Supporting Information for

Decoding Polyethylene Formation in Cr/PNP Catalyzed Ethylene Oligomerization via Experimentally Guided Machine Learning

Youcai Zhu, Yue Mu, Xiaoke Shi, Long Chen, Shu Yang, Li Sun, Zhen Liu*

School of Chemical Engineering, East China University of Science and Technology, Shanghai 200237, China.

*Corresponding author:

Zhen Liu, Email: liuzhen@ecust.edu.cn

Table of Contents

1. General Information

1.1 Ligand database

All PNP ligands included in this study were collected from experimental reports on chromium-catalyzed ethylene oligomerization (Figure S1), with the corresponding polyethylene (PE) formation data extracted directly from the literature. The ligand set spans four major classes defined by the substituents on phosphorus, namely P-N ligands with nitrogen substituents (https://doi.org/10.1002/aoc.6454, https://doi.org/10.1039/C6DT01109H), P-C_Ole ligands with olefinic carbon substituents (https://doi.org/10.1039/d4dt01521e), P-C_Al ligands with aliphatic carbon substituents (https://doi.org/10.1016/j.apcata.2018.04.030), and P-C_Ar ligands with aryl carbon substituents (https://doi.org/10.1016/j.jcat.2006.10.020, https://doi.org/10.1016/j.molcata.2007.01.046).

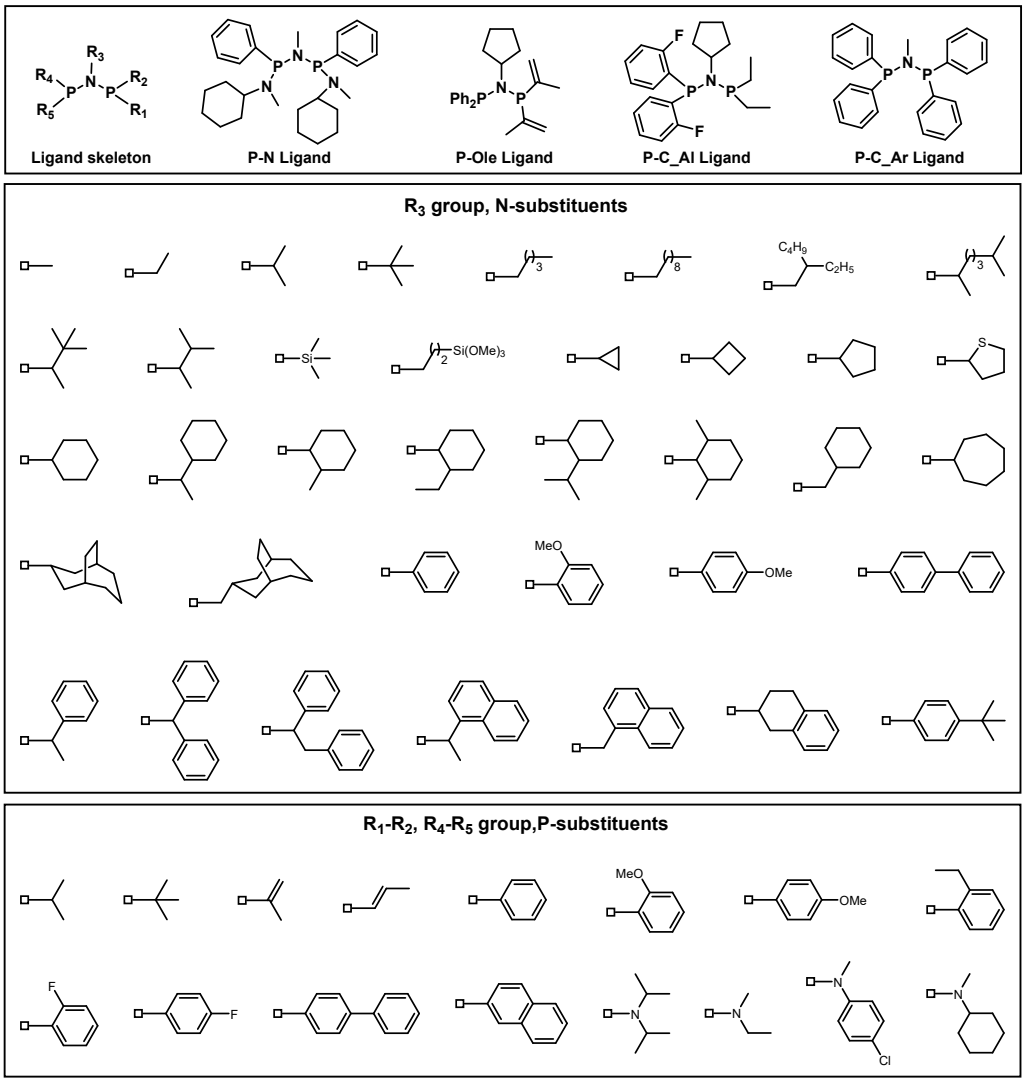


Figure S1. The PNP ligand skeletons and the corresponding substituent libraries used in this work.

Table S1. The substituents in PNP ligands (SMILES)

Ligand	R1	R2	R3	R4	R5
L100001	H	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100002	C	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100003	CC=C	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100004	CCCn1ccocc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100005	CCC[Si](OC)(OC)OC	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100006	CCCCC	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100007	CCCCCCCCCC	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100008	Cc1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100009	CC(CC)CCCC	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100010	C(C)C	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100011	C(C)CCCC(C)C	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100012	CC	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100013	C1CCCCC1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100014	CC1C2CC3CC1CC(C2)C3	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100015	C(C)C(C)C	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100016	C(C)C1CCCCC1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100017	C(C)(C)C	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100018	C1C2CC3CC1CC(C2)C3	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100019	[C@H](C(C)(C)C)(C)	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100020	[C@@H](C(C)(C)C)(C)	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100021	[Si](C)(C)C	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100022	C	c1c(OC)cccc1	c1c(OC)cccc1	c1c(OC)ccccc1	c1c(OC)ccccc1
L100023	C	c1cc(OC)ccc1	c1cc(OC)ccc1	c1cc(OC)cccc1	c1cc(OC)ccc1
L100024	C	c1ccc(OC)cc1	c1ccc(OC)cc1	c1ccc(OC)ccc1	c1ccc(OC)ccc1
L100025	C(C)C	c1ccc(OC)cc1	c1ccc(OC)cc1	c1ccc(OC)ccc1	c1ccc(OC)ccc1
L100026	C(C)C	c1c(OC)ccc	c1ccccc1	c1c(OC)cccc	c1ccccc1

		c1		1	
L100027	C(C)C	c1c(OC)ccc c1	c1ccccc1	c1ccccc1	c1ccccc1
L100028	C(C)C	c1c(CC)cccc 1	c1ccccc1	c1ccccc1	c1ccccc1
L100029	C1CCCC1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100030	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100031	C	c1cccc2cccc c12	c1cccc2cccc c12	c1cccc2cccc c12	c1cccc2cc ccc12
L100032	C	c1ccc(-c2ccccc2)cc 1	c1ccc(-c2ccccc2)cc 1	c1ccc(-c2ccccc2)cc1	c1ccc(-c2ccccc2)c c1
L100033	C(C)C	c1cccs1	c1cccs1	c1cccs1	c1cccs1
L100034	C	CC	CC	CC	CC
L100035	c1c(C)cccc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100036	c1c(CC)cccc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100037	c1c(C(C)C)cccc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100038	c1c(C)cccc1(C)	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100039	C1CCCC1	c1ccccc1	c1ccccc1	N(C)C	N(C)C
L100040	C(C)c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100041	C(c1ccccc1)c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100042	C(Cc1ccccc1)c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100043	c1cccc2ccccc12	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100044	Cc1cccc2ccccc12	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100045	C(C)c1cccc2ccccc12	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100046	C1CCCc2ccccc12	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100047	c1ccc(C(C)(C)C)cc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100048	c1ccc(N(=O)=O)cc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100049	c1ccc(OC)cc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100050	C1CC1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100051	C1CCC1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100052	C1CCCCCC1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1

L100053	C1CCCCCCCCCCCC1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100054	C1CCCCC1(C)	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100055	C1CCCCC1(CC)	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100056	C1CCCCC1(C(C)C)	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100057	C1C(C)CCCC1(C)	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100058	C1CCCC1	c1ccccc1	c1ccccc1	N(CC)CC	N(CC)CC
L100059	CC1CCCCC1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100060	C1CCCC1	c1ccccc1	c1ccccc1	N(C(C)C)C(C)C	N(C(C)C)C(C)C
L100061	C1C(C)C2C(C)(C)C(C2)C1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100062	C1CC2CCC1(C2)	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100063	C1CCCC1	c1ccccc1	c1ccccc1	c1ccccc1	CC
L100064	C1CCCC1	c1ccccc1	c1ccccc1	c1ccccc1	C(C)C
L100065	C1CCCC1	c1ccccc1	c1ccccc1	c1ccccc1	C(C)(C)C
L100066	CCC[Si]12O[Si]3(c4ccccc4)O[Si]4(c5ccccc5)O[Si](c5ccccc5)(O1)O[Si]1(c5ccccc5)O[Si](c5ccccc5)(O2)O[Si](c2ccccc2)(O3)O[Si](c2ccccc2)(O4)O1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccc(C)cc1
L100067	CCC[Si]12O[Si]3(c4ccccc4)O[Si]4(c5ccccc5)O[Si](c5ccccc5)(O1)O[Si]1(c5ccccc5)O[Si](c5ccccc5)(O2)O[Si](c2ccccc2)(O3)O[Si](c2ccccc2)(O4)O1	c1ccccc1	c1ccccc1	c1ccccc1	c1c(OC)ccccc1
L100068	C1CCCC1	c1ccccc1	c1ccccc1	C(C)(C)C	C(C)(C)C
L100069	CCC[Si]12O[Si]3(c4ccccc4)O[Si]4(c5ccccc5)O[Si](c5ccccc5)(O1)O[Si]1(c5ccccc5)O[Si](c5ccccc5)(O2)O[Si](c2ccccc2)(O3)O[Si](c2ccccc2)(O4)O1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccc(OC)cc1
L100070	C1CCCC1	c1ccccc1	c1ccccc1	C(C)C	C(C)C
L100071	CCC[Si]12O[Si]3(c4ccccc4)O[Si]4(c	c1ccccc1	c1ccccc1	c1ccccc1	c1cc(F)ccc

	5ccccc5)O[Si](c5ccccc5)(O1)O[Si]1(c5ccccc5)O[Si](c5ccccc5)(O2)O[Si](c2ccccc2)(O3)O[Si](c2ccccc2)(O4)O1				1
L100072	C1CCCC1	c1ccccc1	c1ccccc1	C	C
L100073	C1CCCC1	c1ccccc1	C(C)C	c1ccccc1	C(C)C
L100074	CCC[Si]12O[Si]3(c4ccccc4)O[Si]4(c5ccccc5)O[Si](c5ccccc5)(O1)O[Si]1(c5ccccc5)O[Si](c5ccccc5)(O2)O[Si](c2ccccc2)(O3)O[Si](c2ccccc2)(O4)O1	c1ccccc1	c1ccccc1	c1ccccc1	c1cc(C(F)(F)F)ccc1
L100075	CCC[Si]12O[Si]3(c4ccccc4)O[Si]4(c5ccccc5)O[Si](c5ccccc5)(O1)O[Si]1(c5ccccc5)O[Si](c5ccccc5)(O2)O[Si](c2ccccc2)(O3)O[Si](c2ccccc2)(O4)O1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccc(C(F)(F)F)cc1
L100076	C1CCCC1	c1ccccc1	CC	c1ccccc1	CC
L100077	C1CCCC1	c1c(F)cccc1	c1c(F)cccc1	c1ccccc1	c1ccccc1
L100078	C1CCCC1	c1c(F)cccc1	c1c(F)cccc1	CC	CC
L100079	C1CCCC1	c1c(F)cccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100080	C1CCCC1	c1c(F)cccc1	c1ccccc1	CC	CC
L100081	C1CCCC1	c1ccc(F)cc1	c1ccc(F)cc1	c1ccccc1	c1ccccc1
L100082	C1CCCC1	c1ccccc1	c1ccccc1	CC	CC
L100083	C(C)C	c1ccccc1	c1ccccc1	C(=C)C	C(=C)C
L100084	C1CCCC1	c1ccccc1	c1ccccc1	C(=C)C	C(=C)C
L100085	C1CCCCC1	c1ccccc1	c1ccccc1	C(=C)C	C(=C)C
L100086	C(C)C1CCCCC1	c1ccccc1	c1ccccc1	C(=C)C	C(=C)C
L100087	C(C)C	c1ccccc1	c1ccccc1	c1ccccc1	C(=C)C
L100088	C(C)C	c1ccccc1	c1ccccc1	CC	C(=C)C
L100089	C(C)C	c1ccccc1	c1ccccc1	C(=C)C	C=CC
L100090	C	c1ccccc1	N(CC)C	c1ccccc1	N(CC)C
L100091	C	c1ccccc1	N(C1CCCCC	c1ccccc1	N(C1CCCC

			1)C		C1)C
L100092	C(C)C	c1ccccc1	N(C1CCCCC1)C	c1ccccc1	N(C1CCCCC1)C
L100093	C1CCCCC1	c1ccccc1	N(C1CCCCC1)C	c1ccccc1	N(C1CCCCC1)C
L100094	C	c1ccccc1	NC(C)(C)C	c1ccccc1	NC(C)(C)C
L100095	C	c1ccccc1	Nc1c(OC)cccc1	c1ccccc1	Nc1c(OC)cccc1
L100096	C	c1ccccc1	N(C)c1c(OC)cccc1	c1ccccc1	N(C)c1c(OC)cccc1
L100097	C	c1ccccc1	N(C)c1ccc(F)cc1	c1ccccc1	N(C)c1ccc(F)cc1
L100098	C	c1ccccc1	Nc1c(Cl)cccc1	c1ccccc1	Nc1c(Cl)cccc1
L100099	C1CCCCC1	c1ccccc1	N(C(C)C)C	c1ccccc1	N(C(C)C)C
L100100	C1CCCCC1	c1ccccc1	N(C(C)C)	c1ccccc1	N(C(C)C)
L100101	c1c(C)c(C)ccc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100102	c1cc(C(F)(F)F)ccc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100103	C(C)C	c1ccccc1	c1ccccc1	c1ccccc1	NC(C)C
L100104	C1CCCCC1	c1ccccc1	c1ccccc1	c1ccccc1	NC1CCCCC1
L100105	C(C)C	c1ccccc1	c1ccccc1	c1ccccc1	NC1CCCCC1
L100106	C1CCCCC1	c1ccccc1	c1ccccc1	c1ccccc1	NC(C)C
L100107	C(C)C	c1ccccc1	c1ccccc1	c1ccccc1	NCCC[Si](OCC)(OCC)OCC
L100108	CCCCCC	c1ccccc1	c1ccccc1	c1ccccc1	NC(C)C
L100109	C(C)C	c1ccccc1	c1ccccc1	c1ccccc1	NCC
L100110	C(C)C	c1ccccc1	c1ccccc1	C	NC(C)C
L100111	C(C)C	C1CCCC1	C1CCCC1	c1ccccc1	NC(C)C
L100112	C(C)C	c1ccccc1	c1ccccc1	c1ccccc1	NC

L100113	CCC[Si]12O[Si]3(c4ccccc4)O[Si]4(c5ccccc5)O[Si](c5ccccc5)(O1)O[Si]1(c5ccccc5)O[Si](c5ccccc5)(O2)O[Si](c2ccccc2)(O3)O[Si](c2ccccc2)(O4)O1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100114	CCCC	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100115	CCC[Si]12O[Si]3(c4ccccc4)O[Si]4(c5ccccc5)O[Si](c5ccccc5)(O1)O[Si]1(c5ccccc5)O[Si](c5ccccc5)(O2)O[Si](c2ccccc2)(O3)O[Si](c2ccccc2)(O4)O1	c1ccccc1(F)	c1ccccc1	c1ccccc1	c1ccccc1
L100116	CCC[Si]12O[Si]3(c4ccccc4)O[Si]4(c5ccccc5)O[Si](c5ccccc5)(O1)O[Si]1(c5ccccc5)O[Si](c5ccccc5)(O2)O[Si](c2ccccc2)(O3)O[Si](c2ccccc2)(O4)O1	c1ccccc1(OC(F)(F)F)	c1ccccc1	c1ccccc1	c1ccccc1
L100117	C(C)CC	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100118	C(CC)c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100119	C(CC)CC	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100120	C(C)CCCCCC	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100121	C(CCCCC)CCCCC	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1
L100122	CCC(CC)CC	c1ccccc1	c1ccccc1	c1ccccc1	c1ccccc1

Table S2. Experimental dataset used for training and testing the machine-learning models.[1-14]

Ligand	Cr precursor	Cr amount (μmol)	ligand amount (μmol)	Al/Cr molar ratio	solvent[a]	solvent volume (mL)	P (bar)	T (°C)	PE (wt%)	Ref.
L100001	$Cr(EH)_3$	3.3	10	300	toluene	100	30	65	2.1	[1]
L100002	$CrCl_3(THF)_3$	3.3	10	300	toluene	100	30	65	1.4	[1]
L100002-1	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	4.9	[1]
L100003	$Cr(EH)_3$	3.3	10	300	toluene	100	30	65	9.8	[1]
L100004	$Cr(AcAc)_3$	3.3	10	300	toluene	100	30	65	23.9	[1]
L100005	$Cr(AcAc)_3$	3.3	10	300	toluene	100	30	65	0.5	[1]
L100006	$CrCl_3(THF)_3$	3.3	10	300	toluene	100	30	65	0.3	[1]
L100007	$Cr(AcAc)_3$	3.3	10	300	toluene	100	30	65	0.7	[1]
L100008	$CrCl_3(THF)_3$	3.3	10	300	toluene	100	30	65	3.7	[1]
L100008-1	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	1.3	[1]
L100009	$Cr(AcAc)_3$	3.3	10	300	toluene	100	30	65	0.8	[1]
L100009-1	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	1.5	[1]
L100010	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	0.9	[1]
L100010-1	$Cr(AcAc)_3$	22	44	300	toluene	100	30	65	1	[2]
L100010-2	$Cr(AcAc)_3$	22	44	300	toluene	100	45	45	1	[2]
L100011	$Cr(AcAc)_3$	3.3	10	300	toluene	100	30	65	1.8	[1]
L100012	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	2.4	[1]
L100013	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	0.8	[1]
L100014	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	0.1	[1]
L100015	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	1.9	[1]
L100016	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	0.6	[1]
L100017	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	1.24	[1]
L100018	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	1.5	[1]
L100019	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	1.5	[1]
L100020	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	1.5	[1]
L100021	$Cr(AcAc)_3$	2.5	3	300	MCH	100	45	65	12.9	[1]
L100022	$Cr(AcAc)_3$	33	66	300	toluene	100	30	65	1	[2]
L100022-1	$Cr(AcAc)_3$	20	40	300	toluene	100	45	45	1	[2]
L100023	$Cr(AcAc)_3$	33	66	300	toluene	100	30	65	11	[2]

L100023-1	$Cr(AcAc)_3$	33	66	300	toluene	100	45	45	19	[2]
L100024	$Cr(AcAc)_3$	33	66	300	toluene	100	45	45	12	[2]
L100024-1	$Cr(AcAc)_3$	33	66	300	toluene	100	30	65	1	[2]
L100025	$Cr(AcAc)_3$	33	66	300	toluene	100	30	65	1	[2]
L100025-1	$Cr(AcAc)_3$	15	18	300	toluene	100	45	45	3	[2]
L100026	$Cr(AcAc)_3$	33	66	300	toluene	100	30	65	1	[2]
L100026-1	$Cr(AcAc)_3$	33	66	300	toluene	100	45	45	7	[2]
L100027	$Cr(AcAc)_3$	33	66	300	toluene	100	30	65	1	[2]
L100027-1	$Cr(AcAc)_3$	33	66	300	toluene	100	45	45	1	[2]
L100028	$Cr(AcAc)_3$	33	66	300	toluene	100	30	65	1	[2]
L100028-1	$Cr(AcAc)_3$	33	66	300	toluene	100	45	45	1	[2]
L100029	$CrCl_3(THF)_3$	33	66	300	toluene	100	30	65	0.3	[3]
L100030	$CrCl_3(THF)_3$	33	66	300	toluene	100	30	65	0.4	[3]
L100031	$CrCl_3(THF)_3$	33	66	300	toluene	100	30	65	1.3	[3]
L100032	$Cr(AcAc)_3$	33	66	300	toluene	100	30	65	1.8	[3]
L100033	$Cr(AcAc)_3$	33	66	300	toluene	100	45	45	2.4	[3]
L100034	$Cr(AcAc)_3$	33	66	300	toluene	100	45	45	13.6	[3]
L100035	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	4.8	[4]
L100036	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	3	[4]
L100037	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	7.8	[4]
L100038	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	3.8	[4]
L100039	$CrCl_3(THF)_3$	3	3	4000	MCH	20	10	45	0.15	[5]
L100039-1	$CrCl_3(THF)_3$	3	3	4000	MCH	20	10	30	0	[5]
L100039-2	$CrCl_3(THF)_3$	3	3	4000	MCH	20	10	45	0.15	[5]
L100039-3	$CrCl_3(THF)_3$	3	3	4000	MCH	20	10	60	0.08	[5]
L100039-4	$CrCl_3(THF)_3$	3	3	2000	MCH	20	10	45	0	[5]
L100039-5	$CrCl_3(THF)_3$	3	3	1000	MCH	20	10	45	0.53	[5]
L100039-6	$CrCl_3(THF)_3$	6	6	2000	MCH	20	10	45	0	[5]
L100039-7	$CrCl_3(THF)_3$	12	12	2000	MCH	20	10	45	0.11	[5]
L100039-8	$CrCl_3(THF)_3$	3	3	4000	MCH	20	50	30	2.81	[5]
L100040	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	0.6	[4]
L100041	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	25.8	[4]
L100042	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	1.7	[4]

L100043	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	3.4	[4]
L100044	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	1	[4]
L100045	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	0.8	[4]
L100046	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	0.8	[4]
L100047	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	1.9	[4]
L100048	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	16.2	[4]
L100049	$Cr(AcAc)_3$	10	10	480	MCH	100	50	60	4.6	[4]
L100050	$Cr(AcAc)_3$	5	7	270	MCH	100	45	60	1.8	[6]
L100051	$Cr(AcAc)_3$	5	7	270	MCH	100	45	60	1.3	[6]
L100052	$Cr(AcAc)_3$	5	7	270	MCH	100	45	60	1.2	[6]
L100053	$Cr(AcAc)_3$	5	7	270	MCH	100	45	60	0.9	[6]
L100054	$Cr(AcAc)_3$	2.5	7	540	cyclohexane	100	45	60	0.8	[6]
L100055	$Cr(AcAc)_3$	2.5	7	270	cyclohexane	100	45	60	0.3	[6]
L100056	$Cr(AcAc)_3$	2.5	7	270	cyclohexane	100	45	60	1.3	[6]
L100057	$Cr(AcAc)_3$	2.5	7	540	cyclohexane	100	45	60	0.3	[6]
L100058	$CrCl_3(THF)_3$	3	3	4000	MCH	20	10	45	0.1	[6]
L100058-1	$CrCl_3(THF)_3$	3	3	2000	MCH	20	10	30	0	[6]
L100058-2	$CrCl_3(THF)_3$	3	3	2000	MCH	20	10	45	0.15	[6]
L100058-3	$CrCl_3(THF)_3$	3	3	2000	MCH	20	10	60	0.09	[6]
L100058-4	$CrCl_3(THF)_3$	3	3	4000	MCH	20	10	45	0.06	[6]
L100058-5	$CrCl_3(THF)_3$	3	3	1000	MCH	20	10	45	0.1	[6]
L100058-6	$CrCl_3(THF)_3$	6	6	2000	MCH	20	10	45	0.04	[6]
L100058-7	$CrCl_3(THF)_3$	12	12	2000	MCH	20	10	45	0.14	[6]
L100058-8	$CrCl_3(THF)_3$	3	3	2000	MCH	20	50	45	2.57	[6]
L100059	Cr(AcAc)3	5	7.5	270	MCH	100	45	60	1.6	[6]
L100060	$CrCl_3(THF)_3$	3	3	4000	MCH	20	10	45	0.49	[5]
L100060-1	$CrCl_3(THF)_3$	3	3	4000	MCH	20	10	30	0.07	[5]
L100060-2	$CrCl_3(THF)_3$	3	3	4000	MCH	20	10	45	0.49	[5]
L100060-3	$CrCl_3(THF)_3$	3	3	4000	MCH	20	10	60	0.25	[5]

L100060-4	$CrCl_3(THF)_3$	3	3	2000	MCH	20	10	45	0.44	[5]
L100060-5	$CrCl_3(THF)_3$	3	3	1000	MCH	20	10	45	0.3	[5]
L100060-6	$CrCl_3(THF)_3$	6	6	4000	MCH	20	10	45	0.32	[5]
L100060-7	$CrCl_3(THF)_3$	12	12	4000	MCH	20	10	45	0.19	[5]
L100060-8	$CrCl_3(THF)_3$	3	3	4000	MCH	20	50	45	3.57	[5]
L100061	$Cr(AcAc)_3$	5	7.5	270	MCH	100	45	60	0.8	[6]
L100062	$Cr(AcAc)_3$	5	7.5	270	MCH	100	45	60	3	[6]
L100063	$CrCl_3(THF)_3$	4.8	4.8	500	MCH	20	10	50	0.5	[7]
L100064	$CrCl_3(THF)_3$	4.8	4.8	500	MCH	20	10	50	0.8	[7]
L100065	$CrCl_3(THF)_3$	4.8	4.8	500	MCH	20	10	50	1.9	[7]
L100066	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	0.6	[8]
L100067	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	0	[8]
L100068	$CrCl_3(THF)_3$	4.8	4.8	500	MCH	20	10	50	9.4	[7]
L100069	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	0.7	[8]
L100070	$CrCl_3(THF)_3$	4.8	4.8	500	MCH	20	10	50	0.4	[7]
L100071	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	0.8	[8]
L100072	$CrCl_3(THF)_3$	4.8	4.8	500	MCH	20	10	50	9	[7]
L100073	$CrCl_3(THF)_3$	4.8	4.8	500	MCH	20	10	50	0.7	[7]
L100074	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	1.1	[8]
L100075	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	0.9	[8]
L100076	$CrCl_3(THF)_3$	4.8	4.8	500	MCH	20	10	50	0.7	[7]
L100077	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	50	0.2	[9]
L100077-1	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	30	0.5	[9]
L100077-2	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	70	0.2	[9]
L100077-3	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	110	0.3	[9]
L100078	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	50	0.4	[9]
L100078-1	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	30	1.1	[9]
L100078-2	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	90	0.5	[9]
L100078-3	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	110	0.6	[9]
L100079	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	50	0.2	[9]
L100080	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	50	0.1	[9]
L100081	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	50	0.1	[9]
L100082	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	50	0.5	[9]

L100083	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	60	0.44	[10]
L100084	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	60	1	[10]
L100085	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	60	0.4	[10]
L100085-1	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	40	0.34	[10]
L100085-2	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	60	0.4	[10]
L100085-3	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	80	0.66	[10]
L100086	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	60	0.51	[10]
L100087	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	60	1	[10]
L100088	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	60	0.25	[10]
L100088-1	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	40	1.3	[10]
L100088-2	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	60	0.25	[10]
L100088-3	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	80	0.25	[10]
L100088-4	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	100	0.86	[10]
L100088-5	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	80	0.25	[10]
L100088-6	$Cr(AcAc)_3$	0.3	0.33	1000	MCH	30	40	80	0.65	[10]
L100088-7	$Cr(AcAc)_3$	0.3	0.33	1500	MCH	30	40	80	0.27	[10]
L100088-8	$Cr(AcAc)_3$	0.3	0.33	2000	MCH	30	40	80	0.32	[10]
L100088-9	$Cr(AcAc)_3$	0.3	0.33	1000	MCH	30	40	60	0.64	[10]
L100088-10	$Cr(AcAc)_3$	0.3	0.33	1500	MCH	30	40	60	0.45	[10]
L100088-11	$Cr(AcAc)_3$	0.3	0.33	1500	MCH	30	30	80	0.61	[10]
L100088-12	$Cr(AcAc)_3$	0.3	0.33	1500	MCH	30	20	80	0.18	[10]
L100089	$Cr(AcAc)_3$	0.5	0.55	1000	MCH	30	40	60	0.47	[10]
L100090	$Cr(AcAc)_3$	10	15	900	toluene	80	40	30	0.85	[11]
L100091	$Cr(AcAc)_3$	20	30	300	toluene	80	30	60	0.28	[11]
L100092	$Cr(AcAc)_3$	20	30	300	toluene	80	30	60	0.48	[11]
L100093	$Cr(AcAc)_3$	20	30	300	toluene	80	30	60	0.4	[11]
L100094	$Cr(AcAc)_3$	20	30	300	toluene	80	30	60	6.2	[11]
L100095	$Cr(AcAc)_3$	20	30	300	toluene	80	30	60	2.1	[11]
L100096	$Cr(AcAc)_3$	20	30	300	toluene	80	30	60	19.9	[11]
L100097	$Cr(AcAc)_3$	20	30	300	toluene	80	30	60	0.4	[11]
L100098	$Cr(AcAc)_3$	20	30	300	toluene	80	30	60	0.7	[11]
L100099	$Cr(AcAc)_3$	20	30	300	toluene	80	30	60	4	[11]
L100100	$Cr(AcAc)_3$	20	30	300	toluene	80	30	60	0.3	[11]

L100101	$Cr(AcAc)_3$	1	1	2000	PhCl	100	45	45	0.8	[12]
L100102	$Cr(AcAc)_3$	1	1	2000	PhCl	100	45	45	0.8	[12]
L100103	$Cr(AcAc)_3$	30	390	25	toluene	100	30	50	0.3	[13]
L100104	$Cr(AcAc)_3$	30	390	25	toluene	100	30	50	0.1	[13]
L100105	$Cr(AcAc)_3$	30	390	25	toluene	100	30	50	0.1	[13]
L100106	$Cr(AcAc)_3$	30	390	25	toluene	100	30	50	0.1	[13]
L100107	$Cr(AcAc)_3$	30	390	25	toluene	100	30	50	0.1	[13]
L100108	$Cr(AcAc)_3$	30	390	25	toluene	100	30	50	0.4	[13]
L100109	$Cr(AcAc)_3$	30	390	25	toluene	100	30	50	0.2	[13]
L100110	$Cr(AcAc)_3$	30	390	25	toluene	100	30	50	0.6	[13]
L100111	$Cr(AcAc)_3$	30	390	25	toluene	100	30	50	3.9	[13]
L100112	$Cr(AcAc)_3$	30	390	25	toluene	100	30	50	0.3	[13]
L100113	$CrCl_3(THF)_3$	0.5	0.5	500	pentane	20	30	45	3.2	[8]
L100113-1	$CrCl_3(THF)_3$	0.5	0.5	500	n-hexane	20	30	45	4.4	[8]
L100113-2	$CrCl_3(THF)_3$	0.5	0.5	500	cyclohexane	20	30	45	3.7	[8]
L100113-3	$CrCl_3(THF)_3$	0.5	0.5	500	heptane	20	30	45	4.1	[8]
L100113-4	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	45	3.4	[8]
L100113-5	$CrCl_3(THF)_3$	0.5	0.25	500	MCH	20	30	45	13.8	[8]
L100113-6	$CrCl_3(THF)_3$	0.5	0.75	500	MCH	20	30	45	5	[8]
L100113-7	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	2.6	[8]
L100113-8	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	45	3.4	[8]
L100113-9	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	2	[8]
L100113-10	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	80	1.8	[8]
L100113-11	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	100	1.8	[8]
L100113-12	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	120	1.7	[8]
L100113-13	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	140	1.7	[8]
L100114	$CrCl_3(THF)_3$	0.5	0.5	500	pentane	20	30	45	1.2	[8]
L100114-1	$CrCl_3(THF)_3$	0.5	0.5	500	n-hexane	20	30	45	1.6	[8]
L100114-2	$CrCl_3(THF)_3$	0.5	0.5	500	cyclohexane	20	30	45	2.4	[8]

L100114-3	$CrCl_3(THF)_3$	0.5	0.5	500	heptane	20	30	45	0.4	[8]
L100114-4	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	45	0.2	[8]
L100114-5	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	0.8	[8]
L100114-6	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	45	0.2	[8]
L100114-7	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	0.8	[8]
L100114-8	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	80	1	[8]
L100114-9	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	100	1.5	[8]
L100114-10	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	120	2.2	[8]
L100114-11	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	140	6.7	[8]
L100115	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	0.4	[8]
L100115-1	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	45	0.3	[8]
L100115-2	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	45	0.4	[8]
L100115-3	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	0.4	[8]
L100115-4	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	80	0.4	[8]
L100115-5	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	100	0.1	[8]
L100115-6	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	120	0.1	[8]
L100115-7	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	140	0.1	[8]
L100116	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	0.8	[8]
L100116-1	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	45	0.2	[8]
L100116-2	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	45	0.2	[8]
L100116-3	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	60	0.4	[8]
L100116-4	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	80	0.5	[8]
L100116-5	$CrCl_3(THF)_3$	0.5	0.5	500	MCH	20	30	140	0.4	[8]
L100117	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	15	0.2	[14]
L100117-1	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	30	0.03	[14]
L100117-2	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	45	0.05	[14]
L100117-3	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	60	0.06	[14]
L100117-4	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	75	0.05	[14]
L100118	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	15	0.04	[14]
L100118-1	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	30	0.02	[14]
L100118-2	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	45	0.01	[14]
L100118-3	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	60	0.02	[14]
L100118-4	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	75	0.02	[14]

L100119	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	30	0.01	[14]
L100120	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	30	0.05	[14]
L100121	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	30	0.02	[14]
L100122	$CrCl_3(THF)_3$	2.4	2.4	500	MCH	20	10	30	0.04	[14]

[a]MCH, methylcyclohexane.

Prior to model training, outliers in the experimental PE content were screened using the 1.5×IQR rule. Flagged entries were cross-checked against the original literature and evaluated for chemical plausibility; only data points that were both statistical outliers and inconsistent with the source report or experimental context were excluded from the final dataset.

The dataset contains 133 entries with PE < 1 wt% and 90 entries with PE ≥ 1 wt%. Classification performance was evaluated using confusion-matrix-based metrics including accuracy, precision, and recall to ensure that performance is assessed for both classes rather than dominated by the larger subset.

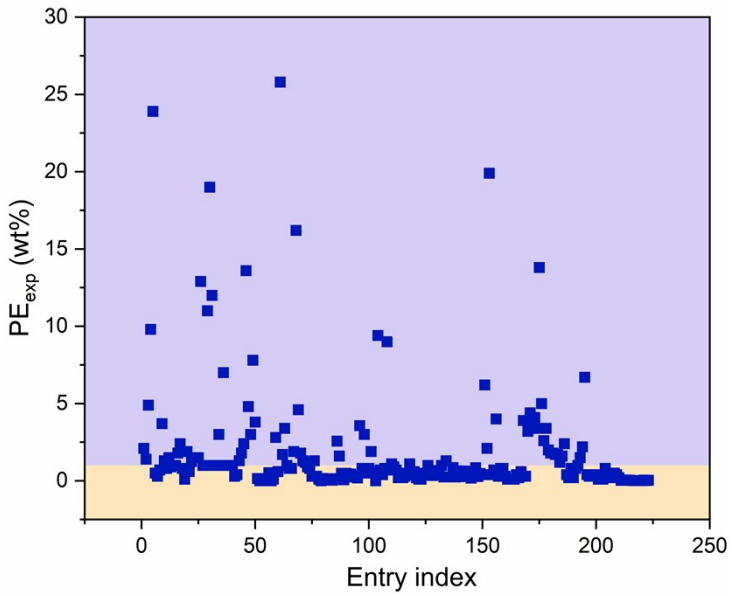


Figure S2. Distribution of experimental PE content (wt%) in the full dataset.

1.2 Catalyst generation

An automated SMILES-based workflow was employed to construct three-dimensional molecular models of the target catalysts. The process began with defining the catalyst skeletons [Cr]1([P](N([P]1(R_1)R_2)R_3)(R_4)R_5)(Cl)(Cl)Cl, where R_1-R_5 represent variable substituents systematically filled by all possible P- and N-substituent combinations.[15] All generated SMILES were standardized to remove redundancy and ensure unique canonical representations for consistent tracking. In the structure generation stage, RDKit[16] was used to add hydrogens and convert SMILES into MOL files with proper bonding and topology, while Open Babel[17] was applied when direct 3D conversion was required. The resulting initial 3D structures were further optimized to remove abnormal or high-energy conformations, and conformer searches were applied to locate low-energy geometries, ensuring physically meaningful and reliable structures for subsequent calculations.

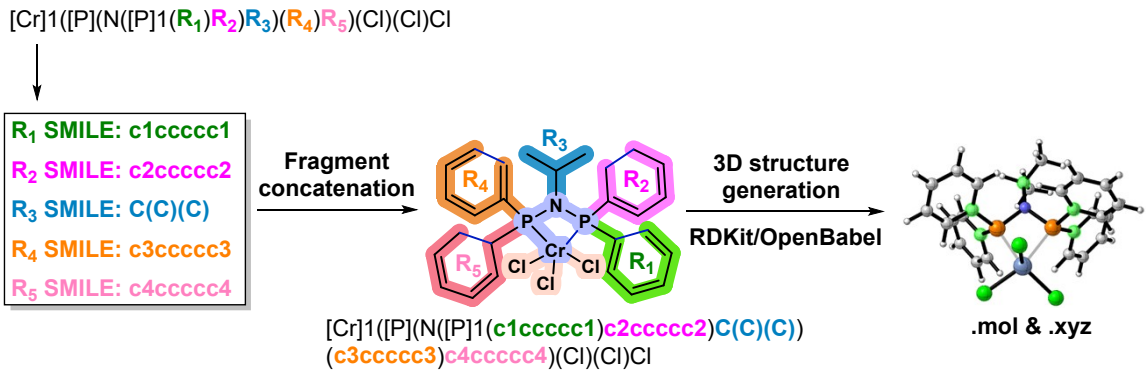
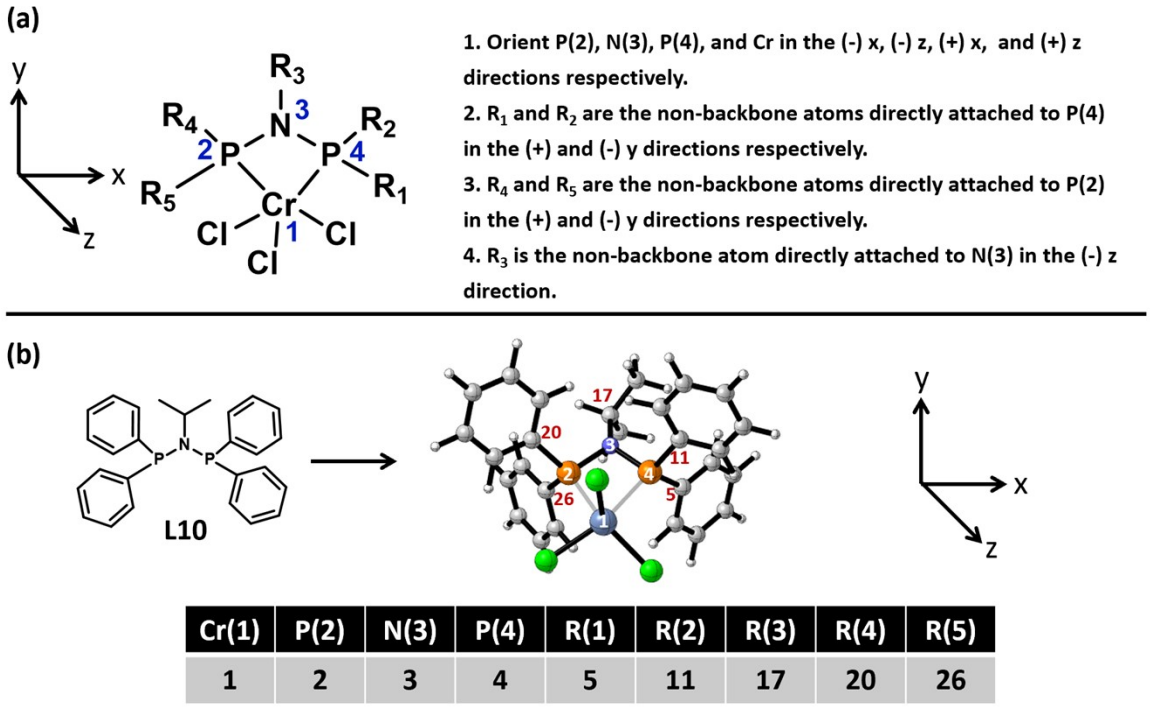


Figure S3. Automated workflow for 3D structure generation of target Cr/PNP complexes based on SMILES strings.

2. Descriptors

2.1 Steric descriptors

In DFT output files, each atom in the structure is assigned a unique index. Since all descriptors were defined based on specific ligand atom numbers (e.g., the P–R bond length requires indices for P and R atoms), it was essential to establish a consistent numbering scheme. This was achieved through the following procedure (Figure S4a). First, P(2), N(3), P(4), and Cr(1) were oriented along the (–) x, (–) z, (+) x, and (+) z directions of an imaginary Cartesian coordinate system, respectively. Substituents R(1) and R(2), directly attached to P(4), were then assigned according to their orientation along the (+) y and (–) y directions, respectively. Similarly, substituents R(4) and R(5), directly attached to P(2), were assigned in the (+) y and (–) y directions, respectively. Finally, R(3) was defined as the non-backbone substituent directly attached to N(3) along the (–) z direction. An example of this numbering scheme applied to ligand **L10** is shown in Figure S4b.



Cr(1)	P(2)	N(3)	P(4)	R(1)	R(2)	R(3)	R(4)	R(5)
1	2	3	4	5	11	17	20	26

Figure S4. PNP ligand numbering convention and representative example.

As this library included disparate PNP ligand skeletons, it was necessary to establish a robust standard for designating which phosphorus atoms should be labelled P(2) and P(4) (Figure S5). For ligands with chemically equivalent phosphorus atoms (e.g. **L10**, **L33** and **L76**), this assignment is inconsequential as both phosphorus atoms are symmetry-related. However, for ligands with non-equivalent phosphorus atoms (e.g. **L27**, **L39** and

L80), a consistent numbering convention was required. To address this, we adopted a parameter-based criterion that relied on the lone pair (LP) occupancies of the phosphorus donor atoms. Across the ligand set, it was consistently observed that P(2) exhibited a higher LP occupancy than P(4). This metric was therefore employed as the universal standard for all PNP ligands with non-equivalent phosphorus atoms, ensuring consistency in descriptor definition and subsequent analysis.

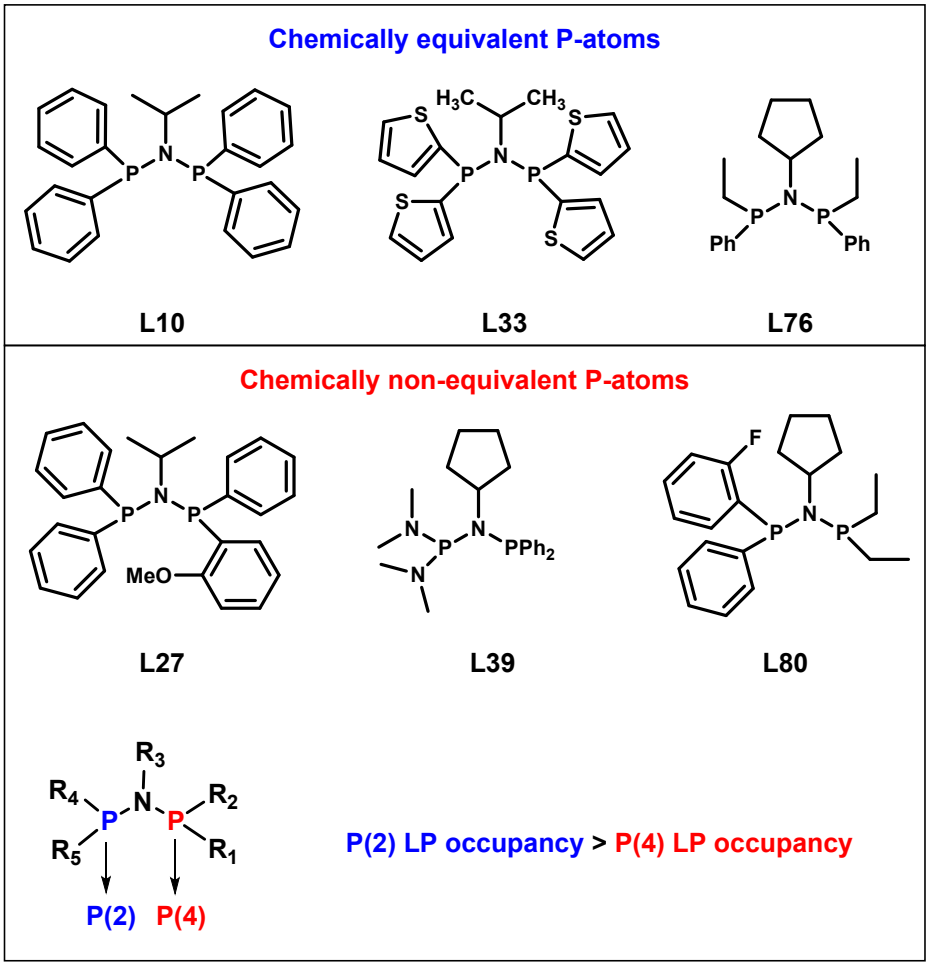
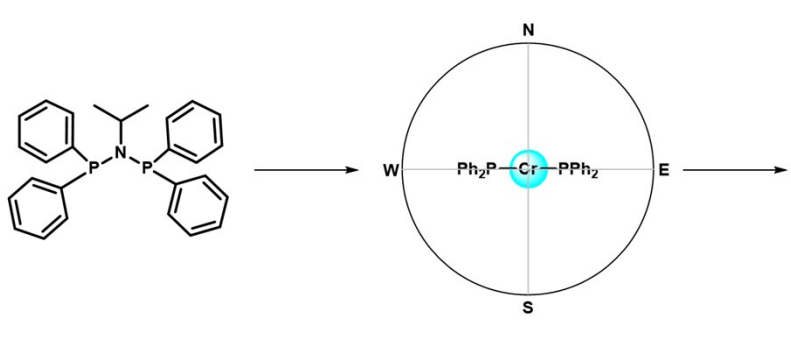
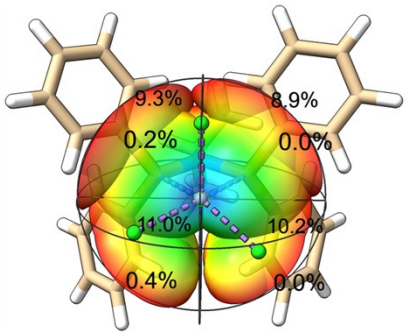


Figure S5. P(2) vs. P(4) designation when phosphine atoms are not symmetry equivalent.

The percent buried volume (%V_{bur}) is a widely used metric to quantify the steric environment imposed by ligands in transition metal complexes.[18] In this work, buried volume parameters were calculated using the Morfeus python library.[19] Spheres with radii ranging from 3.0 to 6.0 Å were centered on the chromium atom, and the fraction of the sphere volume occupied by the ligand was reported. To capture the directional steric features of the PNP ligands, quadrant and octant analyses were performed (Figure S4). For these analyses, the Cr atom was placed at the origin, with P(2) and P(4) aligned along the negative and positive x directions, respectively, and the

y-axis defined as perpendicular to the Cr–P–N plane. Quadrant buried volumes were defined in the xy-plane and extended in both the +z and –z directions, yielding four quadrants (Northeast, Northwest, Southeast, Southwest). Octant buried volumes were further resolved into eight regions, defined by the combinations of x, y, and z coordinates (e.g., Northeast-Front: x+, y+, z+; Northwest-Back: x–, y+, z–). This coordinate system enabled systematic quantification of quadrant- and octant-specific steric effects across the ligand library.





Front view

Quadrant analyses

Quadrant parameter	Cartesian description
Northeast (Q_0)	(x+, y+)
Northwest (Q_1)	(x-, y+)
Southeast (Q_2)	(x+, y–)
Southwest (Q_3)	(x–, y–)

Octant analyses

Octant parameter	Cartesian description
Northeast-Front (O_0)	(x+, y+, z+)
Northwest-Front (O_1)	(x–, y+, z+)
Southeast-Front (O_2)	(x+, y–, z+)
Southwest-Front (O_3)	(x–, y–, z+)
Northeast-back (O_4)	(x+, y+, z–)
Northwest-back (O_5)	(x–, y+, z–)
Southeast-back (O_6)	(x+, y–, z–)
Southwest-back (O_7)	(x–, y–, z–)

Figure S6. Definition of quadrants and octants for buried volume analysis of Cr/PNP complexes

Table S3. Representative examples of steric descriptors

Parameter name	Category	Description of the parameter
3_Q_buried_volume_0	Quadrant buried volume	Buried volume (%) of the northeast quadrant (Q0) within a 3.0 Å sphere around Cr
6_Q_buried_volume_1	Quadrant buried volume	Buried volume (%) of the northwest quadrant (Q1) within a 6.0 Å sphere around Cr
3_Q_free_volume_0	Quadrant buried volume	Free (unoccupied) volume of the northeast quadrant (Q0) at 3.0 Å radius
3_Q_percent_buried_volume_0	Quadrant buried volume	Percent buried volume of the northeast quadrant (Q0) at 3.0 Å radius
3_O_buried_volume_0	Octant buried volume	Buried volume (%) of the northeast–front octant (O0) at 3.0 Å radius
3_%V_{bur}	Buried volume	Total buried volume (%) within a 3.0 Å sphere around Cr
B1 value	Sterimol	Minimum width (B1) of the N-substituent (R_3 group)
B5 value	Sterimol	Maximum width (B5) of the N-substituent (R_3 group)
L value	Sterimol	Axial length (L) of the N-substituent (R_3 group)
Cone angle	Cone angle	Tolman cone angle of the N-substituent (R_3 group)

2.2 Electronic descriptors

Beyond steric parameters, the electronic environment of PNP ligands plays a decisive role in modulating catalytic activity and polyethylene formation. To capture these effects, a set of electronic descriptors was derived from quantum chemical analyses. These include natural bond orbital (NBO) charges on the donor atoms, lone pair occupancies of phosphorus and nitrogen centers, and orbital energy levels that reflect the donor-acceptor characteristics of the ligands. In addition, global electronic features such as HOMO, LUMO, and the V_{min} were incorporated to provide a more comprehensive picture of the electronic structure. Together, these descriptors enable quantitative assessment of both localized donor strength and overall electronic flexibility of the catalyst system.

Table S4. Representative global electronic descriptors

Parameter name	Category	Description of the parameter
Catalyst_Alpha_HOMO	Frontier orbital	Energy of the α-spin highest occupied molecular orbital of the catalyst.
Catalyst_Alpha_LUMO	Frontier orbital	Energy of the α-spin lowest unoccupied molecular orbital of the catalyst
Catalyst_Beta_HOMO	Frontier orbital	Energy of the β-spin highest occupied molecular orbital of the catalyst
Catalyst_Beta_LUMO	Frontier orbital	Energy of the β-spin lowest unoccupied molecular orbital of the catalyst
Ligand_HOMO	Frontier orbital	Energy of the highest occupied molecular orbital of the isolated ligand
Ligand_LUMO	Buried volume	Energy of the lowest unoccupied molecular orbital of the isolated ligand
V_{min}	V_{min}	Minimum value of the molecular electrostatic potential

Table S5. Representative local atomic and bond-specific electronic descriptors

Parameter name	Category	Description of the parameter
Catalyst_N(3)_LP_occ	NBO	Lone-pair occupancy (in electrons) of the backbone nitrogen atom N(3) in the catalyst
Catalyst_N(3)_LP_eng	NBO	Stabilization energy associated with the lone pair on N(3) from NBO analysis
Catalyst_BD_Cr(1)_P(2)_occ	NBO	Occupancy of the Cr(1)-P(2) σ-bonding orbital
Catalyst_BD_Cr(1)_P(2)_eng	NBO	Stabilization energy of the Cr(1)-P(2) σ-bonding orbital
Catalyst_BD*_Cr(1)_P(2)_occ	NBO	Occupancy of the Cr(1)-P(2) σ*-antibonding orbital
Catalyst_BD*_Cr(1)_P(2)_eng	NBO	Stabilization energy of the Cr(1)-P(2) σ*-antibonding orbital
Catalyst_NBO_Charge_Cr(1)	NBO	Natural bond orbital charge on the chromium center Cr(1)
Ligand_P(2)_LP_occ	NBO	Lone-pair occupancy of phosphorus atom P(2) in the free ligand
Ligand_P(2)_LP_eng	NBO	Lone-pair stabilization energy of phosphorus atom P(2) in the free ligand
Ligand_P(4)_LP_occ	NBO	Lone-pair occupancy of phosphorus atom P(4) in the free ligand
Ligand_P(4)_LP_eng	NBO	Lone-pair stabilization energy of phosphorus atom P(4) in the free ligand
Ligand_BD*_P(2)_N(3)_occ	NBO	Occupancy of the P(2)-N(3) σ*-antibonding orbital in the ligand
Ligand_BD*_P(2)_N(3)_eng	NBO	Stabilization energy of the P(2)-N(3) σ*-antibonding orbital
Ligand_NBO_Charge_P(2)	NBO	NBO charge on phosphorus atom P(2) of the free ligand

2.3 Geometric descriptors

Geometric descriptors for the Cr/PNP complexes were categorized into two main groups. The first group comprises backbone-fixed parameters that describe the intrinsic geometry of the ligand framework, including key bond lengths and angles defined by the Cr center, the two phosphorus donors, and the bridging nitrogen atom. The second group involves descriptors that incorporate the variable substituents attached to phosphorus and nitrogen, capturing how changes in R-group identity influence the local geometry around the metal center. Together, these backbone and substituent-related descriptors provide a comprehensive representation of the steric and spatial features relevant to catalyst structure and performance.

Table S6. Backbone-fixed geometric descriptors for Cr/PNP complexes

Parameter name	Category	Description of the parameter
Distance_Cr_P(2)	Distance	Bond length between the chromium center (Cr) and the phosphorus atom P(2)
Distance_Cr_P(4)	Distance	Bond length between Cr and the phosphorus atom P(4)
Distance_P(2)_N(3)	Distance	Bond length between phosphorus P(2) and the backbone nitrogen N(3)
Distance_P(4)_N(3)	Distance	Bond length between phosphorus P(4) and N(3)
Angle_Cr_P(2)_P(4)	Angle	Bite angle defined by Cr-P(2)-P(4)
Angle_Cr_P(2)_N(3)	Angle	Angle formed by Cr-P(2)-N(3), describing backbone geometry
Angle_Cr_P(4)_N(3)	Angle	Angle formed by Cr-P(4)-N(3), describing backbone geometry
Dihedral_Cr_P(2)_N(3)_P(4)	Dihedral	Torsional angle between Cr-P(2)-N(3)-P(4), characterizing ligand backbone twist

Table S7. Substituent-dependent geometric descriptors involving variable R-groups

Parameter name	Category	Description of the parameter
Distance_P(2)_R_4	Distance	Bond length between phosphorus P(2) and the substituent R_4
Distance_P(2)_R_5	Distance	Bond length between phosphorus P(2) and the substituent R_5
Distance_P(4)_R_1	Distance	Bond length between phosphorus P(4) and the substituent R_1
Distance_P(4)_R_2	Distance	Bond length between phosphorus P(4) and the substituent R_2
Distance_N(3)_R_3	Distance	Bond length between nitrogen N(3) and the substituent R_3
Angle_P(2)_R_4_R_5	Angle	Angle formed by P(2)-R_4-R_5, describing the spatial relationship of P(2) substituents
Angle_P(4)_R_1_R_2	Angle	Angle formed by P(4)_R_1_R_2, describing the spatial relationship of P(4) substituents
Angle_N(3)_P(2)_R_3	Angle	Angle formed by N(3)_P(2)_R_3, describing orientation of the N substituent relative to P(2)
Dihedral_Cr_P(2)_N(3)_R_1	Dihedral	Torsional angle Cr_P(2)_N(3)_R_1, characterizing twisting involving the N-substituent
Dihedral_Cr_P(2)_N(3)_R_4	Dihedral	Torsional angle Cr_P(2)_N(3)_R_4, describing the twist involving the P(2) substituent R_4

2.4 Other descriptors

In addition to steric, electronic, and geometric parameters, two additional classes of descriptors were included to capture complementary structural information: RDKit-derived molecular descriptors and atom-centered symmetry functions (ACSF). RDKit descriptors provide a wide range of cheminformatics-based features such as molecular weight, polar surface area, rotatable bond counts, and topological indices, offering a coarse but comprehensive summary of ligand topology and overall shape. ACSF descriptors, on the other hand, encode local atomic environments in a rotationally and translationally invariant manner, ensuring a robust representation of subtle three-dimensional structural differences among ligands. Together, these descriptor sets enrich the feature space and enhance the model's ability to distinguish ligands with similar steric or electronic profiles but distinct spatial or topological characteristics.

Table S8. Representative RDKit and ACSF descriptors used to complement steric, electronic, and geometric features.

Parameter name	Category	Description of the parameter
MolWt	RDKit	Average molecular weight of the ligand, calculated using standard atomic weights
Kappa1	RDKit	Topological descriptor reflecting molecular shape and flexibility based on atom connectivity
FractionCSP3	RDKit	Ratio of sp^3-hybridized carbon atoms to total carbon atoms, indicating molecular 3D character
G2_Cr	ACSF	Radial symmetry G2 function centered on chromium (Cr), encoding pairwise interactions between Cr and its surrounding atoms
G4_Cr	ACSF	Angular symmetry G4 function centered on chromium (Cr), characterizing angular correlations between Cr and its neighboring atoms
G4_P(2)	ACSF	Angular symmetry G4 function centered on phosphorus atom P(2), capturing angular features around P(2)

3. Machine Learning Models

The PE content data show an extremely uneven and non-Gaussian distribution, with a sharp peak near zero and a long heavy tail (Figure S7). Such a distribution is highly unsuitable for regression or similar statistical modeling approaches without appropriate transformation or preprocessing.

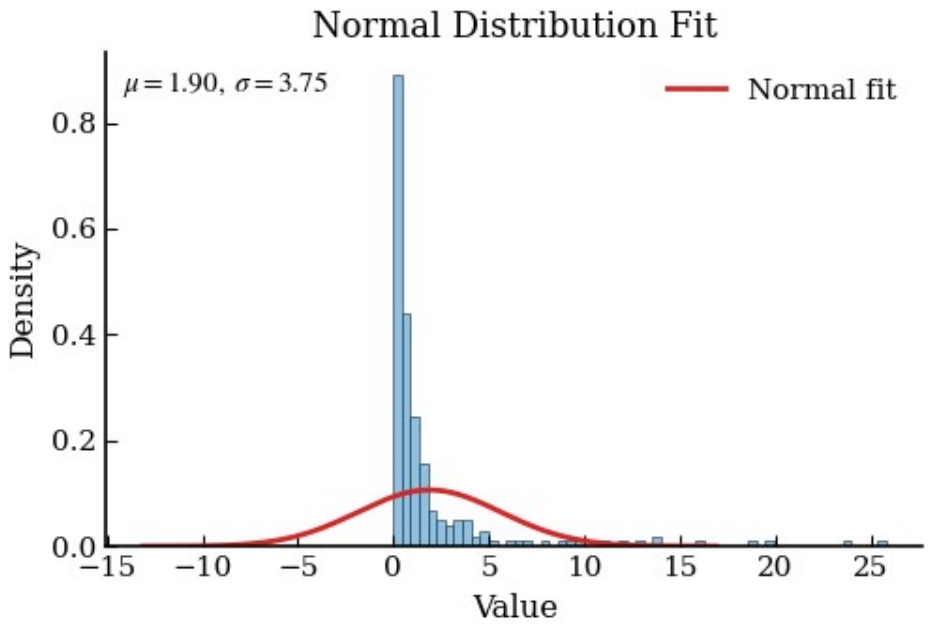


Figure S7. Normal distribution fitting of PE content.

The LASSO + RF combined model yields a moderate performance with R^2 = 0.57 and a relatively large MAE of 2.00 (Figure S8). Although it partially captures the trend of increasing PE_{exp}, significant scatter and deviation are evident across most data points, especially in the low PE region where clustering dominates. This indicates that the model struggles to generalize effectively and fails to achieve robust predictive accuracy across the full range of PE values. Therefore, we adopted a classification + regression approach to improve predictive performance.

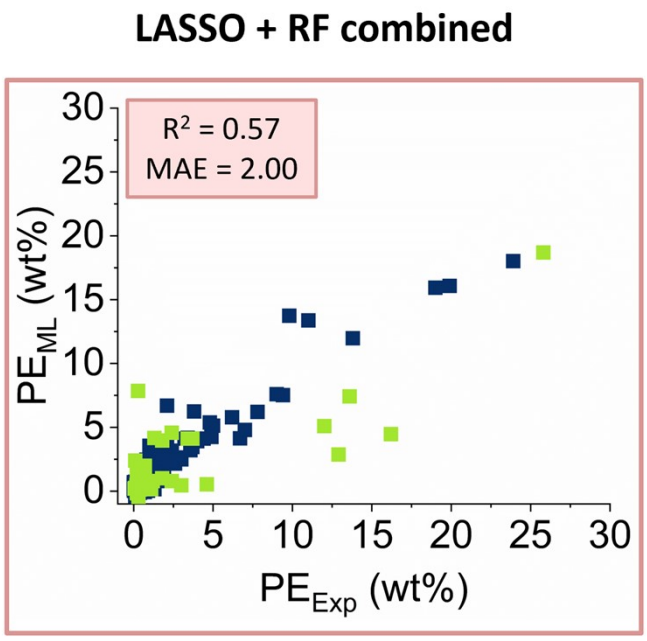


Figure S8. Regression performance for predicting PE content using XGBoost

3.1 Classification models

Logistic regression and random forest models underperformed compared to the decision tree classifier in identifying low-PE and high-PE catalysts. Although the LASSO + RF combined model achieved a relatively high overall accuracy of 0.87, its recall was limited to 0.72, indicating missed detection of a substantial portion of high-PE cases. By contrast, the random forest model yielded perfect precision (Precision = 1), demonstrating its ability to avoid false positives, but it still lagged behind the decision tree in balanced classification performance. These results confirm that decision trees offered the most reliable classification for this dataset.

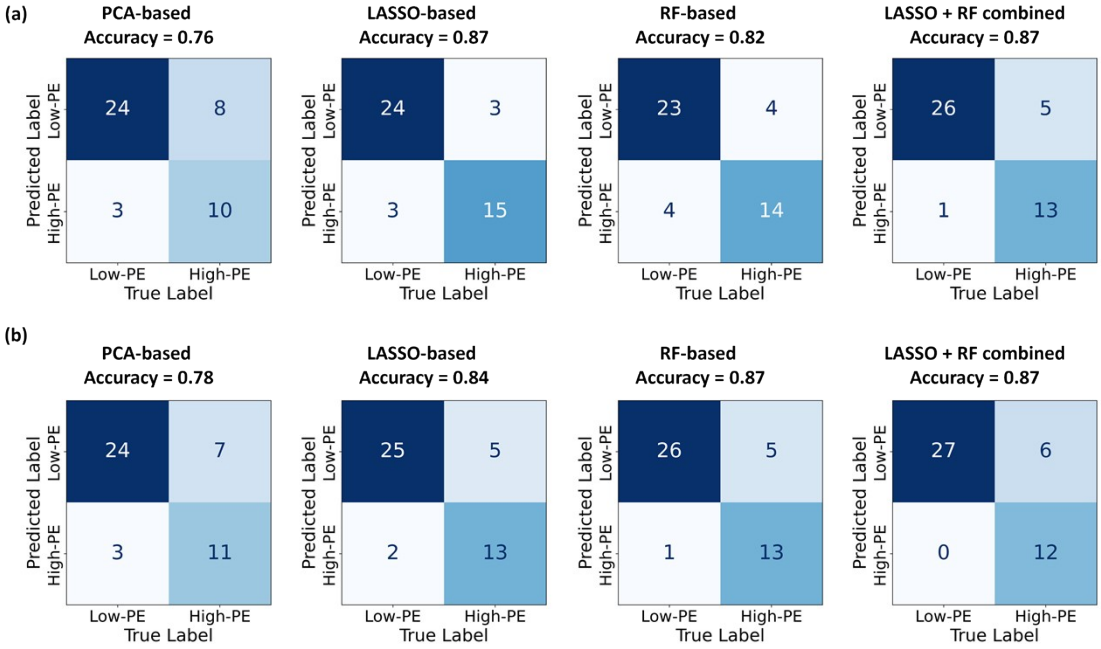


Figure S9. Confusion matrices of classification models for distinguishing low-PE and high-PE catalysts. (a) logistic regression; (b) random forest.

3.2 Regression models

Table S9. Performance comparison of various regression algorithms applied to different descriptor sets.

	Train R^2	Test R^2	Train MAE	Test MAE
		PCA-based		
linear regression	0.34	0.42	0.18	0.15
ridge regression	0.34	0.42	0.18	0.15
random forest	0.69	0.54	0.12	0.14
support vector	0.27	0.40	0.18	0.14
XGBoost	0.74	0.37	0.09	0.14
LASSO	0.00	0.00	0.23	0.22
		LASSO-based		
linear regression	0.57	0.56	0.14	0.13
ridge regression	0.57	0.56	0.14	0.13
random forest	0.71	0.61	0.11	0.13
support vector	0.53	0.49	0.15	0.15
XGBoost	0.75	0.54	0.09	0.13
LASSO	0.09	0.15	0.22	0.20
		RF-based		
linear regression	0.54	0.67	0.15	0.12
ridge regression	0.54	0.67	0.15	0.12
random forest	0.84	0.70	0.09	0.11
support vector	0.50	0.53	0.15	0.14
XGBoost	0.91	0.61	0.05	0.13
LASSO	0.09	0.15	0.22	0.20
		LASSO + RF combined		
linear regression	0.61	0.57	0.15	0.12
ridge regression	0.61	0.58	0.15	0.12
random forest	0.84	0.68	0.09	0.11
support vector	0.58	0.54	0.14	0.14
XGBoost	0.92	0.88	0.06	0.07
LASSO	0.09	0.15	0.22	0.20

Table S9 summarizes the performance of multiple regression algorithms (linear regression, ridge regression, random forest, support vector regression, XGBoost, and LASSO) applied to three descriptor sets (LASSO-based, RF-based, and LASSO + RF combined). Among these, XGBoost generally achieves the highest R^2 values and the lowest MAE across most descriptor sets, demonstrating its strong predictive capability. Random forest also performs competitively, particularly for the RF-based and LASSO + RF combined descriptors, while linear and ridge regression provide moderate accuracy. In contrast, LASSO alone shows poor predictive performance, highlighting the advantage of ensemble methods and combined descriptors for accurate property prediction.

4. Feature Importance

As shown in Figure 7 of the main text, SHAP importance analysis was employed to identify the most influential molecular descriptors. The detailed interpretations of the top six descriptors from both the classification and regression models are provided in Tables S9-S10.

Table S10. The top six descriptors identified by classification model, and their physical meaning.

	Parameter name	Description of the parameter
1	VSA_EState7	A combined descriptor reflecting the distribution of electronic states over molecular surface area. Represents the proportion of highly polarized surface regions
2	4_O_Vbur_7_lec	Buried-volume parameter for octant 7 of radius 4 Å, calculated for the lowest-energy conformer
3	Dihedral_1_2_3_R1_min	Minimum dihedral angle defined by atoms 1_2_3 and the atom bonded to the R_1 substituent
4	pressure	Reaction pressure
5	4_O_Vbur_5_min	Minimum octant-buried-volume value within a 4 Å sphere for octant 5
6	Angle_3_4_17_min	Minimum bond angle defined by atoms 3_4_17

Table S11. The top six descriptors identified by classification model, and their physical meaning.

	Parameter name	Description of the parameter
1	pressure	Reaction pressure
2	R_5_G4_bol	Boltzmann-weighted angular (G4) symmetry function centered on the atom bonded to the R_5 substituent
3	temperature	Reaction temperature
4	Cr amount	Chromium content or loading in catalyst preparation.
5	Al/Cr	Co-catalyst-to-catalyst molar ratio
6	L_NBO_R3_bol	Boltzmann-averaged natural bond orbital charge or occupancy on the atom bonded to the R_3 substituent of Ligand

Table S12. SHAP-based feature importance for the classification model.

Descriptor	Mean |SHAP|	Contribution (%)
VAS_Estate7	0.20	38.5
4_O_Vbur_7_lec	0.10	19.2
Dihedral_1_2_3_R_1_min	0.09	17.3
pressure	0.06	11.5
4_O_Vbur_5_min	0.04	7.7
Angle_3_4_17_min	0.03	5.7

Table S13. SHAP-based feature importance for the regression model.

Descriptor	Mean |SHAP|	Contribution (%)
pressure	0.09	30.0
R_5_G4_bol	0.05	16.7
temperature	0.05	16.7
Cr amount	0.04	13.3
Al/Cr ratio	0.04	13.3
L_NBO_R_3_bol	0.03	10.0

5. Benchmark of Computational Methods

In this work, all intermediates and transition states were fully optimized and characterized by frequency calculations at the B3LYP-D3/def2-SVP level. The use of a dispersion-corrected hybrid functional with a split-valence basis set for geometry optimization is widely adopted for transition-metal complexes, as it provides reliable structures at a moderate computational cost.[20] In the present Cr coordination system, we explicitly validated the suitability of B3LYP-D3/def2-SVP by comparing the optimized structure with the available X-ray crystallographic data.[21] Experimental bond lengths and angles are reported in parentheses in Figure S10. Key structural parameters are well reproduced: Cr-P1 = 2.45 Å (2.48 Å), Cr-Cl1 = 2.28 Å (2.24 Å), Cr-Cl2 = 2.33 Å (2.29 Å), Cr-N1 = 2.09 Å (2.04 Å), and the P1-N2 bond length is 1.73 Å (1.69 Å). The maximum deviation among these critical bond lengths is ≤ 0.06 Å. Importantly, the chelating bite angle ∠P-Cr-P is reproduced as 66.6° (66.4°), with an error of only 0.2°. These results indicate that B3LYP-D3/def2-SVP provides sufficiently accurate geometries for this class of Cr complexes and is therefore an appropriate and cost-effective level for geometry optimizations in this work.

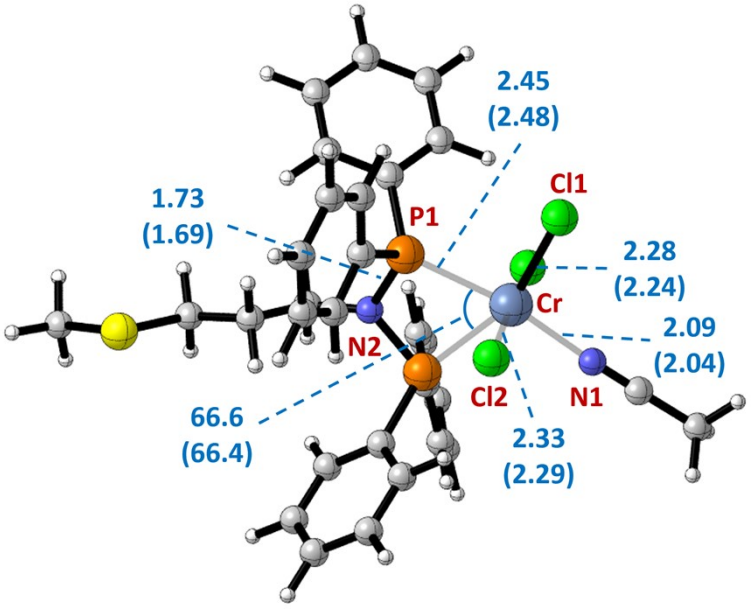


Figure S10. Optimized structure of the Cr(III) bis(phosphino)amine complex at the B3LYP-D3/def2-SVP level. Selected bond lengths (Å) and the P-Cr-P bite angle (deg) are shown; experimental X-ray values are given in parentheses.

Electronic energies were evaluated using the M06-L functional in combination with the larger def2-TZVP basis set. M06-L has been extensively benchmarked and shown to provide reliable reaction energetics for transition-metal systems.[22] To further reduce method dependence, we carried out additional single-point tests against higher-level reference data for the present Cr system. Specifically, the ethylene coordination to $[LCr\text{-}Me]^+$ to form the η^2-ethylene adduct (**1-2**; Figure 8) was chosen as a representative step. DLPNO-CCSD(T) single-point energies

were computed with def2-TZVP and def2-QZVP and extrapolated to the complete-basis-set (CBS) limit, which was used as the reference benchmark. Relative to this DLPNO-CCSD(T)/CBS reference, M06-L/def2-TZVP reproduces the key reaction energetics with deviations within 1 kcal/mol (Figure S11), and shows smaller errors than the alternative functionals examined (e.g., B3LYP, M06, ωB97X-D, PBE and BP86). Therefore, the combination of B3LYP-D3/def2-SVP optimized geometries with M06-L/def2-TZVP single-point energies represents a reasonable and well-founded composite scheme for the present Cr-catalyzed ethylene oligomerization study.

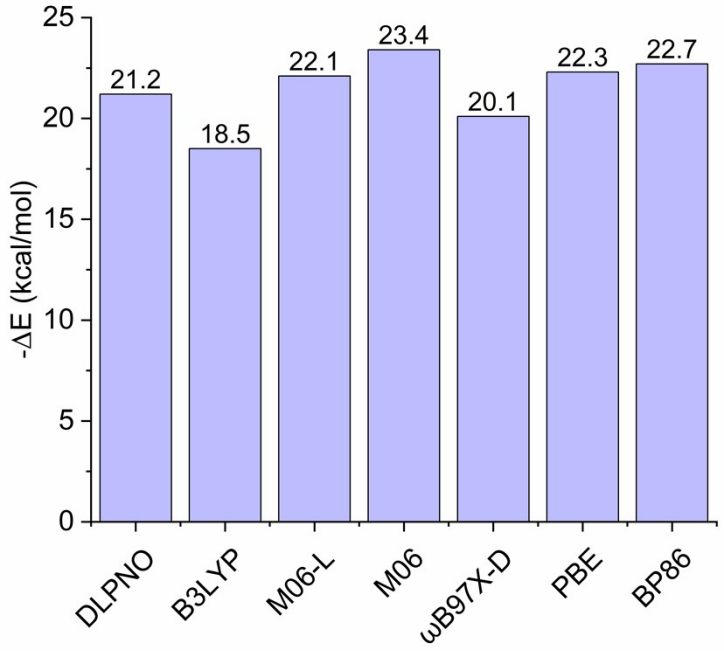


Figure S11. Relative electronic energies (ΔE, kcal/mol) for ethylene binding to $[LCr\text{-}Me]^+$. DLPNO-CCSD(T)/CBS energies were obtained by def2-TZVP/def2-QZVP extrapolation.

6. References

1. K. Blann, A. Bollmann, H. de Bod, J. T. Dixon, E. Killian, P. Nongodlwana, M. C. Maumela, H. Maumela, A. E. McConnell, D. H. Morgan, M. J. Overett, M. Prétorius, S. Kuhlmann and P. Wasserscheid, *J Catal.*, 2007, **249**, 244-249.
2. M. J. Overett, K. Blann, A. Bollmann, J. T. Dixon, F. Hess, E. Killian, H. Maumela, D. H. Morgan, A. Neveling and S. Otto, *Chem Commun (Camb)*, 2005, 622-624.
3. A. Bollmann, K. Blann, J. T. Dixon, F. M. Hess, E. Killian, H. Maumela, D. S. McGuinness, D. H. Morgan, A. Neveling, S. Otto, M. Overett, A. M. Slawin, P. Wasserscheid and S. Kuhlmann, *J. Am. Chem. Soc.*, 2004, **126**, 14712-14713.
4. E. Killian, K. Blann, A. Bollmann, J. T. Dixon, S. Kuhlmann, M. C. Maumela, H. Maumela, D. H. Morgan, P. Nongodlwana, M. J. Overett, M. Pretorius, K. Höfener and P. Wasserscheid, *J. Mol. Catal. A: Chem.*, 2007, **270**, 214-218.
5. J. Zhang, F. Alam, H. Fan, J. Ma and T. Jiang, *Appl. Organomet. Chem.*, 2021, **36**.
6. S. Kuhlmann, K. Blann, A. Bollmann, J. Dixon, E. Killian, M. Maumela, H. Maumela, D. Morgan, M. Pretorius and N. Taccardi, *J Catal.*, 2007, **245**, 279-284.
7. F. Alam, H. Fan, C. Dong, J. Zhang, J. Ma, Y. Chen and T. Jiang, *J Catal.*, 2021, **404**, 163-173.
8. H. Lee and S. H. Hong, *Applied Catalysis A: General*, 2018, **560**, 21-27.
9. Y. Wang, Y. Zhang, H. Fan, Z. Wang, H. Shao, R. Shao, X. Zhang and T. Jiang, *Fuel*, 2024, **375**.
10. T. Zhou, J. Zuo, H. Xie, X. Zhao, M. X. Zhao and J. Zhang, *Dalton Trans*, 2024, **53**, 14011-14017.
11. N. Peulecke, B. H. Muller, A. Spannenberg, M. Hohne, U. Rosenthal, A. Wohl, W. Muller, A. Alqahtani and M. Al Hazmi, *Dalton Trans*, 2016, **45**, 8869-8874.
12. S. Barman, E. A. Jaseer, N. Garcia, M. Elanany, M. Khawaji, W. Xu, S. Lin, H. Alasiri, M. N. Akhtar and R. Theravalappil, *Chem Commun (Camb)*, 2022, **58**, 10044-10047.
13. S. Heinig, A. Wöhl, W. Müller, M. H. Al‐Hazmi, B. H. Müller, N. Peulecke and U. Rosenthal, *ChemCatChem*, 2013, **5**, 3107-3113.
14. H. Shao, Y. Zhang, L. Zhang, X. Zhang, L. Guo, Z. Wang, Y. Wang and T. Jiang, *Applied Catalysis A: General*, 2025, **697**.
15. D. Weininger, *J. Chem. Inf. Comput. Sci.*, 2002, **28**, 31-36.
16. A. P. Bento, A. Hersey, E. Felix, G. Landrum, A. Gaulton, F. Atkinson, L. J. Bellis, M. De Veij and A. R. Leach, *J Cheminform*, 2020, **12**, 51.
17. N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J Cheminform*, 2011, **3**, 33.
18. L. Falivene, R. Credendino, A. Poater, A. Petta, L. Serra, R. Oliva, V. Scarano and L. Cavallo, *Organometallics*, 2016, **35**, 2286-2293.
19. https://github.com/digital-chemistry-laboratory/morfeus.
20. K. D. Vogiatzis, M. V. Polynski, J. K. Kirkland, J. Townsend, A. Hashemi, C. Liu and E. A. Pidko, *Chem Rev*, 2019, **119**, 2453-2523.
21. Z. Weng, S. Teo and T. S. Andy Hor, *Dalton Trans*, 2007, 3493-3498.
22. C. J. Cramer and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2009, **11**, 10757-10816.