Electronic Supplementary Material (ESI) for Dalton Transactions

Screening of substituent and backbone effects on the properties of bidentate P,P-donor ligands (LKB-PP$_{screen}$)

Jesús Jover and Natalie Fey

Supporting Information

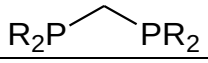
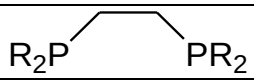
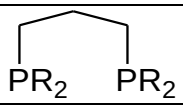
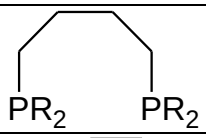
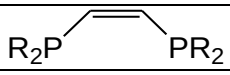
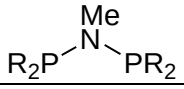
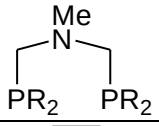
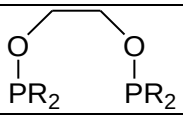
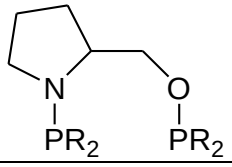
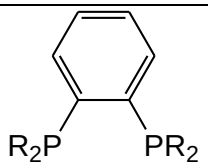
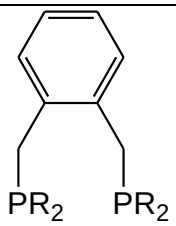
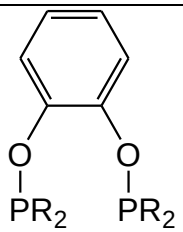
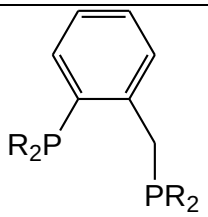
Index of Supporting Information

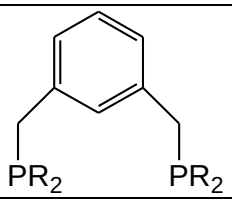
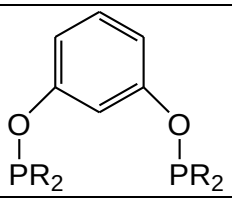
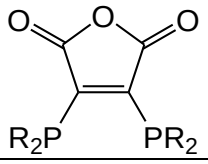
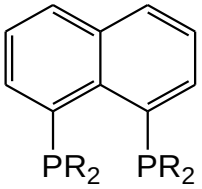
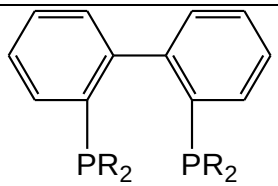
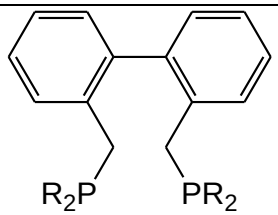
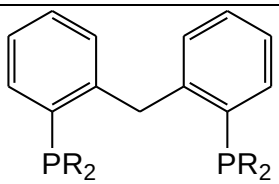
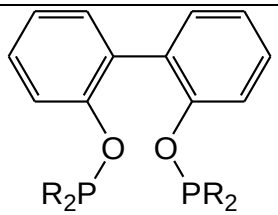
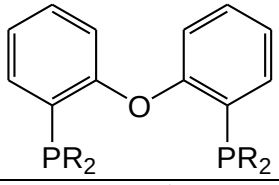
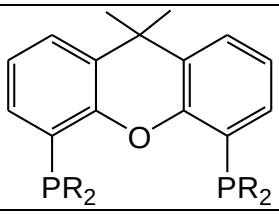
Note: The calculated descriptors for all LKB-PP_{screen} and new LKB-PP ligands reported in this work, as well as principal component scores for LKB-PP_{screen} are available as a separate *.xls file in the supporting information for this manuscript.

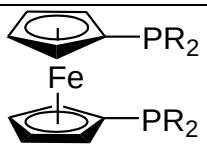
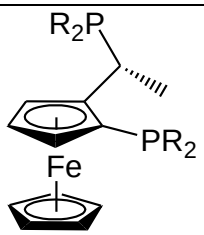
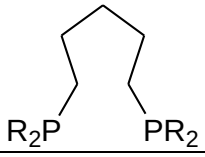
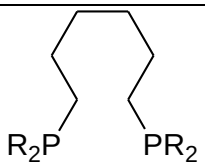
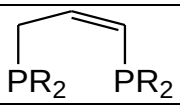
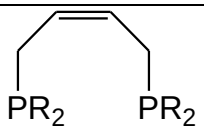
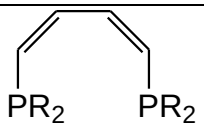
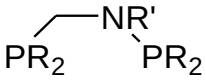
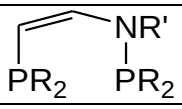
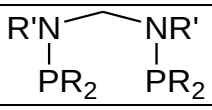
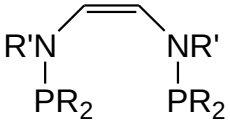
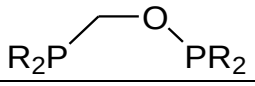
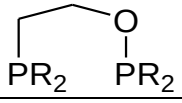
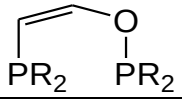
Table S1: Substituents chosen and representative references for their uses in homogeneous organometallic catalysis

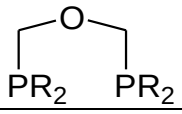
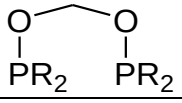
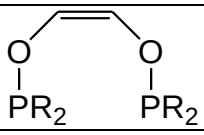
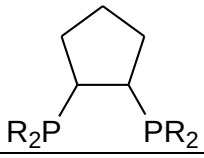
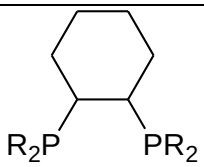
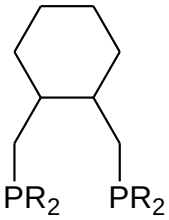
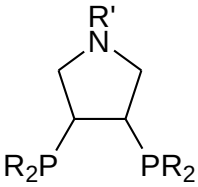
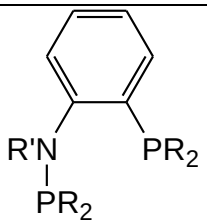
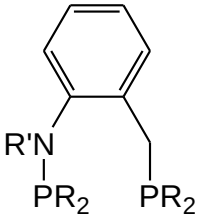
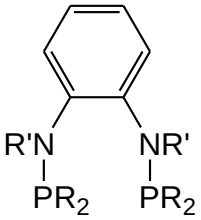
Substituent	**No.**	**Examples and References**
Methyl (Me)	(i)	symmetric catalysis,[1] cross-coupling reactions,[2] synthesis of luminescent compounds[3] and others.[4]
Trifluoromethyl (CF_3)	(ii)	Free[5] and coordinated [6] examples of ligands with this substituent have been published.
Tert-Butyl (tBu)	(iii)	α–Arylation of ketones,[7] ring opening metathesis polymerization[8] as well as in C–C and C–H activations.[9]
Phenyl (Ph)	(iv)	Most common substituent.[10]
Pentafluorophenyl (C_6F_5)	(v)	Alkyne cyclotrimerization[11] and transfer hydrogenation.[12]
Ortho-Tolyl (o-Tol)	(vi)	Rhodium-catalysed asymmetric hydrogenation of esters,[13] alkene oligomerisation[14] and copper catalysed C–N bond formation.[15]
Para-Tolyl (p-Tol)	(vii)	Nickel catalysed enantioselective Michael reactions,[16] ruthenium[17] and iridium catalysed asymmetric hydrogenations,[18] palladium catalysed C–N bond formation,[19] reductive amination of ketones[20] or platinum catalysed asymmetric Baeyer–Villiger oxidations.[21]
Para-trifluoromethylphenyl (p-CF_3-Ph)	(viii)	Iridium catalysed enantioselective hydrogenation of N-aryl imines,[22] platinum catalysed glyoxylate-ene reaction[23] and asymmetric Pauson–Khand reaction using rhodium.[24]
Methoxy (OMe)	(ix)	Oxidation of ketones,[25] asymmetric hydrogenation[26] and styrene hydroformylation.[27]
Phenoxy (OPh)	(x)	
Dimethylamino (NMe_2)	(xi)	Few examples for NMe_2.[28] Nevertheless, similar ligands have been designed with other alkyl (tBu or cyclohexyl) or aryl (Ph, N-Pyrrolyl) groups on the nitrogen atom and those were used in catalytic reactions such as C–H activation (Pt),[29] rhodium catalysed enantioselective hydrogenation of ketones,[26] Pt–Sn-catalysed styrene hydroformylation[27] and CO hydrogenation and transfer hydrogenation (Ru).[30]

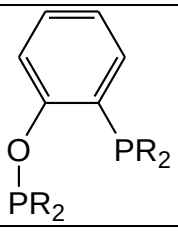
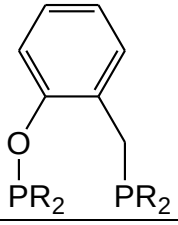
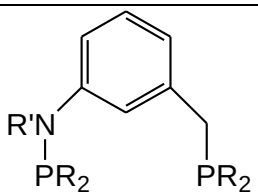
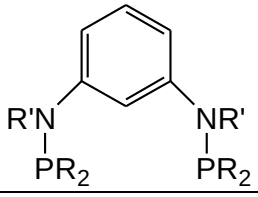
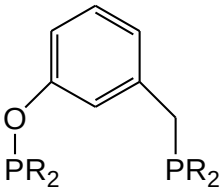
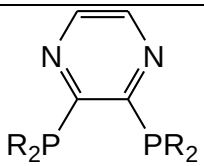
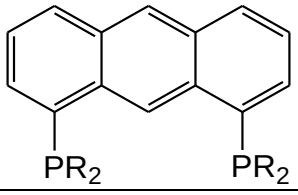
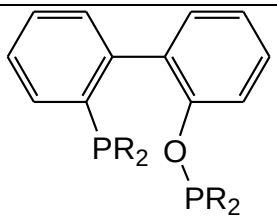
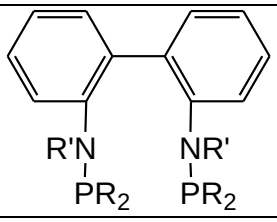
Table S2: Search structures used in CSD mining

No.	Structure	Number of hits	Selected
1		801	Yes
2		4356	Yes
3		1423	Yes
4		254	Yes
5		301	Yes
6		77	Yes
7		10	Yes
8		76	Yes
9		11	Yes
10		377	Yes
11		47	Yes
12		1	Yes, although only one hit can be found in the CSD several ligands bearing this backbone are known [31]
13		9	Yes

14		8	Yes
15		6	Yes, for comparison with 14.
16		21	Yes
17		27	Yes
18		237	Yes
19		13	Yes
20		7	Yes, it was considered to be relevant for comparison with 22, despite the low number of hits.
21		37	Yes
22		71	Yes
23		66	Yes

24		494	Yes
25		59	Yes
26		59	No, favours *trans*
27		1	No, too few
28		7	No, too few
29		2	No, too few
30		29	No, similar to 18
31		4	No, too few
32		7	No, too few
33		2	No, too few
34		9	No, too few
35		0	No
36		5	No, too few
37		10	No, too few
38		4	No, too few

39		0	No
40		0	No
41		1	No, too few
42		0	No
43		70	No, similar to 2
44		9	No, similar to 2
45		6	No, similar to 4
46		18	No, similar to 2
47		4	No, too few
48		0	No
49		3	No, too few

50		8	No, too few
51		0	No
52		0	No
53		9	No, too few
54		1	No, too few
55		5	No, too few
56		3	No, too few
57		2	No, too few
58		0	No

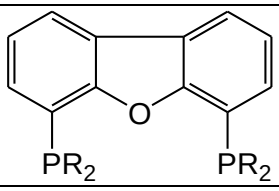
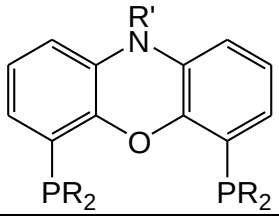
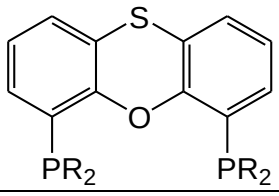
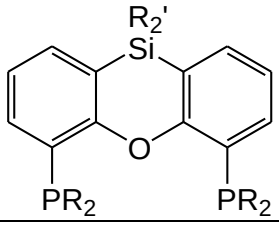
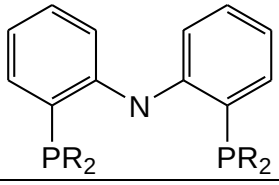
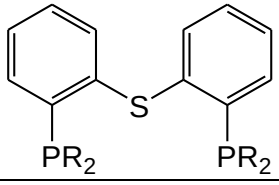
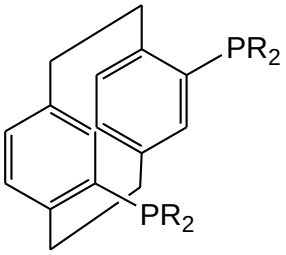
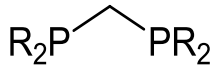
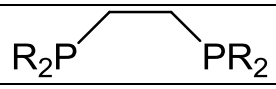
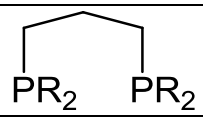
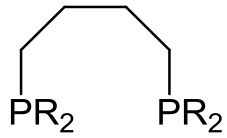
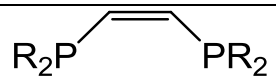
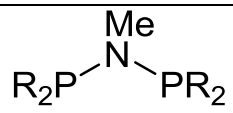
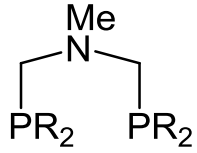
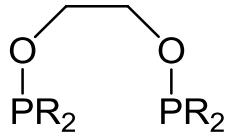
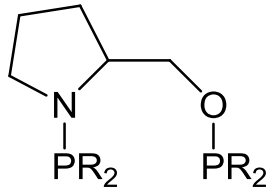
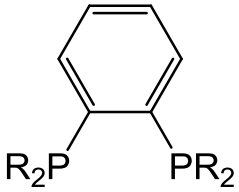
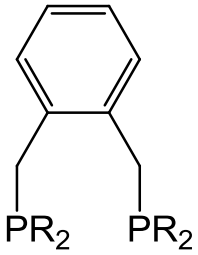
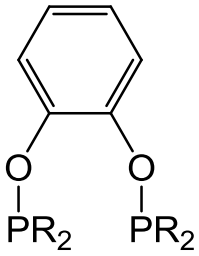
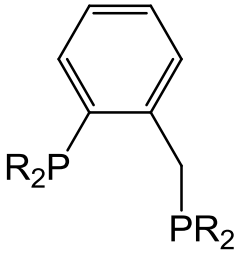
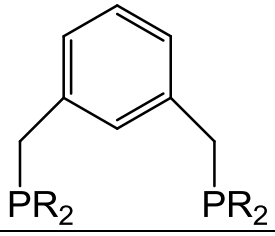
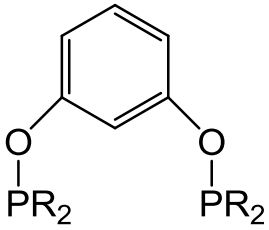
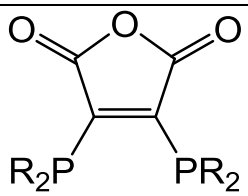
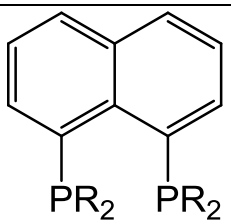
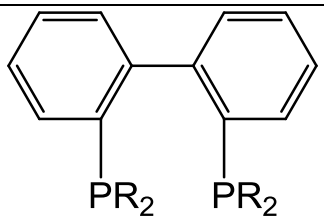
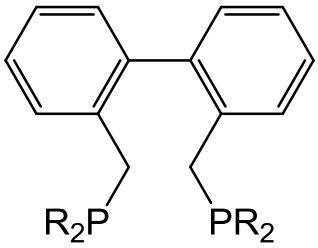
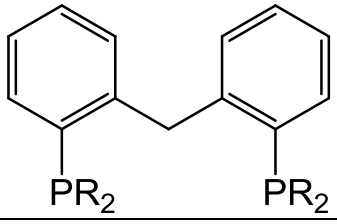
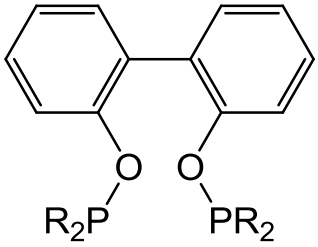
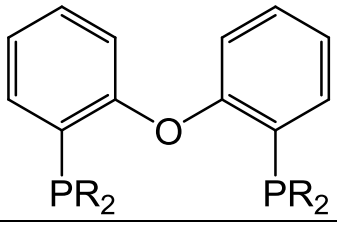
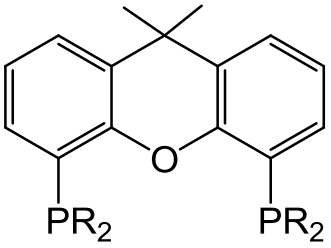
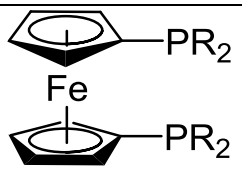
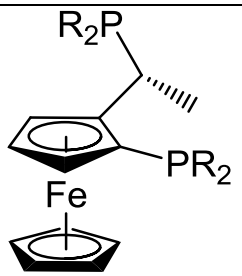
59		4	No, too few
60		2	No, too few
61		6	No, too few
62		3	No, too few
63		138	No, tridentate
64		5	No, too few
65		2	No, too few

Table S3: Backbones chosen and representative references for their uses in homogeneous organometallic catalysis

Backbone	Structure	Examples and References
A		**A**-**D** are members of the most used families in coordination studies[10a,32] but they have also been employed in a variety of cross-coupling,[33] allyllation,[34] hydroformylation,[35] carbonylation,[36] or Suzuki-Miyaura C–C bond formation[10g] catalytic reactions.
B		
C		
D		
E		The ethylene backbone (**E**) enforces small bite angles as well as high structural rigidity; ligands bearing this backbone have been used in catalysis,[34] as models for hydrogenases[37] or in hydrogen activation complexes.[38]
F		A considerable number of ligands containing the PNP backbone **F** have been developed; many different substituents can be found on the nitrogen atom, from alkyl groups to substituted carbon or nitrogen aromatic rings, but we have chosen methyl here because it will help to speed up the calculations as well as avoiding conformational problems. The ligands based on this backbone produce very small bite angles upon metal coordination[39] and their catalytic activity seems to focus mostly on the selective chromium-catalyzed tetramerization of ethylene.[40]
G		At first sight, backbone **G** could seem to be suitable to produce tridentate ligands but, in practice, it rarely does. Ligands based on this structure, where the methyl group has been selected as the substituent on the nitrogen atom for the same reasons as for **F** above, can form interesting complexes with transition metals.[41] In addition, these ligands can be employed in catalytic processes such as the Sonogashira coupling [42] or as models for hydrogenases. [43]
H		Depending on the nature of the substituents of the phosphorus atoms, bidentate phosphites, phosphonites or phosphinites can be obtained using backbone **H.** Ligands bearing this structure are mainly used in metal-catalyzed asymmetric reactions like palladium allylic alkylation[44] or rhodium hydrogenation.[45]
I		Backbone **I** is found mainly in chiral aminophosphine phosphonite bidentate ligands; many different substituents e.g.: alkyl, aryl, can be attached to the nitrogen or the two carbon atoms in the backbone, and here we selected the one bearing a five-membered ring between the nitrogen and the adjacent carbon atom as an average of the whole family. These ligands can be used in catalytic processes such as palladium allylic alkylation, platinum hydroformylation of styrene, ruthenium and iridium asymmetric hydrogenations or rhodium hydroboration.[46]

J		The 1,2-bis(diphosphino)benzene backbone **J** has been used regularly in coordination chemistry as can be deduced from the high number of hits (377) found in the CSD. In addition, it can be employed in iron[47] or nickel[48] catalysts of cross-coupling reactions involving halogenated aromatic rings.
K		Many catalytic applications have been found for ligands containing this backbone, e.g. ethene methoxycarbonylation,[49] methanol carbonylation,[50] rhodium asymmetric hydrogenation[51] and ethylene – carbon monoxide copolymerization.[52]
L		The main application of the ligands with this backbone seems to be coordination chemistry[31a,31d] but they can also be used in other processes such as chiral recognition[31b] or copper-catalyzed enantioselective conjugate addition.[31c]
M		Ligands bearing this backbone can be found mainly in hydrogenation catalysts,[53] some of them as a part of a ferrocene or chromium–arene systems.[54]
N		**N** and **O** tend to behave as pincer tricoordinated ligands, with the two P atoms adopting *trans* coordination positions around the metal and the central C deprotonated, but some examples can be found where they are cis coordianted.[55]
O		
P		Several examples of the coordination chemistry of different metal cations with ligands bearing the maleic anhydride backbone **P** have been reported.[56] More recently, phospholane derivatives of these ligands have been used in asymmetric homogeneous catalysis, mainly in hydrogenation[57] or hydroformylation[58] processes.
Q		Ligand with this backbone have been used in different catalytic reactions, e.g. amination of aryl bromides,[59] ethylene trimerisation and tetramerisation,[4d] as well as in coordination chemistry of transition metals.[60]

R		Numerous derivatives of ligands containing backbone **R**, e.g. BINAP, have proven to be good catalysts for asymmetric Bayer–Villiger oxidation,[21] hydrogenation,[17b] hydrosilylation[61] and hydroxylation[62] processes. The simplest phenyl derivative of this kind of ligands (DIPHEP) is often used in chirality control and discrimination studies.[51,63]
S		They have been widely used in homogeneous catalysis in hydroformylation[64] and hydrogenation[65] catalyzed reactions.
T		It is possible to find ligands with backbone , but, coordinated to metals, they are almost always tridentate unless the carbon between the two benzene rings is substituted.[66] There are some examples of coordination chemistry with nickel, palladium and iron where the C–H bond remains,[67] though.
U		The flexible ligands constructed using the bisphenol backbone (**U**) have been studied thoroughly at Union Carbide Corp. during the 1980s in the rhodium-catalyzed hydroformylation of terminal and internal alkenes.[68] More recently, they have been used for other purposes such as developing new reagents for the asymmetric Pauson–Khand reaction[69] or for carrying out intramolecular biphenyl metathesis.[70]
V		The catalytic applications of both ligand families tend to be very similar; they have been used in alkene hydroformylation,[71] hydrocyanation of styrene,[68] cross-coupling,[72], allylic alkylation,[73], amination of aryl arenes,[74] and ruthenium transfer hydrogenation.[75]
W		
X		Ligands based on backbone **X** have many catalytic applications such as palladium cross-coupling,[72] rhodium-catalyzed hydroformylation,[76] allylic alkylation,[73] and dehalogenation of arenes.[77]
Y		The JOSIPHOS family **Y**, discovered in the early 1990s, has proven to be a very useful source of ligands with a large range of applications in catalytic processes, e.g. hydroboration, hydrogenation, hydroformylation and carbonylation.[78]

T**able S4**: Matrix of ligands studied in this work, showing ligand numbers for each combination screened.

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)	(xi)
A	001	002	003	004	005	006	007	008	009	010	011
B	012	013	014	015	016	017	018	019	020	021	022
C	023	024	025	026	027	028	029	030	031	032	033
D	034	035	036	037	038	039	040	041	042	043	044
E	045	046	047	048	049	050	051	052	053	054	055
F	056	057	058	059	060	061	062	063	064	065	066
G	067	068	069	070	071	072	073	074	075	076	077
H	078	079	080	081	082	083	084	085	086	087	088
I	089	090	091	092	093	094	095	096	097	098	099
J	100	101	102	103	104	105	106	107	108	109	110
K	111	112	113	114	115	116	117	118	119	120	121
L	122	123	124	125	126	127	128	129	130	131	132
M	133	134	135	136	137	138	139	140	141	142	143
N	144	145	146	147	148	149	150	151	152	153	154
O	155	156	157	158	159	160	161	162	163	164	165
P	166	167	168	169	170	171	172	173	174	175	176
Q	177	178	179	180	181	182	183	184	185	186	187
R	188	189	190	191	192	193	194	195	196	197	198
S	199	200	201	202	203	204	205	206	207	208	209
T	210	211	212	213	214	215	216	217	218	219	220
U	221	222	223	224	225	226	227	228	229	230	231
V	232	233	234	235	236	237	238	239	240	241	242
W	243	244	245	246	247	248	249	250	251	252	253
X	254	255	256	257	258	259	260	261	262	263	264
Y	265	266	267	268	269	270	271	272	273	274	275

Table S5: Truncated ligands used.

Truncation	Symmetrical ligand ID	Unsymmetrical ligand ID
TR1, PMe_3	1, 12, 23, 34, 67, 111, 144, 199	133, 265
TR2, $PMe_2(CHCH_2)$	45, 166	
TR3, PMe_2Ph	100, 177, 188, 210, 232, 243, 254	133, 265
TR4, $PMe_2(OMe)$	78, 122, 155, 221	89
TR5, $PMe_2(NMe_2)$	56	89
TR6, $PtBu_2Me$	3, 14, 25, 36, 69, 113, 146, 201	135, 267
TR7, $PtBu_2(CHCH_2)$	47, 168	
TR8, $PtBu_2Ph$	102, 179, 190, 212, 234, 245, 256	135, 267
TR9, $PtBu_2(OMe)$	80, 124, 157, 223	91
TR10, $PtBu_2(NMe_2)$	58	91
TR11, $P(CF_3)_2Me$	2, 13, 24, 35, 68, 112, 145, 200	134, 266
TR12, $P(CF_3)_2(CHCH_2)$	46, 167	
TR13, $P(CF_3)_2Ph$	101, 178, 189, 211, 233, 244, 255	134, 266
TR14, $P(CF_3)_2(OMe)$	79, 123, 156, 222	90
TR15, $P(CF_3)_2(NMe_2)$	57	90
TR16, PPh_2Me	4, 15 26, 37, 114, 147, 202	136, 268
TR17, $PPh_2(CHCH_2)$	48, 169	
TR18. PPh_3	103, 180, 191, 213, 235, 246, 257	136, 268
TR19, $PPh_2(OMe)$	81, 125, 158, 224	92
TR20, PPh2(NMe_2)	59	92
TR21, P(*o*-Tol)$_2$Me	6, 17, 28, 39, 72, 116, 149, 204	138, 270
TR22, P(*o*-Tol)$_2$($CHCH_2$)	50, 171	
TR23, P(*o*-Tol)$_2$Ph	105, 182, 193, 215, 237, 248, 259	138, 270
TR24, P(*o*-Tol)$_2$(OMe)	83, 127, 160, 226	94
TR25, P(*o*-Tol)$_2$(NMe_2)	61	94
TR26, P(*p*-Tol)$_2$Me	7, 18, 29, 40, 73, 117, 150, 205	139, 271
TR27, P(*p*-Tol)$_2$($CHCH_2$)	51, 172	
TR28, P(*p*-Tol)$_2$Ph	106, 183, 194, 216, 238, 249, 260	139, 271
TR29, P(*p*-Tol)$_2$(OMe)	84, 128, 161, 227	95
TR30, P(*p*-Tol)$_2$(NMe_2)	62	95
TR31, P(*p*-CF_3-Ph)$_2$Me	8, 19, 30, 41, 74, 118, 151, 206	140, 272
TR32, P(*p*-CF_3-Ph)$_2$($CHCH_2$)	52, 173	
TR33, P(*p*-CF_3-Ph)$_2$Ph	107, 184, 195, 217, 239, 250, 261	140, 272
TR34, P(*p*-CF_3-Ph)$_2$(OMe)	85, 129, 162, 228	96
TR35, P(*p*-CF_3-Ph)$_2$(NMe_2)	63	96
TR36, $P(C_6F_5)_2Me$	5, 16, 27, 38, 71, 115, 148, 203	137, 269
TR37, $P(C_6F_5)_2(CHCH_2)$	49, 170	
TR38, $P(C_6F_5)_2Ph$	104, 181, 192, 214, 236, 247, 258	137, 269
TR39, $P(C_6F_5)_2(OMe)$	82, 126, 159, 225	93
TR40, $P(C_6F_5)_2(NMe_2)$	60	93
TR41, $P(OMe)_2Me$	9, 20, 31, 42, 75, 119, 151, 207	141, 273
TR42, $P(OMe)_2(CHCH_2)$	53, 174	
TR43, $P(OMe)_2Ph$	108, 185, 196, 218, 240, 251, 262	141, 273
TR44, $P(OMe)_3$	86, 130, 163, 229	97
TR45, $P(OMe)_2(NMe_2)$	64	97
TR46, $P(OPh)_2Me$	10, 21, 32, 43, 76, 120, 153, 208	142, 274
TR47, $P(OPh)_2(CHCH_2)$	54, 175	
TR48, $P(OPh)_2Ph$	109, 186, 197, 222, 241, 252, 263	142, 274

TR49, $P(OPh)_2(OMe)$	87, 131, 164, 230	98
TR50, $P(OPh)_2(NMe_2)$	65	98
TR51, $P(NMe_2)_2Me$	11, 22, 33, 44, 77, 121, 154, 209	143, 275
TR52, $P(NMe_2)_2(CHCH_2)$	55, 176	
TR53, $P(NMe_2)_2Ph$	110, 187, 198, 220, 242, 253, 264	143, 275
TR54, $P(NMe_2)_2(OMe)$	88, 132, 165, 231	99
TR55, $P(NMe_2)_3$	66	99

Input file generation

The combination of 25 selected backbones with 11 different substituents produced the 275 ligand set used to sample the bidentate phosphorus ligand space, Table S4. Building the ligands from the different backbones and substituents is reasonably straightforward and could be achieved in any standard builder. In previous work on LKB-PP,[79] Molecular Mechanics (MM) conformational searches were then employed to find reliable starting structures for the free ligand and its corresponding $Zn(PP)Cl_2$ coordinated species; as we used a default force field, manual selection of input geometries was still required. However, the number of files required here led us to consider further simplifications, with a view to facilitating more extensive automation in the future.

Generation of descriptor data required 6 optimisations[1] per ligand with DFT for 275 ligands, as well as optimisation of a total of 55 truncated ligands (Table S5), i.e. 1705 calculations in total. In the standard LKB-PP, we would also include conformational searches on both free ligands and ligands complexed to the $ZnCl_2$ fragment with molecular mechanics (MM), which here would have meant an additional 550 conformational searches of 500 steps each. While the computational cost of these searches might become increasingly acceptable, especially with the much faster MM approach, manual analysis and selection of suitable input geometries (or alternatively the development of a better force field) hamper automation and thus make this a substantial undertaking. We decided instead to apply a molecular fragment orientation methodology that allows us to combine backbones and substituents in a rapid and effective way.

MM conformational searches were initiated to obtain appropriately aligned backbones and substituents in low energy conformations. Backbones for free and coordinated bidentate ligands were extracted from PPh_2-backbone-PPh_2 (PP) and $Zn(PP)Cl_2$; the phenyl substituent can be considered, in terms of electronics and sterics, as an average of the eleven substituents selected. The substituent conformations were obtained in a similar way, *i.e.* by carrying out conformational searches on free and $ZnCl_2$ coordinated PR_2-CH_2CH_2-PR_2 (backbone **B**) species. In this case the ethyl bridge seemed an appropriate choice, as it can be considered a representative intermediate of the selected backbones in terms of strain/flexibility and bite angle, while its small size will improve the speed of the MM conformational searches. Both sets of backbones and substituents can be oriented using the alignment options available in Maestro[80] to produce the final fragment geometries, from which all other atoms can then be deleted. These fragments can then be combined with appropriate input sections (including metal/He_8 fragments where relevant) to build the complete set of free and chelated ligand structures. This procedure requires only a total of 72 MM conformational searches (50 for the backbones and 22 for the

[1] Free ligand, ligand in chelating conformation (see below), zinc and palladium complexes as well as two He_8 wedge adducts with different constraints as described above.

substituents, free and chelating), instead of 550 that would have been needed if we used the same methodology as described for LKB-PP.[79]

This improves the speed of setting up calculations significantly and is very amenable to automation, but it can compromise the quality of the data due to not reliably using low energy starting conformers for some ligands. Nevertheless, the data included in LKB-PP can be used to check the goodness of the results obtained for ligands in both databases and to measure the conformational noise of this combinatorial approach, see below. The results required for descriptors were extracted with in-house scripts and processed with standard spreadsheet software. Details of calculation settings and software references can be found in the Computational Details section.

Calculation of steric descriptors

In the He_8_wedge calculation, four of the helium atoms are positioned to mimic the interactions between bidentate ligands and the donor atoms of other ligands in an octahedral complex (Scheme 2), with four further helium atoms inserted between these positions to create a more even "steric surface" in the absence of extended fragment bulk. In these systems, the ligand geometry is copied from the optimization of [$ZnCl_2${LL}], "metal"-donor distances are adjusted to idealized values (X-P = 2.28 Å), the dummy atom is removed, the positions of both donor atoms and all helium atoms are frozen, and the rest of the ligand geometry is optimized. The interaction energy is then calculated with reference to the lowest energy conformation found for the free ligand.

For ligands which adopted very large bite angles in the zinc complex used, where the donor-donor distances exceeded 2 x 2.28 Å, X-P distances were set as close to the idealized value as possible when calculating He_8_wedge.

Calculations of nHe_8 are derived from the same starting geometry (Scheme 2) as He_8_wedge. Instead of freezing the positions of donor atoms, a dummy "metal" atom (X) is placed at the metal position and the dummy-donor atom distances are constrained to the same distances as used for He_8_wedge (X-P = 2.28 Å). Helium and dummy atom positions are frozen. The nHe_8 descriptor is calculated as the interaction energy between the optimized system and the free ligand in the lowest energy conformer found.

Representative input files for both calculations are shown below, where # indicates a frozen coordinate:

He_8_wedge

```
&gen
igeopt=1
dftname=bp86
basis=6-31g*
numd=5
iaccg=3
ip11=2
isymm=0
&
&zmat
P3       1.7179010000000#    0.0000000000000#   -1.4990720000000#
P4      -1.7178340000000#    0.0000010000000#   -1.4998110000000#
C5      -0.6665260000000   -0.3255480000000   -3.0221170000000
C6      -3.0302980000000   -1.3002550000000   -1.6290720000000
C7      -2.5987750000000    1.5785110000000   -1.9110440000000
C8       2.4862350000000   -1.6173070000000   -1.9802570000000
C9       3.1162200000000    1.2107020000000   -1.5885420000000
C10      0.6750020000000    0.4453950000000   -2.9949000000000
```

```
H11    -0.4808750000000  -1.4160910000000  -3.0499620000000
H12    -1.2377030000000  -0.0752000000000  -3.9366940000000
H13    -3.5026700000000  -1.3246750000000  -2.6272450000000
H14    -3.8052300000000  -1.1075620000000  -0.8682650000000
H15    -2.5792520000000  -2.2818630000000  -1.4063000000000
H16    -3.1315290000000   1.5263180000000  -2.8774210000000
H17    -1.8728020000000   2.4085390000000  -1.9321460000000
H18    -3.3223960000000   1.8002640000000  -1.1084010000000
H19     3.0082050000000  -1.5622670000000  -2.9525380000000
H20     1.7067150000000  -2.3964250000000  -2.0211350000000
H21     3.2032140000000  -1.9187290000000  -1.1981880000000
H22     3.5897070000000   1.2385820000000  -2.5859390000000
H23     3.8757070000000   0.9345010000000  -0.8375910000000
H24     2.7368460000000   2.2132620000000  -1.3291460000000
H25     1.2453290000000   0.2614430000000  -3.9257240000000
H26     0.4915860000000   1.5353780000000  -2.9401960000000
He      0.000#    2.28#     0.000#
He      0.000#   -2.28#     0.000#
He      1.612#    0.00#     1.612#
He     -1.612#    0.00#     1.612#
He      0.806#    1.14#     0.806#
He     -0.806#    1.14#     0.806#
He     -0.806#   -1.14#     0.806#
He      0.806#   -1.14#     0.806#
&
```

nHe$_8$

```
&gen
igeopt=1
dftname=bp86
basis=6-31g*
numd=5
iaccg=3
ip11=2
isymm=0
&
&zmat
X1      0.0#    0.0#    0.0#
P2      1.7179010000000   0.0000000000000  -1.4990720000000
P3     -1.7178340000000   0.0000010000000  -1.4998110000000
C5     -0.6665260000000  -0.3255480000000  -3.0221170000000
C6     -3.0302980000000  -1.3002550000000  -1.6290720000000
C7     -2.5987750000000   1.5785110000000  -1.9110440000000
C8      2.4862350000000  -1.6173070000000  -1.9802570000000
C9      3.1162200000000   1.2107020000000  -1.5885420000000
C10     0.6750020000000   0.4453950000000  -2.9949000000000
H11    -0.4808750000000  -1.4160910000000  -3.0499620000000
H12    -1.2377030000000  -0.0752000000000  -3.9366940000000
H13    -3.5026700000000  -1.3246750000000  -2.6272450000000
H14    -3.8052300000000  -1.1075620000000  -0.8682650000000
H15    -2.5792520000000  -2.2818630000000  -1.4063000000000
H16    -3.1315290000000   1.5263180000000  -2.8774210000000
H17    -1.8728020000000   2.4085390000000  -1.9321460000000
H18    -3.3223960000000   1.8002640000000  -1.1084010000000
H19     3.0082050000000  -1.5622670000000  -2.9525380000000
H20     1.7067150000000  -2.3964250000000  -2.0211350000000
H21     3.2032140000000  -1.9187290000000  -1.1981880000000
H22     3.5897070000000   1.2385820000000  -2.5859390000000
H23     3.8757070000000   0.9345010000000  -0.8375910000000
H24     2.7368460000000   2.2132620000000  -1.3291460000000
H25     1.2453290000000   0.2614430000000  -3.9257240000000
H26     0.4915860000000   1.5353780000000  -2.9401960000000
He      0.000#    2.28#     0.000#
He      0.000#   -2.28#     0.000#
He      1.612#    0.00#     1.612#
He     -1.612#    0.00#     1.612#
He      0.806#    1.14#     0.806#
```

```
He      -0.806#    1.14#      0.806#
He      -0.806#   -1.14#      0.806#
He       0.806#   -1.14#      0.806#
&
&coord
X1    P2    #   2.28
X1    P3    #   2.28
&
```

The optimisation for the nHe_8 adducts of ligands 52, 74, 140, 184, 217, 228, 250, 260 did not converge when P-X distances were constrained. Instead, a linear regression model was fitted to the remaining 267 ligands and this model was used to predict the nHe_8 descriptor for these ligands. Model details are shown below:

R^2	0.868
Adj. R^2	0.866
Standard error	4.498
Intercept	-1124.46
BE (Pd)	-0.5171
Pd-Cl	500.4469
ΔP2-R(Pd)	120.4895
ΔPd–P2	81.8691

Conformers

Conformational freedom is considerable, especially for the free ligands, and we were initially concerned that the approach used to generate input geometries from fragments as described above could give rise to poor input geometries, which would be more likely to lead to high energy conformers. Having a low energy conformer for the free ligand is particularly important as it is the reference used to derive the binding energies of the metallic fragments and most of the structural descriptors.

Two different conformers have been optimised for the free ligand: one obtained from the combination of backbone and substituent fragments as described above, and the other derived from the chelating conformation as optimised for the zinc complex, which had been built from fragments and then optimised. In our LKB-PP investigation, we showed that exploring these two different starting structures for the free ligands is able to identify low-energy conformers.[14] In most cases the chelating conformer is likely to be higher in energy than the free ligand minimum, but some calculations show this to be the other way around here, suggesting problems with the MM calculations, possibly due to a lack of force field parameters for electronic stabilisation of certain conformers[27] or incomplete conformational searching. Where optimisation from the chelating starting geometry gave lower energies, this conformer has been used to derive descriptors. 161 ligands favour the free ligand conformation and for 59 of those the chelating geometry is more than 2 kcal mol^{-1} higher in energy. Of the 114 ligands that favour the chelating input geometry, 51 had a free ligand geometry which was more than 2 kcal mol^{-1} higher in energy (20 with more than 5 kcal mol^{-1}), suggesting that there are indeed potential problems with this approach and conformational noise levels are likely to be higher. Figure S1 shows the energy difference between the two possible starting geometries after DFT optimisation.

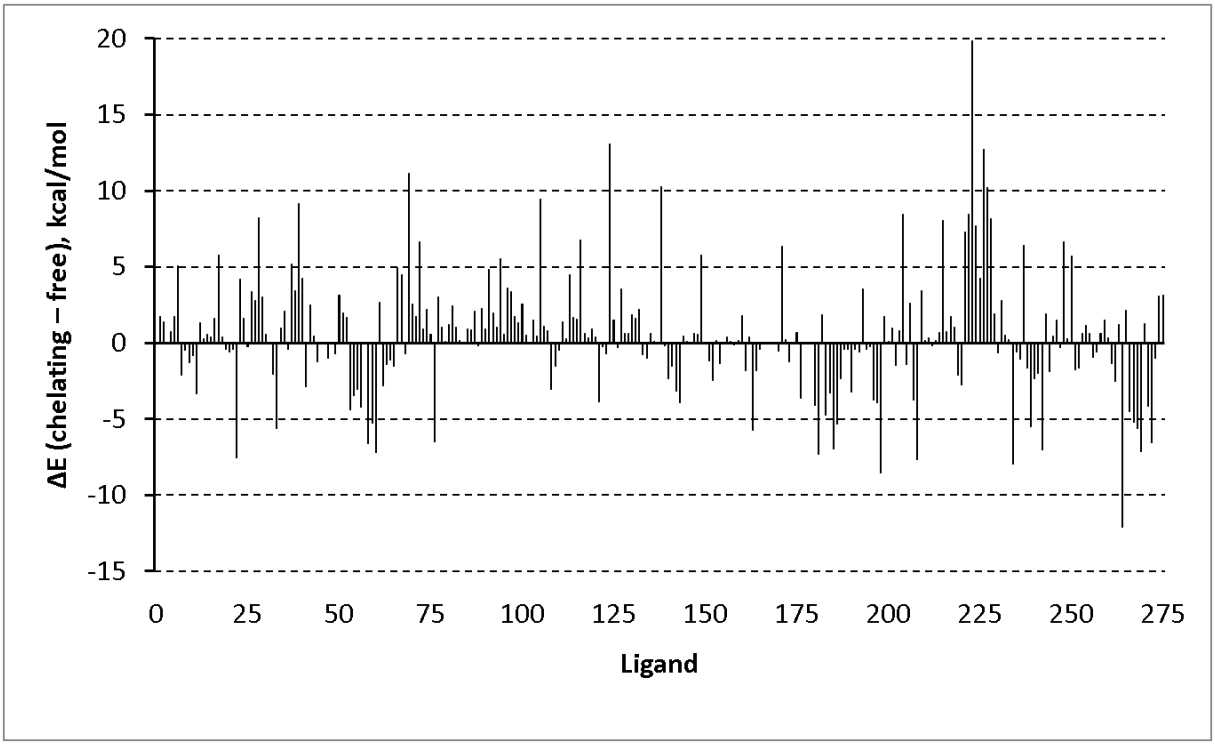


Figure S1. DFT-calculated energy differences between the free and chelating ligand conformations built from fragments (see section on input file generation above); negative values favour the chelating conformer.

We expected that the fragment building procedure used in this work, with only limited conformational searches, might have a negative impact on the quality of the descriptors generated. Where ligands were present in both this new database, LKB-PP$_{screen}$, and LKB-PP,[14] the energies of different fragments could be compared to determine whether this is indeed a problem. The average absolute differences in energy between the structures in LKB-PP$_{screen}$ and LKB-PP are less than 2 kcal mol^{-1} for all the structures optimised in both knowledge bases, and Table S6 summarises these results in detail.

Table S6. Error statistics for DFT-calculated energy differences (in kcal mol^{-1}) between data for ligands reported previously in LKB-PP[14] (full conformational searches) and LKB-PP$_{screen}$ (structures built from fragments, see section on input file generation above).

Complex	Mean error (signed)	Mean absolute error	Max. error	Min. error	STD
Free ligand	0.24	0.43	2.69	-2.53	0.81
[$ZnCl_2$PP]	1.02	1.26	14.13	-1.10	2.88
[$PdCl_2$PP]	0.17	1.06	12.97	-4.16	2.70
He_8_wedge.PP	0.81	1.68	8.50	-5.84	2.74
nHe_8.PP	0.71	0.94	11.20	-1.40	2.21

For LKB-PP, the average conformational noise was estimated to be around 1-2 kcal mol^{-1}.[14] This suggests that most of the starting geometries built in this work are not significantly worse than the structures obtained after more extensive MM conformational searches of bidentate ligands, although some very large deviations are observed as well. These are for a small subset of ligands and it is worth bearing in mind here that structural parameters and their derived descriptors are less sensitive than ligand binding energies to conformational changes, so trends in ligand effects should be captured reasonably well, even if a higher energy conformer is used.

Table S7: Descriptor loadings for PC1-4.

Descriptor	Label in Fig. S1	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Cumulative contribution, %		29.5	55.7	65.3	73.7	79.6	84.9	88.1	90.5
R^2X		0.295	0.262	0.096	0.085	0.059	0.053	0.031	0.024
Q^2		0.209	0.326	0.011	0.135	0.041	0.184	-0.102	-0.003
Q^2 (cum)		0.209	0.467	0.473	0.544	0.562	0.643	0.607	0.606
E_{HOMO_P1}	homo1	0.203	0.251	0.077	-0.003	0.034	0.286	-0.102	0.067
E_{HOMO_P2}	homo2	0.204	0.253	0.092	-0.012	0.037	0.284	-0.092	-0.005
E_{LUMO_P1}	lumo1	0.150	0.116	-0.119	-0.190	-0.556	0.045	0.184	-0.123
E_{LUMO_P2}	Lumo2	0.157	0.114	-0.124	-0.174	-0.556	0.039	0.174	-0.161
PA_{P1}	pa1	0.171	0.270	0.179	0.052	0.127	0.224	-0.109	0.037
PA_{P2}	pa2	0.169	0.270	0.189	0.045	0.128	0.222	-0.100	0.003
He_8_wedge	he.w.pn	-0.182	0.258	-0.089	-0.210	0.029	0.011	0.031	-0.033
nHe_8	nhe	-0.145	0.290	0.071	-0.120	-0.015	0.065	0.074	0.224
BE(Zn)	be.gla	0.179	0.229	0.056	0.230	-0.005	-0.123	-0.059	-0.261
Zn-Cl	ml.gla	-0.022	0.008	0.034	0.169	0.282	0.230	0.882	-0.176
<P1-Zn-P2	dmd.gla	-0.137	0.278	-0.108	-0.126	0.121	-0.294	0.061	-0.018
ΔP1-R(Zn)	d.d1r.gla	-0.203	0.029	0.319	0.136	-0.243	0.018	0.033	0.336
ΔP2-R(Zn)	d.d2r.gla	-0.198	0.035	0.333	0.157	-0.243	-0.069	0.041	-0.068
ΔR-P1-R(Zn)	drd1rgla	0.042	-0.028	0.372	-0.373	0.148	-0.158	-0.018	-0.409
ΔR-P2-R(Zn)	drd2rgla	0.096	-0.055	0.220	-0.460	0.121	-0.114	0.132	0.340
ΔZn-P1	dpzn.d1	-0.206	-0.095	-0.219	-0.154	-0.039	0.396	-0.015	0.189
ΔZn-P2	dpzn.d2	-0.240	-0.085	-0.149	-0.108	-0.060	0.383	-0.030	-0.217
Q(Zn)	nbo.glaf	-0.190	-0.168	0.259	0.226	-0.008	0.202	-0.047	-0.035
BE(Pd)	be.pd	0.289	-0.023	0.019	0.215	-0.054	-0.192	0.025	-0.081
Pd-Cl	ml.pd	0.193	0.236	0.074	0.052	-0.018	-0.057	0.182	0.244
<P1-Pd-P2	dmd.pd	-0.171	0.229	-0.157	-0.111	0.124	-0.336	0.115	0.076
ΔP1-R(Pd)	d.d1r.pd	-0.250	0.116	0.263	0.093	-0.164	-0.057	0.049	0.136
ΔP2-R(Pd)	d.d2r.pd	-0.252	0.097	0.270	0.112	-0.185	-0.101	0.041	-0.096
ΔR-P1-R(Pd)	d.rd1rpd	0.084	-0.207	0.327	-0.256	0.010	0.057	-0.028	-0.299
ΔR-P2-R(Pd)	d.rd2rpd	0.128	-0.209	0.205	-0.333	-0.030	0.085	0.097	0.218
ΔPd-P1	dppdd1	-0.180	0.277	-0.038	-0.144	0.045	0.078	-0.073	-0.102
ΔPd-P2	dppdd2	-0.193	0.275	0.003	-0.102	0.054	0.070	-0.102	-0.232
Q(Pd)	nbo.pd.f	-0.332	-0.013	0.000	-0.043	0.076	0.018	-0.048	-0.131

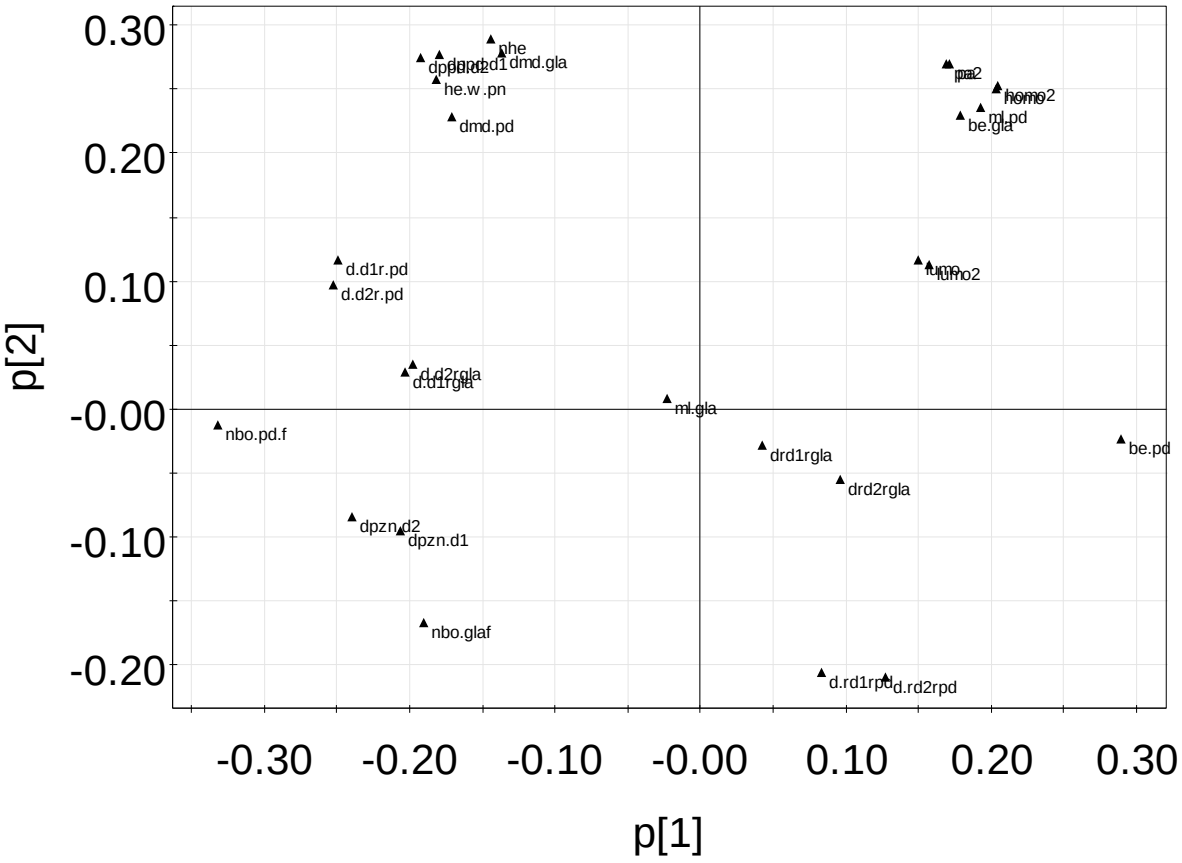


Figure S2: PC loadings plot

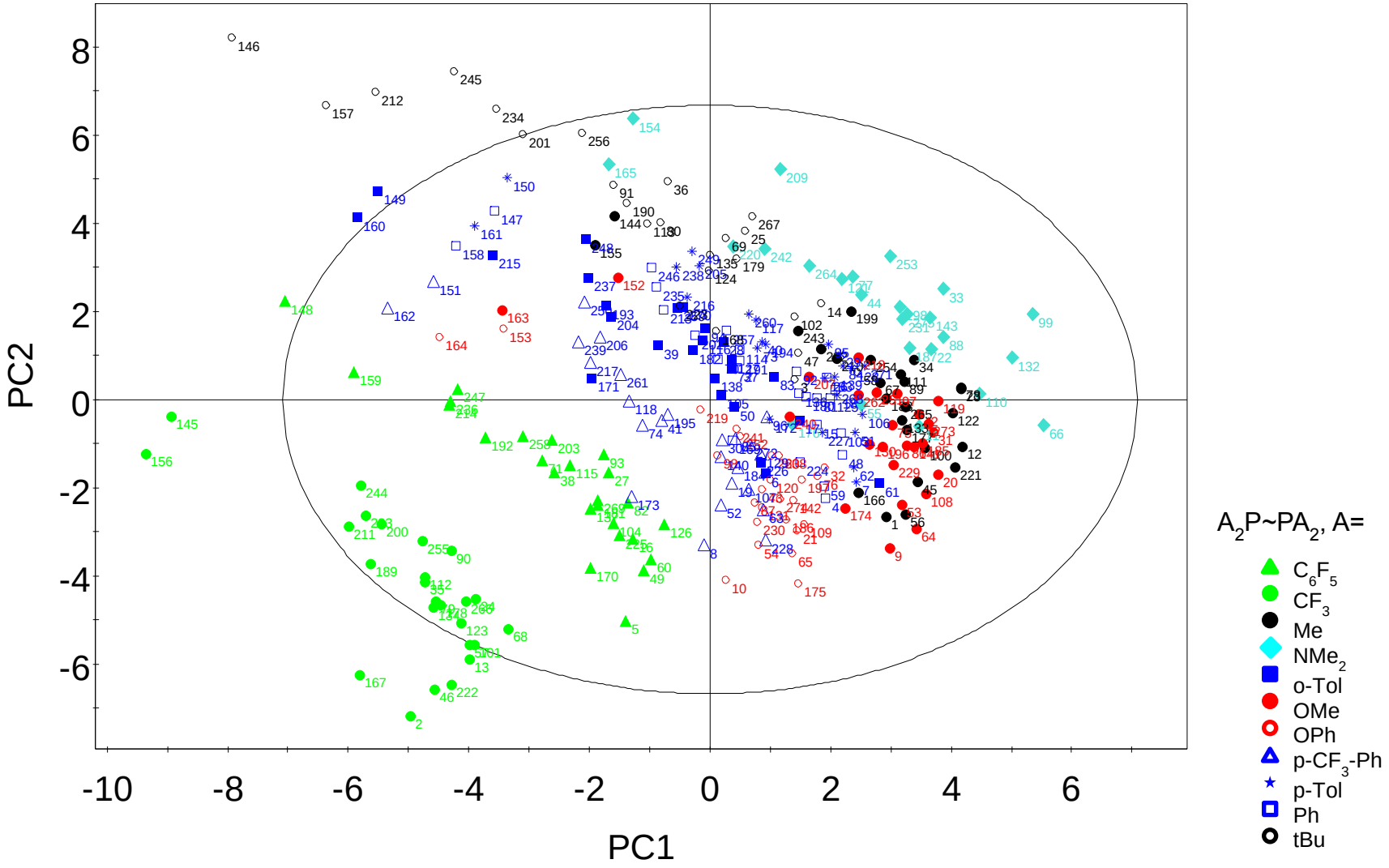


Figure S3: Principal component plot (PC1 and PC2) with colour coding by substituent, capturing ca. 56 % of variation in data.

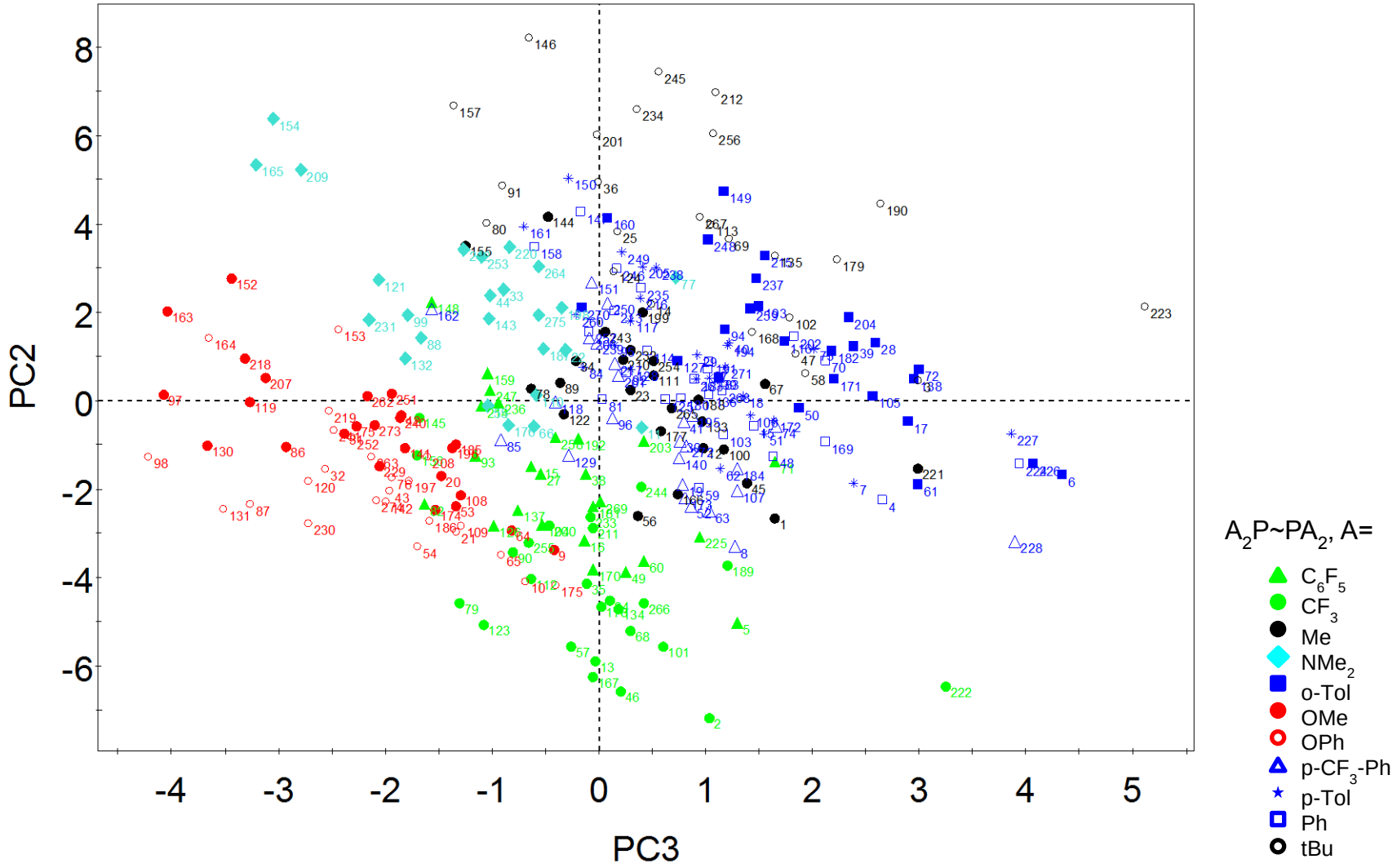


Figure S4: Principal component plot (PC2 and PC3) with colour coding by substituent.

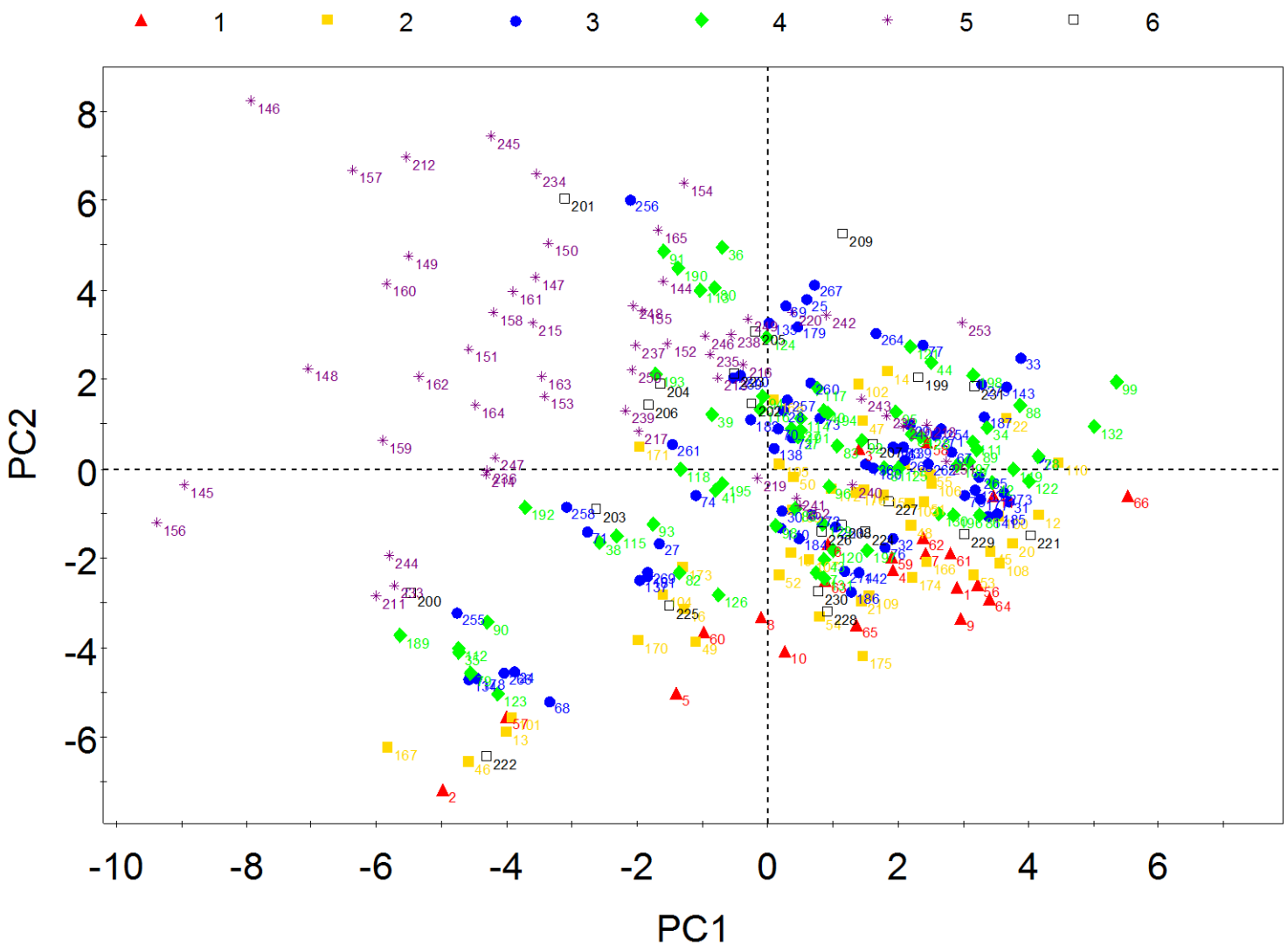


Figure S5: Principal component plot (PC1, PC2) with colour coding by backbone length (see legend at top of plot).

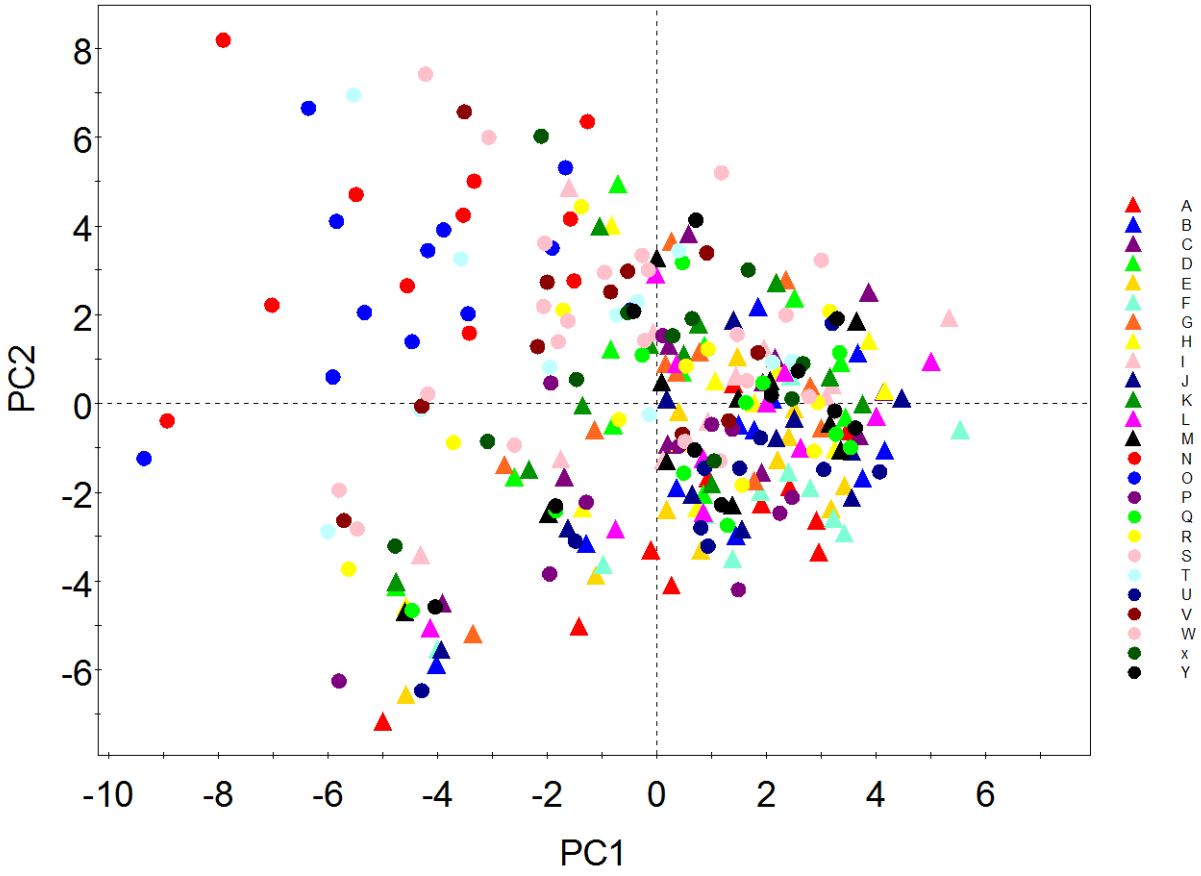


Figure S6: Principal component plot (PC1, PC2) with colour coding by backbone identity (see Scheme 1).

Table S8: Descriptor loadings LKB-PP as published and after inclusion of nHe_8 descriptor, showing first four PCs only.

Var ID (Primary)	PC1 (LKB-PP)	PC2 (LKB-PP)	PC3 (LKB-PP)	PC4 (LKB-PP)	PC1 (incl. nHe_8)	PC2 (incl. nHe_8)	PC3 (incl. nHe_8)	PC4 (incl. nHe_8)
% contribution	33.1	28.0	10.3	6.2	32.8	31.1	10.0	5.7
E_{HOMO_P1}	0.303	-0.057	0.000	-0.037	0.279	0.122	-0.006	-0.105
E_{HOMO_P2}	0.308	-0.062	-0.004	0.001	0.282	0.126	-0.010	-0.045
E_{LUMO_P1}	0.078	-0.096	0.438	0.174	0.062	0.095	0.431	0.276
E_{LUMO_P2}	0.061	-0.090	0.459	0.113	0.047	0.088	0.454	0.168
PA_{P1}	0.298	-0.033	-0.069	-0.034	0.281	0.096	-0.070	-0.111
PA_{P2}	0.304	-0.039	-0.075	0.016	0.285	0.102	-0.077	-0.031
He_8_wedge	0.149	0.269	0.212	0.031	0.198	-0.218	0.193	0.052
nHe_8					0.209	-0.222	0.118	0.015
BE(Zn)	0.286	-0.019	-0.131	0.108	0.267	0.077	-0.149	0.185
Zn-Cl	0.309	0.076	-0.103	0.066	0.311	-0.008	-0.123	0.136
<P1-Zn-P2	0.180	0.255	0.101	-0.053	0.223	-0.195	0.073	0.022
ΔP1-R(Zn)	-0.145	0.159	-0.143	0.187	-0.111	-0.188	-0.148	0.253
ΔP2-R(Zn)	-0.141	0.147	-0.221	0.236	-0.112	-0.176	-0.226	0.422
ΔR-P1-R(Zn)	0.021	-0.158	0.203	0.068	-0.006	0.142	0.200	0.166
ΔR-P2-R(Zn)	0.078	-0.185	0.199	0.301	0.044	0.179	0.202	0.283
ΔZn-P1	-0.170	0.187	0.257	0.013	-0.125	-0.205	0.265	-0.118
ΔZn-P2	-0.188	0.166	0.279	-0.099	-0.144	-0.190	0.287	-0.215
Q(Zn)	-0.259	0.022	-0.218	0.051	-0.248	-0.076	-0.205	0.053
BE(Pd)	0.132	-0.276	-0.164	0.026	0.071	0.278	-0.169	0.118
Pd-Cl	0.280	-0.134	-0.053	0.077	0.243	0.177	-0.062	0.141
<P1-Pd-P2	0.150	0.270	0.090	-0.108	0.197	-0.213	0.064	-0.036
ΔP1-R(Pd)	0.020	0.211	-0.127	0.527	0.009	-0.279	-0.095	0.299
ΔP2-R(Pd)	0.023	0.209	-0.133	0.544	0.019	-0.260	-0.111	0.414
ΔR-P1-R(Pd)	-0.132	-0.253	0.159	0.248	-0.174	0.198	0.181	0.236
ΔR-P2-R(Pd)	-0.106	-0.264	0.162	0.281	-0.152	0.216	0.186	0.203
ΔPd-P1	0.161	0.293	0.142	0.009	0.214	-0.233	0.128	-0.033
ΔPd-P2	0.166	0.282	0.122	0.040	0.216	-0.223	0.106	0.040
Q(Pd)	-0.104	0.326	0.061	-0.042	-0.037	-0.316	0.053	-0.063

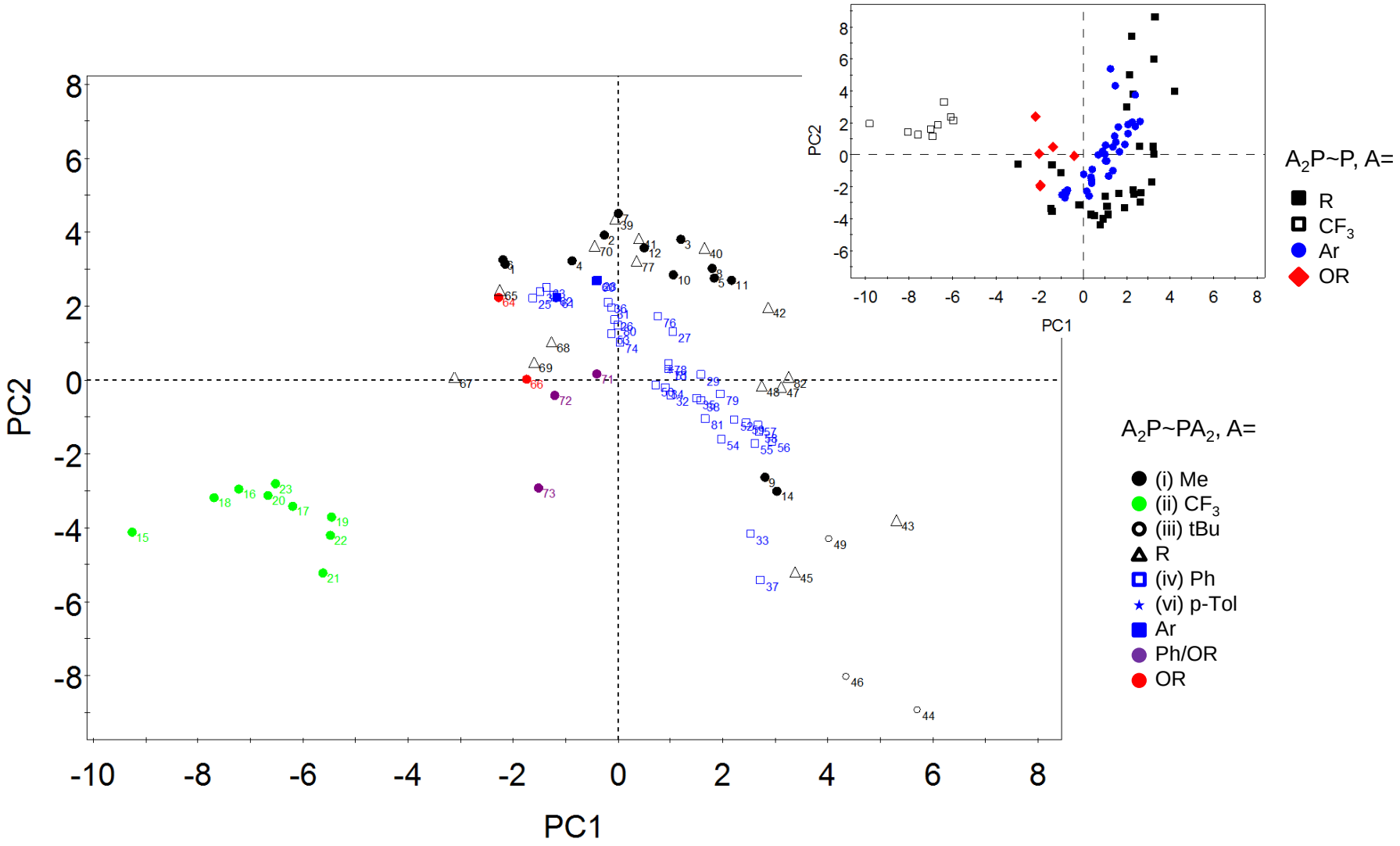
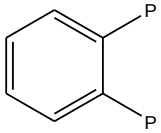
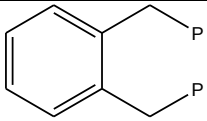
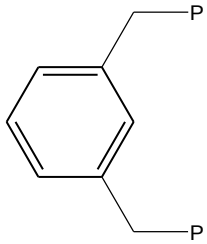
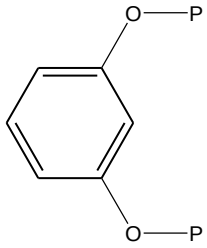
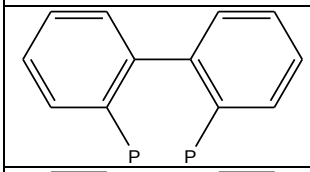
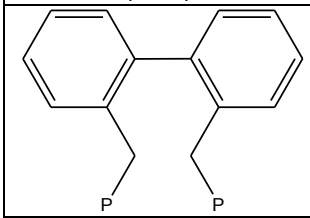
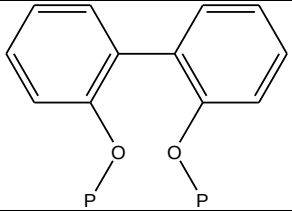
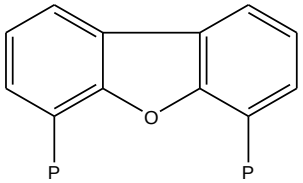
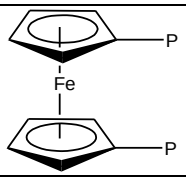


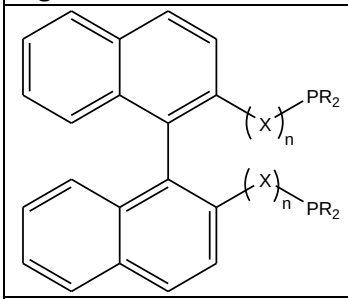
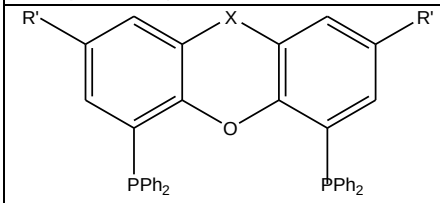
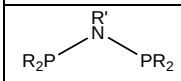
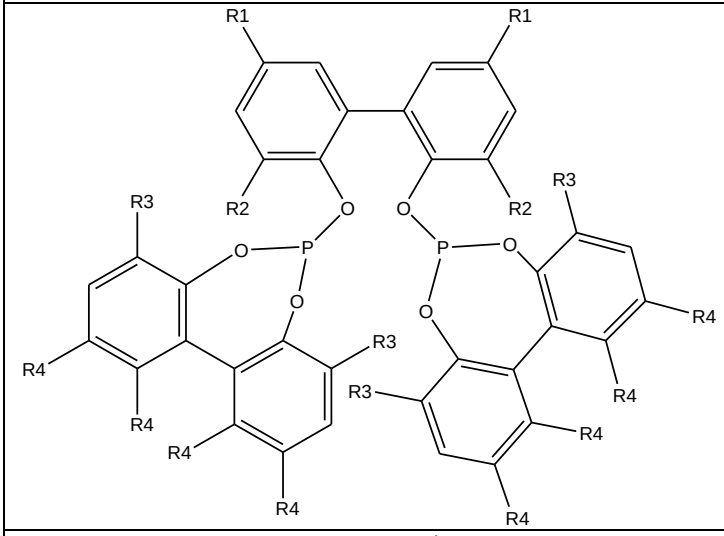
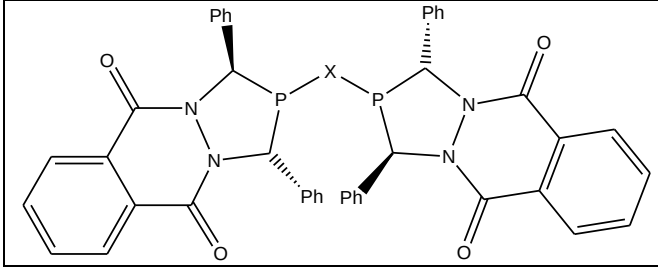
Figure S7: Principal component score plot (PC1, PC2, capturing 64 % of variation in data) for LKB-PP, including nHe_8 parameter and data as previously published.[79] Colour coding by substituent. The inset shows the PC plot for LKB-PP excl. nHe_8, which inverts the sign of PC2 ligand scores but maintains the general shape and spatial relationships between ligands. See Table S9 for ligand numbering.

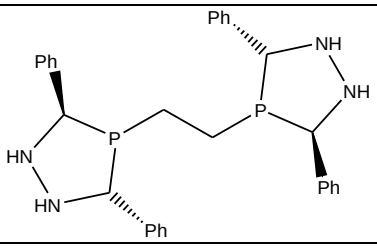
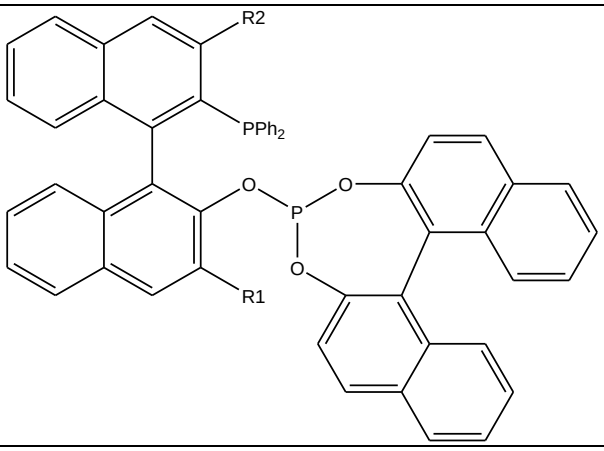
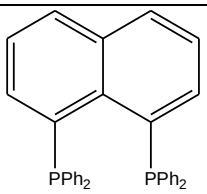
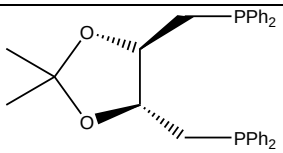
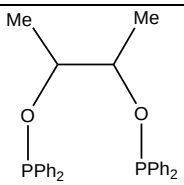
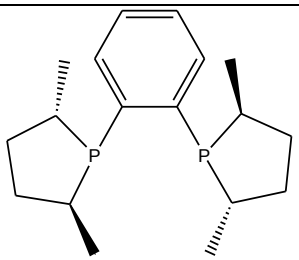
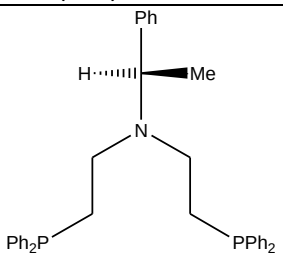
Table S9: Ligands as numbered in LKB-PP.[79]

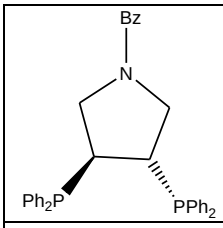
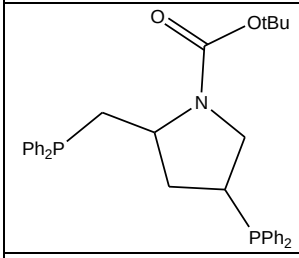
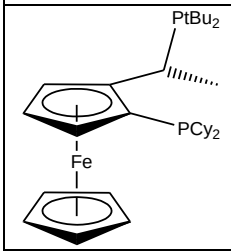
P⌒P	**Me**	**CF_3**	**Ph**	**Other**
PCH_2P	1	15	25	Cy 39
$P(CH_2)_2P$	2	16	26	Cy 40, Et 41
$P(CH_2)_3P$	3	17	27	Cy 42
PCH=CHP	4	18	28	
$PCH_2CH=CHCH_2P$	5	19	29	
P(NH)P	6		30	
	7	20	31	
	8		32	
				Cy 43, ^{t}Bu 44
	9		33	Cy 45, ^{t}Bu 46
	10	21	34	
	11	22	35	
	12	23	36	
	14		37	

			38	Cy 47, ^{i}Pr 48, ^{t}Bu 49

(Ligands 13 and 24 are persistent outliers due to a rigid backbone structure and have been removed from consideration and hence this list.)

Ligand Structure	Numbers
	50 n=0, R=Ph 51 n=0, R=p-Me-C_6H_4 52 n=1, R=Ph 53 n=1, X=O, R=Ph
	54 no X, R'=H 55 X=CMe_2, R'=H 56 X=CMe_2, R'=^{t}Bu 57 X=S, R'=Me 58 X=C=CMe_2, R'=H 59 X=NH, R'=H
	60 R=(o-MeO-C_6H_4), R'=Me 61 R=(o-Et-C_6H_4), R'=Me 62 R=Ph, R'=^{i}Pr 63 R=Ph, R'=Ph
	64 R1, R2, R3, R4 = H 65 R1=OMe, R2=^{t}Bu, R3, R4=H 66 R1, R2=H, R3=^{t}Bu, R4=Me
	67 X=CH_2 68 X=$(CH_2)_2$ 69 X=Ph

	70
	71 R1, R2=H 72 R1=SO_2CF_3, R2=H 73 R1, R2=SO_2CF_3
	74
	75
	76
	77
$Ph_2P(CH_2)_4PPh_2$	78
	79

	80
	81
	82

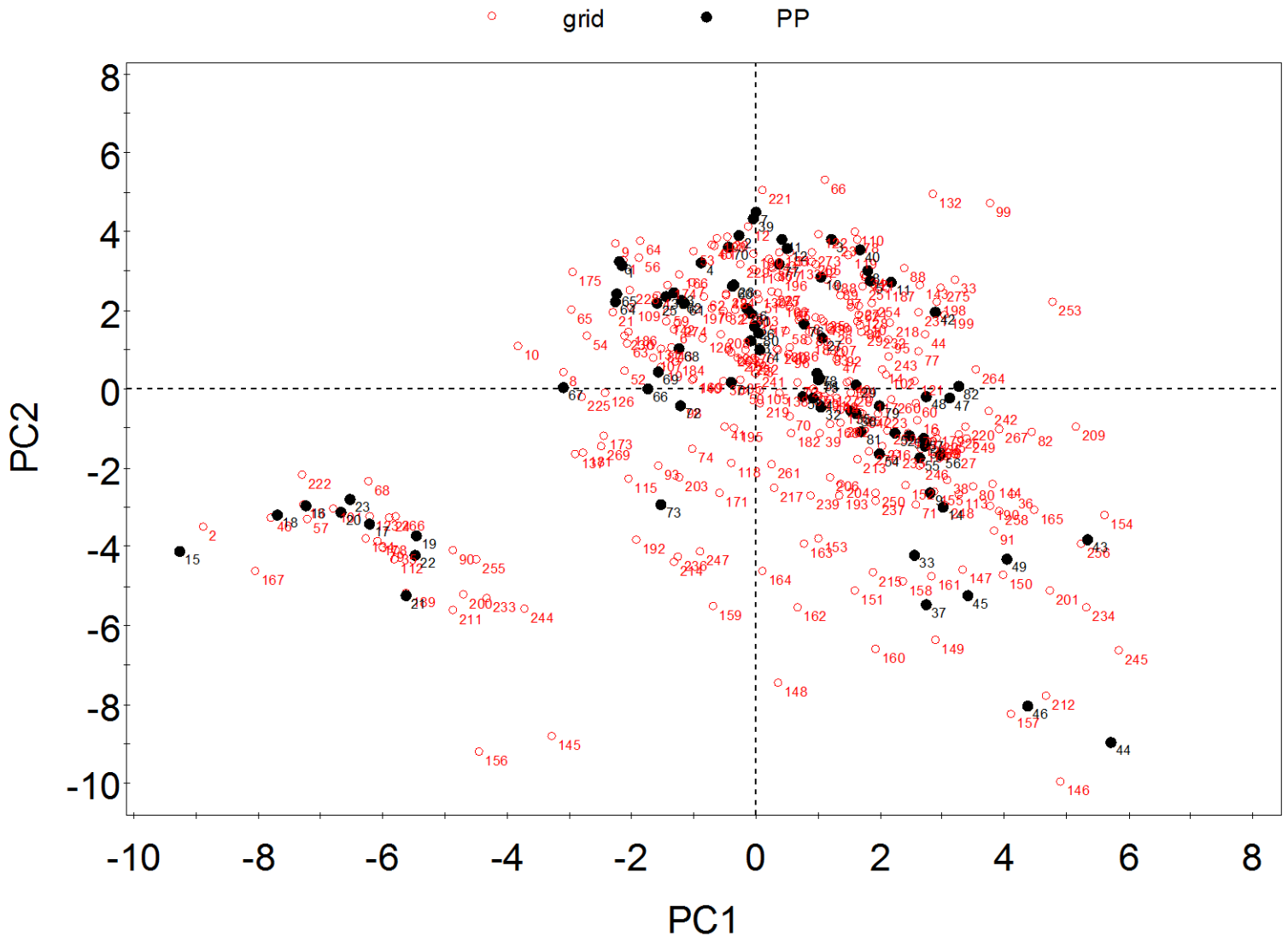


Figure S8: Principal component score plot (PC1, PC2), PCA for LKB-PP as Fig. 7 used to predict scores for LKB-PP$_{screen}$ ligands. See Table S9 for LKB-PP numbering (black dots) and Table S4 for LKB-PP$_{screen}$ numbering (red circles).

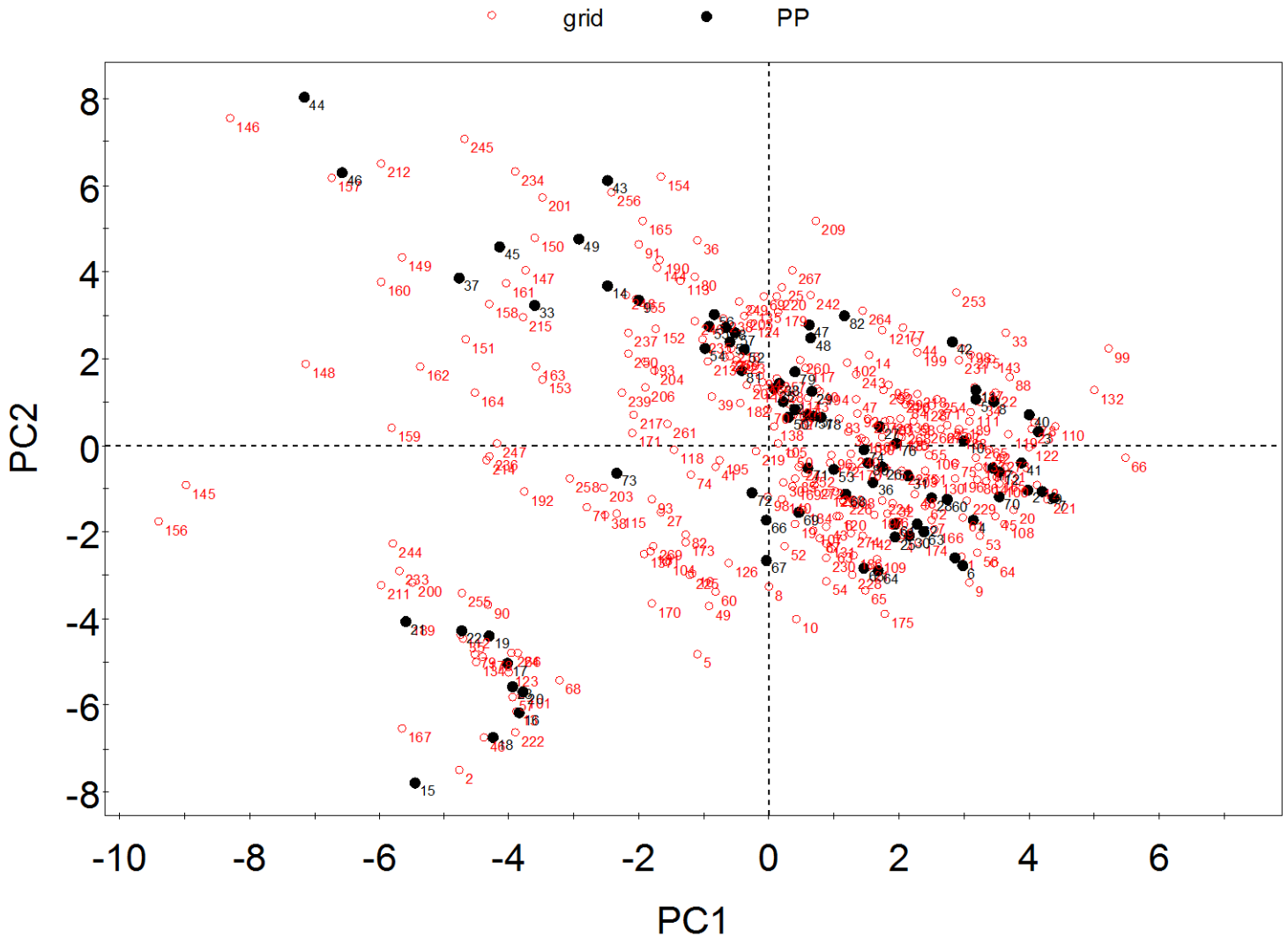


Figure S9: Principal component score plot (PC1, PC2), PCA on combined dataset. See Table S9 for LKB-PP numbering (black dots) and Table S4 for LKB-PP$_{screen}$ numbering (red circles).

Table S10: Descriptor loadings for PCA on combined dataset, showing first 8 PCs (11 PCs selected by auto-fit).

Descriptor	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Cumulative contribution, %	29.4	56.0	65.0	73.5	79.3	84.5	87.6	90.2
R^2X	0.294	0.266	0.090	0.085	0.058	0.052	0.032	0.026
Q^2	0.201	0.338	0.060	0.084	0.072	0.182	-0.130	0.017
Q^2 (cum)	0.201	0.471	0.503	0.545	0.577	0.654	0.620	0.626
E_{HOMO_P1}	0.191	0.263	0.064	-0.014	0.016	0.290	-0.104	0.044
E_{HOMO_P2}	0.193	0.264	0.080	-0.028	0.026	0.279	-0.090	-0.028
E_{LUMO_P1}	0.134	0.100	-0.139	-0.294	-0.524	-0.011	0.189	-0.111
E_{LUMO_P2}	0.138	0.095	-0.154	-0.279	-0.531	0.000	0.172	-0.119
PA_{P1}	0.161	0.277	0.168	0.019	0.094	0.253	-0.112	0.018
PA_{P2}	0.160	0.278	0.182	0.007	0.104	0.239	-0.098	-0.031
He_8_wedge	-0.190	0.250	-0.126	-0.169	0.056	-0.005	0.027	0.029
nHe_8	-0.162	0.278	0.027	-0.123	0.015	0.064	0.037	0.204
BE(Zn)	0.160	0.243	0.133	0.158	-0.071	-0.116	-0.020	-0.250
Zn-Cl	-0.016	0.036	0.063	0.179	0.224	0.221	0.910	-0.112
<P1-Zn-P2	-0.148	0.275	-0.123	-0.060	0.123	-0.299	0.052	-0.003
ΔP1-R(Zn)	-0.203	-0.005	0.357	-0.009	-0.233	0.046	0.012	0.379
ΔP2-R(Zn)	-0.199	-0.001	0.394	0.011	-0.205	-0.076	0.026	-0.089
ΔR-P1-R(Zn)	0.066	-0.029	0.238	-0.424	0.282	-0.170	-0.012	-0.399
ΔR-P2-R(Zn)	0.120	-0.037	0.089	-0.441	0.260	-0.115	0.102	0.398
ΔZn-P1	-0.210	-0.097	-0.253	-0.119	-0.031	0.381	-0.006	0.168
ΔZn-P2	-0.233	-0.098	-0.186	-0.126	-0.049	0.363	-0.017	-0.211
Q(Zn)	-0.171	-0.193	0.303	0.121	-0.043	0.219	-0.023	-0.096
BE(Pd)	0.291	-0.004	0.083	0.182	-0.105	-0.186	0.040	-0.062
Pd-Cl	0.191	0.237	0.091	0.038	-0.034	-0.064	0.139	0.231
<P1-Pd-P2	-0.178	0.232	-0.168	-0.022	0.122	-0.336	0.093	0.056
ΔP1-R(Pd)	-0.255	0.092	0.297	-0.033	-0.161	-0.049	0.042	0.170
ΔP2-R(Pd)	-0.256	0.078	0.312	-0.022	-0.146	-0.116	0.039	-0.115
ΔR-P1-R(Pd)	0.111	-0.207	0.221	-0.341	0.091	0.050	-0.001	-0.262
ΔR-P2-R(Pd)	0.150	-0.203	0.100	-0.360	0.060	0.075	0.081	0.231
ΔPd-P1	-0.189	0.271	-0.071	-0.125	0.053	0.080	-0.055	-0.073
ΔPd-P2	-0.196	0.269	-0.018	-0.109	0.067	0.061	-0.069	-0.222
Q(Pd)	-0.329	-0.015	-0.016	-0.044	0.088	0.011	-0.025	-0.149

Table S11: Descriptor loadings for PCA on LKB-PP expanded by 22 ligands selected in this work, showing all 8 PCs selected by software.

Descriptor	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Cumulative contribution, %	31.9	59.0	67.6	74.5	80.2	85.3	88.6	91.3
R^2X	0.319	0.270	0.086	0.070	0.057	0.051	0.033	0.027
Q^2	0.209	0.342	-0.003	0.038	0.038	0.107	-0.199	0.007
Q^2 (cum)	0.209	0.479	0.477	0.497	0.516	0.568	0.524	0.528
E_{HOMO_P1}	0.158	-0.281	0.031	0.074	0.186	-0.238	-0.029	-0.101
E_{HOMO_P2}	0.158	-0.286	0.022	0.093	0.198	-0.168	0.009	-0.154
E_{LUMO_P1}	0.041	-0.114	0.426	0.264	-0.370	-0.060	0.246	0.029
E_{LUMO_P2}	0.035	-0.102	0.446	0.228	-0.379	-0.144	0.192	0.072
PA_{P1}	0.178	-0.260	-0.050	0.080	0.226	-0.219	-0.050	-0.085
PA_{P2}	0.178	-0.267	-0.061	0.106	0.240	-0.131	0.000	-0.150
He_8_wedge	0.293	0.055	0.189	-0.039	0.032	0.079	-0.092	0.004
nHe_8	0.302	0.038	0.101	0.067	0.013	-0.042	0.001	-0.080
BE(Zn)	0.175	-0.239	-0.137	0.057	-0.049	0.129	0.073	0.062
Zn-Cl	0.103	-0.059	-0.075	0.008	0.312	0.087	0.662	0.578
<P1-Zn-P2	0.292	0.022	0.045	-0.161	-0.114	0.262	-0.090	0.070
ΔP1-R(Zn)	0.083	0.196	-0.128	0.418	-0.020	-0.139	-0.286	0.169
ΔP2-R(Zn)	0.070	0.183	-0.238	0.419	-0.072	0.120	0.105	-0.287
ΔR-P1-R(Zn)	-0.104	-0.080	0.197	0.082	0.220	0.525	0.159	-0.431
ΔR-P2-R(Zn)	-0.084	-0.141	0.186	0.199	0.261	0.338	-0.393	0.374
ΔZn-P1	0.023	0.254	0.283	0.017	0.193	-0.269	-0.182	0.075
ΔZn-P2	-0.022	0.256	0.308	-0.089	0.190	-0.177	0.095	-0.168
Q(Zn)	-0.134	0.229	-0.203	0.221	0.133	-0.158	0.246	-0.044
BE(Pd)	-0.107	-0.283	-0.168	0.031	-0.232	0.064	-0.041	0.024
Pd-Cl	0.123	-0.284	-0.075	0.135	-0.123	0.033	-0.079	0.020
<P1-Pd-P2	0.281	0.047	0.019	-0.213	-0.181	0.249	-0.065	0.039
ΔP1-R(Pd)	0.206	0.198	-0.094	0.318	-0.050	0.019	-0.095	0.163
ΔP2-R(Pd)	0.203	0.180	-0.138	0.309	-0.025	0.222	0.072	-0.087
ΔR-P1-R(Pd)	-0.259	-0.030	0.205	0.205	0.177	0.169	0.058	-0.146
ΔR-P2-R(Pd)	-0.254	-0.060	0.195	0.206	0.169	0.107	-0.164	0.206
ΔPd-P1	0.307	0.059	0.143	-0.054	0.118	-0.013	-0.041	0.021
ΔPd-P2	0.304	0.047	0.110	-0.027	0.152	0.095	0.033	-0.093
Q(Pd)	0.154	0.294	0.040	-0.093	0.084	0.082	0.069	0.011

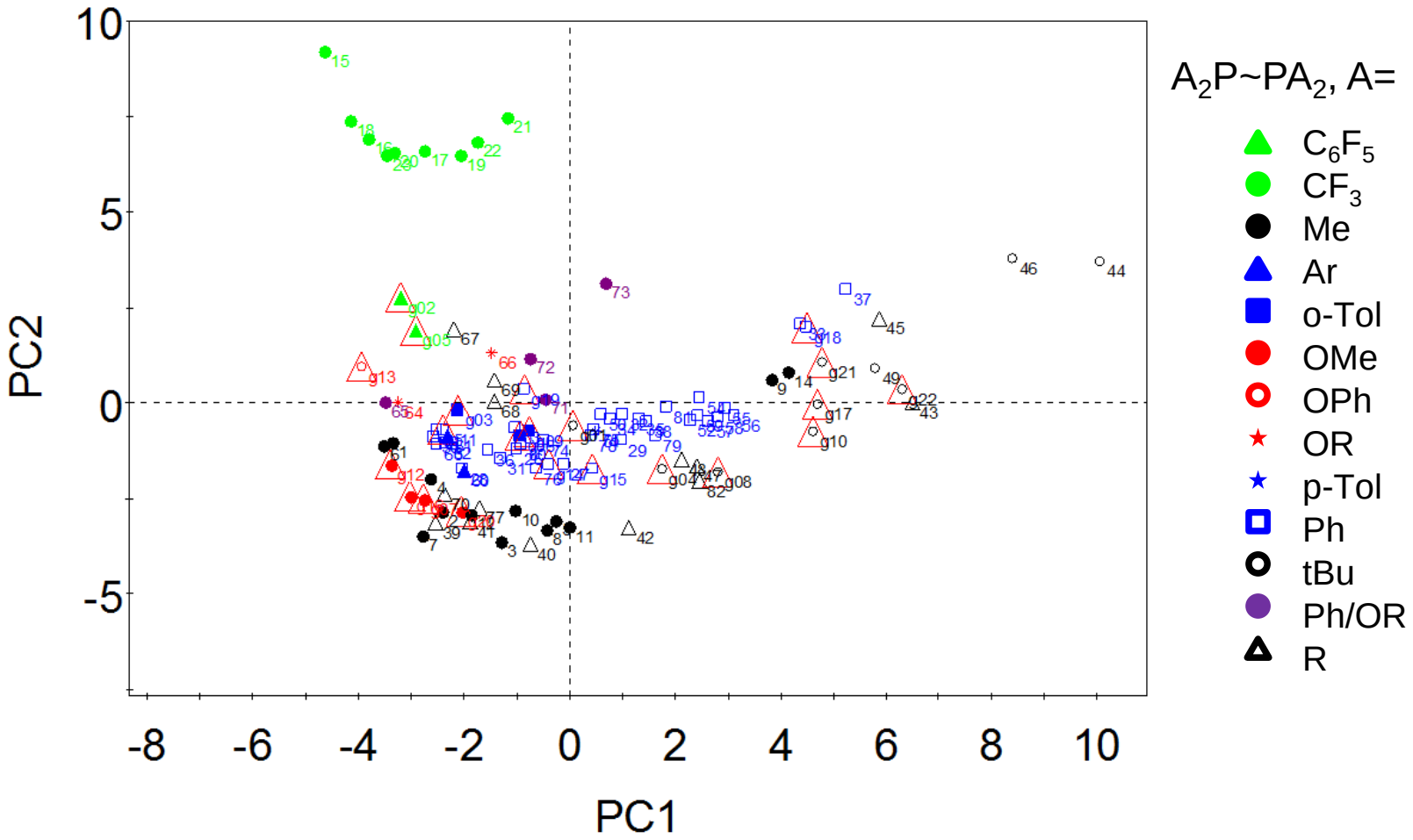


Figure S10. Principal component score plot (PC1, PC2, capturing 59% of variation in data) for LKB-PP after addition of 22 ligands selected from LKB-PP_{screen} results and CSD mining. Showing colours according to substituents, with new ligands inside red triangles.

References

1. V. S. Chan, M. Chiu, R. G. Bergman and F. D. Toste, *J. Am. Chem. Soc.*, 2009, **131**, 6021-6032.
2. M. Park, J. R. Buck and C. J. Rizzo, *Tetrahedron*, 1998, **54**, 12707-12714.
3. S. J. Messersmith, K. Kirschbaum and J. R. Kirchhoff, *Inorg. Chem.*, 2010, **49**, 3857-3865.
4. a) S. S. M. Ling, R. J. Puddephatt, L. Manojlovic-Muir and K. W. Muir, *J. Organomet. Chem.*, 1983, **255**, C11-C14; b) F. Wu and R. F. Jordan, *Organometallics*, 2006, **25**, 5631-5637; c) A. Yahav-Levi, I. Goldberg, A. Vigalok and A. N. Vedernikov, *J. Am. Chem. Soc.*, 2007, **130**, 724-731; d) M. J. Overett, K. Blann, A. Bollmann, R. de Villiers, J. T. Dixon, E. Killian, M. C. Maumela, H. Maumela, D. S. McGuinness, D. H. Morgan, A. Rucklidge and A. M. Z. Slawin, *J. Mol. Cat. A*, 2008, **283**, 114-119; e) A. Vigalok, *Chem. Eur. J.*, 2008, **14**, 5102-5108; f) B. A. Dougan and Z.-L. Xue, *Organometallics*, 2009, **28**, 1295-1302.
5. B. Hoge and C. Thösen, *Inorg. Chem.*, 2001, **40**, 3113-3116.
6. a) I. G. Phillips, R. G. Ball and R. G. Cavell, *Inorg. Chem.*, 1988, **27**, 4038-4045; b) L. D. Field and M. P. Wilkinson, *Organometallics*, 1997, **16**, 1841-1845.
7. G. A. Grasa and T. J. Colacot, *Org. Lett.*, 2007, **9**, 5489-5492.
8. M. A. O. Volland, S. M. Hansen, F. Rominger and P. Hofmann, *Organometallics*, 2004, **23**, 800-816.
9. a) G. A. Chotana, I. I. B. A. Vanchura, M. K. Tse, R. J. Staples, J. R. E. Maleczka and I. I. I. M. R. Smith, *Chem. Commun.*, 2009, 5731-5733; b) A. Gunay, C. Müller, R. J. Lachicotte, W. W. Brennessel and W. D. Jones, *Organometallics*, 2009, **28**, 6524-6530.
10. a) F. A. Cotton and G. Wilkinson, *Advanced Inorganic Chemistry: A Comprehensive Text*, 4th edn., Wiley-Interscience Publications: New York, NY, 1980; b) N. Miyaura, T. Ishiyama, H. Sasaki, M. Ishikawa, M. Sato and A. Suzuki, *J. Am. Chem. Soc.*, 1989, **111**, 314-321; c) S. Huh, Y. Cho, M.-J. Jun, D. Whang and K. Kim, *Polyhedron*, 1994, **13**, 1887-1894; d) S. Subongkoj, S. Lange, W. Chen and J. Xiao, *J. Mol. Cat. A*, 2003, **196**, 125-129; e) V. Percec, G. M. Golding, J. Smidrkal and O. Weichold, *J. Org. Chem.*, 2004, **69**, 3447-3452; f) N. Jiang and A. J. Ragauskas, *Tet. Lett.*, 2006, **47**, 197-200; g) C. Cai, J. Y. L. Chung, J. C. McWilliams, Y. Sun, C. S. Shultz and M. Palucki, *Org. Process Res. Dev.*, 2007, **11**, 328-335; h) L. Ackermann, S. Barfüsser and J. Pospech, *Org. Lett.*, 2010, **12**, 724-726, doi: 10.1021/ol9028034.
11. G. Onodera, M. Matsuzawa, T. Aizawa, T. Kitahara, Y. Shimizu, S. Kezuka and R. Takeuchi, *Synlett*, 2008, **5**, 755-758.
12. a) H. Nakamura, S. Onagi and T. Kamakura, *J. Org. Chem.*, 2005, **70**, 2357-2360; b) A. C. Marr, M. Nieuwenhuyzen, C. L. Pollock and G. C. Saunders, *Organometallics*, 2007, **26**, 2659-2671.
13. V. Blandin, J. F. Carpentier and A. Mortreux, *Eur. J. Org. Chem.*, 1999, **1999**, 1787-1793.
14. L. E. Bowen, M. Charernsuk, T. W. Hey, C. L. McMullin, A. G. Orpen and D. F. Wass, *Dalton Trans.*, 2010, **39**, 560-567.
15. S. Daly, M. F. Haddow, A. G. Orpen, G. T. A. Rolls, D. F. Wass and R. L. Wingad, *Organometallics*, 2008, **27**, 3196-3202.
16. D. A. Evans, R. J. Thomson and F. Franco, *J. Am. Chem. Soc.*, 2005, **127**, 10816-10817.
17. a) W. S. Lam, S. Kok, T. L. Au-Yeung, J. Wu, H. Y. Cheung, F. L. Lam, C. H. Yeung and A. Chan, *Adv. Synth. Catal.*, 2006, **348**, 370-374; b) J. M. Praetorius, R. Wang and C. M. Crudden, *Organometallics*, 2010, **29**, 554-561.
18. T. Yamagata, H. Tadaoka, M. Nagata, T. Hirao, Y. Kataoka, V. Ratovelomanana-Vidal, J. P. Genet and K. Mashima, *Organometallics*, 2006, **25**, 2505-2513.
19. Q. Shen and J. F. Hartwig, *Journal of the American Chemical Society*, 2007, **129**, 7734-7735.
20. L. Rubio-Pérez, F. J. Pérez-Flores, P. Sharma, L. Velasco and A. Cabrera, *Org. Lett.*, 2008, **11**, 265-268.
21. A. Cavarzan, G. Bianchini, P. Sgarbossa, L. Lefort, S. Gladiali, A. Scarso and G. Strukul, *Chem. Eur. J.*, 2009, **15**, 7930-7939.
22. H.-U. Blaser, H.-P. Buser, R. Häusel, H.-P. Jalett and F. Spindler, *J. Organomet. Chem.*, 2001, **621**, 34-38.
23. J. H. Koh, A. O. Larsen and M. R. Gagné, *Org. Lett.*, 2001, **3**, 1233-1236.

24. D. E. Kim, C. Choi, I. S. Kim, S. Jeulin, V. Ratovelomanana-Vidal, J. P. Genêt and N. Jeong, *Adv. Synth. Catal.*, 2007, **349**, 1999-2006.
25. P. Sgarbossa, A. Scarso, E. Pizzo, S. M. Sbovata, A. Tassan, R. A. Michelin and G. Strukul, *J. Mol. Cat. A*, 2007, **261**, 202-206.
26. L. Dahlenburg and C. Eckert, *J. Organomet. Chem.*, 1998, **564**, 227-232.
27. L. Dahlenburg and S. Mertel, *J. Organomet. Chem.*, 2001, **630**, 221-243.
28. a) M. Fild, W. Handke and W. S. Sheldrick, *Zeitschrift fuer Naturforschung*, 1980, **35B**, 838-842; b) M. Fild and K. H. Reichert, *Chemiker-Zeitung*, 1987, **111**, 176-177.
29. L. Dahlenburg, C. Becker, J. Höck and S. Mertel, *J. Organomet. Chem.*, 1998, **564**, 155-166.
30. L. Dahlenburg and C. Kühnlein, *J. Organomet. Chem.*, 2005, **690**, 1-13.
31. a) G. A. Bell, D. W. H. Rankin and P. F. Reinisch, *J. Chem. Soc. Dalton Trans.*, 1987, 3023-3027; b) A. C. Dros, A. Meetsma and R. M. Kellogg, *Tetrahedron*, 1999, **55**, 3071-3090; c) M. Yan and A. S. C. Chan, *Tet. Lett.*, 1999, **40**, 6645-6648; d) P. Bergamini, V. Bertolasi and F. Milani, *Eur. J. Inorg. Chem.*, 2004, **2004**, 1277-1284.
32. A. Maspero, I. Kani, A. A. Mohamed, M. A. Omary, R. J. Staples and J. P. Fackler, *Inorg. Chem.*, 2003, **42**, 5311-5319.
33. a) K. Wakabayashi, H. Yorimitsu and K. Oshima, *Journal of the American Chemical Society*, 2001, **123**, 5374-5375; b) O. Kobayashi, D. Uraguchi and T. Yamakawa, *J. Fluorine Chem.*, 2009, **130**, 591-594.
34. H. Chen and M.-Z. Deng, *J. Organomet. Chem.*, 2000, **603**, 189-193.
35. P. Dierkes and P. W. N. M. van Leeuwen, *J. Chem. Soc. Dalton Trans.*, 1999, 1519-1530.
36. A. Brennführer, H. Neumann and M. Beller, *Angew. Chem. Int. Ed.*, 2009, **48**, 4114-4133.
37. B. E. Barton, G. Zampella, A. K. Justice, L. De Gioia, T. B. Rauchfuss and S. R. Wilson, *Dalton Trans.*, 2010, **39**, 3011-3019.
38. M. T. Olsen, B. E. Barton and T. B. Rauchfuss, *Inorg. Chem.*, 2009, **48**, 7507-7509.
39. M. S. Balakrishna, V. S. Reddy, S. S. Krishnamurthy, J. F. Nixon and J. C. T. R. B. S. Laurent, *Coord. Chem. Rev.*, 1994, **129**, 1-90.
40. a) K. Blann, A. Bollmann, J. T. Dixon, F. M. Hess, E. Killian, H. Maumela, D. H. Morgan, A. Neveling, S. Otto and M. J. Overett, *Chem. Commun.*, 2005, 620-621; b) S. Kuhlmann, John T. Dixon, M. Haumann, David H. Morgan, J. Ofili, O. Spuhl, N. Taccardi and P. Wasserscheid, *Adv. Synth. Catal.*, 2006, **348**, 1200-1206; c) T. Jiang, S. Zhang, X. Jiang, C. Yang, B. Niu and Y. Ning, *J. Mol. Cat. A*, 2008, **279**, 90-93.
41. a) M. R. J. Elsegood, M. B. Smith and P. M. Staniland, *Inorg. Chem.*, 2006, **45**, 6761-6770; b) G. M. Brown, M. R. J. Elsegood, A. J. Lake, N. M. Sanchez-Ballester, M. B. Smith, T. S. Varley and K. Blann, *Eur. J. Inorg. Chem.*, 2007, **2007**, 1405-1414.
42. T. Posset and J. Blümel, *J. Am. Chem. Soc.*, 2006, **128**, 8394-8395.
43. C. J. Curtis, A. Miedaner, R. Ciancanelli, W. W. Ellis, B. C. Noll, M. Rakowski DuBois and D. L. DuBois, *Inorg. Chem.*, 2002, **42**, 216-227.
44. R. K. Sharma, M. Nethaji and A. G. Samuelson, *Tet. Asymmetry*, 2008, **19**, 655-663.
45. a) Y. Chen, X. Li, S.-k. Tong, M. C. K. Choi and A. S. C. Chan, *Tet. Lett.*, 1999, **40**, 957-960; b) V. I. Tararov, R. Kadyrov, T. H. Riermeier, J. Holz and A. Börner, *Tet. Asymmetry*, 1999, **10**, 4009-4015.
46. F. Agbossou-Niedercorn and I. Suisse, *Coord. Chem. Rev.*, 2003, **242**, 145-158.
47. T. Hatakeyama, Y. Kondo, Y.-i. Fujiwara, H. Takaya, S. Ito, E. Nakamura and M. Nakamura, *Chem. Commun.*, 2009, 1216-1218.
48. G. g. Landelle, P. A. Champagne, X. Barbeau and J.-F. o. Paquin, *Org. Lett.*, 2009, **11**, 681-684.
49. W. Clegg, G. R. Eastham, M. R. J. Elsegood, B. T. Heaton, J. A. Iggo, R. P. Tooze, R. Whyman and S. Zacchini, *Organometallics*, 2002, **21**, 1832-1840.
50. a) C. Jimenez-Rodriguez, P. J. Pogorzelec, G. R. Eastham, A. M. Z. Slawin and D. J. Cole-Hamilton, *Dalton Trans.*, 2007, 4160-4168; b) G. Lamb, M. Clarke, A. M. Z. Slawin, B. Williams and L. Key, *Dalton Trans.*, 2007, 5582-5589.
51. J. J. Becker, P. S. White and M. R. Gagné, *J. Am. Chem. Soc.*, 2001, **123**, 9478-9479.

52. S. Doherty, G. R. Eastham, R. P. Tooze, T. H. Scanlan, D. Williams, M. R. J. Elsegood and W. Clegg, *Organometallics*, 1999, **18**, 3558-3560.
53. J. López-Serrano, S. B. Duckett, S. Aiken, K. Q. Almeida Leñero, E. Drent, J. P. Dunne, D. Konya and A. C. Whitwood, *J. Am. Chem. Soc.*, 2007, **129**, 6513-6527.
54. a) W. Braun, A. Salzer, H.-J. Drexler, A. Spannenberg and D. Heller, *Dalton Trans.*, 2003, 1606-1613; b) W. Braun, B. Calmuschi, J. Haberland, W. Hummel, A. Liese, T. Nickel, O. Stelzer and A. Salzer, *Eur. J. Inorg. Chem.*, 2004, **2004**, 2235-2243.
55. a) P. Dani, T. Karlen, R. A. Gossage, W. J. J. Smeets, A. L. Spek and G. van Koten, *J. Am. Chem. Soc.*, 1997, **119**, 11317-11318; b) A. Vigalok, O. Uzan, L. J. W. Shimon, Y. Ben-David, J. M. L. Martin and D. Milstein, *J. Am. Chem. Soc.*, 1998, **120**, 12539-12544; c) D. G. Gusev, M. Madott, F. M. Dolgushin, K. A. Lyssenko and M. Y. Antipin, *Organometallics*, 2000, **19**, 1734-1739.
56. a) F. Mao, C. E. Philbin, T. J. R. Weakley and D. R. Tyler, *Organometallics*, 1990, **9**, 1510-1516; b) M. Fei, S. K. Sur and D. R. Tyler, *Organometallics*, 1991, **10**, 419-423; c) J. S. Lewis, S. L. Heath, A. K. Powell, J. Zweit and P. J. Blower, *J. Chem. Soc. Dalton Trans.*, 1997, 855-862; d) N. W. Duffy, R. R. Nelson, M. G. Richmond, A. L. Rieger, P. H. Rieger, B. H. Robinson, D. R. Tyler, J. C. Wang and K. Yang, *Inorganic Chemistry*, 1998, **37**, 4849-4856.
57. a) J. Almena, A. Monsees, R. Kadyrov, Thomas H. Riermeier, B. Gotov, J. Holz and A. Börner, *Adv. Synth. Catal.*, 2004, **346**, 1263-1266; b) M. E. Fox, M. Jackson, I. C. Lennon, J. Klosin and K. A. Abboud, *J. Org. Chem.*, 2007, **73**, 775-784; c) J. Holz, B. Schäffner, O. Zayas, A. Spannenberg and A. Börner, *Adv. Synth. Catal.*, 2008, **350**, 2533-2543.
58. A. T. Axtell, J. Klosin, G. T. Whiteker, C. J. Cobley, M. E. Fox, M. Jackson and K. A. Abboud, *Organometallics*, 2009, **28**, 2993-2999.
59. B. C. Hamann and J. F. Hartwig, *Journal of the American Chemical Society*, 1998, **120**, 3694-3703.
60. a) H. S. Huh, Y. K. Lee and S. W. Lee, *J. Mol. Str.*, 2006, **789**, 209-219; b) S. E. Kabir, F. Ahmed, S. Ghosh, M. R. Hassan, M. S. Islam, A. Sharmin, D. A. Tocher, D. T. Haworth, S. V. Lindeman, T. A. Siddiquee, D. W. Bennett and K. I. Hardcastle, *J. Organomet. Chem.*, 2008, **693**, 2657-2665; c) T. Weisheit, H. Petzold, H. Görls, G. Mloston and W. Weigand, *Eur. J. Inorg. Chem.*, 2009, **2009**, 3545-3551.
61. B. H. Lipshutz, K. Noson, W. Chrisman and A. Lower, *J. Am. Chem. Soc.*, 2003, **125**, 8779-8789.
62. A. M. R. Smith, H. S. Rzepa, A. J. P. White, D. Billen and K. K. Hii, *J. Org. Chem.*, 2010, **75**, 3085-3096.
63. a) K. Mikami, H. Kakuno and K. Aikawa, *Angew. Chem. Int. Ed.*, 2005, **44**, 7257-7260; b) J. W. Faller and J. C. Wilt, *J. Organomet. Chem.*, 2006, **691**, 2207-2212.
64. a) C. P. Casey, G. T. Whiteker, M. G. Melville, L. M. Petrovich, J. A. Gavney and D. R. Powell, *J. Am. Chem. Soc.*, 1992, **114**, 5535-5543; b) C. P. Casey and L. M. Petrovich, *J. Am. Chem. Soc.*, 1995, **117**, 6007-6014.
65. a) R.-X. Li, X.-J. Li, N.-B. Wong, K.-C. Tin, Z.-Y. Zhou and T. C. W. Mak, *J. Mol. Cat. A*, 2002, **178**, 181-190; b) L. Zhang, Y. Zhang, X.-G. Zhou, R.-X. Li, X.-J. Li, K.-C. Tin and N.-B. Wong, *J. Mol. Cat. A*, 2006, **256**, 171-177.
66. S. Liu, J.-H. Xie, W. Li, W.-L. Kong, L.-X. Wang and Q.-L. Zhou, *Org. Lett.*, 2009, **11**, 4994-4997.
67. W. Lesueur, E. Solari, C. Floriani, A. Chiesi-Villa and C. Rizzoli, *Inorg. Chem.*, 1997, **36**, 3354-3362.
68. P. W. N. M. van Leeuwen, P. C. J. Kamer, J. N. H. Reek and P. Dierkes, *Chem. Rev.*, 2000, **100**, 2741-2770.
69. P. Maji, W. Wang, A. E. Greene and Y. Gimbert, *J. Organomet. Chem.*, 2008, **693**, 1841-1849.
70. K. Ruhland, P. Gigler and E. Herdtweck, *J. Organomet. Chem.*, 2008, **693**, 874-893.
71. M. Kranenburg, Y. E. M. van der Burgt, P. C. J. Kamer, P. W. N. M. van Leeuwen, K. Goubitz and J. Fraanje, *Organometallics*, 1995, **14**, 3081-3089.
72. M. Kranenburg, P. C. J. Kamer and P. W. N. M. v. Leeuwen, *Eur. J. Inorg. Chem.*, 1998, 155-157.
73. M. Kranenburg, P. C. J. Kamer and P. W. N. M. v. Leeuwen, *Eur. J. Inorg. Chem.*, 1998, 25-27.
74. a) J. P. Sadighi, M. C. Harris and S. L. Buchwald, *Tet. Lett.*, 1998, **39**, 5327-5330; b) Y. Guari, D. S. van Es, J. N. H. Reek, P. C. J. Kamer and P. W. N. M. van Leeuwen, *Tet. Lett.*, 1999, **40**, 3789-3790.

75. B. Deb, B. J. Borah, B. J. Sarmah, B. Das and D. K. Dutta, *Inorg. Chem. Commun.*, 2009, **12**, 868-871.
76. a) J. D. Unruh and J. R. Christenson, *J. Mol. Cat.*, 1982, **14**, 19-34; b) P. W. N. M. Van Leeuwen and C. F. Roobeek, *J. Mol. Cat.*, 1985, **31**, 345-353.
77. L. Lassova, H. K. Lee and T. S. A. Hor, *J. Org. Chem.*, 1998, **63**, 3538-3543.
78. H.-U. Blaser, W. Brieden, B. Pugin, F. Spindler, M. Studer and A. Togni, *Topics in Catalysis*, 2002, **19**, 3-16.
79. N. Fey, J. N. Harvey, G. C. Lloyd-Jones, P. Murray, A. G. Orpen, R. Osborne and M. Purdie, *Organometallics*, 2008, **27**, 1372-1383.
80. Maestro 5.1.020, Schrödinger-LLC, New York, NY, 1999-2002.