms the basis of partial least squares regression (PLSR).⁹⁴⁻⁹⁶

Interpretation of the composition of principal components can be challenging, in part because this approach is not statistically robust, *i.e.* the composition, order and descriptor loadings change as the ligand set is changed.^{35, 95} More importantly, perhaps, the approach highlights the largest contributions to variation in the data set in the first few PCs, and these are often a combination of

steric and electronic effects, defying the more familiar use of separate steric and electronic dimensions. This notwithstanding, we have noted that the spatial relationship between ligands rarely changes once a varied set of ligands has been captured, and have begun to attach tentative meaning to the first few PCs, as shown for LKB-P in reference 96. In this context, it is also worth noting that PCA analysis is particularly good at identifying fundamental differences between ligands, such as the differences in coordination behaviour and electronic structures of different types of carbenes (Figure 7).¹⁶⁶ This can make it harder to compare different ligand types, as discussed for P,P- and P,N-donor ligands in LKB-PP, as PCA is designed to highlight such differences.¹²⁵ However, it can also lead to a helpful separation of substituent effects as illustrated in LKB-PP_{screen} (Figure 8).¹²⁴

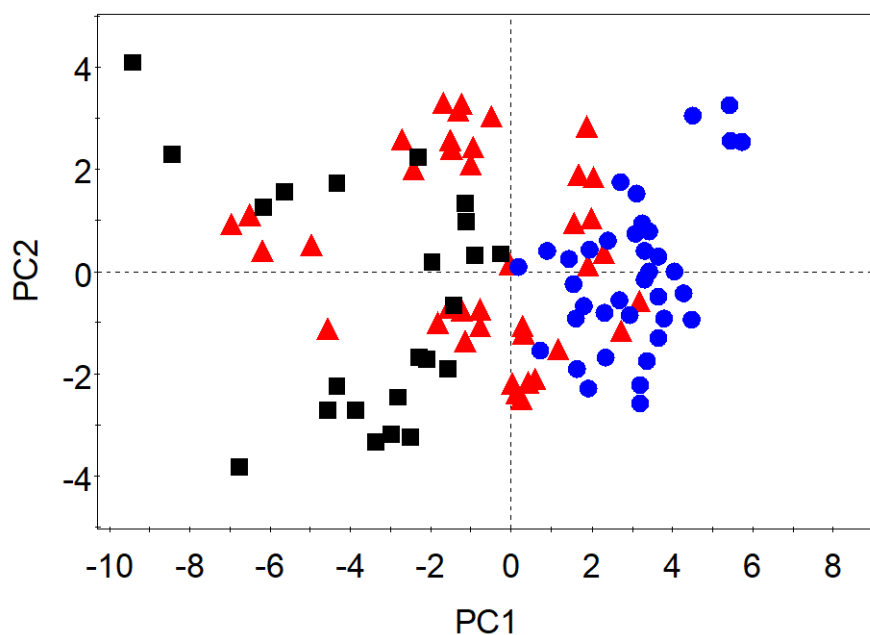


Figure 7: Principal component score plot (PC1 and PC2) for ligands in LKB-C,¹⁶⁶ capturing 58 % of variation in data. Colours and shapes relate to substitution pattern, where red triangle = Schrock-type, black square = Fischer type, blue dot = NHC/Arduengo.

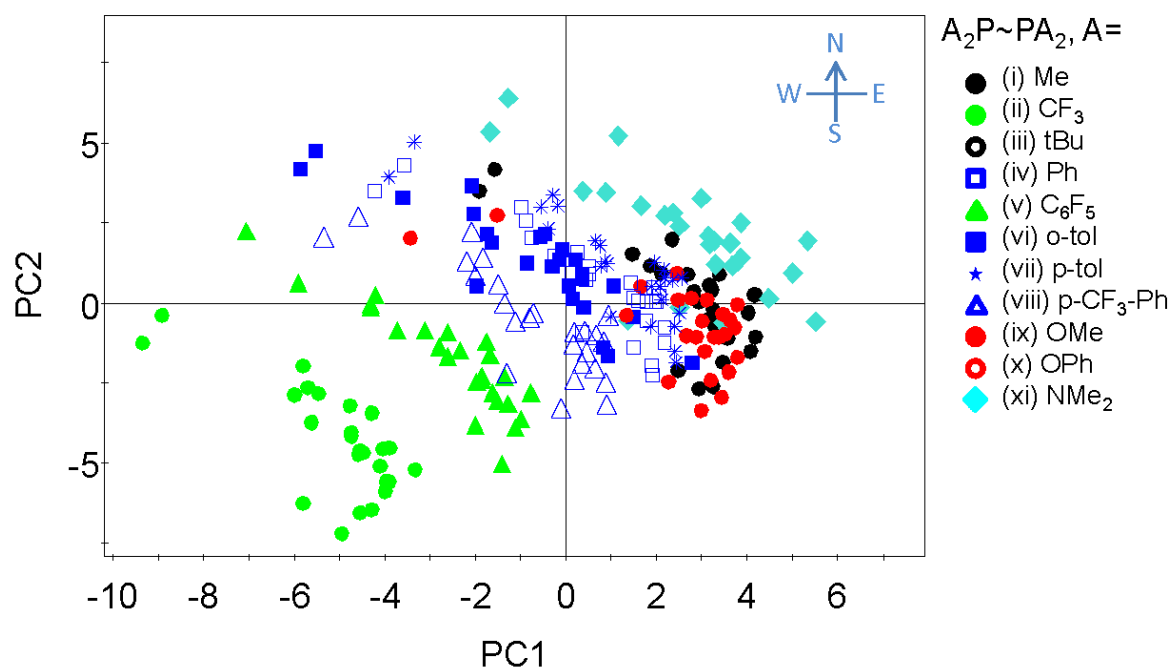


Figure 8: LKB-PP_{screen} ligand map.¹²⁴ Principal component score plot showing the first two principal components (PC1 and PC2) generated by analysis of the full LKB-PP_{screen} database of 28 steric and electronic parameters, calculated for 275 ligands. Each symbol corresponds to one ligand, with colour/shape representing different substituents as shown, and the first two PCs capture ca. 56 % of variation in data. Reproduced from reference 124 with permission from the Royal Society of Chemistry.

Ligand maps can be used in their own right to select alternative ligands and set novel designs into context, as demonstrated for fluorophosphines¹⁰⁶ and a range of unusual ligand designs.^{97, 127} Going beyond a comparison of ligand properties, such maps can also be used to explore the sampling of ligand space by a set of ligands, either, as is the case in this review, driven by commercial and data availability, or for the Design of Experiments (DoE)^{56, 200} and the identification of areas of ligand space that correspond to favourable catalyst performance, as described by both us⁹⁶ and others²⁰⁰ for LKB-P.⁹⁶ Such applications crucially depend on the availability of suitable experimental data. Here, catalyst screening with designed ligand sets, followed by several iterations of data analysis and further screening, are perhaps most promising.^{12, 200}

Figure 9 illustrates the distribution of the P-donor ligand set considered here on the latest published version of the LKB-P map,⁹⁷ with similar maps for LKB-PP¹²⁷ and LKB-C¹⁶⁶ included in the ESI (Figures S1 and S2). Some areas of ligand space are sampled more thoroughly than others (towards the right hand (Eastern) side of the map, and chemically-biased towards alkyl- and aryl-substituted ligands), which can ultimately affect the predictive performance of models, especially where models begin to extrapolate, but may also reflect chemical stability, areas that contain privileged ligands for a wide range of reactions, and indeed biases introduced by commercial availability and the preferences of many research groups.

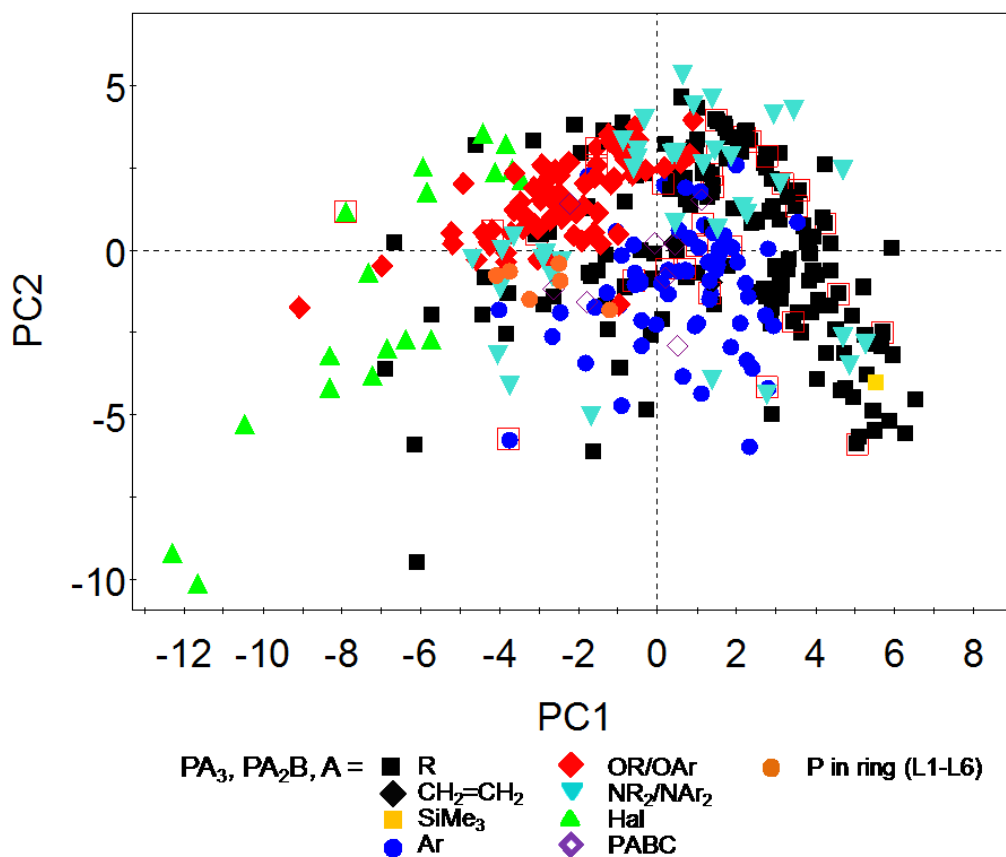


Figure 9: Ligand map generated by principal component analysis of 28 ligand parameters capturing the structures and energies of 366 P-donor ligands through DFT-calculated parameters, collected in LKB-P.⁹⁷ The principal components shown capture 62 % of the variation. Each symbol corresponds to a ligand, and shape and colour are determined by substituents as shown in the legend. Ligands considered here (Table 1a) are marked by red boxes.

We have also used LKB descriptors to fit multivariate regression models for the interpretation and prediction of ligand effects on catalyst properties, which is discussed in section 3.2.

3.2 Analysis of Catalyst Performance

As set out in the introduction, researchers in organometallic and coordination chemistry tend to be reasonably comfortable with using steric and electronic parameters to analyse, interpret and predict catalyst properties, regardless of whether such parameters have been calculated, or measured experimentally. Perhaps the most familiar applications of ligand descriptors are thus in the detection of linear free energy relationships (LFERs) and quantitative structure-property/activity/selectivity relationships (QSPR/QSAR/QSSR). In this context, relationships between a response variable capturing catalyst performance and a single descriptor, or a relatively small number of descriptors, are generally most accessible and intuitive. Most of the approaches discussed here contain at least some scatter plots to illustrate such a relationship, and many rely on linear correlations/trendlines, while some hint at a more complicated relationship described by a curve. As noted above, correlations between different descriptors are also often demonstrated by scatter plots and relatively simple mathematical equations. Correlation and regression coefficients for a single variable are thus a familiar sight, and can indeed lead to improved catalysts, *e.g.* if a higher yield shows a strong linear correlation with a

single descriptor. As we have noted before,^{17, 54} correlation does not necessarily imply causation and interesting new discoveries may well arise from the failure of a simple model.¹⁰⁶

In line with our understanding of metal-ligand interactions, steric and electronic effects often contribute to an experimentally-observed outcome in homogeneous catalysis, necessitating more complicated, multivariate analyses, and, as noted above, correlation between descriptors can mean that several models with seemingly comparable performance can be fitted. In such instances, model evaluation becomes crucial, and a number of criteria need to be considered:

a) *How well the model captures the data available.* This is usually assessed by a regression coefficient (R^2), which should be close to 1 when the relationship is described well by the regression model fitted. Regression coefficients can be low for good reasons (*e.g.* due to a wide range of data, well-understood outliers), and a more nuanced evaluation of model fit may be necessary.

b) *How large and chemically varied the training set is.* This could be assessed by inspection of a map of chemical space (section 3.1, *e.g.* Fig. 8), but more commonly this is done by visual inspection of the systems considered, or it is based on the range and spread of values for a single descriptor. Defining a desirable criterion for assessing the training set will be determined by the intended use of a model (what one might term the “Domain of Applicability”, *i.e.* whether a local or global model will be fitted), but, even for a relatively limited chemical space, the statistical approaches used generally assume that the training data will be a random and representative sample of the global population, something that may not be true for chemical data.

c) *How many variables one is comfortable with including in the model.* As noted above, a small number of variables can be easier to interpret and visualise, while additional descriptors are likely to improve model fit, at least up to the point where noise is fitted. There is thus a trade-off between interpretation and prediction, as well as a risk of overfitting.

d) *What an acceptable performance in terms of prediction errors might look like.* If there are enough data, splitting a database into training and test sets and providing an independent measure of predictive performance will be most desirable, but this is not always feasible, especially not if datasets are small, or sampling between different types of compound is uneven, as models may end up extrapolating by accident, due to such a split. Cross-validation and bootstrapping approaches to estimating prediction errors for model evaluation and comparison can provide alternative measures of the reliability of predictions.

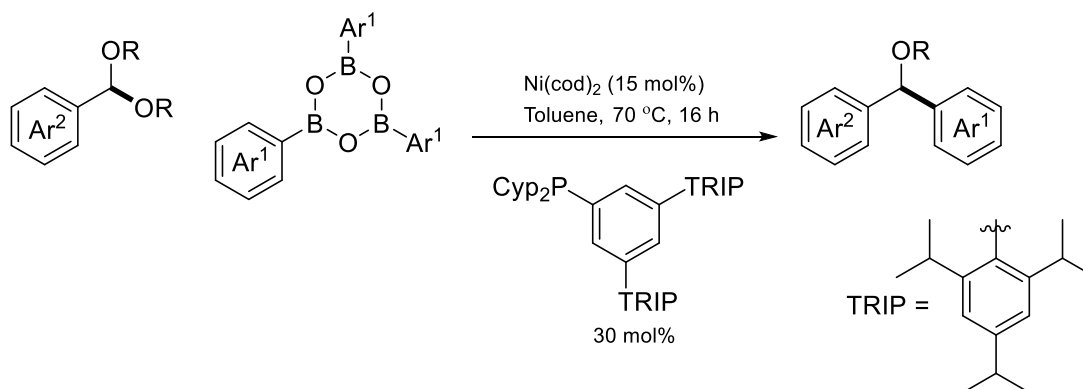
Whether the model is likely to have any transferability to other types of chemistry, and the ability to consider more than one type of ligand/catalyst, might also be useful considerations. The case studies considered in this section have placed different emphases on these criteria, and their grouping is guided largely by criterion c, *i.e.* the number of variables in the model, with a further split according to how many classes of ligands were treated together by the same approach.

3.2.1 Single Class of Ligand

The most common approach to data analysis relevant to organometallic catalysis is focussed on a single type of ligand. Few studies in this context attempt prediction of an experimental response based on its relationship with a single descriptor, although the study by Coll and collaborators showed the individual correlations between experimental pK_b data and their I_{\min} and V_{\min} descriptors, albeit without attempting pK_b prediction.⁶⁶

Most authors allow for a multivariate approach, exploring individual correlations to reduce the number of descriptors considered, using mechanistic insights to guide their descriptor selection, or using more sophisticated data analysis techniques such as PCA on the descriptors to identify key effects.

Scheme 11



Doyle and Wu have investigated Csp³ Suzuki coupling of acetals with boronic acids (Scheme 11) to afford benzylic ethers, using nickel complexes with phosphine ligands.³⁹ Using a small number of descriptors (V_{min} , $\%V_{\text{bur}}$ and θ), they investigated a set of 17 bulky phosphine ligands. With a view to understanding which structural features of these ligands were important, they initially explored the correlations of individual descriptors with yield, finding steric effects to be more important than electronic effects. By considering both cone angles (θ) and buried volume ($\%V_{\text{bur}}$), they established that remote steric effects, marked by small buried volumes and large cone angles, are important in the development of successful catalysts for this reaction. Their best quantitative model, fitted to consider the number of parameters as well as the regression coefficient of models, involved all three descriptors, as well as a cross term, and achieved both a high R^2 (0.96) and good predictive performance as measured by leave-one-out crossvalidation ($Q^2=0.88$). Additional discussion in their ESI addressed the problem of cross-validation if one system is an outlier, as well as the cross term. Their investigation suggested a new catalyst, bearing the novel (P(Cyp)₂(3,5-TRIP-Ph)) ligand (Scheme 11), which was found to give good yields for the desired reaction across a wide range of substrates. In addition, they noted that $\%V_{\text{bur}}$ and θ are not always directly correlated, with the correlation diverging once $\%V_{\text{bur}}$ was high enough to prevent any catalytic activity.

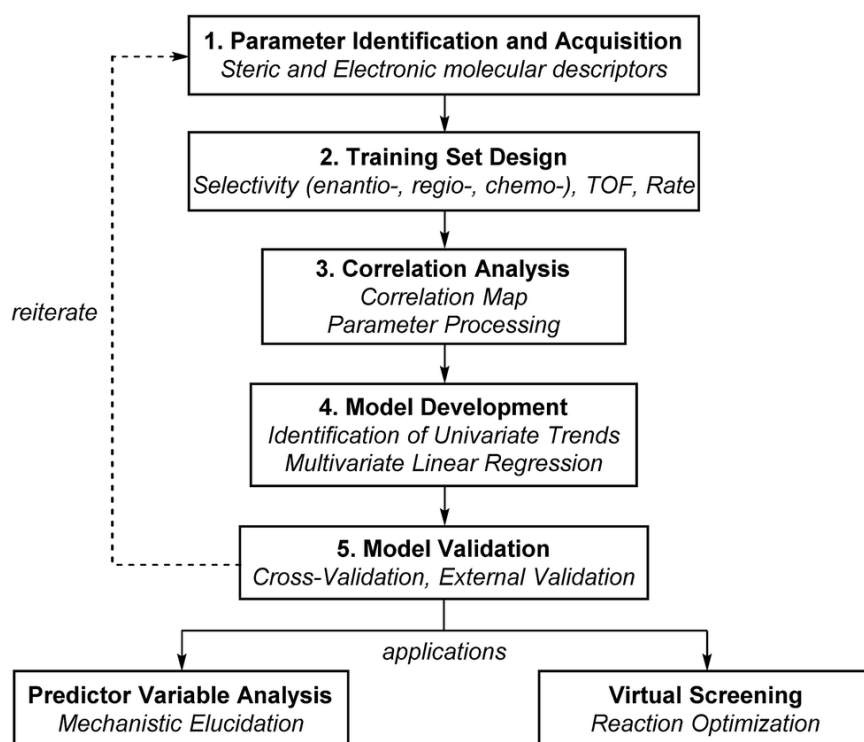


Figure 10: Model development workflow applied by Sigman and co-workers. Reprinted with permission from reference 14. Copyright 2018, American Chemical Society.

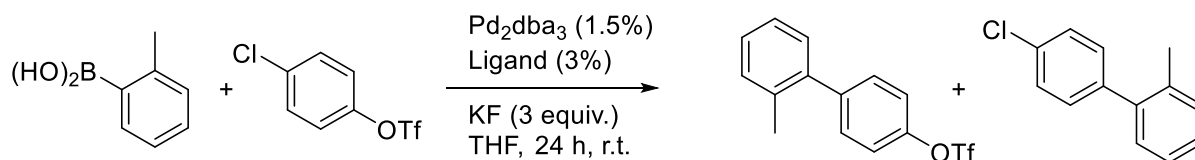
The Sigman group have become well known for their use of LFER and MLR in combination with extensive experimental screening data. While their earlier work focussed on asymmetric catalysis and the control of selectivity (reviewed in reference 13), more recently they have used ligand descriptors to elucidate mechanistic information, optimise ligand structures for specific reactions and produce models to predict the performance of experimental studies (reviewed in reference 14). Their workflow for the latter is reproduced in Figure 9 and generally focuses on the identification of descriptors which correlate to experimental results, yielding mechanistic insights and thus guiding catalyst optimisation. Depending on the problem considered, their models can utilise just one or several descriptors and model evaluation typically involves the separation of the ligands/complexes under study into training, test and validation sets, allowing the development of models which are more robust and not reliant on a singular data set.

Generally,¹⁴ descriptors of interest are identified by inspection of large libraries of steric and electronic ligand descriptors, either computed, often with some truncation/simplification, or the initial parameter set is chosen based on previous investigations and mechanistic information. Descriptors are then normalised and descriptor correlations are identified, helping to reduce the number of descriptors before MLR as the presence of over-correlated descriptors is considered likely to amplify their underlying random noise, or lead to overfitting. Recently, the group have also begun to use correlation maps to improve the identification of high correlations between descriptors.¹⁴

In the next step, simple correlations between response data and individual descriptors are identified, using similar ligands which differ structurally in a single, interpretable way. While this is not always possible, it can help in the development of their final multivariate model. MLR using the chosen descriptors is followed by model evaluation through cross-validation and external testing, with a view to determining whether the model can be used for prediction. This workflow has been applied to a

wide range of reactions,^{13, 14, 36, 37, 90, 123, 154, 198, 199, 201, 202} and we have noted the use of different ligand descriptors in such studies throughout section 2.

Scheme 12



A representative example of this data analysis approach is the parameterisation of 38 P-donor ligands for Suzuki reactions (Scheme 12),³⁶ where they calculated the lowest energy conformers with the highest and lowest cone angles for each ligand. Descriptors were then computed for the two conformers of each ligand, with an MLR model, utilising five of their descriptors ($P-C_{\text{bend}(v)}$, $^{31}\text{P}_{\text{shift}}$, $^{31}\text{P}-\text{Se}_{\text{shift}}$, $P-C_{\text{bend}(l)}$ and cone angle), able to predict experimentally-measured $\Delta\Delta G^\ddagger$ reasonably well ($R^2=0.90$ for predicted vs. experiment). They noted, however, that this global model is difficult to understand due to the presence of four cross terms, and that interpreting mechanistic information is not possible. Drawing on extensive mechanistic studies by others, they separated ligands according to whether it is the $L_2\text{Pd}$ or LPd complex that undergoes oxidative addition. This allowed them to develop simpler models which were easier to interpret. A univariate model utilising the $^{31}\text{P}-\text{Se}_{\text{shift}}$ achieved prediction of $\Delta\Delta G^\ddagger$ with an $R^2=0.86$ and the Buchwald biaryl ligands, as well as smaller phosphines considered, were found to be better described by the descriptors calculated using the minimum cone angle conformer ($R^2=0.84$), rather than the maximum used in the global model ($R^2=0.78$).

While this approach depends on the availability of relatively extensive experimental data, other studies from this group have shown how the chemical insights gained from smaller-scale analyses of ligand effects can guide synthetic work towards promising targets,^{37, 123} neatly bridging LFER/physical organic chemistry approaches with large-scale descriptor calculations.

A similar approach has been used by Paton and co-workers, applied, for example, to an exploration of correlations and regression models for cyclopentadienyl ligands in rhodium-catalysed C-H activations¹⁹⁵ and for the analysis of copper-catalysed asymmetric conjugate additions.⁹² While the latter study, considering ligand screening data for more than 30 chiral phosphoramidite ligands (Scheme 4),⁹² cited Sigman's approach as an influence, their analyses have been supplemented by DFT optimisations of the transition states in the selectivity-determining step of the reaction, adding further mechanistic insights which supported the design of additional ligands. The ESI for this study provided additional details of the process used to derive and evaluate regression models, which included forward selection, cross-validation and splitting the data into training and test sets. As noted above, their initial descriptor database included descriptors for both the whole ligands and truncated aromatic substituents, and their final model included aromatic HOMO energies along with Sterimol steric parameters. This approach demonstrates how data analysis can guide not just experimental screening but also computational mechanistic studies towards the most promising targets, but also discussed the need for further experimental optimisation to achieve a successful catalyst after initial screening failed to match predicted yields (although good selectivities).

Potential applications of the LKB descriptors in models for both interpretation and prediction have been discussed extensively, and we have explored both the direct use of calculated descriptors in multivariate linear regression,^{35, 125} and the use of derived variables in PCR and PLSR, fitting models for experimental^{135, 96, 125} and calculated data,^{96, 125, 166} as well as exploring the relationship between our

descriptors and the TEP.³⁵ In collaboration with Mansson and Welsh,⁹⁵ we have also explored more sophisticated statistical approaches for modelling TEP, including robust linear regression, Least Angle Regression (LAR) and the Least Absolute Shrinkage and Selection Operator (LASSO). Potential issues around sampling of ligand space and model robustness have been explored, highlighting that for TEP at least, a multivariate linear regression model can achieve reasonably good performance, with robust regression also worthy of consideration. Later work on an expanded version of LKB-P⁹⁶ explored the analysis and modelling of high-throughput screening data on palladium-catalysed amination reactions reported by Hartwig's group,²⁰³ which allowed us to compare multivariate linear regression (MLR) and PLSR. Both models were quite poor, but PLSR captured the overall trends across the ligand map better, while MLR gave a better fit to the response data, at the cost of likely overfitting. In this case, overlap between published experimental data and ligands in LKB-P was rather limited, making more extensive analysis difficult; we have also reported the statistical analysis of larger datasets of calculated data, using the binding energy of CO *trans* to ligands in [Cr(CO)₅L] complexes as the response.^{96, 125, 166} This allowed us to compare different approaches on a bigger dataset, illustrating that both MLR and PLSR can achieve satisfactory model performance as measured by R² and prediction error estimates.

LKB-P descriptors have also been used in a small-scale computational study of the palladium-catalysed Suzuki-Miyaura coupling reaction,²⁰⁴ allowing us to quantify ligand effects on calculated barriers for each step along the reaction pathway. In this case, only 4 ligands were considered, with models focussing on just 3 descriptors (E_{HOMO}, E_{LUMO} and He₈_steric). Standardised coefficients in these models helped with interpretation of ligand effects on each reaction step and were chemically plausible, but these models were too limited to attempt prediction.

Grotjahn, Rothenberg and collaborators have investigated the development of descriptor-led predictive modelling for ruthenium-catalysed alkene isomerisation catalysts with P,N mixed-donor bidentate ligands.¹⁹⁶ In their approach,¹¹⁵ a large database of semi-empirically (PM3) calculated descriptors (308 for each metal-ligand complex) for 11 catalysts were calculated. Their descriptors were then ranked and reduced in number by assessing their relationship with 4 figures of merit (FOMs) for the catalysts. The correlation between existing experimental data and the descriptor values allowed selection of those with the highest correlation to the FOMs, which included the yield (%) for either the 2-E-alkene or 3-E-alkene, turnover frequency (TOF) and turnover number (TON). For this subset of 6 descriptors, PCA was used to build a correlation model between the descriptors and experimental FOMs. Using a PCA biplot, four trends relating to catalyst structure and performance were observed. Using these, two new catalyst structures were proposed, synthesised and their descriptors calculated. PLSR modelling of the descriptors and FOMs predicted experimental FOM values for the two new catalysts. Adding these to diagnostic plots for the FOMs suggested generally quite successful predictions could be achieved, with particularly good results for TOF and TON. Their study highlights that descriptor-led prediction in transition metal catalysis can be achieved without a high-throughput screening (HTS) approach to generate experimental data, and they note that the insights gained exceed those from structural analyses of catalysts.

However, with a large experimental dataset from more than 4600 reactions, obtained by ultra-high-throughput screening at Merck, Dreher, Doyle and co-workers have been able to use calculated descriptors together with a range of machine-learning (ML) approaches to make predictions about a range of components in palladium-catalysed cross-coupling of aryl halides, including some exploration of ligand effects (Scheme 3, section 2.1.2.3). They were able to confirm their postulate that ML could produce better models than regression analysis, while avoiding the need for descriptor selection. Among a broad range of linear regression and supervised ML approaches considered, random forest models were found to be most successful. Limitations of this algorithm were discussed, and the use

of a sparse training set (5% of the data), showed that the model derived did indeed give superior performance for the prediction of yields, albeit with an erosion of accuracy. The authors conceded that model interpretation is difficult, but were able to relate descriptor contributions to mechanistic rationales. The ESI for this publication provides substantial further information on the approaches considered as well as the descriptor calculation and data analysis workflows developed. While only a small number of ligands have been considered, the screening of their interactions with other reaction variables is powerful in this case, and this presents significant challenges to many of the familiar, ligand-focussed descriptors reviewed here. We note that this study prompted some controversy and debate while the present contribution was already in review,¹⁰⁷⁻¹¹⁰ highlighting that ML in homogeneous catalysis is still in its infancy.

3.2.2 Comparison of different ligand classes

As noted earlier, different ligand classes can afford similar or complementary reactivity in catalysis, and their comparison relies on the transferability of descriptors, as well as mechanistic consistency. Gusev⁵⁹ presented a large-scale comparison of the donor properties of different ligand classes across a range of calculated experimental datasets, providing a quantitative comparison of their net donor properties, but we are not aware of applications of such data to make predictions relevant to catalysis.

While predictions are likely to fail if changing the ligand accesses a different reaction pathway, problems with fitting a simple linear relationship across ligand classes can serve as a diagnostic for structural and mechanistic differences, as shown for copper-catalysed boracarboxylation of styrene with CO₂ (Scheme 1).⁶⁰ In this study, mono- and bidentate P-donor ligands were considered alongside NHCs, and the correlation of calculated reaction barriers with different descriptors highlighted differences between these ligand classes, prompting further structural and mechanistic analyses.

Ligand effects in ruthenium-catalysed alkene metathesis have been considered by several authors,^{5, 11, 100, 104, 154, 187} and here the comparison of different ligand families aligns well with the development of this chemistry. While so-called first generation catalysts relied on a P-donor supporting ligand, NHCs later replaced this design paradigm, with more recent developments also targeting the supporting ligands *trans* to the NHC.^{5, 11}

Jensen and co-workers have presented a number of computational and experimental studies in this area, with their QSAR model of ligand effects¹¹ reviewed previously.^{12, 54} More recently, this group have worked towards automating the assembly of novel ligands from suitable fragments,²⁰⁵ and they have used their expertise around the mechanism of ruthenium-catalysed metathesis to test the resulting ligand library. The genetic algorithm used for catalyst evolution relied on an indirect fitness criterion, from a QSAR model fitted to DFT-calculated barriers, to evaluate each catalyst design, and this approach showed a clear preference for NHC ligands over phosphines.

The group of Suresh have applied their molecular electrostatic potential (MESP) parameter to the analysis of stereoelectronic effects in both 1st¹⁰⁴ and 2nd generation¹⁰⁰ Grubbs catalysts for alkene metathesis, *i.e.* catalysts supported by phosphines and NHCs. These investigations revealed that the MESP parameters can be used to determine valuable descriptors for these complexes, which describe the steric (V_S), electronic (V_E) and stereo-electronic (V_{SE}) effect of a given complex, compared to a reference system (PH₃ and ImH₂NH₂ respectively for phosphines and NHCs). In these two studies, the active form [Cl₂(L)Ru=CH₂] **2** and ethene bound [Cl₂(L) (CH₂CH₂)Ru=CH₂] **3** form of the catalyst species were investigated.

Calculation of V_S , V_E and V_{SE} for NHCs and phosphines in these studies was not entirely consistent, with NHCs requiring the combined value of the MESP of both the carbene carbon (V_{C1}) and the Ru-CH₂

carbon (V_{C2}),¹⁰⁰ whereas phosphines used only the phosphorus atom (V_P).¹⁰⁴ This makes direct comparison of the ligands difficult, and may explain why these results were published in two separate reports. Nonetheless, once one considers the requirement to use the combined MESP values for carbenes, the method to calculate the three parameters is identical. The methodology for their calculation is shown below, with the PH'_3 and $ImH'_2NH'_2$ systems representing structures where the fixed geometry of any given PR_3 or ImR_2NR_2 ligand has had the R groups replaced with H, and the bond lengths of the N-H or P-H bonds changed to those of the respective PH_3 or ImH_2NH_2 complex. However, the geometry is not modified or reoptimized, so the steric effect of the R groups is preserved without the electronic effect, allowing for calculation of V_S .

Phosphines

$$V_{SE} = V_P(\text{active or ethylene}) - V_P(PH_3)$$

$$V_S = V_P(PH'_3) - V_P(PH_3)$$

$$V_E = V_{SE} - V_S$$

NHCs

$$V_{SE} = (V_{C1} + V_{C2})(\text{active or ethylene}) - (V_{C1} + V_{C2})(ImH_2NH_2)$$

$$V_S = (V_{C1} + V_{C2})(ImH'_2NH'_2) - (V_{C1} + V_{C2})(ImH_2NH_2)$$

$$V_E = V_{SE} - V_S$$

Their data are shown in Table 19, and these reports illustrated that it is possible to interpret steric and electronic effects of ligand modification on catalysts using MESP descriptors specific to this reaction, which show good correlation with steric descriptors external to this study, such as the Tolman cone angle and $S4'$ (section 2.1.1.2). Having thus established that the MESP descriptors captured the desired properties, the authors then explored how well the relationship between either steric or electronic effects and calculated activation energies for key steps of the catalytic cycle could be fitted by a simple linear regression equation. While no prediction was attempted, the authors also noted that such a study could be useful for catalyst discovery or optimisation studies. No comparison of ligand classes was attempted, but we note that the relationship between activation energies and ligand descriptors appeared similar, and that NHCs generally gave rise to lower activation energies than phosphines, in line with their experimental performance.

More recently, this group have looked at the palladium-catalysed oxidative addition of aryl halides to palladium catalysts.¹⁰¹ In this case, reaction-specific MESP data were considered for a varied set of supporting ligands, including phosphines, NHCs, alkenes and alkynes (Scheme 2). All calculated activation barriers correlate highly with the ΔV_{Pd} descriptors (The MESP value at the Pd nucleus of PdL, Table 19), with separate relationships fitted not just for each ligand class, but also for each aryl halide. Trends within each subset are similar, and the regression equations could be used to predict activation barriers without calculating the relevant transition states and intermediates. However, aside from noting these trends and drawing some general conclusions about different ligand classes from ΔV_D descriptors, ligands were not compared in detail in this study.

Table 19: Reaction-specific MESP data for ruthenium-catalysed metathesis^{100, 104} and palladium cross-coupling.¹⁰¹

Ligand No.	Phosphine Ligand	Cl ₂ (PR ₃)Ru=CH ₂ ^a			Cl ₂ (PR ₃)Ru(CH ₂ CH ₂) ^a			ΔV _{Pd1} (PdL ₂) (kcal mol ⁻¹) ^b	ΔV _{Pd2} (PdL) (kcal mol ⁻¹) ^b
		V _{SE1} (kcal mol ⁻¹)	V _{S1} (kcal mol ⁻¹)	V _{E1} (kcal mol ⁻¹)	V _{SE2} (kcal mol ⁻¹)	V _{S2} (kcal mol ⁻¹)	V _{E2} (kcal mol ⁻¹)		
P1	PH ₃	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0
P2	P(Me) ₃	-6.39	3.47	-9.86	-5.40	3.31	-8.70	-11.1	-5.9
P3	P(Et) ₃	-12.71	4.82	-17.53	-10.94	4.46	-15.41	-12.9	-7.4
P4	P(iPr) ₃	-14.54	7.76	-22.31	-12.59	7.86	-20.45	-14.1	-8.5
P5	P(nBu) ₃	-	-	-	-	-	-	-	-
P6	P(tBu) ₃	-11.96	14.89	-26.85	-9.93	14.49	-24.41	-15.7	-10.0
P7	P(Cy) ₃	-	-	-	-	-	-	-16.0	-9.5
P8	P(Ph) ₃	-7.05	6.14	-13.19	-3.87	6.11	-9.97	-5.9	-1.7
P21	P(Ph)(Me) ₂	-6.23	4.06	-10.29	-4.23	4.37	-8.59	-	-
P22	P(Me)(Ph) ₂	-6.51	5.13	-11.65	-4.05	4.56	-8.61	-	-
P23	P(Ph)(Et) ₂	-	-	-	-	-	-	-	-
P24	P(Et)(Ph) ₂	-8.39	5.57	-14.14	-5.47	5.85	-11.33	-	-
	NHC Ligand (Acronym)	Cl ₂ (NHC)Ru=CH ₂ ^c			Cl ₂ (NHC)Ru(CH ₂ CH ₂) ^c				
C1	SimN(H) ₂	0.00	0.00	0.00	0.00	0.00	0.00	-	-
C2	SimN(Me) ₂ (SIMe)	-2.64	4.55	-7.19	6.73	5.60	1.14	-	-
C3	SimN(iPr) ₂ (SiPr)	-5.58	5.02	-10.59	4.74	6.47	-1.73	-	-
C4	SimN(Ph) ₂ (SIPh)	2.53	8.30	-5.78	16.06	18.08	-2.03	-	-

C8	ImN(H)₂	-	-	-	-	-	-	0.0	0.0
C9	ImN(Me)₂ (IMe)	-	-	-	-	-	-	-0.9	-1.7
C20	Im(F)₂N(Me)₂ (^fIMe)	-	-	-	-	-	-	11.7	6.0
C21	Im(Cl)₂N(Me)₂ (^{cl}IMe)	-	-	-	-	-	-	11.1	5.5
C22	Im(NO₂)₂N(Me)₂ (^{NO₂}IMe)	-	-	-	-	-	-	38.3	24.1

^a Reference 104; ^b reference 101; ^c reference 100.

4. Summary and Outlook

This review has focussed on calculated descriptors, and most recent studies in this area have either used atomic coordinates for ligands and their complexes from a variety of calculations to derive a quantitative, usually steric, measure, or descriptors (and response data, in some cases) have been extracted from DFT calculations. Due to the advent of wide-spread high performance computing capabilities in academic departments, the field has matured considerably and we have been able to focus our review predominantly on work from the last decade or so, with some extensions to earlier studies to set these into context. We have compiled data for representative ligand sets, including 30 monodentate P(III)-donors, 23 bidentate P,P-donor ligands and 30 carbenes, facilitating the future application of calculated ligand descriptors in the analysis of catalysis. Some of these datasets cover more extensive ligand databases and we have included some indication of these, as well as noting a selection of studies on other ligand types. Both individual descriptors and databases have been considered, the latter often curated and compiled over long time-spans, with a view to providing a more comprehensive collection of ligand property descriptors.

By reviewing a series of recent application examples, we have illustrated how such descriptors have been used to inform catalyst discovery, optimisation and design, either through maps of chemical space, or using regression models and machine learning to help interpret and, in some cases, predict experimental and calculated observations. In this context, challenges for model building have been discussed, highlighting descriptor correlations, the effect of incomplete sampling of ligand chemical space, as well as the need to evaluate predictive performance through either cross-validation/resampling, or, where possible, by splitting data into training and test sets. Most data analysis approaches are focussed on a single class of ligands, with comparison of several classes rare, and difficult; indeed, many descriptors are not transferable to other ligands, while others are reaction-specific, again potentially limiting their applications, albeit improving data fits and predictions.

With such an extensive survey in hand, we can identify some suggestions for best practice:

- a) Data: While there may be restrictions on data-sharing in some cases, we have found studies that included a summary table of descriptors, responses and predictions more accessible than those that have relied solely on scatter plots to highlight model performance. In addition, multiple versions of a descriptors may have been reported previously (especially for TEP, cone angles and %V_{bur}), and it would be helpful to identify which of these were used clearly.
- b) Philosophy: Descriptor calculations and data analyses are undertaken with a broad range of goals, ranging from the illustration of a computational or data analysis approach *via* mechanistic interpretation and all the way to prediction and design. Such choices can affect model size and evaluation, as well as transferability of descriptors and models. Some discussion of the choices made might help non-specialists decide on the most suitable approach for their area of interest.
- c) Quality of models: In a data-rich, digital research environment, inclusion of regression equations and coefficients (R^2) with figures, as well as access to a broader range of diagnostic plots (residuals vs. predictions) in the ESI, along with the (chemical) discussion of persistent outliers, should become increasingly feasible. This would be akin to including coordinates for all calculated geometries in computational mechanistic studies and cif files for crystallography, and should again facilitate access of non-specialists to data-led approaches.
- d) Model testing: If predictions are made from a statistical model, their quality can be assessed in a range of ways, including diagnostic plots, cross-validation/resampling and the prediction of external test set data. Making these accessible, *e.g.* in the ESI, can help model evaluation.

- e) Sampling: Catalytic activity can be limited to a small range of catalyst structures/ligands, with other systems inactive. From a data analysis perspective, the inactive systems are still worth knowing about, and some discussion of the sampling of chemical space achieved would be valuable in this context.
- f) Experimental testing: The uptake of data analysis in catalyst design can be facilitated by making experimental data available alongside descriptors and predictions. As set out for sampling, reporting of inactive systems and failed predictions are also helpful in this context.

Most of the case studies reviewed here demonstrate these practices very well, and we will readily concede that there can be compelling reasons (such as journal restrictions on length and data sharing, as well as editorial decisions about clarity, and indeed a lack of time for data curation) for compromise. However, in both academic and industrial settings, considerable computing power and data analysis expertise are now available, and, in homogeneous organometallic catalysis, practitioners are moving towards applying this routinely, so it is perhaps timely to consider such issues.

Indeed, there has been a notable (and exciting) increase in published studies using some form of correlation/regression analysis based on calculated descriptors over the last few years, which may well encourage other research groups to enter this field in the near future. Perhaps the data tabulated in this review for a core set of ligands can help facilitate entry to such newcomers, and provide a standard for setting new descriptors into context by relating them to other approaches. At the very start of such data-led studies of catalysis, it is worth considering whether interpretation of experimental data or prediction are the main focus, and also whether the chosen descriptors are meaningful for the project in hand and can be calculated easily, yet with appropriate accuracy.

Taking each of these in turn, data analysis for the **interpretation** of experimental data often relies on small numbers of well-understood descriptors. While high correlations may be desirable, outliers can often be dealt with quite easily in such a setting, simply by data inspection, and more detailed computational study of structures and mechanism provides a feasible backup.

Models for **prediction** generally benefit from bigger datasets (for both descriptors and responses) and better sampling of chemical space, as discussed in section 3.2. Here it may make sense to sacrifice some of the intuitive chemical understanding afforded by simpler models of isolated steric and electronic effects for improved model performance. However, multivariate approaches, with ML as their extremes, can produce a plethora of models with subtly different performances, making it difficult to pick “the best”; in such cases, models with fewer variables which make chemical sense are at least worthy of consideration, even if a black box model can achieve better predictions.

Related to this is the question of statistical robustness, affecting projection techniques such as PCA, along with multivariate models. These approaches are sensitive to the dataset used, such that even small changes can alter the map or model generated, especially if so-called outliers are present. The underlying statistical assumption is that one is dealing with a random and representative subset of a global population, a premise which is frequently difficult to prove, and probably untrue, in chemistry. Some of these outliers are chemically meaningful, others just an accident of poor sampling, and it is worth reflecting on this when assessing model quality, rather than removing them from consideration to achieve a better fit.

Finally, there is a trade-off between computational cost/convenience and the resolution of ligand properties which can be achieved by different **descriptors**, *e.g.* variations in electronic structure require some form of quantum mechanical calculation while steric bulk may be captured well by lower levels of theory. It is often worth investing time and effort not just in descriptor calculations, but also

in setting them into context, *i.e.* understanding whether they capture a single effect (steric or electronic), or a mixture. This investment may run counter to a desire for automation, but facilitates model evaluation down the line. Neither our collective understanding of mechanism in homogeneous organometallic catalysis, nor of applications of data analysis in chemistry are sufficient to use these as a black box, but great strides are being made to improve on this, with ligand descriptors and large-scale experimentation and computational studies set to make important contributions.

Acknowledgements

PhD project funding (DJD) from the EPSRC Centre for Doctoral Training in Catalysis (EP/L016443/1) is gratefully acknowledged.

Supporting Information

Additional PCA plots for LKB-PP and LKB-C. Data tables for all ligand data collated, as well as all references.

Author Biographies

Derek Durand obtained his undergraduate MChem degree from the University of Southampton in 2017. He then joined the EPSRC Centre for Doctoral Training (CDT) in Catalysis, shared between the Universities of Bath, Bristol and Cardiff, for a combined MRes + PhD programme, being awarded an MRes in Catalysis by Cardiff University in 2018. In July 2018, he joined the Fey group at the University of Bristol for his PhD project, focussing on new workflows to computationally predict catalysts for homogeneous organometallic processes from calculated descriptors and mechanistic studies.

Natalie Fey is a Senior Lecturer (\approx Associate Professor, tenured) at the University of Bristol. After completing a BSc in Chemistry and Economics in 1997 and a PhD focussed on “Molecular Modelling of Ferrocenes and Arylphosphines” with James Howell and Paul Yates in 2001, both at Keele University, she worked as a postdoctoral researcher with Rob Deeth at Warwick University. She joined the University of Bristol in 2003, initially as a postdoctoral researcher with Guy Orpen and Jeremy Harvey. After an industry-funded postdoctoral position with Guy Lloyd-Jones, Guy Orpen and Jeremy Harvey (2005-2007), she was awarded a prestigious EPSRC Advanced Research Fellowship in 2007, starting her independent career. She was appointed to a temporary lectureship in 2015, made permanent in 2017, and promoted to Senior Lecturer in 2018. Her current focus is on using computational approaches as a driver for scientific discovery, with a particular interest in applying computational and structural chemistry to the large-scale prediction and design of organometallic catalysts.

References

- (1) Sperger, T.; Sanhueza, I. A.; Schoenebeck, F. Computation and Experiment: A Powerful Combination to Understand and Predict Reactivities. *Acc. Chem. Res.* **2016**, *49*, 1311-1319.
- (2) Sperger, T.; Sanhueza, I. A.; Kalvet, I.; Schoenebeck, F. Computational Studies of Synthetically Relevant Homogeneous Organometallic Catalysis Involving Ni, Pd, Ir, and Rh: An Overview of Commonly Employed DFT Methods and Mechanistic Insights. *Chem. Rev.* **2015**, *115*, 9532-9586.

- (3) Liu, Z.; Patel, C.; Harvey, J. N.; Sunoj, R. B. Mechanism and reactivity in the Morita–Baylis–Hillman reaction: the challenge of accurate computations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30647-30657.
- (4) Kwon, D.-H.; Fuller, J. T.; Kilgore, U. J.; Sydora, O. L.; Bischof, S. M.; Ess, D. H. Computational Transition-State Design Provides Experimentally Verified Cr(P,N) Catalysts for Control of Ethylene Trimerization and Tetramerization. *ACS Catal.* **2018**, *8*, 1138-1142.
- (5) Luo, S.-X.; Engle, K. M.; Dong, X.; Hejl, A.; Takase, M. K.; Henling, L. M.; Liu, P.; Houk, K. N.; Grubbs, R. H. An Initiation Kinetics Prediction Model Enables Rational Design of Ruthenium Olefin Metathesis Catalysts Bearing Modified Chelating Benzylidenes. *ACS Catal.* **2018**, *8*, 4600-4611.
- (6) Rohmann, K.; Holscher, M.; Leitner, W. Can Contemporary Density Functional Theory Predict Energy Spans in Molecular Catalysis Accurately Enough To Be Applicable for in Silico Catalyst Design? A Computational/Experimental Case Study for the Ruthenium-Catalyzed Hydrogenation of Olefins. *J. Am. Chem. Soc.* **2016**, *138*, 433-443.
- (7) Minenkov, Y.; Occhipinti, G.; Jensen, V. R. Complete Reaction Pathway of Ruthenium-Catalyzed Olefin Metathesis of Ethyl Vinyl Ether: Kinetics and Mechanistic Insight from DFT. *Organometallics* **2013**, *32*, 2099-2111.
- (8) Park, Y.; Ahn, S.; Kang, D.; Baik, M. H. Mechanism of Rh-Catalyzed Oxidative Cyclizations: Closed versus Open Shell Pathways. *Acc. Chem. Res.* **2016**, *49*, 1263-1270.
- (9) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science* **2018**, *360*, 186-190.
- (10) Raugei, S.; DuBois, D. L.; Rousseau, R.; Chen, S.; Ho, M.-H.; Bullock, R. M.; Dupuis, M. Toward Molecular Catalysts by Computer. *Acc. Chem. Res.* **2015**, *48*, 248-255.
- (11) Occhipinti, G.; Bjorsvik, H. R.; Jensen, V. R. Quantitative structure-activity relationships of ruthenium catalysts for olefin metathesis. *J. Am. Chem. Soc.* **2006**, *128*, 6952-6964.
- (12) Jover, J.; Fey, N. The Computational Road to Better Catalysts. *Chem. Asian J.* **2014**, *9*, 1714-1723.
- (13) Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A. The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res.* **2016**, *49*, 1292-1301.
- (14) Santiago, C. B.; Guo, J.-Y.; Sigman, M. S. Predictive and mechanistic multivariate linear regression models for reaction development. *Chem. Sci.* **2018**, *9*, 2398-2412.
- (15) Janet, J. P.; Gani, T. Z. H.; Steeves, A. H.; Ioannidis, E. I.; Kulik, H. J. Leveraging Cheminformatics Strategies for Inorganic Discovery: Application to Redox Potential Design. *Ind. Eng. Chem. Res.* **2017**, *56*, 4898-4910.
- (16) Tolman, C. A. Steric effects of phosphorus ligands in organometallic chemistry and homogeneous catalysis. *Chem. Rev.* **1977**, *77*, 313-348.
- (17) Fey, N.; Orpen, A. G.; Harvey, J. N. Building ligand knowledge bases for organometallic chemistry: Computational description of phosphorus(III)-donor ligands and the metal–phosphorus bond. *Coord. Chem. Rev.* **2009**, *253*, 704-722.
- (18) Gibb, B. C. Big (chemistry) data. *Nature Chem.* **2013**, *5*, 248-249.

- (19) Grzybowski, B. A.; Bishop, K. J. M.; Kowalczyk, B.; Wilmer, C. E. The 'wired' universe of organic chemistry. *Nature Chem.* **2009**, *1*, 31-36.
- (20) Fey, N. Lost in chemical space? Maps to support organometallic catalysis. *Chem. Cent. J.* **2015**, *9*:38 (article no.).
- (21) Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem. Int. Ed.* **2016**, *55*, 5904-5937.
- (22) Yada, A.; Nagata, K.; Ando, Y.; Matsumura, T.; Ichinoseki, S.; Sato, K. Machine Learning Approach for Prediction of Reaction Yield with Simulated Catalyst Parameters. *Chem. Lett.* **2018**, *47*, 284-287.
- (23) Janet, J. P.; Chan, L.; Kulik, H. J. Accelerating Chemical Discovery with Machine Learning: Simulated Evolution of Spin Crossover Complexes with an Artificial Neural Network. *J. Phys. Chem. Lett.* **2018**, *9*, 1064-1071.
- (24) Coley, C. W.; Green, W. H.; Jensen, K. F. Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **2018**, *51*, 1281-1289.
- (25) von Lilienfeld, O. A. Quantum Machine Learning in Chemical Compound Space. *Angew. Chem. Int. Ed.* **2017**, *57*, 4164-4169.
- (26) Skoraczyński, G.; Dittwald, P.; Miasojedow, B.; Szymkuć, S.; Gajewska, E. P.; Grzybowski, B. A.; Gambin, A. Predicting the outcomes of organic reactions via machine learning: are current descriptors sufficient? *Sci. Rep.* **2017**, *7*, 3582.
- (27) Livingstone, D. *A Practical Guide to Scientific Data Analysis*; John Wiley & Sons Ltd.: Chichester, West Sussex, PO19 8SQ, UK, 2009.
- (28) Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.* **2017**, *50*, 539-543.
- (29) Sameera, W. M. C.; Maeda, S.; Morokuma, K. Computational Catalysis Using the Artificial Force Induced Reaction Method. *Acc. Chem. Res.* **2016**, *49*, 763-773.
- (30) Guan, Y.; Ingman, V. M.; Rooks, B. J.; Wheeler, S. E. AARON: An Automated Reaction Optimizer for New Catalysts. *J. Chem. Theory Comput.* **2018**, DOI:10.1021/acs.jctc.8b00578 10.1021/acs.jctc.8b00578.
- (31) Fey, N.; Ridgway, B. M.; Jover, J.; McMullin, C. L.; Harvey, J. N. Organometallic reactivity: the role of metal–ligand bond energies from a computational perspective. *Dalton Trans.* **2011**, *40*, 11184-11191.
- (32) McMullin, C. L.; Fey, N.; Harvey, J. N. Computed ligand effects on the oxidative addition of phenyl halides to phosphine supported palladium(0) catalysts. *Dalton Trans.* **2014**, *43*, 13545-13556.
- (33) McMullin, C. L.; Jover, J.; Harvey, J. N.; Fey, N. Accurate modelling of Pd(0) + PhX oxidative addition kinetics. *Dalton Trans.* **2010**, *39*, 10833-10836.
- (34) Rush, L. E.; Pringle, P. G.; Harvey, J. N. Computational Kinetics of Cobalt-Catalyzed Alkene Hydroformylation. *Angew. Chem. Int. Ed.* **2014**, *53*, 8672-8676.

- (35) Fey, N.; Tsipis, A. C.; Harris, S. E.; Harvey, J. N.; Orpen, A. G.; Mansson, R. A. Development of a Ligand Knowledge Base, Part 1: Computational Descriptors for Phosphorus Donor Ligands. *Chem. Eur. J.* **2006**, *12*, 291-302.
- (36) Niemeyer, Z. L.; Milo, A.; Hickey, D. P.; Sigman, M. S. Parameterization of phosphine ligands reveals mechanistic pathways and predicts reaction outcomes. *Nature Chem.* **2016**, *8*, 610-617.
- (37) Christian, A. H.; Niemeyer, Z. L.; Sigman, M. S.; Toste, F. D. Uncovering Subtle Ligand Effects of Phosphines Using Gold(I) Catalysis. *ACS Catal.* **2017**, *7*, 3973-3978.
- (38) Jiao, Y.; Torne, M. S.; Gracia, J.; Niemantsverdriet, J. W.; van Leeuwen, P. W. N. M. Ligand effects in rhodium-catalyzed hydroformylation with bisphosphines: steric or electronic? *Catal. Sci. Tech.* **2017**, *7*, 1404-1414.
- (39) Wu, K.; Doyle, A. G. Parameterization of phosphine ligands demonstrates enhancement of nickel catalysis via remote steric effects. *Nature Chem.* **2017**, *9*, 779-784.
- (40) Gillespie, J. A.; Dodds, D. L.; Kamer, P. C. J. Rational design of diphosphorus ligands - a route to superior catalysts. *Dalton Trans.* **2010**, *39*, 2751-2764.
- (41) Fleming, J. T.; Higham, L. J. Primary phosphine chemistry. *Coord. Chem. Rev.* **2015**, *297*, 127-145.
- (42) Lavoie, C. M.; Stradiotto, M. Bisphosphines: A Prominent Ancillary Ligand Class for Application in Nickel-Catalyzed C-N Cross-Coupling. *ACS Catal.* **2018**, *8*, 7228-7250.
- (43) Lundgren, R. J.; Stradiotto, M. Addressing Challenges in Palladium-Catalyzed Cross-Coupling Reactions Through Ligand Design. *Chem. Eur. J.* **2012**, *18*, 9758-9769.
- (44) Martin, R.; Buchwald, S. L. Palladium-Catalyzed Suzuki-Miyaura Cross-Coupling Reactions Employing Dialkylbiaryl Phosphine Ligands. *Acc. Chem. Res.* **2008**, *41*, 1461-1473.
- (45) Ruiz-Castillo, P.; Buchwald, S. L. Applications of Palladium-Catalyzed C-N Cross-Coupling Reactions. *Chem. Rev.* **2016**, *116*, 12564-12649.
- (46) Birkholz, M.-N.; Freixa, Z.; van Leeuwen, P. W. N. M. Bite angle effects of diphosphines in C-C and C-X bond forming cross coupling reactions. *Chem. Soc. Rev.* **2009**, *38*, 1099-1118.
- (47) Fleckenstein, C. A.; Plenio, H. Sterically demanding trialkylphosphines for palladium-catalyzed cross coupling reactions-alternatives to PtBu₃. *Chem. Soc. Rev.* **2010**, *39*, 694-711.
- (48) Bariwal, J.; Van der Eycken, E. C-N bond forming cross-coupling reactions: an overview. *Chem. Soc. Rev.* **2013**, *42*, 9283-9303.
- (49) Erre, G.; Enthaler, S.; Junge, K.; Gladiali, S.; Beller, M. Synthesis and application of chiral monodentate phosphines in asymmetric hydrogenation. *Coord. Chem. Rev.* **2008**, *252*, 471-491.
- (50) Ager, D. J.; de Vries, A. H. M.; de Vries, J. G. Asymmetric homogeneous hydrogenations at scale. *Chem. Soc. Rev.* **2012**, *41*, 3340-3380.
- (51) Franke, R.; Selent, D.; Börner, A. Applied Hydroformylation. *Chem. Rev.* **2012**, *112*, 5675-5732.

- (52) Schmitz, C.; Holthusen, K.; Leitner, W.; Francio, G. Highly Regio- and Enantioselective Hydroformylation of Vinyl Esters Using Bidentate Phosphine,P-Chiral Phosphorodiamidite Ligands. *ACS Catal.* **2016**, *6*, 1584-1589.
- (53) Franke, R.; Selent, D.; Borner, A. Applied Hydroformylation. *Chem. Rev.* **2012**, *112*, 5675-5732.
- (54) Fey, N. The contribution of computational studies to organometallic catalysis: descriptors, mechanisms and models. *Dalton Trans.* **2010**, *39*, 296-310.
- (55) Kuhl, O. Predicting the net donating ability of phosphines - do we need sophisticated theoretical methods? *Coord. Chem. Rev.* **2005**, *249*, 693-704.
- (56) Murray, P. M.; Tyler, S. N. G.; Moseley, J. D. Beyond the Numbers: Charting Chemical Reaction Space. *Org. Proc. Res. Dev.* **2013**, *17*, 40-46.
- (57) Perrin, L.; Clot, E.; Eisenstein, O.; Loch, J.; Crabtree, R. H. Computed ligand electronic parameters, from quantum chemistry and their relation to Tolman parameters, Lever parameters, and Hammett constants. *Inorg. Chem.* **2001**, *40*, 5806-5811.
- (58) Cooney, K. D.; Cundari, T. R.; Hoffman, N. W.; Pittard, K. A.; Temple, M. D.; Zhao, Y. A priori assessment of the stereoelectronic profile of phosphines and phosphites. *J. Am. Chem. Soc.* **2003**, *125*, 4318-4324.
- (59) Gusev, D. G. Donor Properties of a Series of Two-Electron Ligands. *Organometallics* **2009**, *28*, 763-770.
- (60) Lv, X.; Wu, Y.-B.; Lu, G. Computational exploration of ligand effects in copper-catalyzed boracarboxylation of styrene with CO₂. *Catal. Sci. Tech.* **2017**, *7*, 5049-5054.
- (61) Diebolt, O.; Fortman, G. C.; Clavier, H.; Slawin, A. M. Z.; Escudero-Adán, E. C.; Benet-Buchholz, J.; Nolan, S. P. Steric and Electronic Parameters Characterizing Bulky and Electron-Rich Dialkylbiarylphosphines. *Organometallics* **2011**, *30*, 1668-1676.
- (62) Ciancaleoni, G.; Scafuri, N.; Bistoni, G.; Macchioni, A.; Tarantelli, F.; Zuccaccia, D.; Belpassi, L. When the Tolman Electronic Parameter Fails: A Comparative DFT and Charge Displacement Study of [(L)Ni(CO)₃]0⁻ and [(L)Au(CO)]0⁺. *Inorg. Chem.* **2014**, *53*, 9907-9916.
- (63) Kalescky, R.; Kraka, E.; Cremer, D. New Approach to Tolman's Electronic Parameter Based on Local Vibrational Modes. *Inorg. Chem.* **2014**, *53*, 478-495.
- (64) Setiawan, D.; Kalescky, R.; Kraka, E.; Cremer, D. Direct Measure of Metal-Ligand Bonding Replacing the Tolman Electronic Parameter. *Inorg. Chem.* **2016**, *55*, 2332-2344.
- (65) Cremer, D.; Kraka, E. Generalization of the Tolman electronic parameter: the metal-ligand electronic parameter and the intrinsic strength of the metal-ligand bond. *Dalton Trans.* **2017**, *46*, 8323-8338.
- (66) Coll, D. S.; Vidal, A. B.; Rodríguez, J. A.; Ocando-Mavárez, E.; Añez, R.; Sierraalta, A. A simple method for the determination of the Tolman electronic parameter of different phosphorus containing ligands, by means of the average local ionization energy. *Inorg. Chim. Acta* **2015**, *436*, 163-168.

- (67) Ardizzoia, G. A.; Brenna, S. Interpretation of Tolman electronic parameters in the light of natural orbitals for chemical valence. *Phys. Chem. Chem. Phys.* **2017**, *19*, 5971-5978.
- (68) Ziegler, T.; Rauk, A. Calculation of Bonding Energies by Hartree-Fock Slater Method. 1. Transition-State Method. *Theor. Chem. Acc.* **1977**, *46*, 1-10.
- (69) Mitoraj, M.; Michalak, A. Donor-acceptor properties of ligands from the natural orbitals for chemical valence. *Organometallics* **2007**, *26*, 6576-6580.
- (70) Mitoraj, M.; Michalak, A. Natural orbitals for chemical valence as descriptors of chemical bonding in transition metal complexes. *J. Mol. Model.* **2007**, *13*, 347-355.
- (71) Bilbrey, J. A.; Allen, W. D. In *Annual Reports in Computational Chemistry*; Wheeler, R. A., Ed.; Elsevier, 2013; Vol. 9.
- (72) Clavier, H.; Nolan, S. P. Percent buried volume for phosphine and N-heterocyclic carbene ligands: steric properties in organometallic chemistry. *Chem. Commun.* **2010**, *46*, 841-861.
- (73) Bunten, K. A.; Chen, L. Z.; Fernandez, A. L.; Poe, A. J. Cone angles: Tolman's and Plato's. *Coord. Chem. Rev.* **2002**, *233*, 41-51.
- (74) Guan, Y.; Ingman, V. M.; Rooks, B. J.; Wheeler, S. E. AARON: An Automated Reaction Optimizer for New Catalysts. *J. Chem. Theory Comput.* **2018**, *14*, 5249-5261.
- (75) Smith, J. M.; Coville, N. J. Steric Parameters of Conformationally Flexible Ligands from X-ray Structural Data. 2. P(OR)₃ Ligands in Multiple Ligand Environments. *Organometallics* **2001**, *20*, 1210-1215.
- (76) Guzei, I. A.; Wendt, M. An improved method for the computation of ligand steric effects based on solid angles. *Dalton Trans.* **2006**, 3991-3999.
- (77) Molecular Structure Laboratory, Resources, University of Wisconsin, <http://xray.chem.wisc.edu/Resources.html>, accessed on 17th September 2018.
- (78) Bilbrey, J. A.; Kazez, A. H.; Locklin, J.; Allen, W. D. Exact ligand cone angles. *J. Comp. Chem.* **2013**, *34*, 1189-1197.
- (79) Software from our Publications, CCQC, <https://www.ccqc.uga.edu/references/software.php>, accessed on 17th September 2018.
- (80) Petitjean, M. Analytical algorithms for ligand cone angles calculations. Application to triphenylphosphine palladium complexes. *Comptes Rendus Chimie* **2015**, *18*, 678-684.
- (81) Michel Petitjean / softwares / freewares, Petitjean, M., <http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html>, accessed on 17th September 2018.
- (82) Bilbrey, J. A.; Kazez, A. H.; Locklin, J.; Allen, W. D. Exact Ligand Solid Angles. *J. Chem. Theory Comput.* **2013**, *9*, 5734-5744.
- (83) Dunne, B. J.; Morris, R. B.; Orpen, A. G. Structural systematics. Part 3. Geometry deformations in triphenylphosphine fragments: a test of bonding theories in phosphine complexes. *Dalton Trans.* **1991**, DOI:10.1039/DT9910000653 10.1039/DT9910000653, 653-661.

- (84) Viciu, M. S.; Navarro, O.; Germaneau, R. F.; Kelly, R. A.; Sommer, W.; Marion, N.; Stevens, E. D.; Cavallo, L.; Nolan, S. P. Synthetic and structural studies of (NHC)Pd(allyl)Cl complexes (NHC = N-heterocyclic carbene). *Organometallics* **2004**, *23*, 1629-1635.
- (85) Poater, A.; Cosenza, B.; Correa, A.; Giudice, S.; Ragone, F.; Scarano, V.; Cavallo, L. SambVca: A Web Application for the Calculation of the Buried Volume of N-Heterocyclic Carbene Ligands. *Eur. J. Inorg. Chem.* **2009**, *2009*, 1759-1766.
- (86) SambVca: A Web Application for the Calculation of the Buried Volume of Organometallic Ligands, Poater, A.; Cosenza, B.; Correa, A.; Giudice, S.; Ragone, F.; Scarano, V.; Cavallo, L., <https://www.molnac.unisa.it/OMtools/SambVca-Manual.html>, accessed on 17th September 2018.
- (87) Falivene, L.; Credendino, R.; Poater, A.; Petta, A.; Serra, L.; Oliva, R.; Scarano, V.; Cavallo, L. SambVca 2. A Web Tool for Analyzing Catalytic Pockets with Topographic Steric Maps. *Organometallics* **2016**, *35*, 2286-2293.
- (88) SambVca 2.0, <https://www.molnac.unisa.it/OMtools/sambvca2.0/>, accessed on 17th September 2018.
- (89) Kendall, A. J.; Zakharov, L. N.; Tyler, D. R. Steric and Electronic Influences of Buchwald-Type Alkyl-JohnPhos Ligands. *Inorg. Chem.* **2016**, *55*, 3079-3090.
- (90) Harper, K. C.; Vilardi, S. C.; Sigman, M. S. Prediction of Catalyst and Substrate Performance in the Enantioselective Propargylation of Aliphatic Ketones by a Multidimensional Model of Steric Effects. *J. Am. Chem. Soc.* **2013**, *135*, 2482-2485.
- (91) Verloop, A.; Hoogenstraten, W.; Tipker, A. In *Drug Design*; Ariens, E. J., Ed.; Academic Press: New York, 1976; Vol. VII. Cited in <https://www.degruyter.com/downloadpdf/j/znc.1996.51.issue-1-2/znc-1996-1-202/znc-1996-1-202.pdf>.
- (92) Ardkhean, R.; Mortimore, M.; Paton, R. S.; Fletcher, S. P. Formation of quaternary centres by copper catalysed asymmetric conjugate addition to β -substituted cyclopentenones with the aid of a quantitative structure–selectivity relationship. *Chem. Sci.* **2018**, *9*, 2628-2632.
- (93) Maldonado, A. G.; Rothenberg, G. Predictive modeling in homogeneous catalysis: a tutorial. *Chem. Soc. Rev.* **2010**, *39*, 1891-1902.
- (94) Rothenberg, G. Data mining in catalysis: Separating knowledge from garbage. *Catalysis Today* **2008**, *137*, 2-10.
- (95) Mansson, R. A.; Welsh, A. H.; Fey, N.; Orpen, A. G. Statistical Modeling of a Ligand Knowledge Base. *J. Chem. Inf. Model.* **2006**, *46*, 2591-2600.
- (96) Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M. Expansion of the Ligand Knowledge Base for Monodentate P-Donor Ligands (LKB-P). *Organometallics* **2010**, *29*, 6245-6258.
- (97) Fey, N.; Papadouli, S.; Pringle, P. G.; Ficks, A.; Fleming, J. T.; Higham, L. J.; Wallis, J. F.; Carmichael, D.; Mézailles, N.; Müller, C. Setting P-Donor Ligands into Context: An Application of the Ligand Knowledge Base (LKB) Approach. *P, S, Si, Rel. Elem.* **2015**, *190*, 706-714.
- (98) Mathew, J.; Thomas, T.; Suresh, C. H. Quantitative Assessment of the Stereoelectronic Profile of Phosphine Ligands. *Inorg. Chem.* **2007**, *46*, 10800-10809.

- (99) Mathew, J.; Suresh, C. H. Use of Molecular Electrostatic Potential at the Carbene Carbon as a Simple and Efficient Electronic Parameter of N-heterocyclic Carbenes. *Inorg. Chem.* **2010**, *49*, 4665-4669.
- (100) Mathew, J.; Suresh, C. H. Assessment of Steric and Electronic Effects of N-Heterocyclic Carbenes in Grubbs Olefin Metathesis Using Molecular Electrostatic Potential. *Organometallics* **2011**, *30*, 3106-3112.
- (101) Anjali, B. A.; Suresh, C. H. Interpreting Oxidative Addition of Ph-X (X = CH₃, F, Cl, and Br) to Monoligated Pd(0) Catalysts Using Molecular Electrostatic Potential. *ACS Omega* **2017**, *2*, 4196-4206.
- (102) Suresh, C. H.; Koga, N. Quantifying the Electronic Effect of Substituted Phosphine Ligands via Molecular Electrostatic Potential. *Inorg. Chem.* **2002**, *41*, 1573-1578.
- (103) Suresh, C. H. Molecular Electrostatic Potential Approach to Determining the Steric Effect of Phosphine Ligands in Organometallic Chemistry. *Inorg. Chem.* **2006**, *45*, 4982-4986.
- (104) Mathew, J.; Suresh, C. H. Assessment of Stereoelectronic Effects in Grubbs First-Generation Olefin Metathesis Catalysis Using Molecular Electrostatic Potential. *Organometallics* **2011**, *30*, 1438-1444.
- (105) Orpen, A. G.; Connelly, N. G. Structural Evidence for the Participation of P-X-Sigma-Star Orbitals in Metal-PX₃ Bonding. *Chem. Commun.* **1985**, 1310-1311.
- (106) Fey, N.; Garland, M.; Hopewell, J. P.; McMullin, C. L.; Mastroianni, S.; Orpen, A. G.; Pringle, P. G. Stable Fluorophosphines: Predicted and Realized Ligands for Catalysis. *Angew. Chem. Int. Ed.* **2012**, *51*, 118-122.
- (107) Chuang, K. V.; Keiser, M. J. Comment on "Predicting reaction performance in C-N cross-coupling using machine learning". *Science* **2018**, *362*, eaat8603.
- (108) Estrada, J. G.; Ahneman, D. T.; Sheridan, R. P.; Dreher, S. D.; Doyle, A. G. Response to Comment on "Predicting reaction performance in C-N cross-coupling using machine learning". *Science* **2018**, *362*, eaat8763.
- (109) Dispute over reaction prediction puts machine learning's pitfalls in spotlight, Kraemer, K., <https://www.chemistryworld.com/news/dispute-over-reaction-prediction-puts-machine-learning-pitfalls-in-spotlight/3009912.article>, accessed on 22/01/2019.
- (110) Machine Learning: Be Careful What You Ask For, Lowe, D., <http://blogs.sciencemag.org/pipeline/archives/2018/11/20/machine-learning-be-careful-what-you-ask-for>, accessed on 22/01/2019.
- (111) van Leeuwen, P. W. N. M.; Kamer, P. C. J.; Reek, J. N. H.; Dierkes, P. Ligand Bite Angle Effects in Metal-catalyzed C-C Bond Formation. *Chem. Rev.* **2000**, *100*, 2741-2770.
- (112) Birkholz, M. N.; Freixa, Z.; van Leeuwen, P. Bite angle effects of diphosphines in C-C and C-X bond forming cross coupling reactions. *Chem. Soc. Rev.* **2009**, *38*, 1099-1118.
- (113) Maldonado, A. G.; Hageman, J. A.; Mastroianni, S.; Rothenberg, G. Backbone Diversity Analysis in Catalyst Design. *Adv. Synth. Catal.* **2009**, *351*, 387-396.

- (114) Hageman, J. A.; Westerhuis, J. A.; Fruhauf, H. W.; Rothenberg, G. Design and assembly of virtual homogeneous catalyst libraries - Towards in silico catalyst optimisation. *Adv. Synth. Catal.* **2006**, *348*, 361-369.
- (115) Burello, E.; Rothenberg, G. Topological mapping of bidentate ligands: A fast approach for screening homogeneous catalysts. *Adv. Synth. Catal.* **2005**, *347*, 1969-1977.
- (116) Dierkes, P.; W. N. M. van Leeuwen, P. The bite angle makes the difference: a practical ligand parameter for diphosphine ligands. *Dalton Trans.* **1999**, 1519-1530.
- (117) Freixa, Z.; van Leeuwen, P. W. N. M. Bite angle effects in diphosphine metal catalysts: steric or electronic? *Dalton Trans.* **2003**, 1890-1901.
- (118) Kuhl, O. The natural bite angle - Seen from a ligand's point of view. *Can. J. Chem.* **2007**, *85*, 230-238.
- (119) Mansell, S. M. Catalytic applications of small bite-angle diphosphorus ligands with single-atom linkers. *Dalton Trans.* **2017**, *46*, 15157-15174.
- (120) Niksch, T.; Görls, H.; Weigand, W. The Extension of the Solid-Angle Concept to Bidentate Ligands. *Eur. J. Inorg. Chem.* **2010**, *2010*, 95-105.
- (121) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Cryst. B* **2016**, *72*, 171-179.
- (122) Al-Ghamdi, M.; Vummaleti, S. V. C.; Falivene, L.; Pasha, F. A.; Beetstra, D. J.; Cavallo, L. Structure–Activity Relationship To Screen Ni–Bisphosphine Complexes for the Oxidative Coupling of CO₂ and Ethylene. *Organometallics* **2017**, *36*, 1107-1112.
- (123) Keylor, M. H.; Niemeyer, Z. L.; Sigman, M. S.; Tan, K. L. Inverting Conventional Chemoselectivity in Pd-Catalyzed Amine Arylations with Multiply Halogenated Pyridines. *J. Am. Chem. Soc.* **2017**, *139*, 10613-10616.
- (124) Jover, J.; Fey, N. Screening substituent and backbone effects on the properties of bidentate P,P-donor ligands (LKB-PPscreen). *Dalton Trans.* **2013**, *42*, 172-181.
- (125) Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Murray, P.; Orpen, A. G.; Osborne, R.; Purdie, M. Computational Descriptors for Chelating P,P- and P,N-Donor Ligands¹. *Organometallics* **2008**, *27*, 1372-1383.
- (126) Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M. Expansion of the Ligand Knowledge Base for Chelating P,P-Donor Ligands (LKB-PP). *Organometallics* **2012**, *31*, 5302-5306.
- (127) Newland, R. J.; Smith, A.; Smith, D. M.; Fey, N.; Hanton, M. J.; Mansell, S. M. Accessing Alkyl- and Alkenylcyclopentanes from Cr-Catalyzed Ethylene Oligomerization Using 2-Phosphinophosphinine Ligands. *Organometallics* **2018**, *37*, 1062-1073.
- (128) Flener Lovitt, C.; Frenking, G.; Girolami, G. S. Donor–Acceptor Properties of Bidentate Phosphines. DFT Study of Nickel Carbonyls and Molecular Dihydrogen Complexes. *Organometallics* **2012**, *31*, 4122-4132.

- (129) Furstner, A.; Ackermann, L.; Gabor, B.; Goddard, R.; Lehmann, C. W.; Mynott, R.; Stelzer, F.; Thiel, O. R. Comparative investigation of ruthenium-based metathesis catalysts bearing N-heterocyclic carbene (NHC) ligands. *Chem. Eur. J.* **2001**, *7*, 3236-3253.
- (130) Trnka, T. M.; Grubbs, R. H. The development of L2X2Ru = CHR olefin metathesis catalysts: An organometallic success story. *Acc. Chem. Res.* **2001**, *34*, 18-29.
- (131) Kumar, A.; Ghosh, P. Studies of the Electronic Properties of N-Heterocyclic Carbene Ligands in the Context of Homogeneous Catalysis and Bioorganometallic Chemistry. *Eur. J. Inorg. Chem.* **2012**, DOI:doi:10.1002/ejic.201200622 doi:10.1002/ejic.201200622, 3955-3969.
- (132) Cheng, J.; Wang, L.; Wang, P.; Deng, L. High-Oxidation-State 3d Metal (Ti–Cu) Complexes with N-Heterocyclic Carbene Ligation. *Chem. Rev.* **2018**, *118*, 9930-9987.
- (133) Riener, K.; Haslinger, S.; Raba, A.; Högerl, M. P.; Cokoja, M.; Herrmann, W. A.; Kühn, F. E. Chemistry of Iron N-Heterocyclic Carbene Complexes: Syntheses, Structures, Reactivities, and Catalytic Applications. *Chem. Rev.* **2014**, *114*, 5215-5272.
- (134) Froese, R. D. J.; Lombardi, C.; Pompeo, M.; Rucker, R. P.; Organ, M. G. Designing Pd–N-Heterocyclic Carbene Complexes for High Reactivity and Selectivity for Cross-Coupling Applications. *Acc. Chem. Res.* **2017**, *50*, 2244-2253.
- (135) Iglesias, M.; Oro, L. A. A leap forward in iridium–NHC catalysis: new horizons and mechanistic insights. *Chem. Soc. Rev.* **2018**, *47*, 2772-2808.
- (136) Lazreg, F.; Nahra, F.; Cazin, C. S. J. Copper–NHC complexes in catalysis. *Coord. Chem. Rev.* **2015**, *293-294*, 48-79.
- (137) Tobisu, M.; Chatani, N. Cross-Couplings Using Aryl Ethers via C–O Bond Activation Enabled by Nickel Catalysts. *Acc. Chem. Res.* **2015**, *48*, 1717-1726.
- (138) Fortman, G. C.; Nolan, S. P. N-Heterocyclic carbene (NHC) ligands and palladium in homogeneous cross-coupling catalysis: a perfect union. *Chem. Soc. Rev.* **2011**, *40*, 5151-5169.
- (139) Froese, R. D. J.; Lombardi, C.; Pompeo, M.; Rucker, R. P.; Organ, M. G. Designing Pd N-Heterocyclic Carbene Complexes for High, Reactivity and Selectivity for Cross-Coupling Applications. *Acc. Chem. Res.* **2017**, *50*, 2244-2253.
- (140) Diez-Gonzalez, S.; Nolan, S. P. Stereoelectronic parameters associated with N-heterocyclic carbene (NHC) ligands: A quest for understanding. *Coord. Chem. Rev.* **2007**, *251*, 874-883.
- (141) Dröge, T.; Glorius, F. The Measure of All Rings—N-Heterocyclic Carbenes. *Angew. Chem. Int. Ed.* **2010**, *49*, 6940-6952.
- (142) Nelson, D. J.; Nolan, S. P. Quantifying and understanding the electronic properties of N-heterocyclic carbenes. *Chem. Soc. Rev.* **2013**, *42*, 6723-6753.
- (143) Munz, D. Pushing Electrons—Which Carbene Ligand for Which Application? *Organometallics* **2018**, *37*, 275-289.
- (144) Dorta, R.; Stevens, E. D.; Hoff, C. D.; Nolan, S. P. Stable, three-coordinate Ni(CO)(2)(NHC) (NHC = N-heterocyclic carbene) complexes enabling the determination of Ni-NHC bond energies. *J. Am. Chem. Soc.* **2003**, *125*, 10490-10491.

- (145) Dorta, R.; Stevens, E. D.; Scott, N. M.; Costabile, C.; Cavallo, L.; Hoff, C. D.; Nolan, S. P. Steric and electronic properties of N-heterocyclic carbenes (NHC): A detailed study on their interaction with Ni(CO)₄. *J. Am. Chem. Soc.* **2005**, *127*, 2485-2495.
- (146) Huynh, H. V. Electronic Properties of N-Heterocyclic Carbenes and Their Experimental Determination. *Chem. Rev.* **2018**.
- (147) Gusev, D. G. Electronic and Steric Parameters of 76 N-Heterocyclic Carbenes in Ni(CO)₃(NHC). *Organometallics* **2009**, *28*, 6458-6461.
- (148) Gusev, D. G.; Peris, E. The Tolman electronic parameter (TEP) and the metal–metal electronic communication in ditopic NHC complexes. *Dalton Trans.* **2013**, *42*, 7359-7364.
- (149) Teng, Q.; Huynh, H. V. A unified ligand electronic parameter based on ¹³C NMR spectroscopy of N-heterocyclic carbene complexes. *Dalton Trans.* **2017**, *46*, 614-627.
- (150) Back, O.; Henry-Ellinger, M.; Martin, C. D.; Martin, D.; Bertrand, G. ³¹P NMR Chemical Shifts of Carbene–Phosphinidene Adducts as an Indicator of the π -Accepting Properties of Carbenes. *Angew. Chem. Int. Ed.* **2013**, *52*, 2939-2943.
- (151) Liske, A.; Verlinden, K.; Buhl, H.; Schaper, K.; Ganter, C. Determining the π -Acceptor Properties of N-Heterocyclic Carbenes by Measuring the ⁷⁷Se NMR Chemical Shifts of Their Selenium Adducts. *Organometallics* **2013**, *32*, 5269-5272.
- (152) Verlinden, K.; Buhl, H.; Frank, W.; Ganter, C. Determining the Ligand Properties of N-Heterocyclic Carbenes from ⁷⁷Se NMR Parameters. *Eur. J. Inorg. Chem.* **2015**, *2015*, 2416-2425.
- (153) Vummaleti, S. V. C.; Nelson, D. J.; Poater, A.; Gómez-Suárez, A.; Cordes, D. B.; Slawin, A. M. Z.; Nolan, S. P.; Cavallo, L. What can NMR spectroscopy of selenoureas and phosphinidenes teach us about the π -accepting abilities of N-heterocyclic carbenes? *Chem. Sci.* **2015**, *6*, 1895-1904.
- (154) Engl, P. S.; Santiago, C. B.; Gordon, C. P.; Liao, W.-C.; Fedorov, A.; Copéret, C.; Sigman, M. S.; Togni, A. Exploiting and Understanding the Selectivity of Ru-N-Heterocyclic Carbene Metathesis Catalysts for the Ethenolysis of Cyclic Olefins to α,ω -Dienes. *J. Am. Chem. Soc.* **2017**, *139*, 13117-13125.
- (155) Guha, A. K.; Borthakur, B.; Phukan, A. K. Spectroscopic Distinction of Different Carbon Bases: An Insight from Theory. *J. Org. Chem.* **2015**, *80*, 7301-7304.
- (156) Phukan, A. K.; Guha, A. K.; Sarmah, S. Ligand Properties of Boron-Substituted Five-, Six-, and Seven-Membered Heterocyclic Carbenes: A Theoretical Study. *Organometallics* **2013**, *32*, 3238-3248.
- (157) Phukan, A. K.; Guha, A. K.; Sarmah, S.; Dewhurst, R. D. Electronic and Ligand Properties of Annulated Normal and Abnormal (Mesoionic) N-Heterocyclic Carbenes: A Theoretical Study. *J. Org. Chem.* **2013**, *78*, 11032-11039.
- (158) Dutta, S.; Maity, B.; Thirumalai, D.; Koley, D. Computational Investigation of Carbene–Phosphinidenes: Correlation between ³¹P Chemical Shifts and Bonding Features to Estimate the π -Backdonation of Carbenes. *Inorg. Chem.* **2018**, *57*, 3993-4008.
- (159) Ramsden, C. A.; Oziminski, W. P. Quantitative Index of the Relative Ease of Formation and σ -Bonding Strength of N-Heterocyclic Carbenes. *J. Org. Chem.* **2016**, *81*, 10295-10301.

- (160) Ramsden, C. A.; Oziminski, W. P. A Quantitative Analysis of Factors Influencing Ease of Formation and σ -Bonding Strength of Oxa- and Thia-N-Heterocyclic Carbenes. *J. Org. Chem.* **2017**, *82*, 12485-12491.
- (161) Gómez-Suárez, A.; Nelson, D. J.; Nolan, S. P. Quantifying and understanding the steric properties of N-heterocyclic carbenes. *Chem. Commun.* **2017**, *53*, 2650-2660.
- (162) Ragone, F.; Poater, A.; Cavallo, L. Flexibility of N-Heterocyclic Carbene Ligands in Ruthenium Complexes Relevant to Olefin Metathesis and Their Impact in the First Coordination Sphere of the Metal. *J. Am. Chem. Soc.* **2010**, *132*, 4249-4258.
- (163) Chu, Y.; Heyndrickx, W.; Occhipinti, G.; Jensen, V. R.; Alsberg, B. K. An Evolutionary Algorithm for de Novo Optimization of Functional Transition Metal Compounds. *J. Am. Chem. Soc.* **2012**, *134*, 8885-8895.
- (164) Ajitha, M. J.; Suresh, C. H. Assessment of Stereoelectronic Factors That Influence the CO₂ Fixation Ability of N-Heterocyclic Carbenes: A DFT Study. *J. Org. Chem.* **2012**, *77*, 1087-1094.
- (165) Cavallo, L. Mechanism of ruthenium-catalyzed olefin metathesis reactions from a theoretical perspective. *J. Am. Chem. Soc.* **2002**, *124*, 8965-8973.
- (166) Fey, N.; Haddow, M. F.; Harvey, J. N.; McMullin, C. L.; Orpen, A. G. A ligand knowledge base for carbenes (LKB-C): maps of ligand space. *Dalton Trans.* **2009**, 8183-8196.
- (167) Durand, D. J.; McMullin, C. L.; Fey, N. *in preparation for Dalton Transactions* **2019**.
- (168) Cases, M.; Frenking, G.; Duran, M.; Sola, M. Molecular structure and bond characterization of the Fischer-type chromium-carbene complexes (CO)₅Cr = C(X)R (X = H, OH, OCH₃, NH₂, NHCH₃) and R = H, CH₃, CH = CH₂, Ph, C₆H₅. *Organometallics* **2002**, *21*, 4182-4191.
- (169) Frenking, G.; Sola, M.; Vyboishchikov, S. F. Chemical bonding in transition metal carbene complexes. *J. Organomet. Chem.* **2005**, *690*, 6178-6204.
- (170) Tonner, R.; Heydenrych, G.; Frenking, G. Bonding Analysis of N-Heterocyclic Carbene Tautomers and Phosphine Ligands in Transition-Metal Complexes: A Theoretical Study†. *Chem. Asian J.* **2007**, *2*, 1555-1567.
- (171) Heydenrych, G.; von Hopffgarten, M.; Stander, E.; Schuster, O.; Raubenheimer, H. G.; Frenking, G. The Nature of the Metal-Carbene Bond in Normal and Abnormal Pyridylidene, Quinolyidene and Isoquinolyidene Complexes. *Eur. J. Inorg. Chem.* **2009**, DOI:10.1002/ejic.200801244 10.1002/ejic.200801244, 1892-1904.
- (172) Bistoni, G.; Belpassi, L.; Tarantelli, F. Advances in Charge Displacement Analysis. *J. Chem. Theory Comput.* **2016**, *12*, 1236-1244.
- (173) Huynh, H. V.; Frison, G. Electronic Structural Trends in Divalent Carbon Compounds. *J. Org. Chem.* **2013**, *78*, 328-338.
- (174) Comas-Vives, A.; Harvey, J. N. How Important Is Backbonding in Metal Complexes Containing N-Heterocyclic Carbenes? Structural and NBO Analysis. *Eur. J. Inorg. Chem.* **2011**, *2011*, 5025-5035.
- (175) Carter, E. A.; Goddard III, W. A. Relation between Singlet-Triplet Gaps and Bond Energies. *The Journal of Physical Chemistry* **1986**, *90*, 998-1001.

- (176) Heinemann, C.; Thiel, W. Ab initio study on the stability of diaminocarbenes. *Chemical Physics Letters* **1994**, *217*, 11-16.
- (177) Poater, A.; Ragone, F.; Giudice, S.; Costabile, C.; Dorta, R.; Nolan, S. P.; Cavallo, L. Thermodynamics of N-Heterocyclic Carbene Dimerization: The Balance of Sterics and Electronics. *Organometallics* **2008**, *27*, 2679-2681.
- (178) Guha, A. K.; Sarmah, S.; Phukan, A. K. Effect of substituents at the heteroatom on the structure and ligating properties of heterocyclic carbene, silylene, germylene and abnormal carbene: A theoretical study. *Dalton Trans.* **2010**, *39*, 7374-7383.
- (179) Guha, A. K.; Das, C.; Phukan, A. K. Heterocyclic carbenes of diverse flexibility: A theoretical insight. *J. Organomet. Chem.* **2011**, *696*, 586-593.
- (180) Sarmah, S.; Guha, A. K.; Phukan, A. K. Donor–Acceptor Complexes of Normal and Abnormal N-Heterocyclic Carbenes with Group 13 (B, Al, Ga) Elements: A Combined DFT and Atoms-in-Molecules Study. *Eur. J. Inorg. Chem.* **2013**, *2013*, 3233-3239.
- (181) Borthakur, B.; Rahman, T.; Phukan, A. K. Tuning the Electronic and Ligand Properties of Remote Carbenes: A Theoretical Study. *J. Org. Chem.* **2014**, *79*, 10801-10810.
- (182) Guha, A. K.; Phukan, A. K. Theoretical Study on the Effect of Annulation and Carbonylation on the Electronic and Ligand Properties of N-Heterocyclic Silylenes and Germylenes: Carbene Comparisons begin To Break Down. *J. Org. Chem.* **2014**, *79*, 3830-3837.
- (183) Borthakur, B.; Silvi, B.; Dewhurst, R. D.; Phukan, A. K. Theoretical strategies toward stabilization of singlet remote N-heterocyclic carbenes. *J. Comp. Chem.* **2016**, *37*, 1484-1490.
- (184) Bharadwaz, P.; Chetia, P.; Phukan, A. K. Electronic and Ligand Properties of Skeletally Substituted Cyclic (Alkyl)(Amino)Carbenes (CAACs) and Their Reactivity towards Small Molecule Activation: A Theoretical Study. *Chem. Eur. J.* **2017**, *23*, 9926-9936.
- (185) Schleyer, P. v. R.; Maerker, C.; Dransfeld, A.; Jiao, H.; van Eikema Hommes, N. J. R. Nucleus-Independent Chemical Shifts: A Simple and Efficient Aromaticity Probe. *J. Am. Chem. Soc.* **1996**, *118*, 6317-6318.
- (186) Bernhammer, J. C.; Frison, G.; Huynh, H. V. Electronic Structure Trends in N-Heterocyclic Carbenes (NHCs) with Varying Number of Nitrogen Atoms and NHC–Transition-Metal Bond Properties. *Chem. Eur. J.* **2013**, *19*, 12892-12905.
- (187) du Toit, J. I.; van Sittert, C. G. C. E.; Vosloo, H. C. M. Towards a better understanding of alkene metathesis: elucidating the properties of the major metal carbene catalyst types. *Monatshefte Chemie* **2015**, *146*, 1115-1129.
- (188) Teixeira, F.; Mosquera, R. A.; Melo, A.; Freire, C.; Cordeiro, M. N. D. S. Principal component analysis of Mn(salen) catalysts. *Phys. Chem. Chem. Phys.* **2014**, *16*, 25364-25376.
- (189) Denny, J. A.; Darensbourg, M. Y. Approaches to quantifying the electronic and steric properties of metallodithiolates as ligands in coordination chemistry. *Coord. Chem. Rev.* **2016**, *324*, 82-89.
- (190) Ioannidis, E. I.; Gani, T. Z. H.; Kulik, H. J. molSimplify: A toolkit for automating discovery in inorganic chemistry. *J. Comp. Chem.* **2016**, *37*, 2106-2117.

- (191) Janet, J. P.; Kulik, H. J. Predicting electronic structure properties of transition metal complexes with neural networks. *Chem. Sci.* **2017**, *8*, 5137-5152.
- (192) Bemowski, R. D.; Singh, A. K.; Bajorek, B. J.; DePorre, Y.; Odom, A. L. Effective donor abilities of E-t-Bu and EPh (E = O, S, Se, Te) to a high valent transition metal. *Dalton Trans.* **2014**, *43*, 12299-12305.
- (193) DiFranco, S. A.; Maciulis, N. A.; Staples, R. J.; Batrice, R. J.; Odom, A. L. Evaluation of Donor and Steric Properties of Anionic Ligands on High Valent Transition Metals. *Inorg. Chem.* **2012**, *51*, 1187-1200.
- (194) Billow, B. S.; McDaniel, T. J.; Odom, A. L. Quantifying ligand effects in high-oxidation-state metal catalysis. *Nature Chem.* **2017**, *9*, 837-842.
- (195) Piou, T.; Romanov-Michailidis, F.; Romanova-Michaelides, M.; Jackson, K. E.; Semakul, N.; Taggart, T. D.; Newell, B. S.; Rithner, C. D.; Paton, R. S.; Rovis, T. Correlating Reactivity and Selectivity to Cyclopentadienyl Ligand Properties in Rh(III)-Catalyzed C–H Activation Reactions: An Experimental and Computational Study. *J. Am. Chem. Soc.* **2017**, *139*, 1296-1310.
- (196) Landman, I. R.; Paulson, E. R.; Rheingold, A. L.; Grotjahn, D. B.; Rothenberg, G. Designing bifunctional alkene isomerization catalysts using predictive modelling. *Catal. Sci. Tech.* **2017**, *7*, 4842-4851.
- (197) Canac, Y.; Lepetit, C. Classification of the Electronic Properties of Chelating Ligands in cis-[LL'Rh(CO)₂] Complexes. *Inorg. Chem.* **2017**, *56*, 667-675.
- (198) Guo, J.-Y.; Minko, Y.; Santiago, C. B.; Sigman, M. S. Developing Comprehensive Computational Parameter Sets To Describe the Performance of Pyridine-Oxazoline and Related Ligands. *ACS Catal.* **2017**, *7*, 4144-4151.
- (199) Park, Y.; Niemeyer, Z. L.; Yu, J.-Q.; Sigman, M. S. Quantifying Structural Effects of Amino Acid Ligands in Pd(II)-Catalyzed Enantioselective C–H Functionalization Reactions. *Organometallics* **2018**, *37*, 203-210.
- (200) Moseley, J. D.; Murray, P. M. Ligand and solvent selection in challenging catalytic reactions. *J. Chem. Techn. Bio.* **2014**, *89*, 623-632.
- (201) Niemeyer, Z. L.; Pindi, S.; Khrakovsky, D. A.; Kuzniewski, C. N.; Hong, C. M.; Joyce, L. A.; Sigman, M. S.; Toste, F. D. Parameterization of Acyclic Diaminocarbene Ligands Applied to a Gold(I)-Catalyzed Enantioselective Tandem Rearrangement/Cyclization. *J. Am. Chem. Soc.* **2017**, *139*, 12943-12946.
- (202) Liao, K.; Liu, W.; Niemeyer, Z. L.; Ren, Z.; Bacsá, J.; Musaev, D. G.; Sigman, M. S.; Davies, H. M. L. Site-Selective Carbene-Induced C–H Functionalization Catalyzed by Dirhodium Tetrakis(triarylcyclopropanecarboxylate) Complexes. *ACS Catal.* **2018**, *8*, 678-682.
- (203) Stauffer, S. R.; Hartwig, J. F. Fluorescence Resonance Energy Transfer (FRET) as a High-Throughput Assay for Coupling Reactions. Arylation of Amines as a Case Study. *J. Am. Chem. Soc.* **2003**, *125*, 6977-6985.
- (204) Jover, J.; Fey, N.; Purdie, M.; Lloyd-Jones, G. C.; Harvey, J. N. A computational study of phosphine ligand effects in Suzuki–Miyaura coupling. *J. Mol. Catal. A* **2010**, *324*, 39-47.

(205) Foscatto, M.; Occhipinti, G.; Venkatraman, V.; Alsberg, B. K.; Jensen, V. R. Automated Design of Realistic Organometallic Molecules from Fragments. *J. Chem. Inf. Model.* **2014**, *54*, 767-780.