Dalton Transactions

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Dalton Trans.*, 2025, **54**, 16380

Electronic effects in early transition metal catalyzed olefin polymerization: challenges in featurization and descriptor strengths and weaknesses

Christian Ehm, *[a] Gianluigi Galasso,[b] Antonio Vittoria,[a] Pietro Oriente,[a] Luigi Maiale,[a] Peter H. M. Budzelaar,[a] Roberta Cipullo[a] and Vincenzo Busico[a]

In olefin polymerization, even seemingly simple concepts like the influence of electronic effects have sometimes eluded qualitative understanding despite 70 years of continuing research in the field. Of course, the intimate coupling of electronic and steric effects that olefin polymerization is so famous for might be simply too complex, with data science approaches – which are rapidly gaining adoption in catalysis – being the only way to solve the puzzle. Data science relies on machine-readable features or descriptors that encode essential aspects of the catalysts, and the accuracy of models depends both on the quality of data and featurization. Here, we show that some of the basic assumptions used so far (in any kind of modelling) may be flawed to the extent that they prevent accurate evaluation (and separation) of steric and electronic effects. We undertake a comprehensive analysis of the suitability of different model structures for data science approaches and analyze the performance, reliability, and data spacing of common electronic descriptors determined thereof for several group 3 and group 4 metal complexes. The insight developed in this work points not only to the complexity of the underlying chemistry being problematic but also to the inefficiency of many commonly employed descriptors in properly capturing electronic effects relevant for olefin polymerization. Recognizing the strengths and weaknesses of various approaches may help researchers select appropriate features/descriptors and better understand the scope of models beyond the initial training data.

Received 9th September 2025,
Accepted 16th October 2025

DOI: 10.1039/d5dt02170g

rsc.li/dalton

Introduction

The precise role and magnitude of electronic effects in olefin polymerization catalysis remain largely subject to ongoing debate.[1–16] Certain mechanisms – in particular enantiomorphic site control in primary insertion of α-olefins – are understood to be dominated by steric effects;[17–19] and the understanding of such effects is clear. However, for example in propene polymerization with C_2-symmetric catalysts for the synthesis of isotactic polypropylene (i-PP), chain epimerization (which is considered to be strongly dependent on electronic effects) can limit stereoregularity and may be the most important contribution to stereoerror formation,[20–22] at higher temperatures. Other catalyst performance indicators that are thought to be influenced (also) by electronic effects are regioselectivity in propene polymerization,[1,23] comonomer affinity in ethene/1-alkene copolymerization,[3,4,7] molar mass capability,[2,5,6] and activity.[10,11,24] In general, electronic effects are undoubtedly important but, with few exceptions, authors usually state a variation of "*the observed differences between the performance of different catalysts are likely the result of a subtle balance of steric and electronic effects*". This uncertainty can be traced to the fact that we do not have universally accepted ways of quantifying electronic and steric effects, much less even separating them.

It is important to note that in olefin polymerization catalysis, the roles of metal, ligand set, monomer(s), desired product and reaction conditions are intimately tied, a phenomenon observed for nearly the entire transition metal series.[14] The only generally accepted prerequisite for polymerization reactivity according to the Cossee–Arlman mechanism is an open coordination site and a metal–carbon bond in mutual *cis* arrangement;[25,26] in practice, an electron count of 14 or lower appears to be required. For industrial applications, group 4 based molecular catalysts are most relevant due to their high

[a]*Dipartimento di Scienze Chimiche, Università di Napoli Federico II, 80126 Napoli, Italy. E-mail: christian.ehm@unina.it*
[b]*Scuola Superiore Meridionale, Via Mezzocannone 4, 80134 Napoli, Italy*

activities, but some of the most inspiring approaches to the synthesis of chain end functionalized polyolefins, with medical application potential, have involved the rare earth metals of group 3.[27]

The active species derived from group 3 (oxidation state +3) are isoelectronic with group 4 (oxidation state +4) analogs and development of ligand backbones has largely proceeded in concert over the years. Even considering only early transition metals, almost identical polymers can be prepared from catalysts with vastly different features (metal, ligand set, substituents, see Fig. S1, SI). However, a number of "perfect" ligand/metal *pairings* exist (Fig. 1).[13,28–35]

Predicting the performance of a given ligand/metal pairing in olefin polymerization by DFT is challenging, in part because of the required accuracy (≪1 kcal mol^{-1}) but also for the large number of transition states needed to predict selectivity, accounting for several olefin and chain orientations. Additionally, solvent, anion, and other effects like scavenger presence can complicate the picture. Data science approaches have been used in olefin polymerization for decades but often target focused datasets,[1,3–5,23,36–39] lacking diversity and depth. Models usually need to be retrained for different metal centers, catalyst classes, or reaction conditions. In part, this can be traced to the fact that olefin polymerization is also extremely sensitive to exact reaction conditions;[18,20,24,40–42] screening conditions vary widely and as a result, literature databases are heavily fragmented. This notwithstanding, data science approaches for the prediction of organic or organometallic selectivity are increasingly powerful,[43–47] even for small datasets.[48,49]

General (or global) predictive models – of any kind, be it quantitative-structure–property relationships (QSPR),[50–52] machine learning (ML),[47,53,54] artificial intelligence (AI) or deep learning (DL) – that incorporate more than one metal and catalyst class and use organically growing standardized experimental databases as training data could have a tremendous appeal in olefin polymerization. Such models could indicate when and under which conditions, for example, a metal change might make sense for a given ligand set. Moreover, we envision that they would easily incorporate new datapoints (requiring only minimal retraining), even if significantly diverse from the previous ones. However, building such models requires a conscious effort, both from the experimental and the modeling side.

All data science approaches rely on machine readable information, *i.e.* a set of features/descriptors that encode relevant aspects of the catalysts (or other relevant reaction variables). The challenge from the data science perspective is that as the generality of a model increases, the number of common recurring features/descriptors,[55] which can quantify electronic and steric differences of the various catalysts, decreases. For steric effects this has been partially addressed through the introduction of 3D-steric descriptors, such as Cavallo's buried volume,[56,57] but it remains for electronic effects. For example, atomic charges on the ligand or chemical shifts for ligand atoms are common electronic descriptors, but evaluating Fig. 1, it is obvious that there is no transferability between different ligand sets. Furthermore, a question arises what species to choose to derive electronic descriptors from; the standard choice in the field is often the catalyst precursor, *i.e.* at a stage where absolute identity is still known, under the implicit assumption that electronic and steric effects transfer well from precursor to the (real) active species.

Herein, we systematically approach the question of electronic descriptors/features for statistical models in olefin polymerization catalysis by early transition metal complexes, specifically for group 3 (Sc and Y), and 4 (Ti, Zr, Hf). We analyzed whether the assumption that electronic descriptors derived at the precursor stage give similar trends to those derived for a putative active species is justified. Moreover, we examined the suitability of different electronic descriptors like charges and frontier molecular orbital energies for broader models. Finally, we show that analyzing the descriptor spacing allows insight into strengths and weaknesses of descriptors before any modeling is even attempted. In particular, we show that the data is often clustered – originating from one effect (metal or ligand change) dramatically outweighing the other – which can limit the application ranges of models derived thereof.

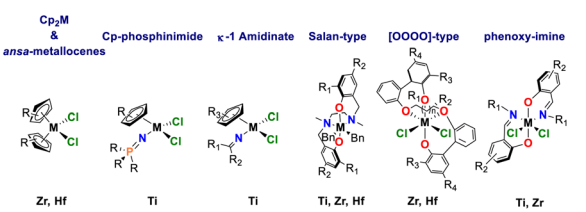

Fig. 1 Selected ligand classes and most important metals for industrially relevant group 4 metals emphasize the importance of pairing a ligand with the "right" metal.

Results and discussion

Initial considerations

Even considering only group 4 metal species, the diversity in catalyst design with respect to ligand backbones is tremendous. The development of molecular catalysts for olefin polymerization started with metallocene catalysts which have pseudotetrahedral, "sandwich", geometry, where two cyclopentadienyl π-ligands occupy two coordination positions.[58] Typical precursor species for industrially relevant group 4 species are dichloride, dimethyl, or dibenzyl species (Cp_2MX_2), giving an overall 16-electron (16e) species (6e each from the Cp fragments, 2e each from the X-type ligands).[14] Activation, that is cationization, produces the polymerization active cationic 14e species (6e each from the Cp fragments, 2e from the metal–carbon bond).[59] In *ansa*-metallocenes, a bridge between the two π-ligands, for example –$SiMe_2$– ensures rigidity.[60] In half-metallocenes one of the π-ligands is replaced by an X-type ligand, as in the case of the so-called constrained geometry catalysts (CGC),[61] originally introduced for group 3 metals by Bercaw[12] and for group 4 metals by Okuda,[62] where one Cp is

replaced by an amide, and at least formally, the electron count is lowered as the amide can donate a maximum of 4 electrons, resulting in a formally 12e species. In post-metallocenes,[63] both π-ligands are replaced and for example McConville's bis (amide) catalysts[64] are formally 10e species. It should be noted here that even Basset's Zr-hydride species immobilized on silica, which are formally 8e species, are polymerization active.[65] Reviews regarding ligand design have been published for example by Mecking[66] (half- and post-metallocene systems) and Cavallo and Resconi[18] (metallocenes).

Certain ligand/metal pairings appear to be privileged. For example, the development of metallocenes and *ansa*-metallocenes focused largely on Zr[18] and, occasionally, on the corresponding Hf species.[1,67,68] The poor success of Ti (the active metal in Ziegler–Natta heterogeneous catalysts) can be traced to the ease with which reduction, and therefore deactivation, to a Ti(III) species occurs in the presence of the typical aluminum alkyl species used as scavengers.[27] Then again, for half-metallocenes, Ti can be the only relevant group 4 metal, *e.g.* for phosphinimide catalysts[69] and CGC catalysts,[66] for which only Ti-based systems show high molar mass capability. Reversing the trend again, post-metallocene [OOOO]-type catalysts are generally Hf- and Zr-based,[31–33] while Ti species are almost inactive.

Group 3 *vs.* group 4. Despite sharing the same propagation mechanism and ligand sets, group 3 and group 4 based catalysts differ in some important aspects, which become even more pronounced when it comes to descriptor evaluation. First, precursor LMX/LMX_2 species are neutral in both cases (L = dianionic ligand or two monoanionic ligands) but group 4 species need to be cationized by Lewis acids (and alkylated in the case of dihalide species) and the active species are ion pairs; instead, LMX species in the case of group 3 are already polymerization active, if X is a carbon based substituent like methyl, benzyl or $-CH_2TMS$.

Ligand sets. For the main body of the analysis, we chose three different ligand sets: a prototypical *ansa*-metallocene (*rac*-Me_2Si(2-Me-indenyl)$_2$,[8] Fig. 2a), a Stephan-type half-metallocene with a Cp phosphinimide with ^{t}Bu substituents on phosphorus (Fig. 2b),[69] and a prototypical minimal Kol-type Salan[70] complex with Me-substituents on the amine linker (Fig. 2c). For these ligands sets, we investigated the complexes with Sc and Y (group 3) and Ti, Zr, and Hf (group 4).

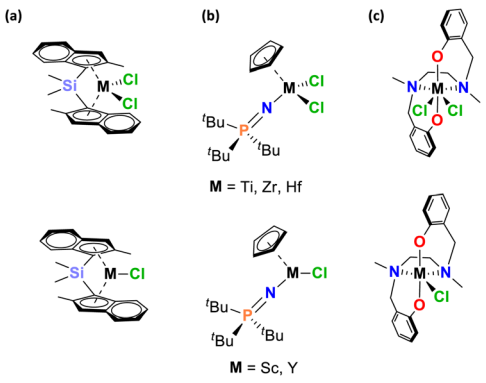

Fig. 2 Ligand sets. (a) *rac*-Me_2Si(2-Me-indenyl)$_2$ *ansa*-metallocene, (b) Cp phosphinimide with N=P(^{t}Bu)$_3$, (c) Kol-type Salan.

Model structures. Traditional data science approaches in olefin polymerization often utilize a single pre-catalyst structure, from which steric and electronic descriptors/features of the catalyst are determined. For group 4 catalysts, Cavallo has popularized the use of neutral dichloride precursor structures, $LMCl_2$, under the explicit assumption that the precursor geometry is representative of the TS.[38] Although being ideal from the perspective of determining steric descriptors – chloride substituents do not allow for hydrogen–heteroatom bonding that could distort the structure – this model structure might be less suited for the evaluation of electronic descriptors, if only because it lacks the metal–carbon bonds necessary for polymerization activity. The most natural strategy to make up for this weak point would be to switch to the neutral dimethyl precursor species.

Translating this approach to group 3 LMX complexes, which only possess one M–X bond, is not trivial and could be problematic from the perspective of evaluating sterics reliably (the MX_2 fragment in group 4 complexes can be used to unequivocally orient the metal complexes in a coordinate system, while a M–X fragment can only define one axis). More importantly, for group 3 (and the lanthanides) the polymerization active species LMMe is uncharged while the isoelectronic group 4 species $LMMe^+$ is an ion pair. How this affects electronic descriptor determination for models aiming at describing d^0 polymerization active systems in general is not clear *a priori*.

Monomer capture[71–74] is an important part of the mechanism and a more general model for the active species might incorporate an olefin. Finally, the TS for the insertion of an olefin into a metal–carbon bond is the only true prerequisite for polymerization activity.

For the purpose of the present study, we chose to consider three levels of structures. Our "best" level uses the fully-optimized insertion TS (TS model, TSM). The next level uses the structure of the olefin complex, *i.e.* the "activated complex just before insertion" (activated complex, ACM). To avoid complications arising from different possible orientations of a realistic growing chain model[73] and/or of a 1-alkene, we chose ethene and a metal-methyl bond for these two model structures. Another step down leads to the precursor model (PM), which uses the $LMCl_x$ or $LMMe_x$ (x = 1 for group 3, x = 2 for group 4) as model structures. The whole set of considered model structures is depicted in Fig. 3. To our best knowledge, there are so far no reports on ACM or TSM models; our results hopefully clarify how sensitive these models are. Importantly, as we will discuss in more detail later, TSM is the only truly universal one when aiming at general models.

Computational details. All structures were optimized at the TPSSh/cc-pVDZ-(PP) level of theory.[75] The absence of imaginary frequencies for minima and one imaginary frequency for TS corresponding to the reaction coordinate was checked to confirm the nature of minima and TS. Single point energy cal-

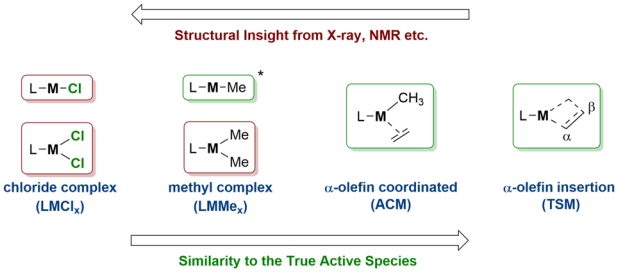

Fig. 3 The four different model structures chosen for electronic descriptor determination in this work. *Often passivated with a Lewis base, *e.g.* PMe_3, in precursor species.

culations were then conducted either within IBOview[76,77] or with Gaussian 16[78] at the MN15-L/TZ level of theory[79] (TZ: cc-pVTZ-(PP) for Gaussain 16, and def2-TZVP for IBOview) and the electronic descriptors were collected from the output files utilizing python scripts detailed in the SI. For the ACM and TSM structures, the isolated cation approximation was employed.[80] For additional details see SI.[95–110]

Targeted electronic descriptors. Considering electronic effects, the choice of "charge" based descriptors, frontier molecular orbital energies, and bonding situation seems a logical one. For charges, there are many different options, hinting at the problem of reliably quantifying them, and it is not given that all possible choices are equally consistent in the context of model building. Specifically, we targeted descriptors that are transferable between the different model complexes and have been used for modeling in the past.[50] We considered metal and ligand charge, q_M and q_L, determined by different methods, *i.e.*, intrinsic bond orbital theory (IBO), natural population analysis (NPA),[81] quantum theory of atoms in molecules (QTAIM) as implemented in IBOview,[82] Mulliken,[83] Charge Model 5 (CM5)[84] and Hirshfeld.[85] Furthermore, Wiberg bond indices[86] and bond composition as determined by IBO and natural bond orbital theory (NBO) as well as HOMO, LUMO and HOMO–LUMO gap energies were considered. For the ACM and TSM model species, some additional descriptors were collected, that cannot be collected from PM models. Overall, 105 different electronic descriptors were collected.

Reproducibility of TS coordination geometries

An overview of key structural parameters of the optimized structures of group 4 complexes is given in Table S1. As mentioned earlier, a key assumption in many statistical modeling approaches is that the geometry of the model structure reflects that of the TS. The TSM consistently produces comparable geometries for both group 3 and group 4 catalysts alike, regardless of the ligand class, reflecting true insertion TS with the olefin in-plane (or nearly in-plane) of the M–Me bond.

For the other model structures, ligand or metal specific structural changes were encountered, as detailed in the following. For group 4 metal dichloride and metal dimethyl complexes, all optimized structures result in *pseudo-tetrahedral* (*ansa*-metallocene and Cp phosphinimide) or *fac–fac* coordination geometries (for octahedral Salan complexes), meaning that the two X-type ligands are in mutual *cis* geometry, in positions where the chain and the vacant site (or olefin) would be.

Initially, we targeted an even simpler model for the Cp phosphinimide system, with a less sterically hindering –N=PMe_3 ligand. The problem of inadvertent geometry changes inducing electronic effects is even more pronounced in this case and therefore worth discussing here. For this ligand, in only a single instance, *i.e.* for the Ti dichloride complex, a change in geometry occurs, not for any of the other model complexes. In the crystal structures reported by Stephan in the literature[69,87] and in all other complexes modeled here, Ti–N–P angles close to 180° are found, implying a sp-hybridized N atom that can donate π-electron density to the metal. In the Ti dichloride complex, the nitrogen of the phosphinimide ligand is much more pyramidalized (Ti–N–P 138.14°), indicative of a substantially sp^2-hybridized N and reduced electron donation to the metal (Fig. 4). Electronic descriptors collected from this structure indicate a substantially more electron poor Ti center compared to the other Ti models, which are arguably much more representative of the active species.

Although switching to the –N=P^{t}Bu$_3$ ligand solved the problem in the precursor species, we encountered a change in coordination geometry in the ACM for this ligand set. While for *ansa*-metallocenes and Salan complexes ethene coordinates in-plane of the M–Me bond, just as for the TSM, for the Sc, Ti, Zr and Hf Cp phosphinimide complexes, ethene is coordinated out-of-plane (Fig. 5).

For descriptor collection pertaining to the two carbons of the coordinated ethene in these ACM species, we defined the carbon, with the longer distance to the methyl substituent as α and the other one as β. This reflects the pathway of easiest rotation from out-of-plane to in-plane coordination upon approaching the TS but is less unambiguous than in the case of in-plane ethene coordination.

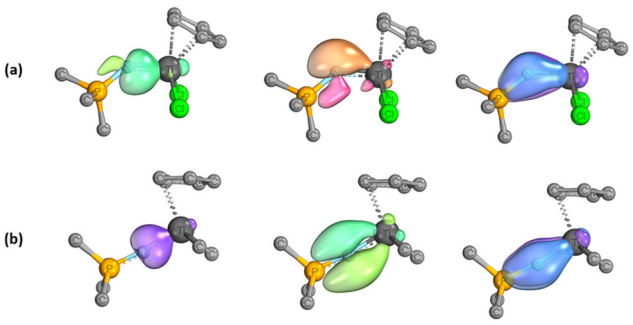

Fig. 4 Simpler Cp phosphinimide Ti complexes $LTiCl_2$ (a) and $LTiMe_2$ (b), illustrating a sudden switch in the ligand structure (rehybridization) that affects electronic descriptors. L = –N=PMe_3. In the $LTiCl_2$ (a), the ligand is a 4e donor, in $LTiMe_2$ (b) a 6e donor (according to the anionic ligand formalism). IBOs visualized with substantial Ti–N contribution evidencing the change of the one of the IBOs (σ to π, center top and bottom) from $LTiCl_2$ to $LTiMe_2$. Pictures generated with IBOview.[76] Hydrogens omitted for clarity.

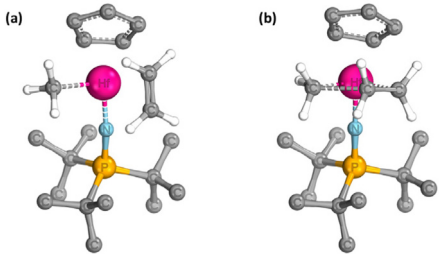

Fig. 5 CpHf(N═P^{t}Bu$_3$) ACM model for the active species (a) and TSM model for the TS (b). In (a) ethene is coordinated out-of-plane in (b) in-plane with respect to the Hf–Me bond. Ligand hydrogens removed for clarity. Pictures generated with IBOview.[76]

For group 3 metals, all the precursor complexes LMCl and LMMe are non-representative for the coordination environment found in the TS (tetrahedral with in-plane ethene coordination for the metallocene and the phosphinimide, *fac–fac* with in-plane ethene coordination for the Salan). Also, for the ACM, structural deviations from the geometry in the TS are observed for group 3 metals, as shown in Table S1 and Fig. 6.

In summary, non-TS model structures have a high chance of at least occasionally producing geometries that are not representative of TS; how this influences modeling remains unknown for the moment, but we believe it is important to be aware of this possibility.

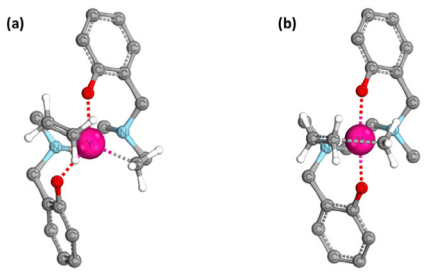

Fig. 6 Yttrium Salan model. (a) ACM in *fac*/*mer* geometry, (b) insertion TS in *fac*/*fac* geometry. Ligand hydrogens removed for clarity. Pictures generated with IBOview.[76]

Comparison of data trends in different model species

We begin our analysis of the electronic descriptor data collected comparing the data trends in the four different model structures (LMCl$_x$, LMMe$_x$, ACM, TSM) for the 15 selected catalysts (3 ligand sets × 5 metals). In general, electronic descriptors evaluated on different model structures can carry the same relative information. For example, the partial charge on the metal q_M determined with the IBO method on the ACM and TSM species reveals nearly identical data trends (R^2 = 0.9995, Fig. 7, left). The overall charges change slightly, *e.g.* for the *ansa*-zirconocene we find $q_{Zr,ACM}$ = 3.05 *vs.* $q_{Zr,TSM}$ = 3.15, indicating that different descriptors derived from different model species should never be mixed, but the data trend, *i.e.* the relative differences between the metals and ligands for the two respective model systems are nearly identical. This indicates that for the IBO derived metal charges q_M, the choice of model is inconsequential: the different model structures (LMCl$_x$, LMMe$_x$, ACM, and TSM) show identical data trends for this descriptor. Therefore, this electronic descriptor could be determined from the simple neutral precursor model commonly used for steric analysis and the assumption that electronic effects transfer from this model to the real active species is true. This finding holds despite the earlier found coordination geometry changes between the different models which are particularly pronounced for group 3 metals.

However, this descriptor insensitivity is not omnipresent, *e.g.* for the partial charge on the ligand, determined with the CM5 method on the LMCl$_x$ model *vs.* the TSM model (Fig. 7, right R^2 = 0.87). Here, we can see that not only do the ligand/metal trends not transfer from one model to another, but the two different metal groups experience different relative changes. In this case, the electronic effects determined from the precursor do not transfer to TSM and for this descriptor the question arises, which of the models to use to evaluate it.

Partial charges. Full details for the sets of 14 charge descriptors can be found in the SI (Table S2). For group 4 metals alone, the choice of model to evaluate metal partial charge trends is inconsequential for several methods: IBO, QTAIM, CM5 and Hirshfeld. On the contrary, Mulliken charges and NPA charges are influenced by the choice of model compound.

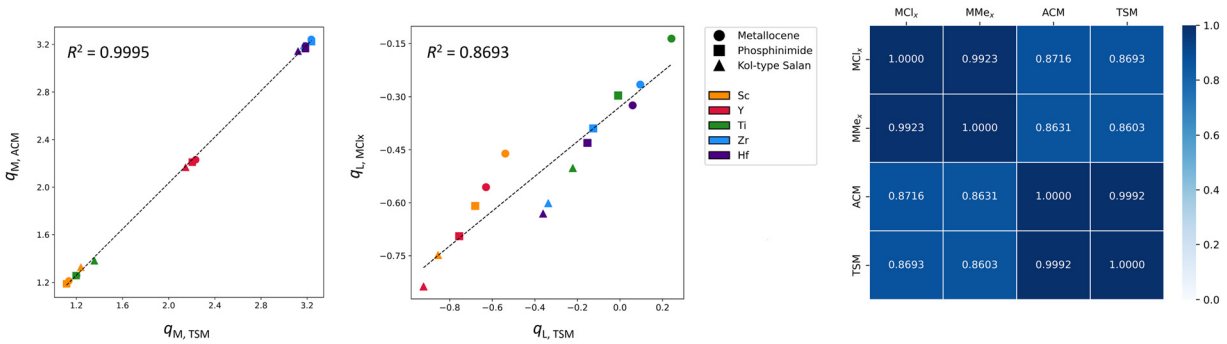

Fig. 7 Comparison of the transferability of electronic descriptor trends between model species. Left: plot of partial charge (IBO) of the metal in the ACM model ($q_{M,ACM}$) *vs.* TSM model ($q_{M,TSM}$). Middle: plot of partial charge (CM5) of the ligand in the LMCl$_x$ model ($q_{L,Cl}$) *vs.* TSM model ($q_{L,TSM}$). Right: heatmap for the correlation (R^2) between different model structures for the partial charge (CM5) of the ligand.

For the charge of the ligand, the situation is somewhat more complex, only for IBO and CM5, model independence is observed, for all other methods, the choice of the model influences the trend that is obtained. For the whole dataset of group 3 and group 4 models, model independence is only observed for metal partial charges determined with IBO and QTAIM, while for all other methods and for the ligand charges, the model influences the observed trend. Almost always, the ACM and TSM models give very similar if not identical trends in the partial charges, and the precursor models give a different trend. This implies that electronic effects in the precursor do not transfer to the models of the active species and the transition state and *vice versa*.

From the perspective of data science, the relevant lesson from this analysis is that for certain descriptors, analyzing one model structure carries all the information that can be gathered, and additional species will not offer further relevant information. However, in most cases surveyed here, the precursor species and the more realistic model structures for the active species yield different datasets for each electronic descriptor. Overall, we included 12 variations of charges in the present work. Half of them are basically consistent between the different model structures (so that one could use any of the model species and obtain basically equivalent data) while the other 6 behave more erratically.

It is seductive to interpret the "consistent set" as somehow more "true" than the others, but we feel this is not the right venue to start a discussion of charge definitions.

Wiberg bond indices and bond composition. Metal–Ligand bonds, which differ between different ligand sets, are not transferable and thus not suitable as descriptors for general models. Therefore, we focused on bonds that characterize the recurring central fragment, *i.e.*, the M–Cl bonds in the $LMCl_x$ models, the M–C_{Me} bonds in the $LMMe_x$ model, and the M–C_{Me}, bonds in the MMe(ethene) ACM and the TSM models. Interestingly, for group 4 model complexes, our analysis indicates that electronic effects, in the sense of metal and ligand changes, on Wiberg bond indices and bond composition as determined by the IBO method are insensitive to the bond identity (Tables S3 and S4). This means that trends in these two descriptors determined for M–Cl and M–Me bonds in the $LMCl_x$ and the $LMMe_x$ precursor models and the cationic ACM model carry the same information. However, differences to the values determined for the TSM model, where bonds are forming and breaking, are large. Therefore, electronic effects affecting WBI can be extrapolated from one bond to the other but not from minima to TS. Instead, this is not the case for WBI determined by the NBO method; in this case, each model compound generates a different dataset with different trends. NBO bond composition analysis yields more detailed information than the bond composition analysis provided by IBO and would therefore be very interesting from a descriptor perspective. However, natural bond orbitals are closely associated with elementary Lewis structure diagrams,[88] which is why NBOs provide a direct link to valency and bonding concepts. If the dominant Lewis structure is ionic, then bond analysis is impossible and not performed. Throughout the dataset this happens irregularly. Manual intervention would be required; while not impossible, this is hard to automatize.

Frontier molecular orbitals. Frontier molecular orbital energies are a common descriptor utilized in various statistical models. Frontier molecular orbital (FMO) theory, pioneered by Nobel prize winner Fukui,[89] provides the basis. Simplified, FMO posits that reactivity can be predicted analyzing the symmetry and energetic similarity of the frontier molecular orbitals: highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), or single occupied molecular orbital (SOMO) of the reactants. Of course, orbital energies and locality (spatial extent and orientation) can change, if the molecule is distorted. A simple example can be found in the higher reactivity of cyclic alkynes over linear alkynes in click chemistry,[90] as the former are already "electronically primed" while the latter need to be significantly distorted to attain the correct FMO symmetry.[91] FMOs are strongly connected with the geometry and reactivity of a molecule and in click chemistry; HOMO–LUMO gap energies are predictive.

Whether or not HOMO–LUMO gap energies are reasonable descriptors for olefin polymerization is not granted. It needs to be recalled here that only reactive bonds are directly transferable between different ligand sets; MOs connected to metal–ligand binding have no meaning whatsoever in a different system. MOs energies and/or energy gaps are meaningful only when the considered orbitals involve the recurring fragment of a model species. However, unambiguous identification of these orbitals is no trivial task, as shown in the following.

For the neutral precursor species, the FMOs carry no similarity as they are connected to the metal–ligand interactions, which is also reflected in the yellow to orange color of some of those complexes. Similarly, the HOMOs in the ACM (Fig. 8) model are not connected to the reactive bonds, *i.e.*, the metal–carbon bond and the olefin π-orbitals, which in fact are HOMO–*X* orbitals (*X* depends on the ligand). Instead, the LUMO can often *but not always* be connected to an empty metal centered d-orbital and the olefin π*-orbital, but a degenerate LUMO+*X* orbital might be observed. This degeneracy cannot always be unequivocally deciphered and narrowed down to the main contribution. Correlation of the energies of the MOs connected to the insertion chemistry is only possible for the ACM complex, as in the TSM complex the relevant orbitals are already mixing.

Regarding IBOs, the IBO method localizes the MOs and the LUMO is generally the π_{ethene}*-orbital (Fig. 9) while the HOMO is still localized on the ligand. We can therefore conclude that HOMO and LUMO energies and the HOMO–LUMO gap are meaningless in the present systems and utilizing them for statistical modeling is at best questionable. Concerning the reactive MOs, the main problem for data collection arises from the fact that ligand centered π-orbitals (in the present case either in aromatic ligand parts or P=N bonds) are lying higher in energy. The problem will be exacerbated by aromatic or heteroatom containing substituents on the ligand. The relative position of the reactive orbitals is a function of the *exact ligand and*

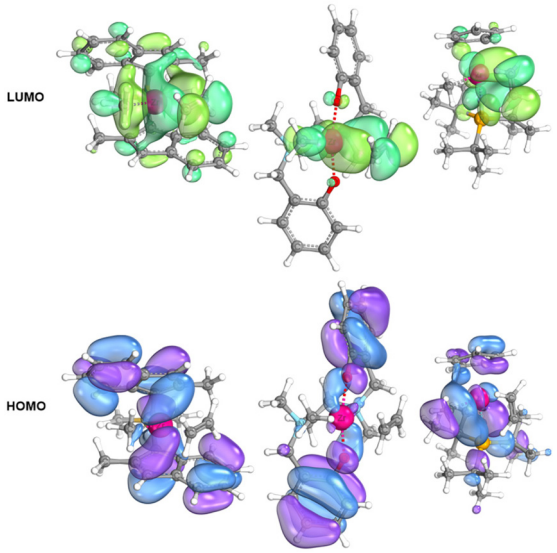

Fig. 8 Visualization of the HOMO and LUMO calculated with MN15-L/cc-pVDZ level of theory for the three LZrMe(ethene)$^+$ ACM models (left: *ansa*-zirconocene, middle: Zr-ONNO, right: CpZrNP^{t}Bu$_3$). HOMOs are generally ligand centered and delocalized, while LUMOs show contributions from an empty d-orbital and the olefin π*-orbital and are degenerate (only one MO shown). Pictures generated with IBOview.[76]

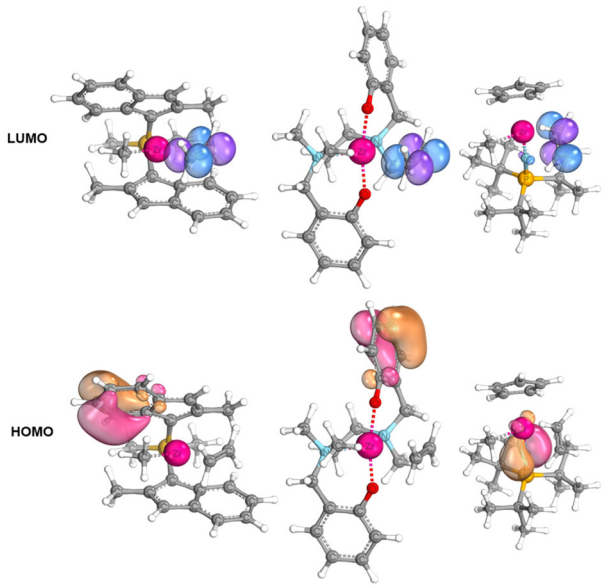

Fig. 9 Visualization of the FMO IBOs calculated with MN15-L/def2-SV(P) level of theory for the three LZrMe(ethene)$^+$ ACM models (left: *ansa*-zirconocene, middle: Zr-ONNO, right: CpZrNP^{t}Bu$_3$). HOMOs are generally ligand centered and delocalized, while LUMOs are centered on the olefin π*-orbital and non-degenerate. Pictures generated with IBOview.[76]

substituent pattern and cannot be predicted *ad-hoc*. Reliably identifying these orbitals is therefore heavily complicated by the fact that their relative ranking is affected by ligand and substituent patterns and the metal.

In light of these observations, we recommend not using automatic HOMO–LUMO data collection; at the same time even human data collection, utilizing orbital visualization to identify the orbitals connected to reactive bonds, is challenging, as MO degeneracy can complicate the picture.

Analysis of data clustering in electronic descriptors

Before beginning this analysis, it is instructive to recall some basic concepts. As outlined in the introduction, olefin polymerization active species are known for almost all the transition metals. Even focusing only on group 4, there is no privileged metal or ligand set, in fact, "perfect pairings" seem to exist, and similar polymers can be made with vastly different catalysts. From a purely theoretical perspective, a descriptor that can describe metal variations, ligand set changes, and substituent effects simultaneously must present a continuum of values rather than small changes in one case *vs.* large case in another (clustering).

Fig. 10, bottom depicts the data spacing for one descriptor (partial charge on the metal determined on the TSM by the IBO analysis). Clearly visible, all the datapoints for the Hafnium and Zirconium species nearly coincide. Ligand changes yield much smaller effects, and three clusters are present (Sc + Ti, Y and Zr + Hf). A different data spacing is found for the partial charge on the ligand as determined on the TSM by the NPA analysis and depicted in Fig. 10, top. Significant overlap occurs between group 3 and group 4 metals, the spacing between the least electron-rich group 3 metal system (Scandium phosphinimide) and the most elec-

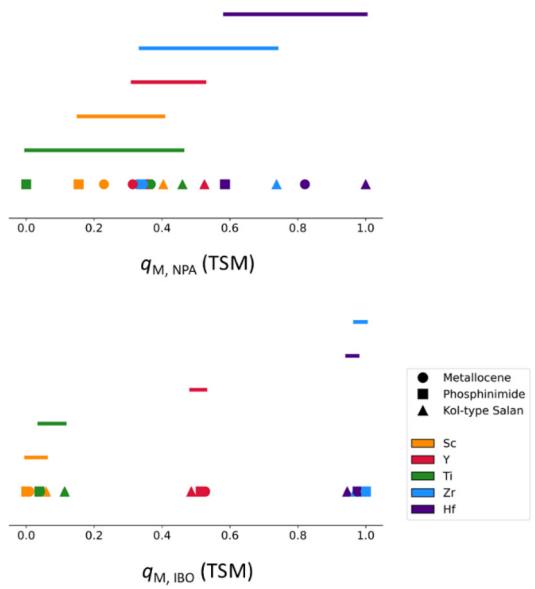

Fig. 10 Comparison of continuously *vs.* clustered data spacing in different descriptors. Top: plot of partial charge (NPA) of the metal in the TSM ($q_{M,TSM}$). Bottom: plot of partial charge (IBO) of the metal in the TSM model ($q_{M,TSM}$). Highlighted regions indicate the descriptor space covered by each metal. Only in the top case electronic effects originating from metal and from ligand changes are of comparable magnitude. Descriptor values were normalized between 0 and 1.

tron-poor group 4 system (Hafnium Salan) is minimal. Moreover, the metal and ligand effect in each group have similar magnitudes, *i.e.* the systems overlap. In fact, the descriptor delivers comparable values for the Titanium Salan, the Hafnium phosphinimide and the Zirconium *ansa*-metallocene.

The two discussed descriptor cases are thus examples for descriptors with a clustered and an evenly spaced dataset, respectively. In the case of the clustered data, it is highly unlikely that a model could describe both different catalyst classes and different metals with a single electronic descriptor, at least in the specific case of Hafnium *vs.* Zirconium. In the second case, a working model should easily be extendable from one catalyst class to another and from one metal to another, as well as from one transition metal group to another.

Tables S5 and S6 list the different descriptors that have been analyzed here and their tendency to deliver either clustered or spaced datapoints. The visualization plots are provided in the SI, Fig. S4–S28. The analysis was performed visually. Mathematical approaches to perform cluster analysis are known,[92] but in this case, we are interested in datapoint distance and overlap between different metals and ligands and the group 3 and group 4 metals. Traditional (mathematical) cluster analysis could deliver a wrong picture, as clustering of datapoints from different species (metal/ligand combinations) differs from segregated clusters of one type.

Finally, the inconsistency of metal and ligand effects going from the precursor species to the active species mentioned earlier strongly suggests that electronic descriptors have to be evaluated at the correct point (model) in catalysis and do not necessarily transfer from one species to another.

Overview of model strengths and weaknesses

Based on the previous discussion, care needs to be taken in assembling a computational database of descriptors regardless of the underlying model species. The following discussion briefly describes an overview of the strengths and weaknesses of the different model structures, for a detailed analysis see Table 1.

$LMCl_x$ models. Optimized geometries might not be representative of the active species. Occasional deviation from the geometry of the putative active species was observed even in the small dataset discussed here, which might not always be obvious. For C_2-symmetric group 4 metal-based catalysts, it has been shown that this model can also be used for the determination of steric descriptors. Several electronic descriptors return questionable results when derived from this model precursor structure, indicating that electronic effects in the precursor species might differ from the active species. Only the IBO and QTAIM charge on the metal can be recommended as an electronic descriptor for models targeting group 3 and 4 metals alike.

$LMMe_x$ models. Optimized geometries might not be representative of the active species. For example, H–F bonding can influence geometries which is relevant for Fluorine containing catalysts.[93,94] Several electronic descriptors return questionable results when derived from this model precursor structure, indicating that electronic effects in the precursor species might differ from the active species. Only the IBO and QTAIM charge on the metal can be recommended as an electronic descriptor for models targeting group 3 and 4 metals alike.

$LMMe(ethene)^+$ ACM models. Optimized geometries of the MMe(ethene) unit are sensitive to the ligand framework. Ethene coordination can be perpendicular or in-plane to the equatorial plane of the catalyst and α-agostic interactions can differ in orientation relative to the ethene (toward or away from). It remains unclear how and if this affects descriptor values. In the coarse screening performed in this work, no dramatic effects were noticed from a change in ethene orientation. Only some electronic descriptors derived from this model structure are clearly not suitable for data collection, such as NPA metal charges and FMO energies.

Table 1 Overview of the strength and weaknesses of different models and derived descriptors depending on whether group 4 or group 3 and 4 models are targeted under the assumption that models target electronic trends of the active species

Model	Geometry	Reliable descriptors[a]	Possibly reliable descriptors comparison of 1 neutral, 1 cationic model recommended	Questionable descriptors[b]
Group 4 only				
$LMCl_x$	Problematic	IBO, QTAIM (M), CM5, HF (M)	QTAIM (L), HF (L), NPA (L), IBO-WBI and BC	Mulliken, NPA (M), NBO-WBI, FMO
$LMMe_x$	Problematic	IBO, QTAIM (M), CM5, HF (M)	QTAIM (L), HF (L), NPA (L), IBO-WBI and BC	Mulliken, NPA (M), NBO-WBI, FMO
ACM	Problematic	IBO, QTAIM, CM5, HF, Mulliken, NPA (L)	IBO-WBI and BC	NPA (M), NBO-WBI, FMO
TSM	OK	IBO, QTAIM, CM5, HF, Mulliken, NPA (L)	IBO-WBI and BC	NPA (M), NBO-WBI, FMO
Groups 3 and 4				
$LMCl_x$	Problematic	IBO (M), QTAIM (M)	IBO (L), QTAIM (L), CM5, HF, IBO-BC	Mulliken, NPA, IBO-WBI, NBO-WBI, FMO
$LMMe_x$	Problematic	IBO (M), QTAIM (M)	IBO (L), QTAIM (L), CM5, HF, IBO-BC	Mulliken, NPA, IBO-WBI, NBO-WBI, FMO
ACM	Problematic	IBO, QTAIM, CM5, Mulliken, HF, NPA	IBO-BC	—
TSM	OK	IBO, QTAIM, CM5, HF, Mulliken, NPA	IBO-BC	—

[a] Identical trends in electronic descriptors compared to the TSM model. [b] Significant model dependence of the trends in electronic descriptors. FMO = frontier molecular orbitals, WBI = Wiberg Bond Index, BC = bond composition, M = metal, L = ligand; when no details are given for IBO, QTAIM, CM5, HF, Mulliken or NPA charges, both M and L are included.

LMMe(ethene)$^+$ TSM models. As stated earlier, in order for a transition metal complex to show polymerization activity, this species *must* exist, which can become important when structures are solely screened *in silico* to identify novel catalysts. Often, trends in electronic descriptors derived from these species coincide with the ones determined from the ACM species. Bond orders are highly sensitive to the early or late nature of the TS and possibly better determined in the ACM model.

Model species closely connected to chemical reactivity. The ACM and TSM models have one additional significant advantage over the precursor models. Being more closely connected to the polymerization reactivity, they allow the collection of more electronic descriptors connected to the incoming alkene and relevant to the insertion mechanism in general (*e.g.* WBI of forming and breaking bonds, charges on the alkene, *etc.*). The spacing analysis for selected descriptors, namely TSM charges on metal and ligand (fragments transferable between different model structures), on methyl and ethene (fragments specific of active species) and on the two methylene groups of ethene, as well as the ACM Wiberg bond indices for the M–Me (ethene) fragment, are summarized in Table 2.

Model complexes should be chosen by what is to be modeled. In the case of selectivity of the active species like stereoselectivity, regioselectivity, comonomer affinity or molar mass capability, electronic descriptors should be determined from a closely connected model. However, for activity, activation might also play an important role. In this case, determining descriptors from a precursor species and an active species model might be advisable.

Data spacing in an expanded dataset

The previous analysis focused on a rather limited ligand set for practical reasons. Curious about how the situation would evolve when more ligands are considered, we also conducted a limited analysis for an expanded ligand set for the TSM model derived electronic descriptors, including minimal models for additional classes such as Exxon-type Salan (Fig. 11d), Dow C4-bridged bis(phenoxy)ether (OOOO, Fig. 11e), CGC (Fig. 11f), and McConville bis(amide) (Fig. 11g). This expanded set now covers 5 metals and 7 ligands (35 complexes). Fig. 12, show data spacing plots for two

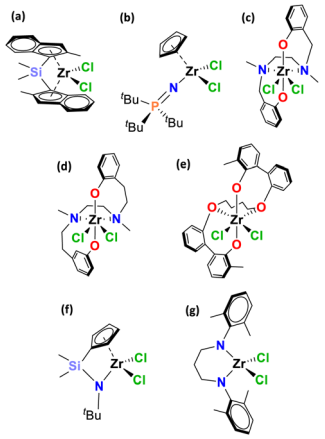

Fig. 11 Expanded ligand sets. (a) *rac*-Me_2Si(2-Me-indenyl)$_2$ *ansa*-metallocene, (b) Cp-phosphinimide with N=P(^{t}Bu)$_3$, (c) Kol-type Salan, (d) Exxon-type Salan, (e) C4-bridged OOOO, (f) CGC, (g) Bis Amide.

Table 2 Detailed performance and clustering analysis for selected descriptors collected from the ACM and TSM models

Method	Descriptor	Total spacing	% ligand spacing	% metal spacing	Group overlap	Series overlap	Dominant influence	Clustering
Partial charges (calculated on TSM)								
IBO	$q_M \vert q_L$	2.122|1.426	7|15	100|100	Yes	No	Metal	**1**: Sc, Ti; **2**: Y; **3**: Zr, Hf **1**: Sc; **2**: Ti; **3**: Y, Zr, Hf
	$q_{Me} \vert q_{Ethene}$	0.352|0.423	17|15	100|100	Yes	No	Metal	**1**: Sc, Ti; **2**: Y, Zr, Hf **1**: Sc; **2**: Ti; **3**: Y, Zr, Hf
	$q_\alpha \vert q_\beta$	0.355|0.087	16|20	100|87	Yes	No|Yes	Metal|Even	**1**: Sc; **2**: Ti; **3**: Y, Zr, Hf|No
QTAIM	$q_M \vert q_L$	0.821|1.070	35|29	83|85	Yes	Yes	Even	No
	$q_{Me} \vert q_{Ethene}$	0.160|0.282	35|22	100|100	Yes	Yes	Metal	No|**1**: Ti; **2**: Sc, Y, Zr, Hf
	$q_\alpha \vert q_\beta$	0.191|0.129	19|35	94|80	Yes	Yes	Metal|Even	**1**: Ti; **2**: Sc, Y, Zr, Hf|No
NPA	$q_M \vert q_L$	0.859|1.046	46|37	60|87	Yes	Yes	Even	No
	$q_{Me} \vert q_{Ethene}$	0.192|0.262	57|14	87|94	Yes|No	Yes	Even|Metal	No|**1**: Sc, Y; **2**: Ti, Zr, Hf
	$q_\alpha \vert q_\beta$	0.197|0.095	20|22	96|85	Yes|No	Yes	Metal	No|**1**: Sc, Y; **2**: Ti, Zr, Hf
Hirshfeld	$q_M \vert q_L$	0.242|0.979	60|25	62|87	Yes|No	Yes	Even|Metal	No|**1**: Sc, Y; **2**: Ti, Zr, Hf
	$q_{Me} \vert q_{Ethene}$	0.140|0.242	52|23	100|79	No	Yes	Metal	**1**: Sc, Y; **2**: Ti, Zr, Hf
	$q_\alpha \vert q_\beta$	0.149|0.101	28|27	84|85	No	Yes	Metal	No|**1**: Sc, Y; **2**: Ti, Zr, Hf
CM5	$q_M \vert q_L$	0.532|1.164	65|40	47|75	Yes|No	Yes	Even	No
	$q_{Me} \vert q_{Ethene}$	0.141|0.270	51|24	100|83	No	Yes	Even|Metal	**1**: Sc, Y; **2**: Ti, Zr, Hf
	$q_\alpha \vert q_\beta$	0.169|0.101	28|27	89|84	No	Yes	Metal	No|**1**: Sc, Y; **2**: Ti, Zr, Hf
Wiberg bond indices and bond compositions (calculated on ACM)								
IBO	WBI_{M-Me}	0.604	19	100	Yes	No	Metal	**1**: Sc; **2**: Ti; **3**: Y, Zr, Hf
	$WBI_{M-\alpha}$	0.201	50	100	Yes	Yes	Metal	**1**: Ti; **2**: Sc, Y, Zr, Hf
	$WBI_{\alpha-\beta}$	0.239	46	100	Yes	Yes	Even	No
	BC_{M-Me}	0.525	15	100			Metal	**1**: Sc, Ti; **2**: Y, Zr, Hf
NBO	WBI_{M-Me}	0.390	23	100	No	Yes	Metal	No
	$WBI_{M-\alpha}$	0.202	49	100	No	Yes	Metal	No
	$WBI_{\alpha-\beta}$	0.228	46	100	No	Yes	Metal	No
	$WBI_{M-\beta}$	0.143	98	99	Yes	Yes	Metal	No

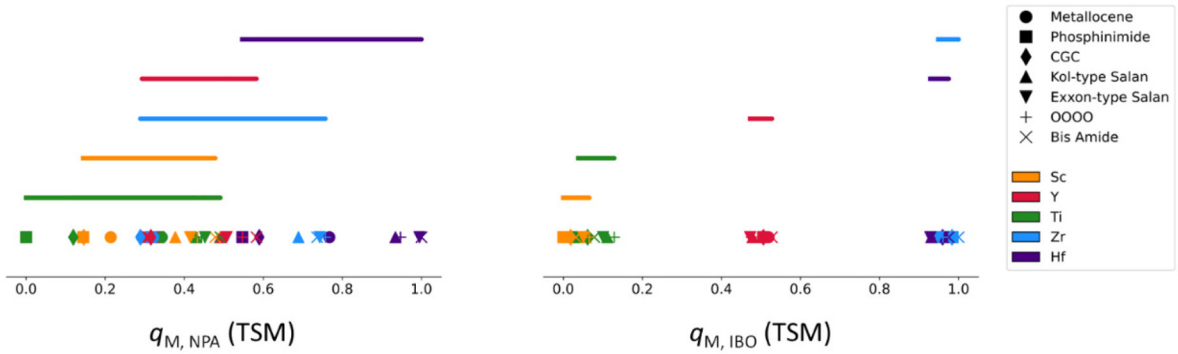

Fig. 12 Example for continuous data spacing *vs.* reinforced clustering in descriptors in the expanded dataset. Left: plot of partial charge (NPA) of the metal in the TSM model ($q_{M,TSM}$). Right: plot of partial charge (IBO) of the metal in the TSM model ($q_{M,TSM}$). Highlighted regions indicate the descriptor space covered by each metal. In the left case, clear overlaps between different metal/ligand combinations indicate continuous modulation of electronics through metal and ligand variation. The right case is characterized by minimal influence of ligand variation and dramatic changes through metal variation. Descriptor values were normalized between 0 and 1.

selected descriptors. For the partial charge (IBO) of the metal in the TSM model (Fig. 12, right) we find a reinforcement of the metal differences, indicating that this descriptor does not pick up ligand differences very well. In contrast, the enhanced coverage of the chemical space through expansion of the dataset leads to a continuous data spacing for the partial charge (NPA) of the metal in the TSM model (Fig. 12, left).

Non-continuous data spacing in descriptors does not mean that descriptors are useless, in fact, data science often aims to go above the fray, ignoring interpretability. In the realm of MLR, one descriptor could pick up ligand differences, while another could describe metal differences. However, if one aims at chemically intuitive models, which use a limited set of descriptors allowing high interpretability, then mixing-and-matching and increasing the number of descriptors, especially electronic ones, becomes problematic. For such models, descriptor space analysis allows to judge core capabilities, even without a model, similar to traditional hardware tools, that are chosen for a specific purpose.

Conclusions & outlook

Machine learning wants to accurately predict catalytic selectivity by using a set of less-than-optimal data, identifying patterns in large sets of data that are incomprehensible to humans.[47] However, is chemistry always too complex to be reduced to simple or at least simplified models? From a practical perspective, especially for a reaction as sensitive to small changes in the reaction conditions as olefin polymerization, the question arises: what happens when a model fails in predictions?

Reinforced learning, the answer of machine learning, runs the risk of leading to negative one directional feedback loops. Bi-directional feedback loops, where the model occasionally informs experiment require testable hypotheses and we believe that this is where simple models could shine. Occasionally, the model might be right and would be confirmed if we repeat the experiment under more controlled conditions.

We have shown that some of the base assumptions usually employed in modeling of olefin polymerization catalysts are likely flawed, including the assumption of transferability of electronic effects measured on a precatalyst to the active species. Metal charges are often ill-suited for global models and FMOs are not connected to the polymerization reactivity. Models closely related to the true active species in catalysis are more promising. Only TS models are likely truly transferable between catalyst classes and metals.

In the end, the very fact that olefin polymerization is a general feature of the transition metals, but electronic effects are still ill-understood could also result from problems in identifying electronic descriptors that are directly connected to the reactivity. Here, we have approached this topic from a purely data centric perspective, laying out strength, weaknesses, and pitfalls of different approaches. Notably, metal charges in precursor complexes often show clustered data spacing already within group 4 and especially when looking at the group 3 and 4 metals combined, yet they have been routinely used to describe substituent effects within small datasets in the field.

Direct computational predictions, for example by DFT, face significant hurdles, *e.g.* required accuracy and complexity of the system, solvent interactions, anions, and the sheer number of TS needed to predict selectivity, accounting for olefin and chain orientations. Statistical models are better suited for this purpose. Every model requires expensive experimental training data. The strength of general models is that past experimental data can inform new experiments, even if they come from a different catalyst class or metal. Moreover, the broader the training set becomes for a general model—covering metal, ligand backbone, and substituent effects—the more one will almost inevitably be forced to decipher the true mechanistic underpinnings.

Author contributions

Conceptualization: C. E.; formal analysis: G. G., A. V., P. O., L. M.; data curation: G. G.; funding acquisition: V. B., investi-

gation: C. E., G. G., A. V., P. O., L. M., P. H. M. B., R. C.; supervision: C. E., V. B.; visualization: G. G., P. O., L. M., writing original draft: C. E., G. G.; writing review and editing: all authors.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the supplementary information (SI). Supplementary information: XYZ files of the model structures, additional computational details, descriptor evaluation and collection, full tables of descriptor correlation, detailed clustering analysis, full tables of extracted descriptors, python scripts for data collection and file conversion. See DOI: https://doi.org/10.1039/d5dt02170g.

Acknowledgements

The authors thank Nic Friederichs (SABIC Europe) for fruitful discussions.

References

1 A. Vittoria, G. P. Goryunov, V. V. Izmer, D. S. Kononovich, O. V. Samsonov, F. Zaccaria, G. Urciuoli, P. H. M. Budzelaar, V. Busico, A. Z. Voskoboynikov, D. V. Uborsky, C. Ehm and R. Cipullo, *Polymers*, 2021, **13**, 2621.
2 K. Heiland and W. Kaminsky, *Makromol. Chem.*, 1992, **193**, 601–610.
3 D. V. Uborsky, D. Y. Mladentsev, B. A. Guzeev, I. S. Borisov, A. Vittoria, C. Ehm, R. Cipullo, C. Hendriksen, N. Friederichs, V. Busico and A. Z. Voskoboynikov, *Dalton Trans.*, 2020, **49**, 3015–3025.
4 C. Ehm, A. Vittoria, G. P. Goryunov, V. V. Izmer, D. S. Kononovich, O. V. Samsonov, P. H. M. Budzelaar, A. Z. Voskoboynikov, V. Busico, D. V. Uborsky and R. Cipullo, *Dalton Trans.*, 2020, **49**, 10162–10172.
5 F. Zaccaria, R. Cipullo, A. Correa, P. H. M. Budzelaar, V. Busico and C. Ehm, *Processes*, 2019, **7**, 384.
6 C. Ehm, P. H. M. Budzelaar and V. Busico, *Eur. J. Inorg. Chem.*, 2017, **2017**, 3343–3349.
7 F. Zaccaria, C. Ehm, P. H. M. Budzelaar and V. Busico, *ACS Catal.*, 2017, **7**, 1512–1519.
8 W. Spaleck, M. Antberg, J. Rohrmann, A. Winter, B. Bachmann, P. Kiprof, J. Behm and W. A. Herrmann, *Angew. Chem., Int. Ed. Engl.*, 1992, **31**, 1347–1350.
9 G. Talarico and P. H. M. Budzelaar, *Organometallics*, 2008, **27**, 4098–4107.
10 P. C. Möhring and N. J. Coville, *Coord. Chem. Rev.*, 2006, **250**, 18–35.
11 J. Klosin, W. J. Kruper, P. N. Nickias, G. R. Roof, P. De Waele and K. A. Abboud, *Organometallics*, 2001, **20**, 2663–2665.
12 P. J. Shapiro, E. Bunel, W. P. Schaefer and J. E. Bercaw, *Organometallics*, 1990, **9**, 867–869.
13 G. van Doremaele, M. van Duin, M. Valla and A. Berthoud, *J. Polym. Sci., Part A:Polym. Chem.*, 2017, **55**, 2877–2891.
14 G. J. P. Britovsek, V. C. Gibson and D. F. Wass, *Angew. Chem., Int. Ed.*, 1999, **38**, 428–447.
15 E. T. Kiesewetter, S. Randoll, M. Radlauer and R. M. Waymouth, *J. Am. Chem. Soc.*, 2010, **132**, 5566–5567.
16 R. Cipullo, V. Busico, N. Fraldi, R. Pellecchia and G. Talarico, *Macromolecules*, 2009, **42**, 3869–3872.
17 V. Busico and R. Cipullo, *Prog. Polym. Sci.*, 2001, **26**, 443–533.
18 L. Resconi, L. Cavallo, A. Fait and F. Piemontesi, *Chem. Rev.*, 2000, **100**, 1253–1346.
19 G. W. Coates, *Chem. Rev.*, 2000, **100**, 1223–1252.
20 V. Busico and R. Cipullo, *J. Am. Chem. Soc.*, 1994, **116**, 9329–9330.
21 V. Busico and R. Cipullo, *J. Organomet. Chem.*, 1995, **497**, 113–118.
22 V. Busico, L. Caporaso, R. Cipullo, L. Landriani, G. Angelini, A. Margonelli and A. L. Segre, *J. Am. Chem. Soc.*, 1996, **118**, 2105–2106.
23 C. Ehm, A. Vittoria, P. G. Goryunov, V. V. Izmer, S. D. Kononovich, V. O. Samsonov, R. Di Girolamo, H. M. P. Budzelaar, Z. A. Voskoboynikov, V. Busico, V. D. Uborsky and R. Cipullo, *Polymers*, 2020, **12**, 1005–1005.
24 V. Busico, R. Cipullo, R. Pellecchia, G. Talarico and A. Razavi, *Macromolecules*, 2009, **42**, 1789–1791.
25 P. Cossee, *Tetrahedron Lett.*, 1960, **1**, 17–21.
26 E. J. Arlman and P. Cossee, *J. Catal.*, 1964, **3**, 99–104.
27 F. Adams, *Macromol. Rapid Commun.*, 2024, **45**, 2400122.
28 M. R. Machat, C. Jandl and B. Rieger, *Organometallics*, 2017, **36**, 1408–1418.
29 F. Guérin, J. C. Stewart, C. Beddie and D. W. Stephan, *Organometallics*, 2000, **19**, 2994–3000.
30 S. Segal, I. Goldberg and M. Kol, *Organometallics*, 2005, **24**, 200–202.
31 A. Cohen, J. Kopilov, M. Lamberti, V. Venditto and M. Kol, *Macromolecules*, 2010, **43**, 1689–1691.
32 E. N. T. Cuthbert, V. Busico, D. E. Herbert and P. H. M. Budzelaar, *Eur. J. Inorg. Chem.*, 2019, **2019**, 3396–3410.
33 E. N. T. Cuthbert, A. Vittoria, R. Cipullo, V. Busico and P. H. M. Budzelaar, *Eur. J. Inorg. Chem.*, 2020, **2020**, 541–550.
34 A. Vittoria, P. S. Kulyabin, G. Antinucci, A. N. Iashin, D. V. Uborsky, E. N. T. Cuthbert, P. H. M. Budzelaar, A. Z. Voskoboynikov, R. Cipullo, C. Ehm and V. Busico, *ACS Catal.*, 2023, **13**, 13151–13155.

35 H. Makio, H. Terao, A. Iwashita and T. Fujita, *Chem. Rev.*, 2011, **111**, 2363–2449.
36 C. Ehm, A. Vittoria, G. P. Goryunov, P. S. Kulyabin, P. H. M. Budzelaar, A. Z. Voskoboynikov, V. Busico, D. V. Uborsky and R. Cipullo, *Macromolecules*, 2018, **51**, 8073–8083.
37 C. Ehm, A. Vittoria, G. P. Goryunov, V. V. Izmer, D. S. Kononovich, P. S. Kulyabin, R. Di Girolamo, P. H. M. Budzelaar, A. Z. Voskoboynikov, V. Busico, D. V. Uborsky and R. Cipullo, *Macromolecules*, 2020, **53**, 9325–9336.
38 A. Poater and L. Cavallo, *Dalton Trans.*, 2009, 8885–8890.
39 B. Maity, Z. Cao, J. Kumawat, V. Gupta and L. Cavallo, *ACS Catal.*, 2021, **11**, 4061–4070.
40 C. Ehm, A. Mingione, A. Vittoria, F. Zaccaria, R. Cipullo and V. Busico, *Ind. Eng. Chem. Res.*, 2020, **59**, 13940–13947.
41 C. Ehm, R. Cipullo, P. H. M. Budzelaar and V. Busico, *Dalton Trans.*, 2016, **45**, 6847–6855.
42 F. Zaccaria, C. Ehm, P. H. M. Budzelaar, V. Busico and R. Cipullo, *Organometallics*, 2018, **37**, 2872–2879.
43 S. H. Newman-Stonebraker, S. R. Smith, J. E. Borowski, E. Peters, T. Gensch, H. C. Johnson, M. S. Sigman and A. G. Doyle, *Science*, 2021, **374**, 301–308.
44 M. H. Samha, L. J. Karas, D. B. Vogt, E. C. Odogwu, J. Elward, J. M. Crawford, J. E. Steves and M. S. Sigman, *Sci. Adv.*, 2024, **10**, eadn3478.
45 N. P. Romer, D. S. Min, J. Y. Wang, R. C. Walroth, K. A. Mack, L. E. Sirois, F. Gosselin, D. Zell, A. G. Doyle and M. S. Sigman, *ACS Catal.*, 2024, **14**, 4699–4708.
46 J. Yano, K. J. Gaffney, J. Gregoire, L. Hung, A. Ourmazd, J. Schrier, J. A. Sethian and F. M. Toma, *Nat. Rev. Chem.*, 2022, **6**, 357–370.
47 A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow and S. E. Denmark, *Science*, 2019, **363**, eaau5631.
48 B. C. Haas, D. Kalyani and M. S. Sigman, *Sci. Adv.*, 2025, **11**, eadt3013.
49 D. Dalmau, M. S. Sigman and J. V. Alegre-Requena, *Chem. Sci.*, 2025, **16**, 8555–8560.
50 M. Karelson, V. S. Lobanov and A. R. Katritzky, *Chem. Rev.*, 1996, **96**, 1027–1044.
51 B. Sanija and P. G. R. Achary, *Int. J. Quant. Struct.-Prop. Relat.*, 2018, **3**, 36–48.
52 P. Willett, *Wiley Interdiscip. Rev.:Comput. Mol. Sci.*, 2011, **1**, 46–56.
53 J. B. O. Mitchell, *Wiley Interdiscip. Rev.:Comput. Mol. Sci.*, 2014, **4**, 468–481.
54 Z. Tu, T. Stuyver and C. W. Coley, *Chem. Sci.*, 2023, **14**, 226–244.
55 D. J. Durand and N. Fey, *Chem. Rev.*, 2019, **119**, 6561–6594.
56 L. Falivene, Z. Cao, A. Petta, L. Serra, A. Poater, R. Oliva, V. Scarano and L. Cavallo, *Nat. Chem.*, 2019, **11**, 872–879.
57 G. Antinucci, B. Dereli, A. Vittoria, P. H. M. Budzelaar, R. Cipullo, G. P. Goryunov, P. S. Kulyabin, D. V. Uborsky, L. Cavallo, C. Ehm, A. Z. Voskoboynikov and V. Busico, *ACS Catal.*, 2022, **12**, 6934–6945.
58 D. S. Breslow and N. R. Newburg, *J. Am. Chem. Soc.*, 1957, **79**, 5072–5073.
59 M. Bochmann, *Organometallics*, 2010, **29**, 4711–4740.
60 W. Kaminsky, K. Külper, H. H. Brintzinger and F. R. W. P. Wild, *Angew. Chem., Int. Ed. Engl.*, 1985, **24**, 507–508.
61 J. C. Stevens, in *Stud. Surf. Sci. Catal.*, ed. J. W. Hightower, W. N. Delgass, E. Iglesia, and A. T. Bell, Elsevier, 1996, vol. 101, pp. 11–20.
62 J. Okuda, *Chem. Ber.*, 1990, **123**, 1649–1651.
63 P. B. Konstantin, *Russ. Chem. Rev.*, 2007, **76**, 253.
64 J. D. Scollard, D. H. McConville, N. C. Payne and J. J. Vittal, *Macromolecules*, 1996, **29**, 5241–5243.
65 C. Lecuyer, F. Quignard, A. Choplin, D. Olivier and J.-M. Basset, *Angew. Chem., Int. Ed. Engl.*, 1991, **30**, 1660–1661.
66 M. C. Baier, M. A. Zuideveld and S. Mecking, *Angew. Chem., Int. Ed.*, 2014, **53**, 9722–9744.
67 A. Schöbel, E. Herdtweck, M. Parkinson and B. Rieger, *Chem. – Eur. J.*, 2012, **18**, 4174–4178.
68 M. R. Machat, D. Lanzinger, A. Pöthig and B. Rieger, *Organometallics*, 2017, **36**, 399–408.
69 D. W. Stephan, J. C. Stewart, F. Guérin, R. E. V. H. Spence, W. Xu and D. G. Harrison, *Organometallics*, 1999, **18**, 1116–1118.
70 E. Y. Tshuva, I. Goldberg and M. Kol, *J. Am. Chem. Soc.*, 2000, **122**, 10706–10707.
71 Z. Flisak and T. Ziegler, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 15338–15342.
72 P. Margl, L. Q. Deng and T. Ziegler, *J. Am. Chem. Soc.*, 1998, **120**, 5517–5525.
73 P. Margl, L. Deng and T. Ziegler, *Organometallics*, 1998, **17**, 933–946.
74 F. Zaccaria, R. Cipullo, P. H. M. Budzelaar, V. Busico and C. Ehm, *J. Polym. Sci., Part A:Polym. Chem.*, 2017, **55**, 2807–2814.
75 J. M. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.
76 G. Knizia, *J. Chem. Theory Comput.*, 2013, **9**, 4834–4843.
77 G. Knizia and J. E. M. N. Klein, *Angew. Chem., Int. Ed.*, 2015, **54**, 5518–5522.
78 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar,

J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision A.03*, Gaussian, Inc., Wallingford CT, 2016.
79 H. S. Yu, X. He and D. G. Truhlar, *J. Chem. Theory Comput.*, 2016, **12**, 1280–1293.
80 A. Laine, B. B. Coussens, J. T. Hirvi, A. Berthoud, N. Friederichs, J. R. Severn and M. Linnolahti, *Organometallics*, 2015, **34**, 2415–2421.
81 E. D. Glendening and F. Weinhold, *J. Comput. Chem.*, 1998, **19**, 593–609.
82 P. Salvador and E. Ramos-Cordoba, *J. Chem. Phys.*, 2013, **139**, 071103.
83 R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833–1840.
84 A. V. Marenich, S. V. Jerome, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2012, **8**, 527–541.
85 F. L. Hirshfeld, *Theor. Chim. Acta*, 1977, **44**, 129–138.
86 K. B. Wiberg, *Tetrahedron*, 1968, **24**, 1083–1096.
87 D. W. Stephan, F. Guérin, R. E. V. H. Spence, L. Koch, X. Gao, S. J. Brown, J. W. Swabey, Q. Wang, W. Xu, P. Zoricak and D. G. Harrison, *Organometallics*, 1999, **18**, 2046–2048.
88 F. Weinhold and C. R. Landis, *Valence and Bonding - A Natural Bond Orbital Donor-Acceptor Perspective*, Cambridge University Press, Cambridge, UK, 2005.
89 K. Fukui, T. Yonezawa and H. Shingu, *J. Chem. Phys.*, 1952, **20**, 722–725.
90 H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2001, **40**, 2004–2021.
91 Y.-H. Shen, A. M. Esper, I. Ghiviriga, K. A. Abboud, K. S. Schanze, C. Ehm and A. S. Veige, *Dalton Trans.*, 2021, **50**, 12681–12691.
92 J. A. Hartigan and M. A. Wong, *J. R. Stat. Soc., C: Appl. Stat.*, 1979, **28**, 100–108.
93 A. Iwashita, M. C. W. Chan, H. Makio and T. Fujita, *Catal. Sci. Technol.*, 2014, **4**, 599–610.
94 S. Dehnen, L. L. Schafer, T. Lectka and A. Togni, *Inorg. Chem.*, 2021, **60**, 17419–17425.
95 D. V. Uborsky, M. I. Sharikov, G. P. Goryunov, K. M. Li, A. Dall'Anese, C. Zuccaccia, A. Vittoria, T. Iovine, G. Galasso, C. Ehm, A. Macchioni, V. Busico, A. Z. Voskoboynikov and R. Cipullo, *Inorg. Chem. Front.*, 2023, **10**, 6401–6406.
96 F. A. Momany, *J. Phys. Chem.*, 1978, **82**, 592–601.
97 J. P. Foster and F. Weinhold, *J. Am. Chem. Soc.*, 1980, **102**, 7211–7218.
98 F. Weinhold and C. R. Landis, *Chem. Educ. Res. Pract.*, 2001, **2**, 91–104.
99 E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, P. Karafiloglou, C. R. Landis and F. Weinhold, *NBO 7.0*, Theoretical Chemistry Institute, University of Wisconsin, Madison, 2018.
100 G. Knizia, http://www.iboview.org/index.html.
101 P. Schwerdtfeger, *ChemPhysChem*, 2011, **12**, 3143–3155.
102 K. A. Peterson, D. Figgen, M. Dolg and H. Stoll, *J. Chem. Phys.*, 2007, **126**, 124101.
103 N. B. Balabanov and K. A. Peterson, *J. Chem. Phys.*, 2005, **123**, 064107.
104 D. Figgen, K. A. Peterson, M. Dolg and H. Stoll, *J. Chem. Phys.*, 2009, **130**, 164108.
105 B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson and T. L. Windus, *J. Chem. Inf. Model.*, 2019, **59**, 4814–4820.
106 M. Feyereisen, G. Fitzgerald and A. Komornicki, *Chem. Phys. Lett.*, 1993, **208**, 359–363.
107 O. Vahtras, J. Almlöf and M. W. Feyereisen, *Chem. Phys. Lett.*, 1993, **213**, 514–518.
108 E. J. Baerends, D. E. Ellis and P. Ros, *Chem. Phys.*, 1973, **2**, 41–51.
109 J. L. Whitten, *J. Chem. Phys.*, 1973, **58**, 4496.
110 The Berny algorithm was never fully published; see the Gaussian documentation for details.