Selection of Low-Dimensional 3-D Geometric Descriptors for Accurate Enantioselectivity Prediction

Giuseppe Antinucci,† Busra Dereli,‡ Antonio Vittoria, Peter H. M. Budzelaar, Roberta Cipullo,*, Georgy P. Goryunov, Pavel S. Kulyabin, Dmitry V. Uborsky,*, Luigi Cavallo,*, Christian Ehm,*, Alexander Z. Voskoboynikov, and Vincenzo Busico

ABSTRACT: This study focuses on building enantioselectivity models using only a few intuitively meaningful descriptors based on the "buried volume" idea. Appropriate dissection of the sphere is used to calculate the buried volume of quadrants and octants, which we name %$V_{BO}$ and %$V_{BO}$. Propene polymerization catalysis to isatotic polypropylene (iPP) and 1,1′-bis-2-naphthol (BINOL)-phosphoric acid-catalyzed thiol addition to N-acyl imines are used to illustrate the approach. For iPP, only a single steric descriptor derived from the comparison of hindrance in differently occupied octants (Δ%$V_{BO}$) is needed, and electronic effects are unimportant. Moreover, the model (mean absolute deviation, MAD, 0.12 kcal/mol) works for more than a single catalyst class, allowing in silico catalyst design. For thiol addition, the best performance is achieved by comparison of hindrance in octants, and one steric descriptor is needed (Δ%$V_{BO}$) in addition to an electronic descriptor, the natural population analysis (NPA) charge on the P atom. In both cases, key ingredients are (a) the use of properly chosen "scanning regions" (e.g., octants or quadrants) and (b) the availability of highly accurate experimental data sets. The low dimensionality of descriptor space and their obvious intuitive meaning naturally provide guidelines for further catalyst optimization.

KEYWORDS: enantioselective catalysis, molecular catalysts, descriptor design, buried volume, olefin polymerization, iPP

INTRODUCTION

Asymmetric catalysis, where selective access to one of two possible mirror-image products (enantiomers) is desired, is a central theme in synthetic organic chemistry¹ and polymer chemistry.²,³ Catalyst optimization for asymmetric reactions relies heavily on tuning of steric factors. Despite the simplicity of the relevant interactions, design is even nowadays still dominated by empirical optimization.

The aim of computer-aided catalyst design⁴–⁶ is to speed up the optimization process and deliver optimized catalytic performance within an expedited timeframe. Direct prediction of catalyst selectivity by quantum mechanical methods is complicated by the high level of accuracy needed: fractions of a kcal/mol. This is an accuracy that even the most sophisticated learning methods have the capacity to identify patterns in large sets of data that are incomprehensible to experimental practitioners and to generate models based on experimental data that can then be used to predict the performance of untested candidates.⁴,⁷

However, such approaches need rather large databases (in the order of thousands or more data points) for training and validation. Modern high-throughput experimentation (HTE) approaches²¹–²⁹ allow, in favorable cases, fast construction of such databases. Unfortunately, organometallic chemistry is often not ideally suited to HTE—especially when air- and moisture-sensitive catalysts are targeted and/or if the catalyst synthesis is complex—and databases of a few dozen catalysts are already considered “huge” in the field.¹⁶,¹⁸–²⁰,³⁰

Received: February 24, 2022
Revised: May 11, 2022
Descriptor design also is not necessarily straightforward. Focusing on steric effects, over the years, different approaches have been developed to measure steric bulk: interference values, A-values, Charton, and Taft parameters, Tolman’s cone angle, AREA (θ), Sterimol parameters, comparative molecular field analysis (CoMFA), percent buried volume (%V\text{Bur}), and average steric occupancy (ASO). The very fact that so many approaches have been reported testifies to the notion that “simple descriptor” does not mean “simple in design”.

CoMFA, %V\text{Bur}, and ASO are voxel approaches: a scanning region is defined, and points within it are assigned a numerical value based on their being occupied by the catalyst (ASO takes catalyst flexibility into account by merging values over low-energy conformations). Selected combinations of voxels can then be used as black-box descriptors (ASO, CoMFA). This is the approach followed in the entirely statistical protocol developed by Denmark, in which the emphasis is on machine learning. Alternatively, voxel values can be summed over well-defined scanning regions of space (e.g., sphere, quadrants, octants), resulting in one or a few descriptors, which in favorable cases can be intuitively connected to the underlying reaction mechanism and thus emphasize human learning. One could argue that for the massive statistical approach, the choice of scanning regions is hardly relevant, but if the results are to be combined in just a few descriptors, the choice of regions becomes crucial. They should include all of the space that—if occupied—affects enantioselectivity, but not any space that would not make a difference. Asymmetry in space occupation between given regions can then be correlated to enantioselectivity. This approach was at the basis of the quadrant model developed by Farina and Ewen to provide a qualitative interpretation of enantioselectivity in propene polymerization.

For quantitative analysis within the %V\text{Bur} framework, a logical starting point is dissecting the whole coordination sphere, properly oriented, into quadrants (Figure 1). Differences in %V\text{Bur} between more and less hindered quadrants can then be used as a descriptor (indicated here by Δ%V\text{BO}). An alternative approach could instead focus only on the hemisphere where the reaction occurs, i.e., the “front-half”, with the “back-half” being considered irrelevant. Dissection of the properly oriented hemisphere would then lead to octants, and differences in %V\text{Bur} between more and less hindered octants can then be used as a descriptor (Δ%V\text{BO}). Considering a typical metal catalyst precursor, e.g., a metal dichloride (LMCl\text{2}; L = ligand), this implies determining %V\text{Bur} only in the hemisphere containing the M–Cl bonds, as indicated in Figure 1.

Using the octants approach (initially called quadrants of the hemisphere) and literature polymerization data, some of us showed that this enabled qualitative prediction of enantioselectivity (R\text{2} between experimental and fitted selectivities equal to 0.74) in propene polymerization promoted by a series of metalocene and non-metalocene catalysts. Of course, this low R\text{2} would prevent developing any model having predicting capability. It should be noted that we here use the term qualitative to indicate models that predict rough trends (mean absolute deviation, MAD, ≥ 1 kcal/mol) and quantitative for models good enough to use for catalyst optimization (MAD ≈ 0.2 kcal/mol).

Octant models (or quadrants of a half-sphere) have been also used recently by the Denmark group in the analysis of asymmetric cyclopropanation, hydrogenation, and thiol addition. Differently, some of us very recently showed that quadrants of the full sphere can also be used for quantitative prediction of isotactic polypropylene (iPP) enantioselectivity (R\text{2} between experimental and fitted data 0.92) of rigid C\text{2}-symmetric indenyl-based ansa-zirconocenes, even allowing moderate extrapolation. However, the step from qualitative to quantitative prediction required an accurate HTE database, a re-benchmarking of the sphere radius used for scanning, and the deletion of some substituents in the ligand backbone and on the outer edges of the catalytic pocket (believed to be irrelevant to enantioselectivity, yet still impacting calculated steric bulk in the scanning sphere). While the latter study allowed us to reach accurate correlation between experimental and fitted values, manipulating the structure of the catalyst to delete irrelevant parts is not a practical approach.

Given this background, the present work aims to test the performance of the quadrant and octant versions of the Δ%V\text{Bur} descriptor in two cases of enantioselective catalysis. The first is isoselectivity in propene polymerization to iPP, using a sizable data set obtained through modern HTE methods; we will show that the octant Δ%V\text{BO} descriptor works across (some) catalyst classes (metalocenes and salans). This will validate the octant model and will clarify that the qualitative performance achieved previously by some of us was essentially due to inconsistency in the experimental data set used for training, as it was assembled using data from different research groups.

Recognizing that iPP catalysis might be a special case, we also reanalyzed 1,1’-bis-2-naphthol (BINOL)-derived chiral phosphoric acid-catalyzed thiol addition to N-acyl imines and will show that a single-conformer method, together with the octant Δ%V\text{BO} descriptor, can be used successfully. For the
sake of completeness, in both cases, we will also discuss the performance of the quadrant $\Delta \% V_{BQ}$ descriptor, showing its potentials and limitations.

## RESULTS AND DISCUSSION

**Stereoselectivity in Olefin Polymerization with C$_2$-Symmetric Catalysts.** The origins of stereoselectivity induction in olefin polymerization (Cossee–Arlman mechanism, $^{48-51}$ Corradini model for enantioselectivity $^{52-54}$) are well known. Furthermore, it is well established that stereoselectivity is only depending on steric factors provided certain experimental conditions are met (see the SI for a short discussion). $^{3,55,56}$ Unsurprisingly, therefore, a long history of predictive modeling both via direct quantum mechanics (QM) prediction $^{57-61}$ and structure-performance-type statistical approaches $^{16,18,20,30,44,46,62}$ exists. However, olefin polymerization is also extremely sensitive to variations in the polymerization conditions, and only recently, HTE has enabled researchers to build catalyst performance databases of sufficient size to allow the generation of quantitative models, employing virtually identical polymerization conditions. $^{16-20,29}$

In the present case, we looked at the performance of 50 catalyst precursors in propene polymerization, belonging to four different catalyst families: 28 ansa-zirconocenes (A1–A28) $^{16,18}$ 10 bis(phenolate-amine) catalysts (salan-type, S1–S10) $^{66,67}$ 11 bis(phenolate-ether) catalysts ([OOOO]-type, O1–O11) $^{68,69}$ and one Mitsui-type catalyst (M1 $^{65a}$ Scheme 1). Different from earlier attempts, our HTE workflow allowed us to assemble a coherent database with results generated under rigorously identical conditions ($T_p=60 \, ^{\circ}C$, $P_{propene} = 6.6 \, \text{bar}$, toluene solvent) where contributions from chain epimerization $^{3,55,56}$ are negligible. Data for the ansa-zirconocenes was previously published by us and taken from refs 16 and 18. Data for the post-metallocenes was generated for the express purpose of this paper (see the Experimental Section and SI Table S1).

### Quadrants and Octants.

Figure 2A illustrates the performance of the plain quadrant model ($\Delta \% V_{BQ}$) on the C$_2$-symmetric ansa-metallocene set of medium to high ($\sigma < 0.9995$) stereoselectivity. To cover the influential 4,4$'$$'$-position substituents, a scanning sphere radius of 5.0 Å was used. $^{16}$ An $R^2$ of 0.73 is achieved, i.e., the model is only qualitatively correct and lacks quantitative predictive power. More concerning is that $R^2$ drops to 0.43 if the four catalysts with low enantioselectivity ($\Delta \Delta G^{\text{enantio}}_{\text{‡}} < 3.5 \, \text{kcal/mol}$) $^{a}$ are excluded.

The ansa-metallocene model can be applied to other catalyst classes, and we tested its performance for the salan catalysts S1–S10 (Table S2) and the Mitsui-type catalyst M1 (i.e., using the slope and intercept of the metallocene model to predict the performance of catalysts of the other two catalyst classes). A severe underestimation of enantioselectivity ($\Delta \Delta G_{\text{enantio}}^{\text{‡}} > 3.5 \, \text{kcal/mol}$) are excluded.

The metallocene model can be applied to other catalyst classes, and we tested its performance for the salan catalysts S1–S10 (Table S2) and the Mitsui-type catalyst M1 (i.e., using the slope and intercept of the metallocene model to predict the performance of catalysts of the other two catalyst classes). A severe underestimation of enantioselectivity can be noted (in terms of $\Delta \Delta G_{\text{enantio}}^{\text{‡}}$, MAD = 1.69 kcal/mol) for all but the single nonstereoselective salan catalyst (S6) in the set, while an
even more dramatic deviation is observed for M1 (4.5 kcal/mol, overestimation). These poor results are due to the quadrant model scanning steric bulk in regions that are irrelevant to catalyst performance. As mentioned above, the deletion of selected substituents before calculating Δ%V
BQ
 offers an avenue to circumvent this problem. Figure 2B illustrates the performance of our published quantitative model for stereoselectivity of the _ansa_-zirconocene set, showing an R^2 of 0.92, still using a scanning sphere of 5.0 Å. However, the selection of atoms for deletion is backbone-dependent and can only be unambiguously defined within a catalyst class: the model cannot be used as is for other catalyst classes such as salan or [OOOO] catalysts.

This result is unsurprising when looking at Figure 3. For _ansa_- (bis)indenyl zirconocenes, the distribution of steric bulk in the catalyst backbone situated in the back hemisphere (i.e., opposite to the active pocket) aligns nearly perfectly with the occupied and formally empty quadrants of the front hemisphere.

On the contrary, for salan catalysts, most of the steric bulk in the back hemisphere is situated in regions formally associated with the empty front quadrants. Therefore, in the evaluation of Δ%V
BQ
, some of the separation of steric bulk found in the front hemisphere is canceled by contributions found in the back hemisphere that are not actually relevant to enantioselectivity. Since this type of compensation will be different for different catalyst classes, a global model applicable to multiple classes and based on a single quadrant descriptor cannot be expected to work. An octant model fares much better. Figure 4 shows the correlation of predicted with experimentally observed enantioselectivity for the metallocene set. The R^2 on the metallocene subset is 0.85, i.e., a significant improvement is observed compared to the quadrant model with no deletions, using the same sphere size of 5.0 Å (variation of the scanning sphere did not improve R^2). Moreover, salan catalysts (and the lone Mitsui-type catalyst) now fit perfectly in the correlation derived from metallocenes. The regression equation changes only minimally. The R^2 for the combined model is 0.95. Leave-one-out cross-validation (LOOCV) analysis results in a Q^2 value of 0.93.

Using the slope and intercept of the metallocene model to predict the performance of the other two catalyst classes, the predicted performance of post-metallocenes S1-S10 and M1 shows a MAD from the experiment of only 0.20 kcal/mol.

Figure 2. Correlation of predicted and experimentally observed stereoselectivity (as ΔΔG
enantio
) using a quadrant model for _ansa_-zirconocenes A1-A28 and employing (A) original input with no deletions and (B) edited input where the SiR₂ bridge and the substituents in position 5, 6, and 7 are deleted. R
sphere = 5.0 Å.

Figure 3. Illustration of why a quadrant model cannot simultaneously describe different catalyst classes. Occupied quadrants gray, “empty” quadrants white. The bulk of the ligand framework in the back hemisphere of a typical _ansa_-zirconocene (A2, left) is concentrated in the same quadrants as the important 4,4’-Ph substituents (front hemisphere). For the generic o-tBu-substituted salan (post-metallocene), smaller _ortho_-substituent chosen for better visualization), the _ortho_-substituent defines the contents of the “occupied” quadrant, but the bulk of the ligand framework in the back hemisphere is concentrated in the formally “empty” quadrants. When quadrants are used to scan the catalyst, this cancels some of the separation of steric bulk in the evaluation of Δ%V
BQ
.

Figure 4. Correlation of predicted and experimentally observed enantioselectivity (as ΔΔG
enantio
) using an octant model for _ansa_-zirconocenes A1-A28 (blue data points), and salan catalysts S1-S10 (red data points), and Mitsui catalyst M1 (violet data point, unselective) using unaltered input (no deletions). Unselective catalysts (ΔΔG
enantio_exp = 0; M1 and S6) and catalysts for which σ > 0.9995 (S1 and S8) were not part of the training set, ΔΔG
enantio_predicted set to 0 for negative predictions. R
sphere = 5.0 Å.
(Table S3), which drops further to 0.10 kcal/mol when the highly enantioselective catalysts S1 and S8 are excluded (σ = 0.9998, large experimental error bars result from σ approaching 1). Interestingly, even the four Hf-based salans (S7–S10) included in the salan data set fit the model (derived from a different catalyst class and from a different central metal) perfectly. Ansa-hafnocene performance is dominated by chain epimerization at 60 °C.22 Consequently, it is impossible to include them in the model. However, post-metallocenes are less prone to the series of β-hydrogen eliminations that need to occur for chain epimerization.

As indicated by these results, a well-chosen single steric descriptor can quantitatively describe enantioselectivity of the active pocket across different catalyst classes (metallocenes and post-metallocenes of salan and Mitsui type). Interestingly, this indicates that while salans are more flexible in a general sense,65b they are essentially similarly “inflexible” compared to ansa-zirconocenes at the olefin insertion state (TS) stage. The ortho-carbons bearing bulky substituents in octahedral post-metallocenes are located in the back hemisphere, while the 4-position carbon in (bis)indenyl ansa-zirconocenes serving the same role is already located in the front hemisphere (to prove the point, a salan with an ortho-H substituent yields atactic PP,66 while Me2Si(2-Me-4-H-indenyl)2ZrCl2 yields iPP). Therefore, larger substituents at the ortho-carbon of salan catalysts are required to achieve high enantioselectivity than are needed for the 4-position carbon of (bis)indenyl ansa-zirconocenes (Figure 5).

Figure 5. (A) Optimized transition-state geometries for stereerror insertion in a basic metallocene (A2, left) and a generic α-tBu-substituted salan (right, smaller ortho-substituent chosen for better visualization). The arrows indicate non-bonded C–C distances where ligand bulk is closest to the methyl group of the inserting monomer: position 4 for metallocenes, the phenolate ortho position for salan. Blue arrow: shortest distance, red arrows: other short distances. (B) Topographic steric maps generated with SambVca 2.1.5

Extension to [OOOO]-Type Catalysts. While the ansa-zirconocene-derived regression equation can be used to predict the performance of salans and Mitsui-type catalysts, the extension to [OOOO]-type catalysts O1–O11 fails and large deviations (MAD 1.50 kcal/mol) from experimental enantioselectivities are observed (Table S4). However, the observed [OOOO] data appear to be systematically offset relative to the predicted values (Figure 6, blue datapoints, the blue trendline serves as a guide to the eye). For the [OOOO] subset O1–O11 on its own, R² is 0.92 and the LOOCV Q² is 0.89 (Figure 6, red trendline). The slopes of the models for the metallocene + salan data set and the [OOOO] data set, respectively, are nearly identical, and the systematic deviation is indicated by a different intercept.

Most likely, this bias can be traced to an increased (and unfavorable) flexibility of the backbone, even at the TS stage (Figure 7). Catalysts O1–O11 have rather flexible backbone linker C₂, C₃, or C₄ groups, as recently demonstrated by Cuthbert et al.68,69

In a broader sense, this suggests that modeling of the active pocket via a single three-dimensional (3D) geometric descriptor determined from a dichloride precursor and a global model can only be expected to work when relatively stiff metal complexes are targeted, i.e., when the precursor resembles the active site. For flexible catalysts, a truly global model must fail. However, the presence of backbone flexibility does not prevent building a catalyst class specific model from a precatalyst structure, as indicated by the nearly unchanged slope for the [OOOO] catalysts compared to the metallocene and salan set in Figure 6.

Stereoselectivity in Thiol Addition to N-Acyl Imines. While accurate description of enantioselectivity in a specific reaction class using the Δ%VBO descriptor is certainly encouraging, it is not obvious that this approach would work for other reactions: iPP might be a special case. Therefore, we also tested the Δ%VBO and Δ%VBO models for an example of organocatalysis, reanalyzing HTE data collected by Denmark for thiol addition to acyl imines (Scheme 2; catalysts P₁–P₁₉, S2–70 in ref 11). The Danish group tested both ASO and quadrant-based descriptors for this reaction.11,19,47 They also reported that electronic effects are significant and included a descriptor based on the electrostatic potential. Multivariate linear regression using ASO and electrostatic potential resulted in excellent correlation of the experimental enantioselectivities, with an R² of 0.95 and a MAD on ΔΔG°enantio, exp equal to 0.16 kcal/mol only.

Herein, we resorted to test the capability of Δ%VBO and Δ%VBO descriptors analogous to the approach employed in iPP modeling. Like Denmark, we find that electronic effects are
Figure 7. Visualization of the backbone flexibility of catalyst O1. Faint dichloride precursor structure in background, overlayed with the TS backbone geometry. Clearly visible is the change in the dihedral angle between the backbone Ph and o-Ph substituent (A + B, indicated by red arrows) leading to reduced steric pressure in the occupied quadrants and the movement of backbone Ph toward the active pocket leading to increased steric pressure in the formally empty quadrants (C, indicated by violet arrow).

Scheme 2. Precatalysts Employed for Predictive Modeling of Enantioselectivity in Thiol Addition to N-Acyl Imines
important, and we use the natural population analysis (NPA) charge on P ($q_P$) to describe this. We initially performed a radius scan to calculate $\Delta V_{BO}$ and $\Delta V_{BO}$ at varying sphere radius and settled on a radius of 3.0 Å, associated with the largest $R^2$ value.

A quadrant model comprising only one electronic parameter ($q_P$) and one steric descriptor ($\Delta V_{BO}$) performs moderately well, with $R^2 = 0.87$ and $Q^2 = 0.78$ (see SI Figure S1 for the regression plot), resulting in reasonable correlation and predictive power. This is a better performance compared to that observed in the case of propene polymerization discussed above. The quadrant model improves slightly with the inclusion of a second steric parameter: the occupancy of the most hindered quadrant, here denoted by $V_{BO}$. This gave a reasonable three-descriptor model, with an $R^2$ value of 0.92, though it resulted in no improvement in LOOCV accuracy, $Q^2 = 0.79$ (Figure 8A).

For the octants case, the two-descriptor model, using only $\Delta V_{BO}$ as a steric parameter along with one electronic descriptor, yields $R^2 = 0.94$ and $Q^2 = 0.90$, which is a clearly better predictive model than its quadrant counterpart (Figure 8B). This is consistent with the case of propene polymerization discussed above (see the SI for the regression plot). In this case, inclusion of $\% V_{BO}$ for the highly occupied octant, resulting in a three-descriptor model, produced no improvement, with $R^2 = 0.94$ and $Q^2 = 0.89$ comparable to the two-descriptor model (Figure S2).

Overall, the multivariate linear regression equations derived from quadrant and octant models, as displayed in regression plots in Figure 8, show good performance, with the two-descriptor octants model giving the best performance, close to the $R^2$ of 0.95 achieved by Denmark with the conformationally dependent ASO model. Furthermore, the two-descriptor octant model also has an appreciable predictive power for this class of organocatalysts, with a MAD value of 0.16 kcal/mol, again consistent with the value achieved by Denmark and co-workers. As for the quadrant models, the reasonable accuracy they achieve can be easily rationalized, considering that all of the precatalysts of Scheme 2 have the same BINOL linker, thus resulting in similar occupancy in the back hemisphere over the whole set of systems. This is at odds with the case of catalysts for iPP production discussed above, where the different linkers yield different occupancies in the back hemisphere, resulting in poor performance of the quadrant-based model.

To further confirm the necessity of using an electronic descriptor, already highlighted by Denmark and co-authors, a regression analysis was also carried out by excluding the $q_P$ descriptor, creating a steric-only octant model based on $\% V_{BO}$. The performance of this model drops to $R^2 = 0.68$ and $Q^2 = 0.59$, emphasizing the significance of electronic effects for this specific class. Finally, using a steric-only octant model based on both $\Delta V_{BO}$ and $\% V_{BO}$, yields $R^2 = 0.83$ and $Q^2 = 0.76$, still clearly worse than the model based on $\Delta V_{BO}$ and the NPA charge on P.

### CONCLUSIONS

The two examples discussed above demonstrate that it is possible to develop truly predictive models based on simple, intuitively meaningful descriptors. The $\% V_{BO}$ approach appears to be a good starting point, leading to the $\Delta V_{BO}$ and $\% V_{BO}$ descriptors, but care should be taken when selecting regions for descriptor evaluation (e.g., quadrants vs octants, choice of radius). The quality of the underlying experimental database is at least as important, as shown by the leap in the quality of iPP modeling between 2009, when the training database was based on different literature sources, and the present work, based on highly accurate and consistent methodologies.

The above work on enantioselectivity in iPP catalysis indicates that a quantitative model relying on a single octant-based steric descriptor, $\% V_{BO}$, can accurately correlate the experimental values, whereas the quadrant-based steric descriptor, $\Delta V_{BO}$, only offers mediocre performance. Moreover, the octant-based model can be based on, and is valid for, more than a single catalyst class, indicating the transferability of the model. Conformational averaging does not appear to be needed in this case. The use of a low-dimensional model for iPP leads to particularly simple guidelines for systematic catalyst improvement: fill the full quadrants/octants further and keep the empty regions empty.

---

**Figure 8.** Correlation of the predicted data with experimentally observed enantioselectivity (as $\Delta \Delta G_{\text{enantio, exp}}$) using quadrant (A) and octant (B) models for BINOL-derived phosphoric acids. $R_{\text{sphere}} = 3.0$ Å.
The work on enantioselectivity in thiol addition to N-acyl imines catalyzed by BINOL-based phosphoric acids indicates that it is possible to build useful models for thiol addition based on a quadrant steric descriptor, \( \Delta V_{Br} \), together with one electronic descriptor. Replacing the quadrant descriptor by the analogous octant-based descriptor \( \Delta V_{Bo} \) results in clearly superior performance, comparable to that achieved via comprehensive sampling of conformational space by Denmark.

Overall, this work has demonstrated that sectioning the original \( \% V_{Bur} \) descriptor into different regions of the first coordination sphere around the active center can lead to a variety of models that are highly accurate in fitting known data and can predict the experimental behavior of systems out of the training data set with good accuracy. Of course, the selection of the most appropriate model depends on the specific case in hand. An apparent feature of both discussed catalytic reactions is that the precursor sufficiently resembles the active site. On the other hand, if the precatalyst geometry is not representative of the actual catalyst at the transition state of the selectivity-determining step, these models, as any regression model, will probably fail. Finally, one cannot expect the selectivity-determining step, these models, as any not representative of the actual catalyst at the transition state of the active site. On the other hand, if the precatalyst geometry is a certain of a special case because so much is already known about geometries and energetics. Still, there are important and well-studied reactions out there that might conceivably benefit from this type of modeling. As demonstrated with the BINOL-based phosphoric acids, enantioselectivity predictions employing low-dimensional intuitive descriptors also work for other \( C_2 \)-symmetric systems. Conceivably, the approach should be easily amenable to predict the enantioselectivity of \( C_2 \)-symmetric catalysts outside of the chosen two examples, provided accurate experimental data is available. Extension to \( C_1 \)-symmetry needs to be tested. Catalyst chemoselectivity, regioselectivity, or molar mass capability are obvious further catalyst performance parameters of interest, and we are actively engaged in extending the intuitive descriptors approach, within the world of olefin polymerization and beyond.

## EXPERIMENTAL SECTION

### Catalyst Synthesis

The synthesis of the precatalysts is detailed in the Supporting information.

### Polymer Synthesis and Characterization

All polymerization experiments were performed in toluene in a Freeslate Parallel Pressure Reactor setup with 48 reaction cells (PPR48), fully contained in a triple MBraun glovebox under nitrogen. The cells (5.0 mL working volume), feature 800 rpm magnetically coupled stirring, and individual online reading/control of temperature, pressure, monomer uptake, and uptake rate. The setup and the operating protocol are described in full detail in refs 70 and 25 and the SI, and have been used successfully in various homogeneous and heterogeneous polymerization studies.\(^{16-18,69,71-73}\) Polymerization conditions were as follows: \( T = 60°C \); \( p(C,H) = 6.6 \) bar; scavenger = Al(iBu)\(_3\); \( \) activator \( \) [trityl]\[B(C,F)\(_3\)\] (TTB), \( \) TTB/\([Zr] \) = 0.9.\(^{74}\) The catalysts were not preactivated prior to injection into the PPR cells. All experiments were performed at least in duplicate. The monomer was fed on demand.

The polymers were characterized by (a) high-temperature gel permeation chromatography (GPC) with a Freeslate Rapid-GPC setup and (b) quantitative \(^{13}\)C NMR with a Bruker DRX 400 spectrometer equipped with a high-temperature cryoprobe (for 5 mm OD tubes) and a preheated robotic sample changer. All results are averages of duplicate polymerization experiments. More details can be found in SI Table S1.

### Computational Details

**Precatalyst Optimization: Olefin Polymerization.** A protocol was chosen that was previously tested extensively for olefin polymerization-related problems, including modeling.\(^{16,18}\) The structures of metal dichlorides were fully optimized using the Gaussian16 software package\(^1\) in combination with the OPTIMIZE routine of Baker\(^{75}\) and the B0pt software package.\(^2\) Following the protocol proposed in ref 79, all precatalysts were optimized at the TPSSTPS(0)/cc-pVDZ(-PP)\(^{11-13}\) level of theory, using a small core pseudopotential on Zr.\(^{94,95}\) Justifying the accuracy of the calculated geometries, the protocol has been successfully used, in combination with M06-2X\(^{46}\) single point energy corrections, to address several polymerization-related problems: absolute barrier heights for propagation,\(^{88,89}\) comonomer reactivity ratios,\(^{88,89}\) metal–carbon bond strengths under polymerization conditions,\(^{90-92}\) electronic and steric tuning of molar mass capability,\(^91\) and modeling.\(^{16,18}\) The density fitting approximation (resolution of identity, RI)\(^{94-97}\) and standard Gaussian16 quality settings were used. All structures represent true minima (as indicated by the absence of imaginary frequencies) or transition states (exactly one imaginary frequency, corresponding to the reaction coordinate).

**BINOL-PA Optimization: Thiol Addition.** The structures of all precursors were fully optimized using the Gaussian16 software package,\(^76\) at the PBE0/TZVP level of theory. Standard Gaussian16 quality and convergence settings were used. All structures represent true minima (as indicated by the absence of imaginary frequencies). Manual conformational search was performed to identify the lowest energy conformation of the systems.

**Descriptor Determination.**Descriptors were calculated based on the optimized geometries of metal dichloride precursors or BINOL-phosphoric acid catalysts, using the SambVca 2.1 standalone fortran code,\(^75\) which can be downloaded from the SambVca 2.1 webserver, as well as with an equivalent standalone python script that facilitated batch processing (AutoSMaps).\(^98\) Probe spheres were centered on the metal or phosphorus atoms, with the positive direction of the z-axis along the axis passing through the metal and the midpoint of the Cl atoms for iPP polymerization catalysts and with the negative direction of the z-axis along P atom and the midpoint of the C–C BINOL bridge for thiol addition catalysts, and the xz plane defined by one of the Cl atoms (iPP catalysts) and by one of the C atoms of the C–C BINOL bridge for thiol addition catalysts. In both cases, the resulting coordinate system defines the front hemisphere as the one containing the Cl atoms (iPP) and P(O)OH fragment, respectively. The octants analyzed are situated in the front hemisphere (see Figure 1).

For the iPP example, optimization of the probe sphere radius was performed previously for the quadrant model;\(^6\) a limited testing confirmed that 5.0 Å was also an optimal choice for the octant model. For the BINOL-phosphoric acids example, the sphere radius was fully optimized within the 3.0–8.0 Å range, with a step size of 0.1 Å, and a final best
radius of 3.0 Å for both the quadrant and octant models. For
the iPP polymerization catalysts, hydrogen atoms were
included, but the MCl₂ fragment was not. For the BINOL-
phosphoric acid catalysts, hydrogen atoms were not included,
but the P(O)OH fragment was; the \( \Delta \% V_{BO} \) values
reported are differences between the averages of the two
most occupied quadrants/octants and the averages of the two
least occupied quadrants/octants.

**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge at
https://pubs.acs.org/doi/10.1021/acscatal.2c00976.

Polymerization conditions and polymer analysis, regres-
sion analysis, and catalyst synthesis (PDF)

LMCl₂ and TS structures for metallocenes and BINOL-
PA structures (XYZ)

**AUTHOR INFORMATION**

Corresponding Authors

Roberta Cipullo — Dipartimento di Scienze Chimiche,
Università di Napoli Federico II, 80126 Napoli, Italy; DPI,
5600 AX Eindhoven, the Netherlands; orcid.org/0000-
0003-3846-1999; Email: rcipullo@unina.it

Dmitry V. Uborsky — Department of Chemistry, Lomonosov
Moscow State University, 119991 Moscow, Russia; DPI,
5600 AX Eindhoven, the Netherlands; orcid.org/0000-
0002-0093-6867; Email: duborsky@med.chem.msu.ru

Luigi Cavallo — Catalysis Research Center, Physical Sciences
and Engineering Division, King Abdullah University of
Science and Technology, Thuwal 23955, Saudi Arabia;
orcid.org/0000-0002-1398-338X; Email: luigi.cavallo@kaust.edu.sa

Christian Ehm — Dipartimento di Scienze Chimiche,
Università di Napoli Federico II, 80126 Napoli, Italy; DPI,
5600 AX Eindhoven, the Netherlands; orcid.org/0000-
0002-2538-5141; Email: christian.ehm@unina.it

Authors

Giuseppe Antinucci — Dipartimento di Scienze Chimiche,
Università di Napoli Federico II, 80126 Napoli, Italy; DPI,
5600 AX Eindhoven, the Netherlands; orcid.org/0000-
0003-1258-6363

Busra Dereli — Catalysis Research Center, Physical Sciences
and Engineering Division, King Abdullah University of
Science and Technology, Thuwal 23955, Saudi Arabia

Antonio Vittoria — Dipartimento di Scienze Chimiche,
Università di Napoli Federico II, 80126 Napoli, Italy; DPI,
5600 AX Eindhoven, the Netherlands; orcid.org/0000-
0003-3179-7427

Peter H. M. Budzelaar — Dipartimento di Scienze Chimiche,
Università di Napoli Federico II, 80126 Napoli, Italy; DPI,
5600 AX Eindhoven, the Netherlands; orcid.org/0000-
0003-0039-4479

Georgy P. Goryunov — Department of Chemistry, Lomonosov
Moscow State University, 119991 Moscow, Russia; DPI,
5600 AX Eindhoven, the Netherlands; orcid.org/0000-
0002-5188-5799

Pavel S. Kulyabin — Department of Chemistry, Lomonosov
Moscow State University, 119991 Moscow, Russia; DPI,
5600 AX Eindhoven, the Netherlands; orcid.org/0000-
0003-2548-2117

Alexander Z. Voskoboynikov — Department of Chemistry,
Lomonosov Moscow State University, 119991 Moscow,
Russia; DPI, 5600 AX Eindhoven, the Netherlands

Vincenzo Busico — Dipartimento di Scienze Chimiche,
Università di Napoli Federico II, 80126 Napoli, Italy; DPI,
5600 AX Eindhoven, the Netherlands

Complete contact information is available at:
https://pubs.acs.org/10.1021/acscatal.2c00976

Author Contributions

G.A. and B.D. contributed equally. The manuscript was
written through contributions of all authors. All authors have
given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This research forms part of the research program of DPI,
project 835. L.C. thanks KAUST for supporting a sabbatical
stay at the University of Naples, and the University of Naples
for hosting him. L.C. thanks the KAUST Supercomputing
Laboratory for providing computational resources on the HPC
platform Shaheen II.

**ABBREVIATIONS**

iPP, isotactic polypropylene; HTE, high-throughput exper-
imentation; \( M_w \), weight-averaged molecular weight; \( M_n \),
number-averaged molecular weight; NMR, nuclear magnetic
resonance; MAD, mean absolute deviation; \( \% V_{BO} \), per-
centage of buried volume; \( \Delta \% V_{BO} \), difference between
the occupied and formally empty octants; \( \Delta \% V_{BO} \), differ-
ence between the occupied and formally empty quadrants; \( \% V_{BO} \),
percentage of buried volume in the most occupied octant; \( \% V_{BO} \),
percentage of buried volume in the most occupied quadrant

**REFERENCES**

(1) Noyori, R. Asymmetric Catalysis: Science and Opportunities

(2) Resconi, L.; Cavallo, L.; Fait, A.; Piemontesi, F. Selectivity in
Propene Polymerization with Metallocene Catalysts. Chem. Rev. 2000,
100, 1253–1346.

(3) Busico, V.; Cipullo, R. Microstructure of Polypropylene. Prog.

(4) Nandy, A.; Duan, C.; Taylor, M. G.; Liu, F.; Steeves, A. H.;
Kulik, H. J. Computational Discovery of Transition-metal Complexes:
2021, 121, 9927–10000.

(5) Falivene, L.; Cao, Z.; Petta, A.; Serra, L.; Poater, A.; Oliva, R.;
Scarano, V.; Cavallo, L. Towards the Online Computer-Aided Design of

(6) Porée, C.; Schoenebeck, F. A Holy Grail in Chemistry:
Computational Catalyst Design: Feasible or Fiction? Acc. Chem. Res. 2017,
50, 605–608.

(7) Lledós, A. Computational Organometallic Catalysis: Where We
2555.

Sci. 2011, 1, 46–56.

(9) Mitchell, J. B. O. Machine Learning Methods in Chemo-

(10) Janet, J. P.; Liu, F.; Nandy, A.; Duan, C.; Yang, T.; Lin, S.;
Kulik, H. J. Designing in the Face of Uncertainty: Exploiting
Electronic Structure and Machine Learning Models for Discovery in


