End-to-End High-Throughput Approach for Data-Driven Internal Donor Development in Heterogeneous Ziegler-Natta Propylene Polymerization

Toshiaki Taniike[*,#,†,‡], *Felicia Daniela Cannavacciuolo*[#,§,‡], *Mostafa Khoshsefat*[†,‡], *Diego De Canditiis*[§,‡], *Giuseppe Antinucci*[*,§,‡], *Patchanee Chammingkwan*[†,‡], *Roberta Cipullo*[§,‡], *Vincenzo Busico*[§,‡]

[†]Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan

[‡]DPI, P.O. Box 902, 5600 AX Eindhoven, the Netherlands

[§]Dipartimento di Scienze Chimiche, Università di Napoli Federico II, Via Cintia, 80126 Napoli, Italy

KEYWORDS: Heterogeneous Ziegler-Natta catalyst; propylene polymerization; internal donor; high-throughput experimentation; machine learning

ABSTRACT: Internal donors (IDs) play a decisive role in shaping the structure and performance of Ziegler-Natta catalyst formulations for the isotactic polypropylene production. Unfortunately, their diverse and intricate functions remain elusive, and rational ID discovery therefore is still problematic. Exploitation of artificial intelligence methods such as machine learning, in turn, has been hindered by the lack of training datasets with adequate quality and size. This study proposes an integrated high-throughput workflow encompassing catalyst synthesis, propylene polymerization, and polypropylene characterization. Its application to an ID library of 35 molecules generated a robust and consistent dataset, which highlighted important and intriguing quantitative structure-properties relations (QSPRs). Furthermore, by fingerprinting ID molecular structure in combination with feature selection, a black-box QSPR model correlating ID molecular structure and catalytic performance was successfully implemented. This study demonstrates that the combination of high-throughput experimentation and machine learning is a promising asset for accelerating the research and development of Ziegler-Natta catalysts.

INTRODUCTION

Moving from Natta's initial invention, heterogeneous Ziegler-Natta (ZN) catalysts have dominated the industrial production of isotactic polypropylene (PP).[1–4] The first two generations were combinations of crystalline $TiCl_3$ in a layered modification, either pure or in mixture with $AlCl_3$, and an alkyl-Al compound.[3] That the addition of an organic Lewis base (e.g., an ether) can be beneficial for certain aspects of catalytic performance (like productivity) is known since the 1960s, but catalyst modification with bases became systematic only from the 1970s with the so-called high-yield formulations.[4] These consist of (a) a solid precatalyst, where a titanium precursor (most typically $TiCl_4$) is adsorbed on high-surface-area $MgCl_2$, (b) a trialkyl-Al

(usually Et_3Al), and (c) one or more Lewis bases, generally referred to as (electron) donors. In the commonly used terminology, a donor utilized at the stage of precatalyst preparation is named internal donor (ID), while one added in polymerization (in combination with the trialkyl-Al) is referred to as external donor (ED). Generally, both are necessary for optimal catalytic performance. ID/ED pairs for third-generation catalysts (early 1970s) were esters of benzoic acid, e.g., ethylbenzoate (EB), as ID and either the same EB or a para-substituted benzoate as ED.[5,6] Donor diversification, with phthalic diester IDs (e.g., dibutylphthalate) and alkoxysilane EDs, marked fourth-generation catalysts (late 1970s),[7,8] which still represent the workhorses of the PP industry. Since then, many more donors belonging in various molecular classes have been proposed; the discovery of novel IDs, in particular, has driven the process of PP catalyst evolution, as they generally have a stronger and more characteristic impact than EDs. For example, 1,3-dimethoxypropane (1,3-diether) IDs (ca. 1990)[9] and succinate IDs (ca. 2000)[10] enabled to produce highly isotactic PP with narrow and broad molecular weight distribution, respectively.

Implementing ID discovery is a highly demanding empirical task. The potential of specific candidate molecules is highly unpredictable, and must be assessed by means of a labor and time intensive process entailing custom synthesis, precatalyst preparation with proper protocols, and full catalyst verification in polymerization. Unfortunately, even IDs belonging in the same molecular class and featuring similar substitution patterns can greatly differ in performance, and quantitative structure-properties relations (QSPRs) are elusive. Behind this unpredictability is a multifaceted role: not only do IDs interact with other components of the catalyst formulation during precatalyst preparation and activation, potentially influencing the generation or elimination of byproducts, but they also affect the growth of solid $MgCl_2$ nanostructures, e.g.,

steering the formation of certain crystallite terminations due to preferential chemisorption.[11–19] In polymerization, IDs modify the steric and electronic properties of the active Ti species by co-adsorbing at adjacent surface locations,[20–24] in some cases dynamically due to reaction(s) with the R_3Al/ED cocatalyst that can liberate part of the surface for ED adsorption, ending up with mixed ID/ED pools.[4,25–29] Also in view of the difficulty to intercept and characterize the catalytic species, present in low amount and extremely reactive,[30,31] it is fair to acknowledge a "black box" character for this catalysis.

Statistical QSPR modeling is certainly part of any "best practices" protocols for searching novel IDs. In principle, machine learning (ML) methods should be considered with high priority; in practice, however, their use is presently hampered by the unavailability of datasets with adequate size and quality, similar to other fields of materials science.[32,33] The field of ZN catalysis is highly competitive, and researchers follow diverse and sometimes partly undisclosed procedures; even the metrics adopted for reporting the results do not conform to generally accepted standards. This makes it difficult or even impossible to collate and compare results from different sources.

In such a context, the introduction of high-throughput experimentation (HTE) represented a breakthrough.[34–38] Throughput intensification in organometallic catalysis is much lower than in pharma; yet, a 100-fold increase relative to conventional methods with similar or even higher quality standards is easily achieved. In our labs, we recently undertook a collaborative program for implementing advanced HTE tools and methods for the rapid exploration of the variables hyperspace in ZN PP catalysis. Our workflow permits an extensive coverage of the knowledge and value chains by means of unique tools for (pre)catalyst preparation and analysis,[39] robust

protocols of catalyst screening under industrially relevant conditions,[23,40–49] and high-end rapid polymer characterizations.[50,51] Integration with ML instruments is now in progress.

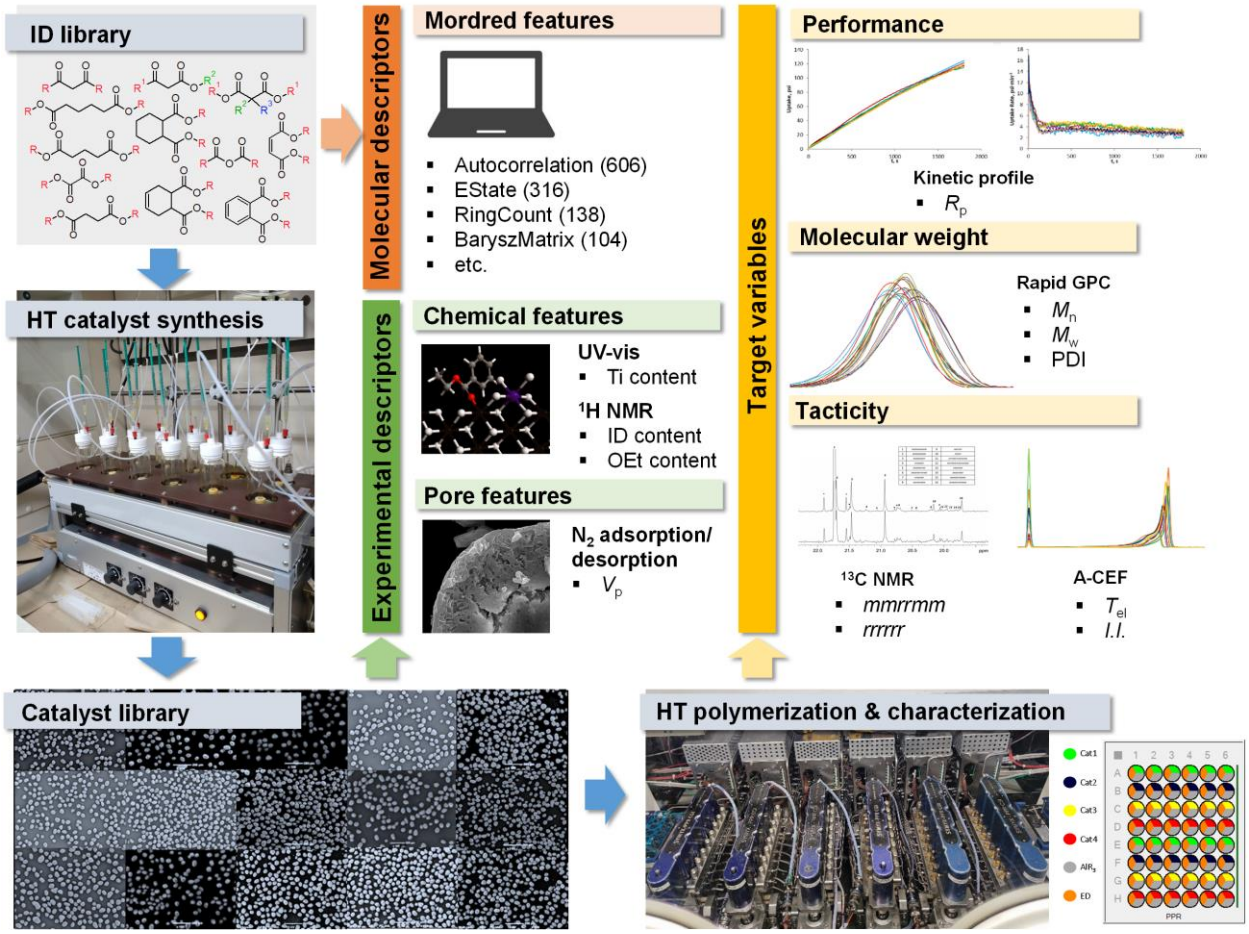

Figure 1. Workflow of high-throughput experimentation and data acquisition.

In the present paper, we illustrate how the workflow, as depicted in Figure 1, can be applied to the generation of a high-quality dataset for fast searches of novel IDs. The paper is primarily methodological, and the database used for exemplification is not meant for predictive application. This notwithstanding, the emerging picture does not fully align with conventional

wisdom, which in our opinion indicates there is still ample room for data-driven innovation—an important fact, considering the need to re-design the PP industry for a circular economy.

EXPERIMENTAL SECTION

ID library

The library comprised 35 commercially available Lewis bases, all featuring two carbonyl groups. Each Lewis base molecule is labeled as X-Y#, as shown in Figure 2. Y qualifies the chemical class (DE for diesters; KE for ketoesters; DK for diketones; AN for anhydrides); X qualifies the spacer between the carbonyls (Ar for aromatic; C for cyclic; Ln for linear with n C atom length, between 0 (L0) and 4 (L4); no X in the case of Y = AN); # qualifies the alkyl substituents of the carbonyls. A (pre)catalyst featuring Lewis base X-Y# as ID was denoted as CAT-X-Y#.

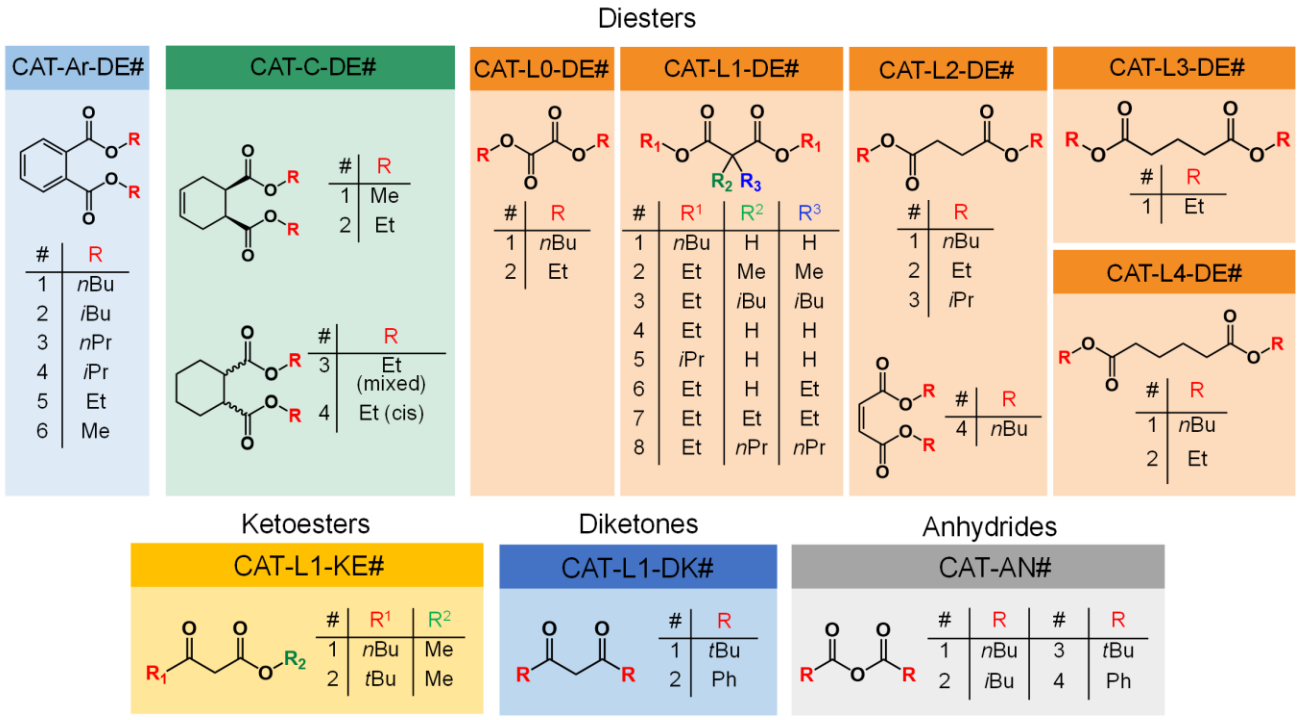

Figure 2. ID library, comprising 35 commercially available Lewis bases with two carbonyl groups. Catalysts produced alongside these IDs are labeled as CAT-X-Y#, where X signifies the carbon skeleton (Ar for aromatic, C for cyclic, Ln for linear), and L0−L4 denotes the linear skeleton's carbon atom length. Anhydrides are coded as CAT-AN# due to uniform skeletal structure. Y represents the compound group (DE for diesters, KE for ketoesters, DK for diketones, AN for anhydrides), while # signifies alkyl substituent variations within each X-Y.

Materials

Mg powder was purchased from Merck KGaA. Ethanol (purity > 99.5%, Kanto Chemical) was dried over a 3Å molecular sieve. Iodine (I_2, purity > 99.9%, FUJIFILM Wako Pure Chemical) was used as received. *n*-Heptane (purity > 99.5%) and toluene (purity > 99.5%) purchased from FUJIFILM Wako Pure Chemical were dried over a 3Å molecular sieve followed by N_2 bubbling prior to use. Titanium tetrachloride ($TiCl_4$, special grade) was purchased from FUJIFILM Wako Pure Chemical. Di-*n*-butyl phthalate (purity > 99%, Sigma-Aldrich), di-*i*-butyl phthalate (purity > 98%, TCI), di-*n*-propyl phthalate (purity > 98%, TCI), di-*i*-propyl phthalate (purity > 98%, TCI), diethyl phthalate (purity > 99%, KANTO chemical), dimethyl phthalate (purity > 99%, TCI), dimethyl *cis*-4-cyclohexene-1,2-dicarboxylate (purity > 97%, TCI), diethyl *cis*-4-cyclohexene-1,2-dicarboxylate (purity > 98%, TCI), diethyl 1,2-cyclohexanedicarboxylate (purity > 95%, TCI), diethyl *cis*-1,2-cyclohexanedicarboxylate (purity > 98%, TCI), di-*n*-butyl oxalate (purity > 99%, TCI), diethyl oxalate (purity > 99%, TCI), di-*n*-butyl malonate (purity > 99%, TCI), diethyl dimethylmalonate (purity > 97%, Sigma-Aldrich), diethyl di-*i*-butylmalonate (purity > 98%, Sigma-Aldrich), diethyl malonate (purity > 99%, Sigma-Aldrich), di-*i*-propyl malonate (purity > 99%, TCI), diethyl ethylmalonate (purity > 99%, Sigma-Aldrich), diethyl

diethylmalonate (purity > 98%, Sigma-Aldrich), diethyl di-*n*-propylmalonate (purity > 97%, TCI), di-*n*-butyl succinate (purity > 99%, TCI), diethyl succinate (purity > 99%, Sigma-Aldrich), di-*i*-propyl succinate (purity > 99%, TCI), di-*n*-butyl maleate (purity > 95%, TCI), diethyl glutarate (purity > 98%, TCI), di-*n*-butyl adipate (purity > 99%, TCI), diethyl adipate (purity > 99%, TCI), methyl 3-oxoheptanoate (purity > 95%, TCI), methyl 4,4-dimethyl-3-oxovalerate (purity > 95%, TCI), dipivaloylmethane (purity > 98%, TCI), 1,3-diphenyl-1,3-propanedione (purity > 98%, TCI), valeric anhydride (purity > 98%, TCI), isovaleric anhydride (purity > 99%, TCI), pivalic anhydride (purity > 98%, TCI), and benzoic anhydride (purity > 95%, TCI) were used as received.

Magnesium ethoxide preparation

Magnesium ethoxide ($Mg(OEt)_2$) as a solid catalyst precursor was synthesized from the reaction between Mg and ethanol using I_2 as an initiator according to the previously reported method.[52,53] To a 500 mL jacket-type three-necked flask equipped with a mechanical stirrer rotating at 180 rpm, 0.68 g of I_2 and 35 mL of ethanol were introduced. After complete dissolution of I_2 at the ethanol reflux temperature, 3.0 g of Mg powder and 35 mL of ethanol were repeatedly added for nine times with the interval time of 10 min. After the last addition, the mixture was kept reacted at the same temperature for 2 h. The obtained solid product was washed with heptane and dried under vacuum at 40 °C.

(Pre)catalyst synthesis

Catalyst samples were synthesized in high-throughput mode using a custom-designed parallel reactor system.[39] The system comprises three main components: reaction vessels, a unit for

temperature control and mixing, and a feeder. The reaction vessel is a 50 mL glass bottle sealed with a custom-designed Teflon cap with the aid of O-ring. The cap is integrated with a magnetic suspended stir bar to facilitate mixing without destructing catalyst macroparticles. It also equips two small open ports: one is used for maintaining the N_2 atmosphere, and the other for introducing reagents into the vessel or extracting the supernatant for decantation. The temperature control and mixing unit integrates a closed oil circulation bath featured with twelve holes for placing the reaction vessels in the upper section, and a belt-drive magnetic stirring system in the lower section. The temperature control is achieved by circulating oil from a separate heating/cooling unit, allowing precise control within the range of –20 °C to 150 °C. The feeder utilizes a peristaltic pump (Masterflex L/S, Cole Parmer) with twelve cartridges for feeding a $TiCl_4$ solution to twelve reaction vessels at a controlled flow rate.

High-throughput catalyst synthesis was implemented as follows: in a N_2 glove bag, 2.0 g of $Mg(OEt)_2$ was added to each reaction vessel, which was then sealed with a Teflon cap and connected to a N_2 line. The vessel was placed onto the temperature control and mixing system. After introducing 14 mL of toluene, the suspension was cooled down to –5 °C under stirring, and 8.0 mL of a $TiCl_4$ solution in toluene (1/1 v/v) was fed with a peristaltic pump. Thereafter, the mixture was gradually heated to 90 °C, followed by the addition of 2.07 mmol of an ID. The mixture was further heated to 110 °C and held at this temperature for 2 h. The solid product was washed with toluene twice by decantation, and further treated with 4.0 mL of $TiCl_4$ in 18 mL of toluene at 110 °C for another 2 h. The resultant catalyst was repetitively washed with toluene and heptane, and dried under vacuum. The protocol was performed on 12 vessels in parallel, where only the ID was swapped.

Catalyst characterization

The catalysts synthesized were subjected to a standard set of characterizations. The Ti content was determined using a colorimetric method with a UV-vis spectrometer (V-670, JASCO). A 50 mg sample was dissolved in an aqueous solution of HCl/H_2SO_4, followed by the addition of H_2O_2 to form a titanium peroxo complex. The Ti concentration was calculated from the absorption intensity of the band at 410 nm based on external calibration. Organic contents were analyzed using ^1H NMR on a Bruker AVANCE III 400 MHz spectrometer, following a previously reported procedure.[54] A sample (10–30 mg) was dissolved in 0.7 mL of DMSO-d^6 containing 40 μg of 1,1,2,2-tetrachloroethane as an internal standard. The protons of the aromatic/aliphatic spacers were used to confirm the inclusion of IDs in the catalyst samples, except for oxalates and anhydrides, whose inclusion was verified using the protons of the side groups. All forms of IDs, including primary and secondary forms, were analyzed when side reactions occurred: transesterification for diesters and ketoesters, isomerization or tautomerization for diketones and ketoesters, and esterification for anhydrides. The protons of –OCH_2– at 3.4 ppm were utilized to quantify the OEt group associated with $TiCl_x(OEt)_{4-x}$ species. N_2 adsorption and desorption experiments were performed at 77 K using a BELSORP−max instrument (BEL JAPAN, Inc.). A sample was outgassed at 80 °C for 3 h in vacuum prior to the measurement. The pore volume and pore size distribution were determined based on the non-local density functional theory (NLDFT) method.

Propylene polymerization

Polymerization experiments were carried out using a HTE platform (Freeslate Parallel Pressure Reactor). This setup, integrally contained in a triple MBraun LabMaster glovebox, features 48

reaction cells arrayed in 6 eight-cell modules. Polymerization reactions are run in a semi-continuous mode. The injection system consists of a dual-arm robot adopting different technologies for catalyst solutions and slurries, with specialized needles and injectors. The slurry needle, in particular, is designed to penetrate the gas cap of the reaction cells and dispense the catalyst slurry directly into the liquid phase, ensuring highly accurate and precise dosing. The PPR software allows the operator to modify the design of experiment (DoE) for the planned set of 48 polymerization experiments during execution.

The polymerization experiments were performed based on a common protocol. In brief, polymerization conditions were as follows: heptane slurry, T_p = 70 °C; $p_{(C3H6)}$ = 4.5 bar; $p_{(H2)}$ = 0.20 bar; [Al]/[Ti] = 160; [ED]/[Al] = 0.10; ED = diisopropyldimethoxysilane. The reactions were left to proceed under stirring (800 rpm) at constant temperature and pressure with a continuous feed of propylene for a desired time (usually 30 or 60 min), and quenched by over-pressurizing the cell with dry air. Once all cells were quenched, the modules were cooled down and vented, the stir-tops were removed, and the glass inserts containing the reaction phase were taken out and transferred to a Martin Christ RVC 2-33 CDplus centrifugal evaporator, where all volatiles were distilled out and the polymers were thoroughly dried overnight. Reaction yields were double-checked against on-line monomer conversion measurements by robotically weighing the dry polymers while still in the reaction vials, subtracting the pre-recorded tare. Polymer aliquots were then sent to the characterizations. A fully detailed description of the protocol was reported elsewhere by some of us.[41,42,47]

PP characterization

The characterization of the PP samples was performed in high-throughput mode. Gel permeation chromatography (GPC) curves were recorded with a Freeslate Rapid GPC setup, equipped with a set of 2 mixed-bed Agilent PLgel 10 μm columns and a Polymer Char IR4 detector. The upper deck of the setup includes a sample dissolution station capable of handling up to 48 samples in 10 mL magnetically stirred glass vials. With robotic operation, predetermined polymer quantities (usually 1 to 4 mg) were dissolved in specific volumes of 1,2-dichlorobenzene with a 0.40 mg/mL concentration of 4-methyl-2,6-di-*tert*-butylphenol (BHT) stabilizer, so as to obtain solutions at a concentration of 0.5 to 1.0 mg/mL. After 2 h at 150 °C under gentle stirring to ensure complete dissolution, the sample array was transferred to a thermostated bay at 145 °C, and the samples were sequentially injected into the column line at 145 °C and a flow rate of 1.0 mL/min. In a post-trigger delay operation mode, the analysis time was 12.5 min per sample. Calibration was carried out using 10 monodisperse polystyrene samples (M_n between 1.3 and 3700 KDa). Before and after each campaign, samples from a known PP batch produced with an *ansa*-zirconocene catalyst were analyzed for a consistency check. Number average molecular weight (M_n), weight average molecular weight (M_w) and polydispersity index (PDI; defined as M_w/M_n) of the produced polymers were collected via GPC.

Analytical crystallization elution fractionation (A-CEF) curves were collected with a Polymer Char A-CEF setup, equipped with an autosampler (42 wells), an IR5 detector and a dual capillary viscometer detector. With robotic operation, predetermined polymer quantities (typically 8–16 mg) were dissolved in 1,2-dichlorobenzene containing 0.40 mg/mL of BHT stabilizer, so as to achieve a concentration of 2.0 mg/mL. After 90 min at 150 °C under vortexing in sealed vials to ensure complete dissolution, the samples were sequentially charged into the injection loop, where they were held at 95 °C for 5 min and then moved into the column. The

crystallization step entailed an 8.0 °C/min cooling ramp down to 35 °C at a flow rate of 0.24 mL/min; 1 min after reaching 35 °C, sample elution was started, with a 4 °C/min heating ramp up to 150 °C at a flow rate of 1.0 mL/min. The resulting analysis time was 60 min per sample. For each sample, the amorphous fraction (*AF*) (i.e., the relative sample amount eluted at 35 °C) was utilized to calculate the index of isotacticity (*I.I.*) as *I.I.* = (100 – *AF*) wt%.

Quantitative ^{13}C NMR spectra were recorded with a Bruker Avance III 400 spectrometer operating at 100 MHz, equipped with a 5 mm high-temperature cryoprobe, and a robotic sample changer with a pre-heated carousel (24 positions). The NMR spectra were acquired with a 45° pulse, 2.7 s acquisition time, 2.0 s relaxation delay, and either 400 or 800 transients (corresponding to analysis times of 30 or 60 min). Broad-band proton decoupling was achieved using a modified WALTZ16 sequence (BI_WALTZ16_32 by Bruker).

Machine learning

A ML approach was adopted to model ID QSPRs. The experimental data (next Section, Table 1) were used without any data preprocessing or removal. ID molecular structures were fingerprinted using Mordred, a software developed for calculating molecular descriptors.[55] This software offers distinct advantages when compared to alternative tools. Notably, it excels in rapid calculation speeds, user-friendliness, capability to handle large molecules, and providing a broader array of descriptors within a lenient licensing framework. The calculations were performed using Python 3.7, along with the Pandas, RD-kit, and Mordred 1.0.0 packages.

Among the total descriptors (1825), zero-variance descriptors as well as three-dimensional descriptors were excluded. The remaining descriptors (1138) primarily comprised autocorrelation (606) and Barysz Matrix (104). These descriptors are based on the atom

connectivity and molecular graph structures, encompassing information about parameters such as size, electronegativity, ionization potential, Van der Waals volume, etc.

Thus, individual IDs were uniquely defined by their molecular descriptors and experimentally acquired chemical compositions and structures of the catalysts, serving as explanatory variables. However, the surplus of these features compared to the number of IDs necessitated feature selection. This was accomplished by a genetic algorithm (GA) to select a predetermined number of features that optimize the score of regression. Given the limited dataset size and the potential for outlying molecular structures, we adopted Huber regression − a robust form of multiple linear regression.[56] The efficacy of each feature set was evaluated using the mean absolute error (MAE) within leave-one-out cross-validation (LOOCV). For each target variable, feature selection by GA, involving 800 populations and 1600 generations, was conducted three times. The feature set and model with the lowest MAE value across these iterations were then chosen. The calculations were performed on a high-performance computing (HPC) system, utilizing a custom Python 3.8 code in conjunction with common libraries such as Pandas, Numpy, Scikit-Learn, and Shap. Further details are read elsewhere as part of the automatic feature engineering technique recently developed by some of us.[57]

RESULTS AND DISCUSSION

ID library, HTE dataset and classical description

Within the constraints of commercial availability, we assembled the ID library in order to ensure an adequate diversity of molecular structures along with practical relevance. Although there exists a wide variety of organic Lewis bases, those suited as ZN IDs are quite limited: small molecules that can effectively bind to coordinatively unsaturated Mg^{2+} ions on $MgCl_2$ surfaces

and do not contain active protons.[58] In particular, donors featuring O-based functional groups, like ethers and esters, are dominant for practical use, while N- or S-containing donors are also possible but far less common. Moreover, molecules with two electron donating groups tend to outperform those with single donating groups, due to their stronger binding to the surface in the chelating or bridging mode; in fact, ZN catalysts from the third generation onward all contain bifunctional IDs. Not surprisingly, ID performance is significantly affected by electronic and steric factors, including (but not limited to) the distance between the two electron donating groups (often referred to as spacer length), the presence of sterically demanding substituents, the rigidity of the molecular structure, etc. We carefully considered all said elements when selecting the 35 molecules in the library, all featuring two O-based functional groups (Figure 2). Diesters were chosen as the major components, because they offer the desired diversity within commercially available compounds. The library was then augmented with some ketoesters, diketones, and anhydrides with comparable spacer length. A number of 35 structures may be regarded as inadequate for advanced ML techniques, but it can suffice when combined with feature engineering (see below). Moreover, it is worth noting that there are very few instances in the existing literature where catalyst preparation and evaluation were accomplished using a unified process for such a number of IDs. In this respect, the present paper represents a pioneering effort in establishing a consistent ID dataset by streamlining high-throughput experimental protocols for catalyst preparation, polymerization, and polymer analysis (Figure 1).

Various methods are available for preparing ZN (pre)catalysts. The formation of catalyst particles from alcohol solutions of $MgCl_2$ in the presence of an ID is a most commonly employed one, but it was not adopted in this study due to potential non-uniform morphology of the catalyst macroparticles depending on ID molecular structure. Instead, we opted for the chemical

conversion of $Mg(OEt)_2$, another widely used method.[59,60] In this approach, $Mg(OEt)_2$ particles are transformed into catalyst macroparticles through $TiCl_4$ treatment in the presence of an ID, while preserving their original particle morphology. Initially, we synthesized a sufficient quantity of $Mg(OEt)_2$ spheroidal particles characterized by a narrow particle size distribution and mechanical robustness against mechanical stirring and the heat generated during catalyst preparation. This synthesis was carried out using a well-established set of processes and conditions. Subsequently, we prepared catalysts containing various IDs in parallel, using the same $Mg(OEt)_2$ source as the starting material and employing the identical process and conditions.[39] A total of 12 catalysts could be prepared in single operation of parallel preparation, that is, the 35 catalysts were obtained in three parallel preparations. The thus prepared catalysts were handled under rigorously inert atmosphere for the characterizations; the resulting dataset is reported in columns 2–5 of Table 1. The characteristics of the catalysts' macroparticles are omitted, as they retained the morphology of $Mg(OEt)_2$ particles regardless of the molecular structure of the IDs, exhibiting nearly identical shapes and size distributions (Figure S1); in other words, the structural features at this scale are entirely unrelated to the observed changes in catalytic performance.

Table 1. Overall dataset of the 35 ZN catalysts prepared and screened in this work.

Code[a]	Ti[b] [mmol/g]	OEt[c] [mmol/g]	ID[c] [mmol/g]	V_p[d] [cm^3/g]	$R_{p,rel}$[e] [%]	M_n[f] [kDa]	M_w[f] [kDa]	PDI[f]	*mmrrmm*[g] [mol%]	*rrrrrr*[g] [mol%]	T_{el}[h] [°C]	*I.I.*[h] [wt%]
CAT-Ar-DE-1	0.46	0.20	0.53	0.331	100	42	313	7.3	0.2	0.1	118.2	97
CAT-Ar-DE-2	0.42	0.11	0.46	0.320	72	53	360	6.8	0.2	0.1	118.3	97
CAT-Ar-DE-3	0.36	0.11	0.51	0.295	79	53	321	6.1	0.1	0.1	118.3	97
CAT-Ar-DE-4	1.15	0.27	0.32	0.410	20	33	201	6.1	0.4	0.2	116.7	95

CAT-Ar-DE-5	0.38	0.16	0.61	0.484	73	53	346	6.5	0.2	0.1	118.5	98
CAT-Ar-DE-6	0.56	0.09	0.81	0.602	20	49	252	5.1	0.2	0.1	117.9	98
CAT-C-DE-1	0.50	0.11	0.41	0.813	17	41	203	5.0	0.5	0.2	117.0	96
CAT-C-DE-2	0.54	0.18	0.34	0.499	40	40	224	5.6	0.4	0.2	117.6	97
CAT-C-DE-3	0.44	0.13	0.59	0.582	63	46	266	5.8	0.4	0.2	118.1	97
CAT-C-DE-4	0.44	0.16	0.55	0.503	43	44	243	5.5	0.6	0.2	117.4	95
CAT-L0-DE-1	1.82	0.04	0.02	0.538	15	30	176	6.0	0.6	0.5	116.0	92
CAT-L0-DE-2	1.34	0.02	0.05	0.726	13	40	180	4.5	0.7	0.4	116.2	94
CAT-L1-DE-1	0.52	0.16	0.60	0.756	27	32	200	6.3	0.4	0.4	117.1	94
CAT-L1-DE-2	0.42	0.22	0.51	0.439	80	48	244	5.1	0.7	0.3	116.8	95
CAT-L1-DE-3	0.79	0.27	0.48	0.308	103	33	163	4.9	0.8	0.2	116.0	94
CAT-L1-DE-4	0.50	0.09	0.79	0.720	19	29	149	5.0	0.5	0.4	117.1	95
CAT-L1-DE-5	1.61	0.20	0.27	0.676	9	35	180	5.2	0.7	0.4	115.4	94
CAT-L1-DE-6	0.44	0.13	0.49	0.503	48	33	182	5.5	0.6	0.3	116.7	95
CAT-L1-DE-7	0.56	0.13	0.47	0.384	49	33	199	6.0	0.6	0.3	116.9	95
CAT-L1-DE-8	0.54	0.20	0.53	0.352	85	33	163	4.9	0.8	0.2	116.2	94
CAT-L2-DE-1	0.50	0.07	0.49	0.639	11	41	210	5.1	0.2	0.2	117.0	97
CAT-L2-DE-2	0.52	0.07	0.44	0.877	4	34	185	5.5	0.4	0.2	116.5	97
CAT-L2-DE-3	1.65	0.36	0.20	0.500	8	28	146	5.2	0.9	0.4	115.2	94
CAT-L2-DE-4	0.44	0.09	0.84	0.669	25	47	274	5.8	0.2	0.1	118.0	97
CAT-L3-DE-1	0.38	0.02	0.72	0.840	8	42	227	5.3	0.4	0.2	117.1	96
CAT-L4-DE-1	0.36	0.11	0.71	0.507	9	40	209	5.2	0.3	0.2	116.9	96
CAT-L4-DE-2	0.71	0.04	0.80	0.633	2	37	180	4.9	0.3	0.2	115.7	97
CAT-L1-KE-1	0.79	0.38	0.34	0.548	75	36	182	5.0	1.0	0.5	115.7	92
CAT-L1-KE-2	1.11	0.36	0.21	0.480	32	22	122	5.6	1.1	0.6	114.9	89
CAT-L1-DK-1	1.42	0.31	0.28	0.493	37	41	189	4.7	1.0	0.3	115.4	94
CAT-L1-DK-2	1.42	0.29	0.27	0.495	42	37	154	4.2	1.2	0.4	114.9	93
CAT-AN-1	1.21	0.40	0.47	0.299	23	37	179	4.9	0.8	0.4	115.9	93
CAT-AN-2	1.13	0.27	0.51	0.245	21	34	185	5.4	0.9	0.5	115.9	92
CAT-AN-3	0.67	0.38	0.40	0.336	67	32	176	5.6	1.1	0.6	115.3	91

CAT-AN-4	0.77	0.29	1.03	0.436	31	29	137	4.7	1.0	0.4	115.0	92

[a] Corresponding to the ID code described in Figure 2.

[b] Determined by UV-vis spectroscopy.

[c] Determined by ^{1}H NMR. It included not only the original form of ID but also secondary forms, typically derived by transesterification (for DE and KE), isomerization or tautomerization (for DK and KE), and esterification (for AN).

[d] Obtained from a N_2 adsorption isotherm at 77 K using NLDFT.

[e] Relative productivity to that of CAT-Ar-DE-1.

[f] Determined by high-temperature GPC.

[g] Determined by ^{13}C NMR.

[h] Obtained from A-CEF. T_{el} and *I.I.* refer to the peak temperature of the elution curve and the weight fraction of the eluates above 35 °C, respectively.

Propylene homopolymerization experiments were conducted under strictly controlled experimental conditions, using a state-of-the-art HTE reaction platform and thoroughly validated protocols (see Experimental Section). The results of these experiments (averages of 35 duplicate experiment pairs) and of the polymer characterizations via Rapid GPC, A-CEF, and ^{13}C NMR spectroscopy are summarized in columns 6–13 of Table 1. The maximum observed error was 15% on catalyst productivity and average MW values; for other properties, experimental uncertainties were within ±1 on the last significant digit. Average catalyst productivity (R_p, expressed in $kg_{(PP)}\ g_{(Ti)}^{-1}\ h^{-1}$) is a convenient kinetic descriptor as long as catalyst deactivation is negligible—a condition met by all catalysts in the library. To facilitate comparison, relative values ($R_{p,rel}$) were reported as percentages (%) of the absolute R_p value measured for CAT-Ar-DE-1, a popular fourth-generation system chosen as a reference. The ^{13}C NMR fractions of *mmrrmm* and *rrrrrr* heptads (% of total methyl integral) are indicators of isolated stereodefects within highly isotactic and stereoblock PP chains, respectively. The amorphous fraction (*AF*)

measured by A-CEF represents the relative amount of polymer produced at poorly stereoselective catalytic sites. This quantity nicely correlates with the so-called xylene-soluble (*XS*) PP fraction, and was used to measure the index of isotacticity (*I.I.*), that is the relative amount of polymer produced at the highly stereoselective catalytic sites (conventionally referred to as "highly isotactic"). The maximum elution temperature in the A-CEF curve (T_{el}), in turn, is another way to estimate the stereoregularity of the insoluble fraction.

The overall dataset in Table 1, which is unprecedented when considering the size and internal consistency, highlights a fascinating and intriguing mix of predictable and surprising QSPRs.

- The negative correlation between Ti and ID contents (Figure S2a) can be understood as a result of the competitive adsorption of $TiCl_4$ and ID on the coordinatively unsaturated $MgCl_2$ surfaces.[3,4,24,61,62] IDs resulting into high Ti contents included those prone to secondary reactions (that is, tautomerization for KE and DK; esterification with Ti-OEt for AN; transesterification with Ti-OEt for isopropyl DE), and oxalates (L0-DE#) that hardly adsorbed on $MgCl_2$. In line with common sense, this indicates that Lewis bases acting effectively as IDs are chemically stable/inert and have high affinity for $MgCl_2$.
- Catalysts with high Ti content (i.e., low ID content) produced PP with comparatively low stereoregularity, as indicated consistently by the relevant polymer features, namely *I.I.*, T_{el} (Figure S2b), *mmrrmm* and *rrrrrr*.
- OEt groups left over from $Mg(OEt)_2$ in the catalyst have a detrimental impact on activity and stereoselectivity (Figure S2c and previous literature).[39,54] Residual OEt amount turned out to depend significantly on ID molecular structure: it was lower in general with diester IDs (and especially so with oxalates) than with all other ID classes.

- Pore volume turned out to be highly sensitive to ID molecular structure: while an overall trend was hard to discern, specific phthalates and malonates were found to contribute to the smaller pore volume. The weak negative correlation of this parameter with catalyst productivity (Figure S2d) suggests that fragmentation of denser particles is important for enhancing activity.
- Features regarding performance in polymerization are strongly intercorrelated. That this holds for M_n and M_w, as well as for *mmrrmm*, *rrrrrr*, T_{el}, and *I.I.*, is evident and requires no further discussion. The fact that higher values of M_n and M_w are associated with lower values of *mmrrmm* and *rrrrrr* and higher values of T_{el} and *I.I.* can be traced to a low propensity of highly isotactic PP chains to undergo chain transfer. In particular, M_w shows strong negative and positive correlation with *mmrrmm* and PDI (Figures S2e and S2f), respectively. This suggests that the factors hampering stereoirregular propylene insertion and β-H transfer to propylene monomer at the most stereoselective catalytic sites are synergic (possibly coincident).
- Looking at the overall catalytic performance, the superiority of phthalate IDs, and especially dibutylphthalate (CAT-Ar-DE1), is outstanding. Comparable productivity was observed only with certain disubstituted malonate IDs, whereas diesters lacking substituents on the spacer generally exhibited low(er) activity. M_n, M_w, and PDI values were also high in general with phthalate IDs, none surpassing the reference in terms of PDI. DK and diethyl oxalate IDs, to the other extreme, showed notably narrow PDI values.

Blackbox QSPR modeling by ML

Notwithstanding the several correlations highlighted in the previous section, the overall QSPRs for the ZN IDs in the library are too intricate for classical (deterministic) modeling attempts. Thus, we aimed to construct models that learned the response of a black box via supervised ML, where the molecular structure of IDs served as input, and the catalyst performance as output. This required fingerprinting the features of IDs (molecular structure and properties). Listing all relevant features based on domain knowledge was proven to be difficult. Hence, we generated a diverse set of features, and performed feature selection to identify crucial features in describing each performance metric.

Specifically, out of 1138 features generated from the IDs' molecular structures using Mordred, we used a genetic algorithm-based method to select five features that minimized the MAE value in LOOCV with Huber regression. Huber regression, a form of multiple regression, uses Huber loss instead of the standard least square, offering robustness against outliers.[56] Such a simple machine learning method was effective in preventing overfitting with a limited observation number—35 molecular structures in this case. The use of the cross-validation score, not the training score, for the feature selection also helped mitigate overfitting.

The feature selection was performed in two cases, depending on whether experimentally obtained features, namely chemical composition and pore volume of the catalysts listed in Table 1, were included or not in the feature selection. In our catalyst preparation, the only variable was ID molecular structure, while all other conditions, e.g., ID amount, chlorination temperature, purification protocol, etc., did not vary. Therefore, the aforementioned experimental features can be considered as consequences of ID molecular structure, and as such dependent variables that ideally may be omitted. Nonetheless, including those experimental features might enhance the model's accuracy when these features directly relate to the performance.

The top section of Figure 3 exhibits parity plots for the models built with $R_{p,rel}$, PDI, T_{el}, and *I.I.* as target variables by employing only the Mordred molecular features. The selected features, as well as training and cross-validation scores, are summarized in Table 2. Reasonable regression results were obtained without including the experimental features, irrespective of the target variables. The MAE_{LOOCV} values were 6.89 for $R_{p,rel}$, 0.18 for PDI, 0.21 °C for T_{el}, and 0.48 wt% for *I.I.*, considerably smaller compared to the standard deviation (SD) of each target variable (28.9, 0.65, 1.06 °C, 2.13 wt%, respectively). Moreover, these values were significantly lower than those obtained via 10-fold feature selection with y-randomization (mean ± 2SD = 13.4 ± 2.4, 0.27 ± 0.07, 0.47 ± 0.16 °C, 0.94 ± 0.19 wt%), indicating a very low possibility of chance correlation typically observed when using a large number of features. Besides, the MAE values were consistent between training and cross-validation, suggesting no tendency to overfitting, owing to the simple machine learning method involving the small number of features.

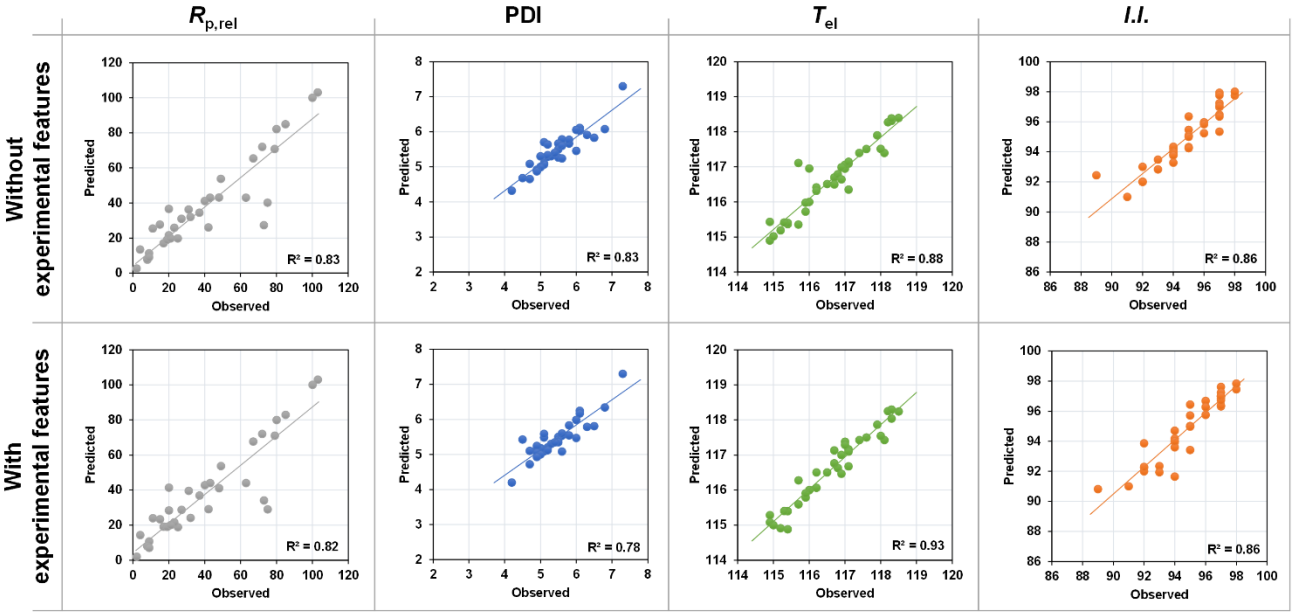

Figure 3. Regression results for different metrics of catalyst performance during propylene polymerization. Molecular descriptors were generated from the structural formulas of IDs using Mordred. A genetic algorithm was used to select five features per target variable that minimize

the mean absolute error of leave-one-out cross validation in Huber regression. The upper and lower panels compare results between models involving only molecular features and those involving at least one experimental feature. The selected features, along with the training and cross-validation scores, are given in Table 2.

The importance of the selected features was analyzed using Shapley Additive Explanation (SHAP) summary plots (Figure S3).[63] These plots visually represent each feature's impact on the model's predictions, displaying the magnitude and direction of influence. Higher SHAP values denote a stronger impact, with positive values indicating an increase in the target variable, while negative values indicate the opposite effect. These visualizations offer a clear understanding of each feature's significance and directionality in shaping the model's behavior. The most important feature and its impact direction were identified for each target variable: lower TiC0 (total information content index with a neighborhood symmetry of zero-order) values increase $R_{p,rel}$, while higher VR2_A (VR2 of the adjacency matrix) values raise PDI. Similarly, higher piPC8 (8-ordered pi-path count on a logarithmic scale) values and higher GATS2are (Geary coefficient of lag 2 weighted by Allred-Rocow electronegativity) values lead to enhancements in T_{el} and *I.I.*, respectively.

Table 2. Results of feature selection for different target variables.[a]

Target variable	Selected features	MAE_{train}	MAE_{LOOCV}
$R_{p,rel}$	AATSC4i, ZMIC3, TIC0, VR2_A, ATS7se	6.66	6.89
PDI	VR2_A, JGI7, MATS8i, MIC0, MATS4d	0.170	0.179
T_{el}	SMR_VSA6, AETA_eta_FL, piPC8, MATS1dv, ATSC1dv	0.201	0.210

I.I.	GATS2c, GATS8c, VSA_EState9, GATS4c, GATS2are	0.463	0.480
$R_{p,rel}$[b]	VR2_A, TIC0, ATS7se, MATS4i, **Ti**	7.18	7.34
PDI[b]	ATSC0dv, VR2_A, V_p, AATS6p, Xp-1d	0.202	0.213
T_{el}[b]	**Ti**, piPC8, MATS2c, AATS7dv, AATS2d	0.198	0.221
I.I.[b]	**OEt**, MDEO-11, AATSC2d, AATSC6i, MATS5c	0.512	0.542

[a] Five features were selected for each target variable to minimize the mean absolute error (MAE) of leave-one-out cross-validation (LOOCV) in Huber regression. The feature selection was carried out three times, and the selected features and scores of the best model are summarized.

[b] Incorporating experimentally obtained catalyst compositions and structures into feature selection, the best model among those including at least one experimental feature is presented. Note that the inclusion of experimental features can result in a decrease in the regression scores.

While the selected features can accurately describe the relationship between the molecular structure of IDs and catalyst performance, they are very intricate, primarily representing the chemical structure of the molecules with a convolution of properties such as the charge and electron count of each atom. Consequently, extracting physical insights directly from the selected features is challenging, indeed a practice that is rarely performed in the literature. Thus, the approach of selecting relevant features from all possible features can provide a high-precision model without relying on specific assumptions, but may sacrifice interpretability of the descriptors.

The lower section of Figure 3 and Table 2 presents the optimal models, out of those including at least one experimental feature, for different target variables. As seen in Table 2, regardless of the target variables, incorporating experimental features as descriptors *worsened* the predictive power of the models, leading to higher MAE_{LOOCV} values. This fact indicates that catalyst chemical composition and pore volume (both known to significantly impact the performance of Ziegler-Natta catalysts) could be advantageously substituted by ID molecular features.

According to the SHAP analysis, the experimental features playing a significant role were the Ti content for T_{el} and OEt content for *I.I.* (Figure S3). The observed directions of the impact, i.e., higher T_{el} with lower Ti contents and lower *I.I.* with higher OEt contents, align with previous findings: a decrease in Ti content improves the effectiveness of donors in enhancing catalyst stereoselectivity, likely by making individual active sites isolated from each other and allowing neighboring surface sites occupied by donor molecules as stereo-controlling "pseudo-ligands"; residual Ti-OEt groups, in turn, seem to greatly increase the production of atactic PP fractions.

In summary, this ML-aided study yielded promising results in (at least) three aspects. Firstly, it represents the first example of utilizing machine to successfully learn the intricate relationship between ID molecular structures and performance of the resulting ZN PP catalysts. The machine was able to accurately describe catalyst performance based on the molecular structure information on IDs, without explicitly describing the complex interplay due to the availability of a reasonably sized and consistently acquired dataset. Secondly, it was evident that high-precision regression is achievable solely using ID molecular features without the necessity of experimentally acquired features. This finding might pave the way for virtual ID screenings without the need for any experimentation. Combining molecular features with experimental features allowed to promote our understanding on the catalysis without significantly compromising the accuracy of regression. Last but not least, the use of a relatively limited-sized dataset made feature selection not-definitive. Models involving different feature combinations did not significantly underperform the best-performing models, indicating their similarity in capturing the properties of known molecular structures. However, their predictive behaviors could be very different for dissimilar molecules. Eliminating alternative models or "alternative

hypotheses" requires active learning—recursively incorporating experimental data into the dataset for model reinforcement, hence our future research focus.

CONCLUSION

Artificial intelligence methodologies such as machine learning (ML) are becoming fundamental tools in all fields of science. An important prerequisite is the availability of very large and robust datasets; this is much easier in certain fields (e.g., pharma) than in others, like classical organometallic chemistry in general, and more specifically in olefin polymerization catalysis.

In this paper we introduced an integrated high-throughput experimentation (HTE) workflow for the rapid generation of reliable quantitative structure-properties relation (QSPR) databases in Ziegler-Natta polypropylene (ZN PP) catalysis, suitable for ML applications. Compared with previous studies, the main aspect of novelty here was the integration of parallel precatalyst synthesis and catalyst polymerization screening tools. In particular, a custom-made precatalyst preparation platform enabled us to prepare ZN precatalysts containing different internal donors (IDs) under otherwise identical experimental conditions, leading to catalyst particles with similar well-controlled morphology. This led to the implementation of a highly consistent and accurate database.

In order to highlight the potential of the proposed approach, a representative training set of 35 catalysts featuring different IDs was exploited. Application of state-of-the-art "black-box" ML feature selection allowed us to build QSPRs very decently interpolating catalyst polymerization performance. As noted in the introduction, the dataset used in the present work was rather small and therefore not meant for predictive application.

All in all, we believe that the present work represents a first step towards a more efficient approach to improved IDs for ZN PP catalysts. We plan to extend the training set, both in terms of number and structural diversity, in order to strengthen further the "black-box" ML tool, and possibly try "clear-box" ML variants, ultimately aiming to predict the behavior of these elusive catalyst formulations.

ASSOCIATED CONTENT

Supporting Information

Catalyst particle morphologies, representative one-to-one correlations, and SHAP summary plots are shown in Figures S1–S3.

AUTHOR INFORMATION

Corresponding Author

*T. Taniike. Email: taniike@jaist.ac.jp, *G. Antinucci. Email: giuseppe.antinucci@unina.it

Author Contributions

The manuscript was written with contributions from all authors. #T.T. and F.D.C. equally contributed as first authors. All authors have reviewed and approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work forms part of the research programme of DPI, project #848. The machine learning part by T.T. was financially supported by the Japan Science and Technology Agency (JST) Mirai Program (Grant Number JPMJMI22G4).

ABBREVIATIONS

ZN, Ziegler-Natta; PP, polypropylene; ID, internal donor; ED, external donor; EB, ethylbenzoate; QSPR, quantitative structure-properties relation; ML, machine learning; HTE, high-throughput experimentation; NLDFT, non-local density functional theory; DoE, design of experiment; GPC, gel permeation chromatography; BHT, 4-methyl-2,6-di-*tert*-butylphenol; PDI, polydispersity index; A-CEF, analytical crystallization elution fractionation; AF, amorphous fraction; *I.I.*, index of isotacticity; GA, genetic algorithm; MAE, mean absolute error; LOOCV, leave-one-out cross-validation; HPC, high-performance computing; SD, standard deviation; SHAP, Shapley additive explanation.

REFERENCES

(1) Busico, V. Giulio Natta and the Development of Stereoselective Propene Polymerization. In: Kaminsky W. (Eds) Polyolefins: 50 Years after Ziegler and Natta I. In *Advances in Polymer Science vol. 257*; Springer: Berlin, Heidelberg, **2013**.

(2) Antinucci, G.; Cipullo, R.; Busico, V. Imagine Polypropylene. *Nat. Catal.* **2023**, *6*, 456–457.

(3) Pasquini, N. *Polypropylene Handbook, 2nd Ed.*; Hanser Publishers: Munich, Germany, **2005**.

(4) Cecchin, G.; Morini, G.; Piemontesi, F.; Seidel, A. *Kirk-Othmer Encyclopedia of Chemical Technology*; Wiley-Interscience: New York, **2007**; p vol. 26.

(5) Giannini, U.; Cassata, A.; Longi, P.; Mazzoch, R. Procédé Pour La Polymérisation Stéréorégulière Des Alpha-Oléfines. BE 785332 and BE 785334, **1972**.

(6) Luciani, L.; Kashiwa, N.; Barbè, P. C.; Toyota, A. Katalysatoren Zur Polymerisation von Alpha-Olefinen. DE 2643143, **1977**.

(7) Luciani, L.; Kashiwa, N.; Barbè, C; Toyota, A. α-Olefin Polymerization Catalyst. JP 151691, **1977**.

(8) Parodi, S.; Nocci, R.; Giannini, U.; Barbè, P. C.; Scata, U. Components and Catalysts for the Polymerization of Olefins. EP 45977, **1982**.

(9) Albizzati, E.; Barbè, P. C.; Noristi, L.; Scordamaglia, R.; Barino, L.; Giannini, U.; Morini, G. Components and Catalysts for the Polymerization of Olefins. EP 361494 A1, **1990**.

(10) Morini, G.; Balbontin, G.; Gulevich, Y. V.; Kelder, R. T.; Duijghuisen, H. P. B.; Klusener, P. A. A.; Kondorffer, F. M. Components and Catalysts for the Polymerization of Olefins. EP 1088009 A1, **2001**.

(11) Andoni, A.; Chadwick, J. C.; Niemantsverdriet, H. J. W.; Thüne, P. C. The Role of Electron Donors on Lateral Surfaces of $MgCl_2$-Supported Ziegler–Natta Catalysts: Observation by AFM and SEM. *J. Catal.* **2008**, *257*, 81–86.

(12) Andoni, A.; Chadwick, J. C.; Niemantsverdriet, H. (J. W); Thune, P. C. A Flat Model Approach to Ziegler-Natta Catalysts for Propylene Polymerization and a Preparation Method of Well-Defined Crystallites of $MgCl_2$-Supported Catalysts. *Macromol. Symp.* **2007**, *260*, 140–146.

(13) Credendino, R.; Liguori, D.; Morini, G.; Cavallo, L. Investigating Phthalate and 1,3-Diether Coverage and Dynamics on the (104) and (110) Surfaces of $MgCl_2$-Supported Ziegler–Natta Catalysts. *J. Phys. Chem. C* **2014**, *118*, 8050–8058.

(14) Da Silveira, J.; Chikuma, H.; Takasao, G.; Wada, T.; Chammingkwan, P.; Taniike, T. Deciphering the Role of Internal Donors in Shaping Heterogeneous Ziegler-Natta Catalysts Based on Non-Empirical Structural Determination. *ACS Catal.* **2024**, in Press.

(15) Busico, V.; Causà, M.; Cipullo, R.; Credendino, R.; Cutillo, F.; Friederichs, N.; Lamanna, R.; Segre, A.; Van Axel Castelli, V. Periodic DFT and High-Resolution Magic-Angle-Spinning (HR-MAS) ^1H NMR Investigation of the Active Surfaces of $MgCl_2$-Supported Ziegler−Natta Catalysts. The $MgCl_2$ Matrix. *J. Phys. Chem. C* **2008**, *112*, 1081–1089.

(16) Credendino, R.; Pater, J. T. M.; Correa, A.; Morini, G.; Cavallo, L. Thermodynamics of Formation of Uncovered and Dimethyl Ether-Covered $MgCl_2$ Crystallites. Consequences in the Structure of Ziegler–Natta Heterogeneous Catalysts. *J. Phys. Chem. C* **2011**, *115*, 13322–13328.

(17) Capone, F.; Rongo, L.; D'Amore, M.; Budzelaar, P. H. M.; Busico, V. Periodic Hybrid DFT Approach (Including Dispersion) to $MgCl_2$-Supported Ziegler–Natta Catalysts. 2. Model Electron Donor Adsorption on $MgCl_2$ Crystal Surfaces. *J. Phys. Chem. C* **2013**, *117*, 24345–24353.

(18) Kuklin, M. S.; Bazhenov, A. S.; Denifl, P.; Leinonen, T.; Linnolahti, M.; Pakkanen, T. A. Stabilization of Magnesium Dichloride Surface Defects by Mono- and Bidentate Donors. *Surf. Sci.* **2015**, *635*, 5–10.

(19) Breuza, E.; Antinucci, G.; Budzelaar, P. H. M.; Busico, V.; Correa, A.; Ehm, C. $MgCl_2$-Supported Ziegler-Natta Catalysts: A DFT-D "Flexible-Cluster" Approach to Internal Donor Adducts. *J. Phys. Chem. C* **2018**, *122*, 9046–9053.

(20) Taniike, T.; Terano, M. Coadsorption Model for First-Principle Description of Roles of Donors in Heterogeneous Ziegler–Natta Propylene Polymerization. *J. Catal.* **2012**, *293*, 39–50.

(21) Correa, A.; Piemontesi, F.; Morini, G.; Cavallo, L. Key Elements in the Structure and Function Relationship of the $MgCl_2$/$TiCl_4$/Lewis Base Ziegler−Natta Catalytic System. *Macromolecules* **2007**, *40*, 9181–9189.

(22) Antinucci, G.; Vittoria, A.; Cipullo, R.; Busico, V. Regioirregular Monomeric Units in Ziegler–Natta Polypropylene: A Sensitive Probe of the Catalytic Sites. *Macromolecules* **2020**, *53*, 3789–3795.

(23) Vittoria, A.; Meppelder, A.; Friederichs, N.; Busico, V.; Cipullo, R. Demystifying Ziegler-Natta Catalysts: The Origin of Stereoselectivity. *ACS Catal.* **2017**, *7*, 4509–4518.

(24) Busico, V.; Cipullo, R. Microstructure of Polypropylene. *Prog. Polym. Sci.* **2001**, *26*, 443–533.

(25) Sacchi, M. C.; Tritto, I.; Shan, C.; Mendichi, R.; Noristi, L. Role of the Pair of Internal and External Donors in $MgCl_2$-Supported Ziegler-Natta Catalysts. *Macromolecules* **1991**,

24, 6823–6826.

(26) Vittoria, A.; Antinucci, G.; Zaccaria, F.; Cipullo, R.; Busico, V. Monitoring the Kinetics of Internal Donor Clean-up from Ziegler–Natta Catalytic Surfaces: An Integrated Experimental and Computational Study. *J. Phys. Chem. C* **2020**, *124*, 14245–14252.

(27) Busico, V.; Corradini, P.; De Martino, L.; Proto, A.; Savino, V.; Albizzati, E. Polymerization of Propene in the Presence of $MgCl_2$-Supported Ziegler-Natta Catalysts, 1. The Role of Ethyl Benzoate as "Internal" and "External" Base. *Die Makromol. Chemie* **1985**, *186*, 1279–1288.

(28) Busico, V.; Corradini, P.; De Martino, L.; Proto, A.; Albizzati, E. Polymerization of Propene in the Presence of $MgCl_2$-Supported Ziegler-Natta Catalysts, 2. Effects of the Co-Catalyst Composition. *Die Makromol. Chemie* **1986**, *187*, 1115–1124.

(29) Noristi, L.; Barbè, P. C.; Baruzzi, G. Effect of the Internal/External Donor Pair in High-Yield Catalysts for Propylene Polymerization, 1. Catalyst-Cocatalyst Interactions. *Die Makromol. Chemie* **1991**, *192*, 1115–1127.

(30) Thakur, A.; Wada, T.; Chammingkwan, P.; Terano, M.; Taniike, T. Development of Large-Scale Stopped-Flow Technique and its Application in Elucidation of Initial Ziegler-Natta Olefin Polymerization Kinetics. *Polymers* **2019**, *11,* 1012.

(31) Piovano, A.; Wada, T.; Amodio, A.; Takasao, G.; Ikeda, T.; Zhu, D.; Terano, M.; Chammingkwan, P.; Groppo, E.; Taniike, T. Formation of Highly Active Ziegler–Natta Catalysts Clarified by a Multifaceted Characterization Approach. *ACS Catal.* **2021**, *11*, 13782–13796.

(32) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent Advances and Applications of Machine Learning in Solid-State Materials Science. *npj Comput. Mater.* **2019**, *5*, 83.

(33) Ling, C. A Review of the Recent Progress in Battery Informatics. *npj Comput. Mater.* **2022**, *8*, 33.

(34) Dar, Y. L. High-Throughput Experimentation: A Powerful Enabling Technology for the Chemicals and Materials Industry. *Macromol. Rapid Commun.* **2004**, *25*, 34–47.

(35) Boussie, T. R.; Diamond, G. M.; Goh, C.; Hall, K. A.; LaPointe, A. M.; Leclerc, M.; Lund, C.; Murphy, V.; Shoemaker, J. A. W.; Tracht, U.; Turner, H.; Zhang, J.; Uno, T.; Rosen, R. K.; Stevens, J. C. A Fully Integrated High-Throughput Screening Methodology for the Discovery of New Polyolefin Catalysts: Discovery of a New Class of High Temperature Single-Site Group (IV) Copolymerization Catalysts. *J. Am. Chem. Soc.* **2003**, *125*, 4306–4317.

(36) Hagemeyer, A., Strasser, P., Volpe, A. F. *High-Throughput Screening in Chemical Catalysis*; Wiley-VCH: Weinheim, Germany, **2004**.

(37) Peil, K. P.; Neithamer, D. R.; Patrick, D. W.; Wilson, B. E.; Tucker, C. J. Applications of High Throughput Research at The Dow Chemical Company. *Macromol. Rapid Commun.* **2004**, *25*, 119–126.

(38) Cannavacciuolo, F. D.; Yadav, R.; Esper, A.; Vittoria, A.; Antinucci, G.; Zaccaria, F.; Cipullo, R.; Budzelaar, P. H. M.; Busico, V.; Goryunov, G. P.; Uborsky, D. V.; Voskoboynikov, A. Z.; Searles, K.; Ehm, C.; Veige A. S. A High-Throughput Approach

to Repurposing Olefin Polymerization Catalysts for Polymer Upcycling. *Angew. Chem. Int. Ed.* **2022**, *61*, e202202258.

(39) Chammingkwan, P.; Khoshsefat, M.; Terano, M.; Taniike, T. Parallel Catalyst Synthesis Protocol for Accelerating Heterogeneous Olefin Polymerization Research. *Polymers* **2023**, *15*.

(40) Busico, V.; Pellecchia, R.; Cutillo, F.; Cipullo, R. High-Throughput Screening in Olefin-polymerization Catalysis: From Serendipitous Discovery towards Rational Understanding. *Macromol. Rapid Commun.* **2009**, *30*, 1697–1708.

(41) Vittoria, A.; Meppelder, A.; Friederichs, N.; Busico, V.; Cipullo, R. Ziegler–Natta Catalysts: Regioselectivity and "Hydrogen Response". *ACS Catal.* **2020**, *10*, 644–651.

(42) Vittoria, A.; Mingione, A.; Abbate, R. A.; Cipullo, R.; Busico, V. High Throughput Experimentation Protocol for Quantitative Measurements of Regioselectivity in Ziegler–Natta Polypropylene Catalysis. *Ind. Eng. Chem. Res.* **2019**, *58*, 14729–14735.

(43) Ehm, C.; Vittoria, A.; Goryunov, G. P.; Kulyabin, P. S.; Budzelaar, P. H. M.; Voskoboynikov, A. Z.; Busico, V.; Uborsky, D. V.; Cipullo, R. Connection of Stereoselectivity, Regioselectivity, and Molecular Weight Capability in *Rac*-R′$_2$Si(2-Me-4-R-Indenyl)$_2$ZrCl$_2$ Type Catalysts. *Macromolecules* **2018**, *51*, 8073–8083.

(44) Ehm, C.; Vittoria, A.; Goryunov, G. P.; Izmer, V. V; Kononovich, D. S.; Samsonov, O. V; Di Girolamo, R.; Budzelaar, P. H. M.; Voskoboynikov, A. Z.; Busico, V.; Uborsky V. D.; Cipullo, R. An Integrated High Throughput Experimentation/Predictive QSAR Modeling Approach to *Ansa*-Zirconocene Catalysts for Isotactic Polypropylene. *Polymers* **2020**, *12*,

1005.

(45) Ehm, C.; Vittoria, A.; Goryunov, G. P.; Izmer, V. V.; Kononovich, D. S.; Samsonov, O. V.; Budzelaar, P. H. M.; Voskoboynikov, A. Z.; Busico, V.; Uborsky, D. V.; Cipullo, R. On the Limits of Tuning Comonomer Affinity of "Spaleck-Type": *Ansa* -Zirconocenes in Ethene/1-Hexene Copolymerization: A High-Throughput Experimentation/QSAR Approach. *Dalt. Trans.* **2020**, *49*, 10162–10172.

(46) Cannavacciuolo, F. D.; Vittoria, A.; Ehm, C.; Cipullo, R.; Busico, V. Polyolefin Chain Shuttling at Ansa-Metallocene Catalysts: Legend and Reality. *Eur. Polym. J.* **2021**, *150*, 110396.

(47) Busico, V.; Cipullo, R.; Mingione, A.; Rongo, L. Accelerating the Research Approach to Ziegler–Natta Catalysts. *Ind. Eng. Chem. Res.* **2016**, *55*, 2686–2695.

(48) Antinucci, G.; Dereli, B.; Vittoria, A.; Budzelaar, P. H. M.; Cipullo, R.; Goryunov, G. P.; Kulyabin, P. S.; Uborsky, D. V; Cavallo, L.; Ehm, C.; Voskoboynikov, A. Z.; Busico, V. Selection of Low-Dimensional 3-D Geometric Descriptors for Accurate Enantioselectivity Prediction. *ACS Catal.* **2022**, *12*, 6934–6945.

(49) Ehm, C.; Mingione, A.; Vittoria, A.; Zaccaria, F.; Cipullo, R.; Busico, V. High-Throughput Experimentation in Olefin Polymerization Catalysis: Facing the Challenges of Miniaturization. *Ind. Eng. Chem. Res.* **2020**, *59*, 13940–13947.

(50) Vittoria, A.; Urciuoli, G.; Costanzo, S.; Tammaro, D.; Cannavacciuolo, F. D.; Pasquino, R.; Cipullo, R.; Auriemma, F.; Grizzuti, N.; Maffettone, P. L.; Busico, V. Extending the High-Throughput Experimentation (HTE) Approach to Catalytic Olefin Polymerizations:

From Catalysts to Materials. *Macromolecules* **2022**, *55*, 5017–5026.

(51) Antinucci, G.; Pucciarelli, A.; Vittoria, A.; Zaccaria, F.; Urciuoli, G.; Ehm, C.; Cannavacciuolo, F. D.; Cipullo, R.; Busico, V. Fast Analytics of High-Impact Polypropylene (HIPP). *ACS Appl. Polym. Mater.* **2023**, *5*, 3894–3897.

(52) Wada, T.; Funako, T.; Chammingkwan, P.; Thakur, A.; Matta, A.; Terano, M.; Taniike, T. Structure-Performance Relationship of $Mg(OEt)_2$-Based Ziegler-Natta Catalysts. *J. Catal.* **2020**, *389*, 525–532.

(53) Chammingkwan, P.; Wannaborworn, M.; Mai, L. T. T.; Terano, M.; Taniike, T.; Phiriyawirut, P. Particle Engineering of Magnesium Ethoxide-Based Ziegler-Natta Catalyst through Post-Modification of Magnesium Ethoxide. *Appl. Catal. A Gen.* **2021**, *626*, 118337.

(54) Thakur, A.; Chammingkwan, P.; Wada, T.; Onishi, R.; Kamimura, W.; Seenivasan, K.; Terano, M.; Taniike, T. Solution-State NMR Study of Organic Components of Industrial Ziegler-Natta Catalysts: Effect of by-Products on Catalyst Performance. *Appl. Catal. A Gen.* **2021**, *611*, 117971.

(55) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *J. Cheminform.* **2018**, *10*, 4.

(56) Huber, P. J. Robust Estimation of a Location Parameter. In *Breakthroughs in Statistics: Methodology and Distribution*; Kotz, S., Johnson, N. L., Eds.; Springer New York: New York, NY, 1992; pp 492–518.

(57) Taniike, T.; Fujiwara, A.; Nakanowatari, S.; García-Escobar, F.; Takahashi, K. Automatic

Feature Engineering for Catalyst Design Using Small Data without Prior Knowledge of Target Catalysis. *Commun. Chem.* **2024**, *7*, 11.

(58) Taniike, T.; Terano, M. The Use of Donors to Increase the Isotacticity of Polypropylene. In *Polyolefins: 50 years after Ziegler and Natta I: Polyethylene and Polypropylene*; Kaminsky, W., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, **2013**; pp 81–97.

(59) Terano, T.; Murai, A.; Inoue, M.; Miyoshi, K. Solid Catalyst Component for the Polymerization of Olefins and an Olefin Polymerization Catalyst. US 4816433 A, **1987**.

(60) Taniike, T.; Funako, T.; Terano, M. Multilateral Characterization for Industrial Ziegler–Natta Catalysts toward Elucidation of Structure–Performance Relationship. *J. Catal.* **2014**, *311*, 33–40.

(61) Potapov, A. G.; Bukatov, G. D.; Zakharov, V. A. DRIFT Study of Internal Donors in Supported Ziegler–Natta Catalysts. *J. Mol. Catal. A Chem.* **2006**, *246*, 248–254.

(62) Corradini, P.; Barone, V.; Fusco, R.; Guerra, G. A Possible Model of Catalytic Sites for the Stereospecific Polymerization of Alpha-Olefins on 1st-Generation and Supported Ziegler-Natta Catalysts. *Gazz. Chim. Ital.* **1983**, *113*, 601–607.

(63) Lundberg, S. M.; Lee, S.-I. Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017); Long Beach, CA, USA, **2017**.