Computational Descriptors for Chelating P,P- and P,N-Donor Ligands[1]

Natalie Fey,*,† Jeremy N. Harvey,*,† Guy C. Lloyd-Jones,† Paul Murray,‡ A. Guy Orpen,*,† Robert Osborne,‡ and Mark Purdie§

School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, U.K., AstraZeneca, Process Research & Development, Avlon Works, Severn Road, Hallen, Bristol BS10 7ZE, U.K., and AstraZeneca, Process Research & Development, Charnwood, Bakewell Road, Loughborough, Leicester LE11 5RH, U.K.

Received August 21, 2007

The ligand knowledge base approach has been extended to capture the properties of 108 bidentate P,P- and P,N-donor ligands. This contribution describes the design of the ligand set and a range of DFT-calculated descriptors, capturing ligand properties in a variety of chemical environments. New challenges arising from ligand conformational flexibility and donor asymmetry are discussed, and descriptors are related to other parameters, such as the ligand bite angle. A novel map of bidentate ligand space, potentially useful in catalyst design and discovery, has been derived from principal component analysis of the resulting LKB-PP descriptors. In addition, a range of multiple linear regression models have been derived for both experimental and calculated data, considering ligand bite angles in square-planar palladium complexes and ligand dissociation energies from octahedral chromium complexes, respectively. These data sets were fitted with models based on LKB descriptors to explore the transferability of descriptors to different coordination environments and to illustrate potential applications of such models in catalyst design, allowing predictions about novel or untested ligands.

Introduction

Ligands with phosphorus(III) donor atoms represent perhaps the most widely used ligand class in homogeneous catalysis, evidence of their synthetic stability and the comparative ease with which their properties can be modified synthetically. By choosing an appropriate ligand, the properties and structures of the organometallic and coordination complexes they form can be modified extensively. This process of identifying a suitable ligand can be put on a more quantitative basis by the determination of appropriate structural and electronic parameters, which can then be used both to map chemical space and to predict the properties of novel ligands and their complexes. Various experimental and computational approaches have been developed that seek to capture and quantify the steric and σ- and π-electronic properties of P-donor ligands; these have been summarized in a number of publications (see, e.g., refs 5–11 and work cited therein).

Complexes of chelating ligands are often more stable than those of related monodentate ligands. The origin of and detailed contributions to this "chelate effect" continue to be investigated.[12] In addition to this enhanced stability, asymmetry can be introduced easily into bidentate ligands and hence their complexes, through use of chiral backbones/substituents and different donor atoms.[7,13,14]While chelating ligands have been applied extensively in reactions such as cross-coupling, allylic alkylation, hydroformylation, hydrogenation, and CO/ethene copolymerization (reviewed, e.g., in refs 9, 13–16), few descriptors

* Corresponding authors. E-mail: Natalie.Fey@Bristol.ac.uk; Jeremy.Harvey@Bristol.ac.uk; Guy.Orpen@Bristol.ac.uk. Tel: (+44) 0117 331 8260 (N.F.). Fax: (+44) 0117 925 1295.

† University of Bristol.

‡ AstraZeneca, Avlon Works.

§ AstraZeneca, Charnwood.

(1) Development of a Ligand Knowledge Base, Part 4. See refs 2–4 for Parts 1–3.

(2) Fey, N.; Tsipis, A.; Harris, S. E.; Harvey, J. N.; Orpen, A. G.; Mansson, R. A. *Chem.−Eur. J.* **2006**, *12*, 291–302.

(3) Fey, N.; Harris, S. E.; Harvey, J. N.; Orpen, A. G. *J. Chem. Inf. Model.* **2006**, *46*, 912–929.

(4) Mansson, R. A.; Welsh, A. H.; Fey, N.; Orpen, A. G. *J. Chem. Inf. Model.* **2006**, *46* (6), 2591–2600.

(5) (a) Sparta, M.; Børve, K. J.; Jensen, V. R. *J. Am. Chem. Soc.* **2007**, *129* (27), 8487–8499. (b) Zuidema, E.; Daura-Oller, E.; Carbó, J. J.; Bo, C.; van Leeuwen, P. W. N. M. *Organometallics* **2007**, *26* (9), 2234–2242. (c) Suresh, C. H. *Inorg. Chem.* **2006**, *45* (13), 4982–4986. (d) Kühl, O. *Coord. Chem. Rev.* **2005**, *249*, 693–704. (e) Burello, E.; Farrusseng, D.; Rothenberg, G. *Adv. Synth. Catal.* **2004**, *346*, 1844–1853. (f) Babij, C.; Poë, A. J. *J. Phys. Org. Chem.* **2004**, *17*, 162–167. (g) Wilson, M. R.; Prock, A.; Giering, W. P.; Fernandez, A. L.; Haar, C. M.; Nolan, S. P.; Foxman, B. M. *Organometallics* **2002**, *21* (13), 2758–2763. (h) Bunten, K. A.; Chen, L.; Fernandez, A. L.; Poe, A. J. *Coord. Chem. Rev.* **2002**, *41*, 233–234. (i) Perrin, L.; Clot, E.; Eisenstein, O.; Loch, J.; Crabtree, R. H. *Inorg. Chem.* **2001**, *40*, 5806. (j) Dias, P. B.; Piedade, M. E. M. d.; Simões, J. A. M. *Coord. Chem. Rev.* **1994**, *135/136*, 737–807. (k) Brown, T. L.; Lee, K. J. *Coord. Chem. Rev.* **1993**, *128*, 89–116.

(6) Kühl, O. *Can. J. Chem.* **2007**, *85*, 230–238.

(7) Cobley, C. J.; Froese, R. D. J.; Klosin, J.; Qin, C.; Whiteker, G. T.; Abboud, K. A. *Organometallics* **2007**, *26* (12), 2986–2999.

(8) (a) Hageman, J. A.; Westerhuis, J. A.; Fruehauf, H.-W.; Rothenberg, G. *Adv. Synth. Catal.* **2006**, *348*, 361–369. (b) Burello, E.; Marion, P.; Galland, J.-C.; Chamard, A.; Rothenberg, G. *Adv. Synth. Catal.* **2005**, *347*, 803–810. (c) Burello, E.; Rothenberg, G. *Adv. Synth. Catal.* **2005**, *347*, 1969–1977.

(9) Freixa, Z.; van Leeuwen, P. W. N. M. *Dalton Trans.* **2003**, 1890–1901.

(10) (a) Cooney, K. D.; Cundari, T. R.; Hoffman, N. W.; Pittard, K. A.; Temple, M. D.; Zhao, Y. *J. Am. Chem. Soc.* **2003**, *125* (14), 4318–4324. (b) Burello, E.; Rothenberg, G. *Adv. Synth. Catal.* **2003**, *345*, 1334–1340.

(11) Tolman, C. A. *Chem. Rev.* **1977**, *77* (3), 313.

(12) (a) Davydova, E. I.; Sevastianova, T. N.; Timoshkin, A. Y.; Suvorov, A. V.; Frenking, G. *Int. J. Quantum Chem.* **2004**, *100* (4), 419–425. (b) Vallet, V.; Wahlgren, U.; Grenthe, I. *J. Am. Chem. Soc.* **2003**, *125* (48), 14941–14950. (c) Emmenegger, F.; Schlaepfer, C. W.; Stoeckli-Evans, H.; Piccand, M.; Piekarski, H. *Inorg. Chem.* **2001**, *40* (16), 3884–3888.

(13) (a) Guiry, P. J.; Saunders, C. P. *Adv. Synth. Catal.* **2004**, *346* (5), 497–537. (b) Arrayas, R. G.; Adrio, J.; Carretero, J. C. *Angew. Chem., Int. Ed.* **2006**, *45* (46), 7674–7715.

(14) Breit, B. *Acc. Chem. Res.* **2003**, *36* (4), 264–275.

10.1021/om700840h CCC: $40.75
Publication on Web 03/13/2008

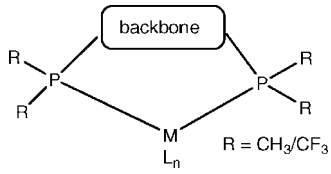


Figure 1. Ligands 1–24.

have been derived for multiple bidentate ligands,[6–9,11,15–19] with the L−M−L bite angle a commonly considered parameter.[6–9,15,16,18–20]

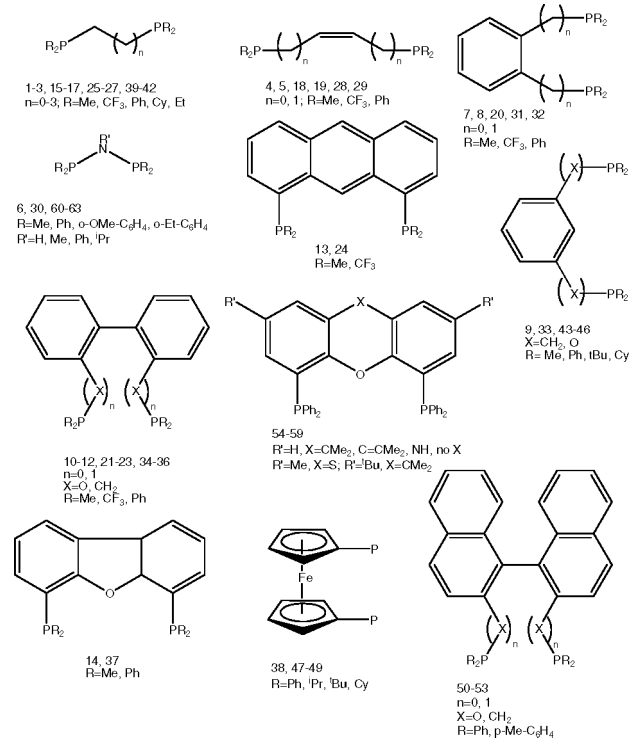
We have recently described the development of computational descriptors for monodentate phosphorus(III) donor ligands, collated in a ligand knowledge base (LKB-P).[2] In this context we have explored potential applications of such descriptors in maps of chemical space and predictive models and discussed suitable statistical analysis techniques.[2,4] Here we apply this approach to the development of a ligand knowledge base for chelating P,P- and P,N-donor ligands (LKB-PP), aimed at exploring bidentate ligand space. We describe the design of novel descriptors to capture metal–ligand binding for bidentate ligands and discuss the challenges arising from capturing conformationally flexible ligands in a knowledge-based approach. In addition, we present the first map of bidentate ligand space and illustrate potential applications of such maps and predictive regression models derived from these descriptors in ligand and catalyst design.

Knowledge Base Design and Development

Ligand Set. For the initial testing of descriptors we wanted to limit conformational noise (vide infra) while introducing some electronic variability to the ligand set. The map of ligand space generated for our prototype monodentate LKB-P[2] suggested that changing from methyl to trifluoromethyl substitution captures considerable electronic variation, and we have combined a range of common chelating ligand backbones with these auxiliary substituents (Me: 1–14, CF_3: 15–24, Figure 1). In contrast, many synthetically relevant ligands have very similar aryl groups as auxiliary substituents, but widely varied backbones. We have thus also included a range of ligand backbones with phenyl substituents (25–38, 50, 52–59, 62, 63, 74–76, 78–81). Furthermore, ligands have been considered to improve overlap with available experimental data sets (39–49, 51, 60, 61, 64–73, 82).

Mixed donor ligands have emerged as a synthetically valuable way of introducing both steric and electronic asymmetry to a range of reactions.[13,21] Furthermore, the different donor atoms introduce electronic asymmetry close to the metal center, which poses a challenging test for our descriptors. We have thus included a number of prototypical P,N-donor ligands to represent this subset (83–108). The full set of ligands considered in this work is represented in Charts 1–3, while detailed ligand numbers are given in the Supporting Information (Table S1).

Chart 1. General Ligand Structures (P,P-Donor)



Descriptors. Our initial work on monodentate ligands (LKB-P) has highlighted a number of desirable properties for descriptors in a knowledge base.[2] From a chemical point of view, descriptors should ideally be representative of and transferable to a variety of chemical environments and sample chemical space extensively and evenly. Statistical data analysis techniques usually implicitly assume a random and representative sample has been drawn from an extensive population, and while this is rarely practical in chemistry, we have calculated a range of descriptors derived from metal complexes to capture metal–ligand bonding in different environments. The (approximately) tetrahedral zinc dichloride [$ZnCl_2${LL}] fragment (where {LL} represents P,P- and P,N-donor, bidentate ligands) was used as an archetypal Lewis acid complex displaying the coordination of two donor atoms. Donor−metal−donor angles in this fragment are quite flexible, allowing for the adoption of a range of ligand bite angles (vide infra). In contrast, the square-planar palladium dichloride [$PdCl_2${LL}] fragment enforces a small bite angle (close to 90°), and this fragment is closely related to the $[PdCl_3L]^-$ complexes used in the monodentate ligand knowledge base LKB-P.[2]

We have also investigated using a tetrahedral [$Pt(PH_3)_2${LL}] complex, again analogous to LKB-P, but Pt(0)−N bonding is weak and most of the P,N-donor ligands partially dissociated on optimization, indicating a lack of computational robustness[2] for these complexes. Calculations on P,N-donor ligands occupying two equatorial positions in a trigonal bipyramidal rhodium(I) complex [$RhH(CO)_2${LL}] frequently led to dissociation of the Rh(I)−N bonds, and these complexes were also not considered further. Bulky ligands, such as 37, 49, 57, and 65, coordinated to an octahedral [$Cr(CO)_4${LL}] complex often led to failed or slow to converge optimizations, so we have excluded this complex from consideration as a source of descriptors as well.

The metal–ligand distances are distinct for phosphorus and nitrogen donor atoms and, if used as descriptors, show a

(15) Dierkes, P.; van Leeuwen, P. W. N. M. *J. Chem. Soc., Dalton Trans.* **1999**, 1519–1529.
(16) van Leeuwen, P. W. N. M.; Kamer, P. C. J.; Reek, J. N. H.; Dierkes, P. *Chem. Rev.* **2000**, *100* (8), 2741–2769.
(17) Jeulin, S.; Duprat de Paule, S.; Ratovelomanana-Vidal, V.; Genet, J.-P.; Champion, N.; Dellis, P. *Proc. Natl. Acad. Sci.* **2004**, *101* (16), 5799–5804.
(18) Kamer, P. C. J.; van Leeuwen, P. W. N. M.; Reek, J. N. H. *Acc. Chem. Res.* **2001**, *34* (11), 895–904.
(19) (a) Casey, C. P.; Whiteker, G. P. *Isr. J. Chem.* **1990**, *30*, 299–304. (b) Casey, C. P.; Whiteker, G. P.; Melville, M. G.; Petrovich, L. M.; Gavney, J. A.; Powell, D. R. *J. Am. Chem. Soc.* **1992**, *114* (14), 5535–5543.
(20) Raebiger, J. W.; Miedaner, A.; Curtis, C. J.; Miller, S. M.; Anderson, O. P.; DuBois, D. L. *J. Am. Chem. Soc.* **2004**, *126* (17), 5502–5514.

Chart 2. Additional Ligands in LKB-PP

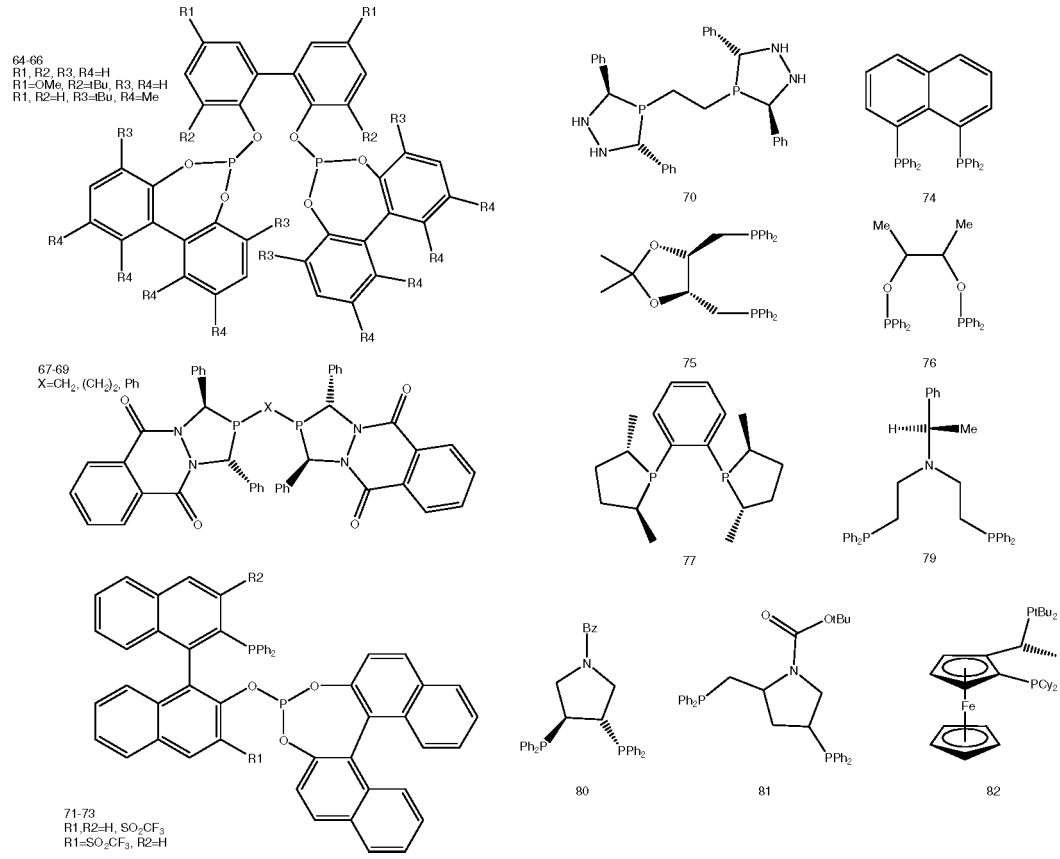
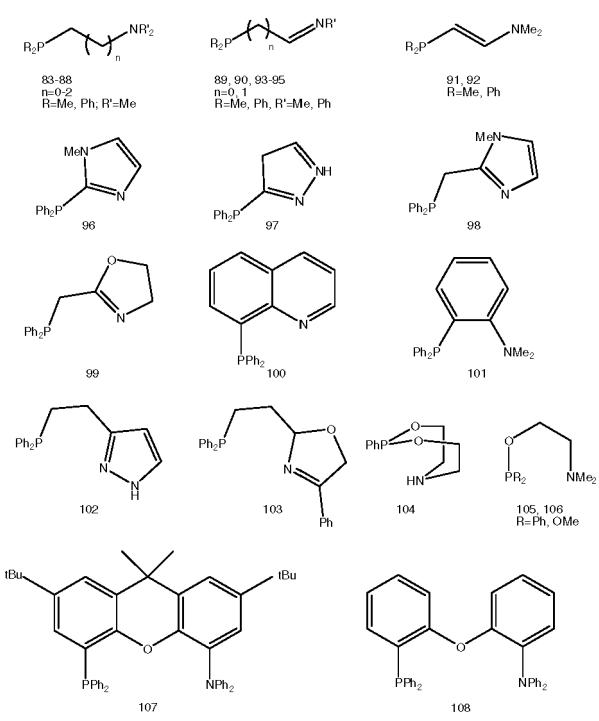
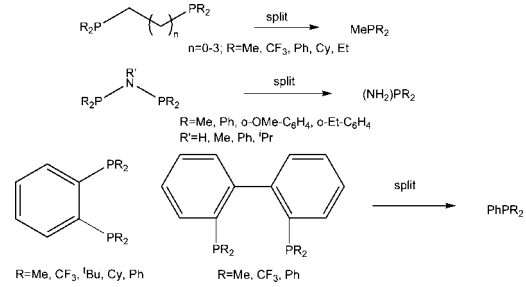


Chart 3. P,N-Donor Ligands in LKB-PP



Scheme 1. Truncation of Ligands



bimodal distribution, i.e., clustering around two different average bond lengths. Principal component analysis (discussed in detail below) seeks to maximize the variation of the data set in few dimensions, and such distinct bond length preferences give rise to clearly separated clusters on ligand maps, which may obscure more subtle similarities. Instead, we have used the *change* in metal–ligand bond lengths compared to those for a standard reference ligand (ligand 1 for P donors and ligand 83 (D2) for N donors) as the metal–ligand bond length descriptors, ΔM−D.

In addition, we have extracted metal−chlorine bond lengths from both metal fragments, which capture differences in ligand steric and electronic influences on other ligands in these complexes. We have also included structural parameters describing geometry changes of the ligands on complexation; these are calculated as the difference in average ligand bond lengths (D−A) and angles (A−D−A) between the complexed and free ligand geometries.

For monodentate phosphorus donor ligands, frontier molecular orbital energies and proton affinities (PA) are useful descriptors of their electronic properties, correlating with the lone pair on phosphorus (E_{HOMO}, PA) and a potentially π-accepting ligand orbital (E_{LUMO}), respectively. For chelating ligands the orbitals on the two donor atoms mix, and this relationship becomes more complicated; likewise the calculation of proton affinities for two donor atoms becomes conceptually more difficult and might be affected by movement of the ligand backbone. However, these

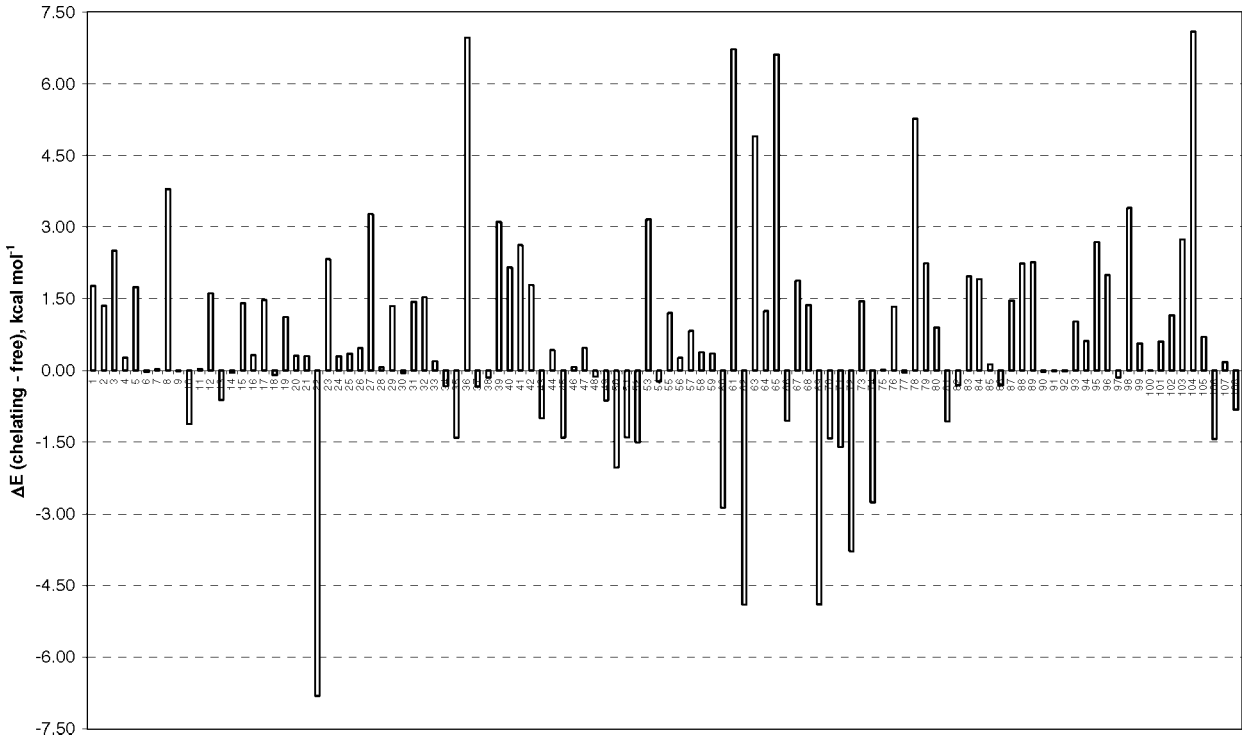


Figure 2. Relative energies of free ligand geometries (chelating − free, kcal mol^{-1}) from DFT optimization of MM conformational search results on free and chelated ligands.

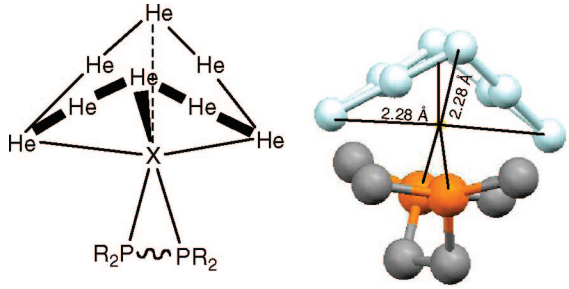


Figure 3. He_8_wedge geometry.

properties are mostly determined by the local environment of the individual donor atom, with the character of the bridging section (backbone) less important. We have therefore used some simple rules to split and truncate the ligands and treat each donor atom individually, allowing us to calculate these descriptors. The truncation generally occurs after the bridge atom connected directly to the donor, and the bridge is replaced with a representative substituent, e.g., methyl for sp^3 carbons, vinyl for nonaromatic sp^2 carbons, phenyl for aromatic groups (including ferrocenes), methoxy for OR links, etc. The other substituents on the donor atoms have usually been left unchanged, with the exception of ligands 64–66, 71–73 (simplified to OMe), 67–69, 77 (Me), and 51, 60, 61 (Ph). Scheme 1 illustrates some representative truncations, and Table S2 in the Supporting Information shows the full list of split ligands used.

Calculations should use a reliable computational approach to minimize the occurrence of spurious outliers. Computational cost is also important, as extensive automation of both input file generation and data extraction would facilitate expansion of the knowledge base to cover a wide range of ligands. While the chosen DFT methodology (BP86/6-31G* and LACV3P on metals, see Computational Details) is well-established and relatively inexpensive, achieving adequate conformational sampling remains a computational challenge at this level of theory. The bidentate ligands considered here display a range of flexible backbones, and many have conformationally flexible substituents on both donor atoms, so multiple low-energy conformers are likely to occur.

Conformational searching remains beyond the reach of pure DFT, and mining of structural databases such as the Cambridge Structural Database (CSD)[22] relies on the availability of suitable crystal structure data for most of the ligands considered, which may not always be possible. Molecular mechanical (MM) conformational searches thus become the most viable option. However, the conformational energy surfaces of DFT and MM calculations are likely to differ substantially.[23] This could be addressed by force field reparameterization, but it would significantly increase the computational effort required, and here we have used default force field parameters only. For large databases such as these ligand knowledge bases, conformational noise thus remains difficult to eliminate; however, in most cases descriptor values remain quite similar for low-energy conformers of the free ligands, and conformational freedom is reduced on coordination to metal fragments.

(21) (a) Speiser, F.; Braunstein, P.; Saussine, W. *Acc. Chem. Res.* **2005**, *38* (10), 784–793. (b) Hou, X. L.; You, S. L.; Tu, T.; Deng, W. P.; Wu, X. W.; Li, M.; Yuan, K.; Zhang, T. Z.; Dai, L. X. *Top. Catal.* **2005**, *35* (1–2), 87–103. (c) Shibasaki, M.; Vogl, E. M.; Ohshima, T. *Adv. Synth. Catal.* **2004**, *346* (13–15), 1533–1552. (d) Atkinson, R. C. J.; Gibson, V. C.; Long, N. J. *Chem. Soc. Rev.* **2004**, *33* (5), 313–328. (e) Chelucci, G.; Orru, G.; Pinna, G. A. *Tetrahedron* **2003**, *59* (48), 9471–9515. (f) Kranich, R.; Eis, K.; Geis, O.; Muhle, S.; Bats, J. W.; Schmalz, H. G. *Chem.−Eur. J.* **2000**, *6* (15), 2874–2894. (g) Helmchen, G.; Pfaltz, A. *Acc. Chem. Res.* **2000**, *33* (6), 336–345. (h) Loiseleur, O.; Hayashi, M.; Keenan, M.; Schmees, N.; Pfaltz, A. *J. Organomet. Chem.* **1999**, *576* (1–2), 16–22.

(22) Orpen, A. G. *Acta Crystallogr.* **2002**, *B58*, 398–406. Allen, F. H. *Acta Crystallogr.* **2002**, *B58*, 380–388.

(23) (a) Balcells, D.; Drudis-Sole, G.; Besora, M.; Doelker, N.; Ujaque, G.; Maseras, F.; Lledos, A. *Faraday Discuss.* **2003**, *124*, 429–441. (b) Buda, C.; Burt, S. K.; Cundari, T. R.; Shenkin, P. S. *Inorg. Chem.* **2002**, *41* (8), 2060–2069. (c) Gillespie, A. M.; Morello, G. R.; White, D. P. *Organometallics* **2002**, *21* (19), 3913–3921. (d) Olsen, P. T.; Jensen, F. *J. Chem. Phys.* **2003**, *118* (8), 3523–3531.

Table 1. Calculated Descriptors in LKB-PP ({LL} represents P,P- and P,N-donor bidentate ligands; note that for P,N ligands, D1 = P, D2 = N)

descriptor[a]	derivation (units)
Split ligands (L_{D1}, L_{D2})	
E_{HOMO}_D1, E_{HOMO}_D2	energy of highest occupied molecular orbital (hartrees)
E_{LUMO}_D1, E_{LUMO}_D2	energy of lowest unoccupied molecular orbital (hartrees)
PA_D1, PA_D2	proton affinity, PA = $E(L_{Dn}) - E([HL_{Dn}]^+)$ (kcal mol^{-1})
Free ligands ({LL})	
He_8_wedge	interaction energy between ligand in chelating conformation and wedge of 8 He atoms,[b] $E_{He8w} = E(He_8.\{LL\}) - [E(He_8) + E(\{LL\})]$ (kcal mol^{-1}) (Figure 3)
Zinc complexes ([$ZnCl_2${LL}])	
BE(Zn)	bond energy for dissociation of {LL} from fragment (kcal mol^{-1})[c]
Zn−Cl	r(Zn−Cl) (Å)
∠D1−Zn−D2	ligand bite angle in complex (deg)
ΔD1−A(Zn), ΔD2−A(Zn)	change in av r(D−A) cf. {LL} (Å)
ΔA−D1−A(Zn), ΔA−D2−A(Zn)	change in av ∠(A−D−A) cf. {LL} (deg)
ΔZn−D1, ΔZn−D2	change in r(Zn−D) cf. PP01 (D1), PN01 (D2 = N) (Å)
Q(Zn)	NBO charge on [$ZnCl_2$] fragment
Palladium complexes ([$PdCl_2${LL}])	
BE(Pd)	bond energy for dissociation of {LL} from fragment (kcal mol^{-1})[c]
Pd−Cl	r(Pd−Cl) (Å)
∠D1−Pd−D2	ligand bite angle in complex (deg)
ΔD1−A(Pd), ΔD2−A(Pd)	change in av r(D−A) cf. free {LL} (Å)
ΔA−D1−A(Pd), ΔA−D2−A(Pd)	change in av ∠(A−D−A) cf. free {LL} (deg)
ΔPd−D1, ΔPd−D2	change in r(Pd−D) cf. PP01 (D1), PN01 (D2 = N) (Å)
Q(Pd)	NBO charge on [$PdCl_2$] fragment

[a] All calculations were performed on isolated molecules. [b] Based on metal−P distance = 2.28 Å, metal−N distance = 2 Å. [c] BE = [E(metal fragment) + E({LL})] − E(complex).

We have used MM stochastic conformational searches (described in computational details) to screen conformational space for free ligands and their tetrahedral zinc complexes ($ZnCl_2$PP). These searches were aimed at eliminating strained, high-energy conformers as input geometries, rather than reliably locating the global minimum for both MM and DFT. The conformer of lowest MM energy was then used as input for DFT optimizations, modifying the metal fragment as required for calculating other descriptors.

In addition to the lowest energy conformer found for the free ligand, we have also optimized free ligands starting from their chelating conformation in the zinc complex. This can serve as a simple and fast test for discrepancies between MM default force fields and DFT, suitable for large databases of ligands. In general, the chelating conformer is likely to be higher in energy than the free ligand minimum. However, if the chelating conformer is considerably (>2–3 kcal mol^{-1}) lower in DFT energy than the free ligand minimum found by MM conformational searching, this suggests problems with the MM calculations (inadequate default force field parameters, e.g., neglecting anomeric stabilization of sterically hindered conformers (P−O, P−N),[24] or incomplete conformational searching). Figure 2 shows the energy difference between DFT-optimized ligand geometries starting from MM conformational search results for the free and $ZnCl_2$-chelated ligand. Where optimization from the chelating starting geometry gave lower energies, we have used this structure for our descriptors.[25]

When comparing the optimized geometries of ligands where the chelating starting geometry gave a considerably lower energy conformer, it is often difficult to identify the origin of this preference and hence whether inadequate parameters or incomplete conformational sampling is to blame, as multiple degrees of freedom can be involved. However, these results suggest that for most ligands this combination of MM conformational searches and DFT optimizations will identify low-energy conformers, and we estimate average conformational noise around 1–2 kcal mol^{-1} in this knowledge base, acceptable for a study of 108 ligands.

In addition to changing the backbone, the steric properties of other substituents on the phosphorus are varied for several examples (e.g., ligands 2, 16, 26, 40, 41 or 38, 47–49), and a purely steric descriptor would be useful. We have therefore adapted the He_8_ring steric parameter described previously[2] for use with bidentate ligands. The original He_8_ring descriptor is an energetic measure of steric bulk and is calculated as the interaction energy between the phosphorus(III) ligand and a ring of eight helium atoms, designed to approximate the nonbonded interactions of the ligand with the *cis* ligands in an octahedral complex. For chelating ligands, a comparable interaction is better represented by a wedge of eight helium atoms positioned to mimic the closest approach of ligands in an octahedral complex (Figure 3). Four of the helium atoms take the positions of other ligands in an octahedral complex at a distance of 2.28 Å from a metal center (X) and with ideal angles (90°/180°) between them, with a further four helium atoms inserted between these positions to prevent meshing with bulky groups on the bidentate ligand. The ligand starting geometry is derived from the tetrahedral zinc complex, with metal−donor distances adjusted to 2.28 Å for P donor atoms and 2 Å for N donors. For ligands with very large bite angles, such short distances cannot be achieved (ligands 8, 13, 14, 24, 32, 37, 43–46, 49, 107, 108), and in those cases the shortest possible metal−donor distances are used. The metal atom is then removed and the positions of the ligand donor atoms and all helium atoms are frozen, while the rest of the ligand is optimized. This procedure changes the bite angle, but maintains the same distance between the donor

(24) Belyakov, A. V.; Dalhus, B.; Haaland, A.; Shorokov, D. J.; Volden, H. V. *J. Chem. Soc., Dalton Trans.* **2002**, 3756–3762.

(25) Out of 108 ligands, only 9 showed a difference of more than 1.5 kcal mol^{-1}, with a further 9 ligands between 1 and 1.5 kcal mol^{-1} in favor of the chelating input geometry. 21 ligands favored the chelating input geometry by less than 1 kcal mol^{-1}. In contrast, the chelating starting geometry was more than 1 kcal mol^{-1} higher in energy for 43 ligands, with a total of 69 ligands favoring the free ligand conformation.

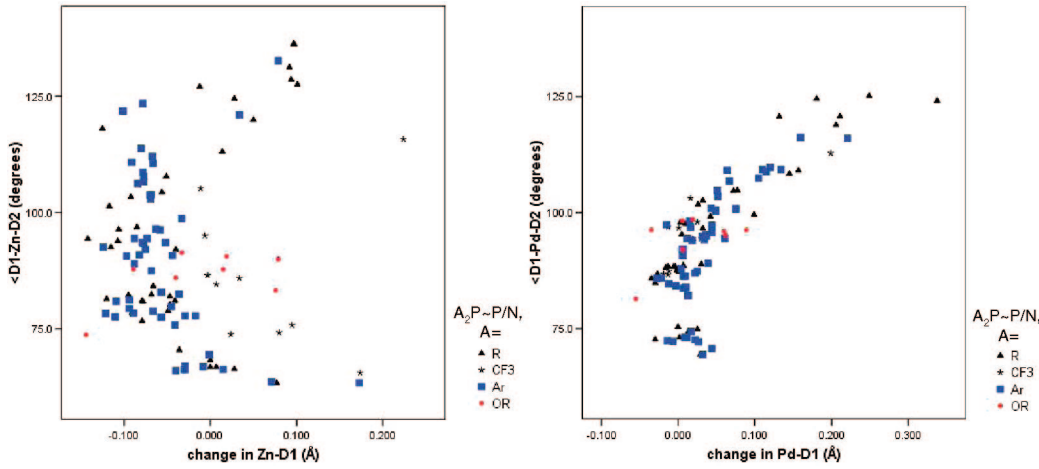


Figure 4. Bond length changes (D1) versus bite angles (∠D−M−D) for zinc and palladium complexes in LKB-PP.

atoms as for the zinc fragment. The interaction energy is then calculated with reference to the lowest energy free ligand conformer.

The calculated descriptors (Table 1) therefore include:

- frontier molecular orbital energies and ligand proton affinities for each donor atom, derived from "split" ligand fragments,
- adduct binding energies,
- a range of structural parameters describing geometry changes of ligands upon complexation, as well as the geometry of the metal fragments and changes in metal–ligand donor atom bond lengths compared to reference ligands,
- ligand bite angles in the metal complexes,
- metal fragment charges,
- a measure of steric bulk, He_8_wedge (Figure 3).

Analysis

Descriptor Contextualization. Bite angles are frequently reported descriptors for chelating ligands[6–9,15,16,18–20] and capture both steric and electronic ligand properties.[9] When bite angles and their accessible ranges are determined in a consistent fashion from calculations[15,16,19] or standardized searches of structural databases, corrected for changes in M−P distances,[15] these give a good indication of the steric and electronic properties of bidentate ligands, although they have recently been criticized.[6]

We have included donor−metal−donor angles (∠D−M−D) for both metal fragments considered here. These are not natural bite angles, which are standardized to account for changes in metal−donor distances (see, e.g., ref 15), but instead relaxation may occur both through bite angle change and changes in the metal−donor bond lengths when compared to a standard reference ligand (ΔM−D). By considering the effect of several coordination environments on metal–ligand binding (∠D−M−D, ΔM−D), ligand flexibility can be captured in an LKB context.

For the zinc fragment we observe a wide range of ∠D−M−D angles for the ligands investigated here (range: 63.4–136.3°, mean 90.9°, standard deviation, std 18.8°), with no pronounced clustering around 109° as might be expected if these complexes showed a strong preference for a tetrahedral coordination geometry. In addition, no pronounced linear relationship between bite angles and the changes in metal−donor distances (ΔZn−D1, ΔZn−D2) can be observed for this fragment (bivariate linear correlation coefficients = 0.033 (D1) and 0.138 (D2), where values close to 1 indicate that the relationship between two variables is described well by a linear equation, $y = ax + b$),

Table 2. Linear Bivariate Correlation Coefficients between LKB-PP Descriptors and External Data

	bite_m (13 ligands)[15,16]	bite_x (9 ligands)[15,16]	bite_Pd (15 ligands)[27]
∠D1−Zn−D2	0.963	0.922	0.969
∠D1−Pd−D2	0.837	0.901	0.991

although a scatter plot of ΔZn−D1 versus ∠D1−Zn−D2 (Figure 4) suggests that larger bond length increases occur for very large and very small bite angles than in the middle of the range. In contrast, the square-planar coordination geometry is enforced in the palladium complex, giving a more limited range of ∠D1−Pd−D2 angles (range: 69.4–125.1°, mean 93.5°, std 13.2°), as well as a clear linear relationship between ∠D1−Pd−D2 and the changes in metal donor distances (ΔPd−D1, ΔPd−D2), as shown in the scatter plot of ΔPd−D1 versus ∠D1−Pd−D2 in Figure 4 (bivariate linear correlation coefficients = 0.758 (D1) and 0.771 (D2)). This suggests that some of the strain introduced by rigid ligand backbones favoring large bite angles is relieved by longer metal−donor bonds. The rigid ligand backbones of ligands 107 and 108, combined with relatively weaker Pd−N bonds, lead to partial dissociation of these ligands for the palladium fragment, in agreement with experimental observations for the xantphosamine ligand, 107.[26] Figure 4 shows scatter plots of bond length changes (D1) versus bite angles for both fragments.

While the ligand bite angle is a popular structural parameter for bidentate ligands, it has only been determined consistently for a limited number of ligands in our LKB-PP. Natural bite angles from MM calculations have been reported for 13 ligands (bite_m),[15,16] whereas only 9 ligands from van Leeuwen's standardized surveys of the CSD[15,16] are represented in the LKB-PP (bite_x). We have also updated a survey of bite angles in [$PdCl_2${LL}] complexes by Ney[27] (see Supporting Information for details), giving results for 15 ligands overlapping the LKB-PP ligand set (bite_Pd). The relationship between these external data sets and LKB-calculated ∠D−M−D bite angles is close to linear, and correlation coefficients are thus high (Table 2), as would be expected, but it is interesting to note that the zinc ∠D1−Zn−D2 data are more highly correlated with natural bite angles than the palladium ∠D1−Pd−D2 results, further

(26) van der Veen, L. A.; Keeven, P. K.; Kamer, P. C. J.; van Leeuwen, P. W. N. M. *Dalton Trans.* **2000**, 2105–2112.
(27) Ney, J. E.; Wolfe, J. P. *J. Am. Chem. Soc.* **2005**, *127* (24), 8644–8651.

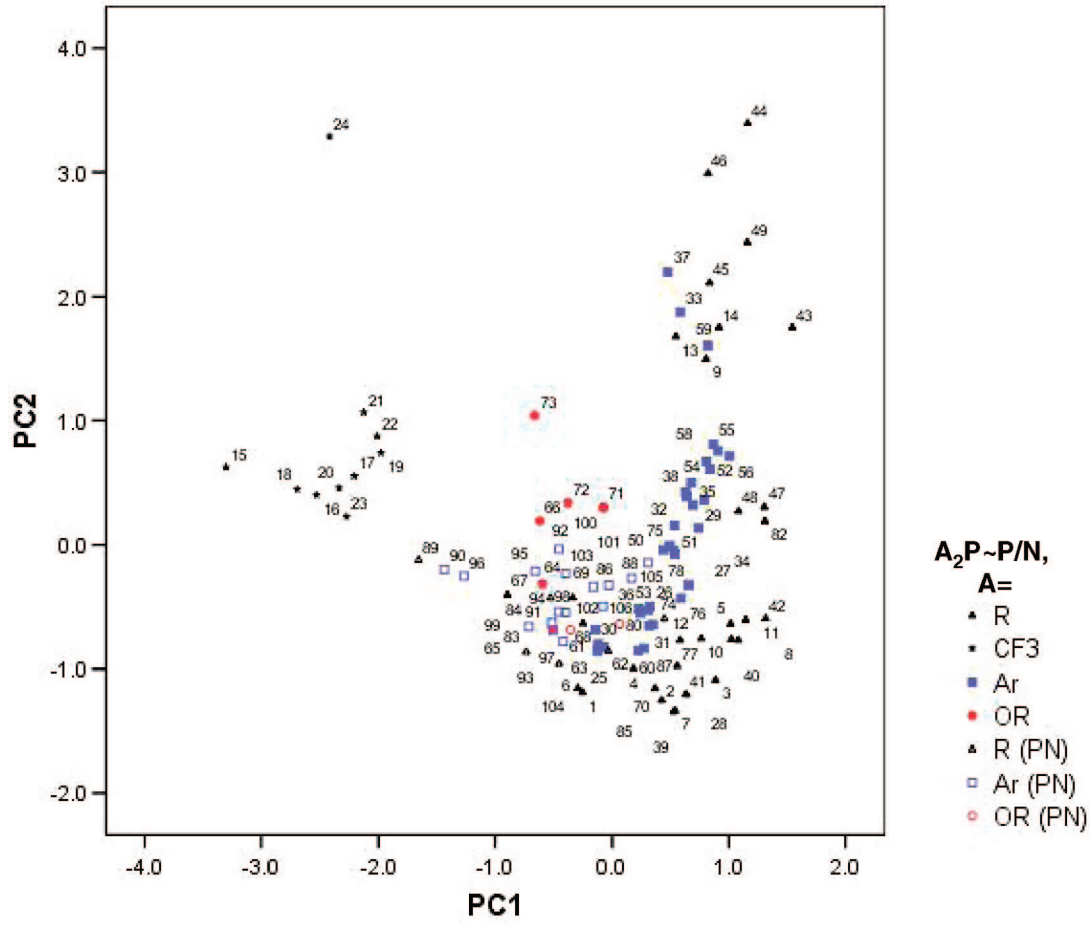


Figure 5. Principal component score plot (PC2 vs PC1) for ligands in LKB-PP.

confirming that ligands are free to adopt a range of bite angles when coordinated to this fragment. Based on comparison with these experimental data sets, our LKB descriptors thus appear suitable for capturing bite angle variations.

Mapping of Chemical Space. We[2] and other authors[10,11] have demonstrated the usefulness of maps of ligand space for monodentate phosphorus(III) donor ligands to locate novel ligands and guide experimental screening and design. To our knowledge, no comparable maps have been derived for bidentate ligands. However, the 27 LKB-PP descriptors shown in Table 1 are difficult to visualize and related terms, e.g., ligand bond lengths in different complexes, are often highly correlated, indicating that visualization could be achieved by suitable simplification. We have thus used principal component analysis (PCA) of the correlation matrix to allow the visualization of ligand space by projecting the descriptor set to fewer dimensions. The resulting principal components (PCs) are linear combinations of the original descriptors which capture large proportions of the variation in the data set in fewer, uncorrelated variables. Pairwise plots of these PCs can be used to illustrate ligand similarities, as these will lead to clusters on the resulting map. In addition, such maps can be used to guide experimental and computational work by highlighting alternative ligands close in ligand space for a given application and by facilitating the design of ligand sets to ensure extensive sampling of chemical space. As discussed previously,[2] this statistical analysis technique is sensitive to outliers and thus not robust to changes in the ligand or descriptor sets, and interpretation of descriptor loadings beyond a simple qualitative description remains dubious. Figure 5 shows a plot of the first two PCs derived from the LKB-PP descriptors (Table 1; see Figure S1 for larger picture).

The first two PCs capture about 52% of the variation in the data set, with PC3 and PC4 adding another 15 and 8%, respectively (Table 3), suggesting that these PCs should also be considered when exploring ligand similarities and in principal component regression (PCR), which uses PCs as variables in linear regression models. A matrix of scatter plots of PC1–4 has been included in the Supporting Information (Figure S2), and the component loadings, i.e., the contribution of each descriptor to individual principal components, are shown in Tables 3 and S3, with the relevant component and loading plots included as Figures S3a and S3b. Principal component scores can then be determined for individual ligands by multiplying the original descriptor values by the appropriate component loading/coefficient. The principal component score plot of PCs 1 and 2 shown in Figure 5 displays chemically intuitive clustering of ligands according to their auxiliary substituents. For P,P-donor ligands we can observe separate "bands" at values of PC1 around –2.4, 0.5, and 0.7 for perfluorinated, aryl, and alkyl substituents respectively. Alkyl-substituted ligands cover the widest range of values, stretching from PC1 = −0.9 to +1.5, while aryl and perfluorinated substituents are more tightly clustered. Alkoxy-substituted ligands (64–66, 71–73) occur mostly around PC1 = −0.5, although there are too few examples in the knowledge base to describe a consistent trend with confidence. Ligands with rigid backbones, such as the anthracene-derived 13 and 24, the DBFphos-derived 14 and 37 and nixantphos 59, as well as those with bulky substituents 9, 33, 43–46, and 49, occur slightly away from the bulk of the

Table 3. Principal Component Loadings (descriptors with contributions < 0.3 are not displayed; see Table S3 for all values)

descriptor	PC1	PC2	PC3	PC4
% of variance explained	29.6	22.4	14.7	8.2
PC eigenvalues	7.99	6.05	3.97	2.20
E_{HOMO}_D1	0.831			
E_{HOMO}_D2	0.767		0.445	
E_{LUMO}_D1				0.711
E_{LUMO}_D2			0.383	0.643
PA_D1	0.821			
PA_D2	0.881			
He_8_wedge	0.404	0.785		
BE(Zn)	0.756		0.335	
Zn−Cl	0.912			
∠D1−Zn−D2	0.514	0.746		
ΔD1−A(Zn)	−0.436	0.426		−0.361
ΔD2−A(Zn)	−0.321		0.897	
ΔA−D1−A(Zn)		−0.448		0.489
ΔA−D2−A(Zn)			−0.886	
ΔZn−D1	−0.501	0.542	−0.311	
ΔZn−D2	−0.523	0.496		0.415
Q(Zn)	−0.780			−0.350
BE(Pd)	0.515	−0.670		
Pd−Cl	0.809		−0.370	
∠D1−Pd−D2	0.449	0.749		
ΔD1−A(Pd)		0.605		−0.310
ΔD2−A(Pd)		0.344	0.692	
ΔA−D1−A(Pd)	−0.381	−0.678		
ΔA−D2−A(Pd)			−0.846	
ΔPd−D1	0.372	0.853		
ΔPd−D2	0.442	0.759		
Q(Pd)	−0.472	0.564	0.499	

data at large positive values of PC2 (>1.4). Mixed P,N-donor ligands (83–106) group mostly around PC1 = –0.5 and PC2 = –0.5, showing some clustering of aryl-substituted ligands, although there are too few examples of other substituents to identify a consistent trend. Since optimization of their [$PdCl_2${LL}] complexes was unsuccessful for ligands 107 and 108, these cannot be shown on the map.

While detailed quantitative interpretation of principal component loadings is difficult due to the sensitivity to outliers and hence lack of robustness of this technique,[2,4,28] the qualitative composition of the first few PCs is of interest for understanding the observed clustering. PC1 may be associated mainly with σ-donor descriptors (e.g., E_{HOMO}, PA, Zn−Cl, Pd−Cl, Zn descriptors), explaining the grouping of ligands according to their auxiliary substituents. PC2 has a strong steric component (He_8_wedge, bite angles), but (π-) electronic factors also play a role (bite angles, Pd descriptors), giving rise to the clustering of large and rigid ligands at high values of PC2. PC3 seems to capture donor asymmetry of the ligands, as descriptors arising from the second donor atom load highly, while the LUMO energies are the main contributors to PC4. However, these results are sensitive to changes in the ligand and descriptors sets, and perhaps the main value of such PCA maps is realized in visualizing ligand space, identifying ligand similarities, and hence guiding ligand screening and design.

Regression Models. Moving beyond a mainly qualitative identification of ligand similarities/differences, we have also explored the application of these knowledge base descriptors in multiple linear regression (MLR) models (see ref 4 for a more detailed discussion). Linear models can provide a useful way of describing the relationship between a range of descriptors and a response variable. Their simple form facilitates the interpretation of individual descriptor contributions to the linear equation, and suitable models can also be used to make predictions for other ligands. In addition, testing knowledge base descriptors in models for a range of "external" data sets (experimental or calculated) allows us to assess their ability to capture ligand properties in different chemical environments and hence their transferability and chemical robustness. We have discussed statistical approaches suitable for the analysis of ligand knowledge bases and the associated challenges for model selection and evaluation elsewhere;[4] here we will focus on simple examples to illustrate this application.

While bidentate ligands are popular in homogeneous catalysis, few data sets have been reported that are suitable for (predictive) regression model testing. Ideally, such data should have been determined under the same experimental conditions and sample a large, representative, and chemically varied ligand set. As discussed above, tables of bite angles have been published for a range of bidentate ligands,[9,15,16,27] but some of these have been collated from a variety of sources, combining crystallographic and molecular mechanics estimates, so the size of useful data sets that overlap with the ligands in LKB-PP is limited. This may be detrimental if models for data prediction are of interest, as some predictions may need to extrapolate outside the data used to fit the model. Nevertheless, we were interested in exploring whether additional descriptors might improve the prediction of bite angles in [$PdCl_2${LL}] complexes (bite_Pd, 15 ligands, discussed above) compared to linear regression models using calculated ligand–metal–ligand angles, ∠D1−Pd−D2, as the only descriptor.

We have derived a range of ordinary least-squares linear regression models with bite_Pd as the response and assessed their performance based on their complexity, i.e., the number of variables in the model,[29] and their prediction errors as determined by cross-validation and bootstrapping (see Computational Details; in addition, ref 4 contains further details of the statistical approaches used).[4,30,31] Resampling methods such as cross-validation and bootstrapping are useful in this case, because splitting such a small data set (15 ligands) into training and test sets would be difficult, so instead prediction errors are estimated by fitting models to different subsets of the data and assessing the quality of predictions for the response data not included in these fits. Variable selection for these models can also be challenging, as there are fewer data than available descriptors and the descriptors are correlated, but these prediction errors allow useful comparisons between models. While methods such as principal component regression (PCR) and partial least-squares (PLS, see, e.g., ref 4 for details of these approaches) would allow us to enter the complete set of variables as derived orthogonal variables, in this case even simple linear regression models show very good performance. Given the known lack of robustness of PCR and PLS[28] and our interest in interpreting descriptor contributions,[4] we have not pursued

(28) (a) Cundari, T. R.; Sârbu, C.; Pop, H. F. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1363–1369. (b) Hadi, A. S.; Ling, R. F. *Am. Statistician* **1998**, *52* (1), 15–19. (c) Ramsey, F. L. *Am. Statistician* **1986**, *40* (4), 323–324.

(29) Adding more variables into a model will generally improve performance as measured by the regression coefficient, but models with fewer variables are usually preferable, as less noise will be introduced. See Hawkins, D. M. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12.

(30) (a) Hawkins, D. M.; Basak, S. C.; Mills, D. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 579. (b) Efron, B.; Tibshirani, R. *J. Am. Stat. Assoc.* **1997**, *92* (438), 548–560. (c) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning; Data Mining, Inference, and Prediction*; Chapman & Hall: New York, 2001. (d) Shao, J. *J. Am. Stat. Assoc.* **1993**, *88* (422), 486–494. (e) Shao, J. *J. Am. Stat. Assoc.* **1996**, *91* (434), 655–665.

(31) (a) Efron, B.; Tibshirani, R., Technical Report **1995**; (b) Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*; Chapman & Hall: San Francisco, 1993.

Table 4. Summary of Regression Models for bite_Pd and Prediction Error Estimates (see Table S4 for model coefficients)

response	N^a	model	descriptors in model	R^2 (adj. R^2)^b	estimated prediction errors^c MLR	CV	boot^d
bite_Pd (°),[27] this work	15	1	∠D1−Pd−D2	0.983 (0.981)	1.30	1.69	1.678
		2	∠D1−Pd−D2, ∠D1−Zn−D2, ΔPd−D1	0.986 (0.982)	1.05	1.99	2.00
		3	∠D1−Pd−D2, ∠D1−Zn−D2, ΔPd−D1, ΔZn−D2	0.987 (0.982)	0.94	2.84	3.14

[a] Number of ligands in sample. [b] Adjusted R^2 takes the number of variables in the model into account when computing the regression coefficient. [c] Mean squared errors. 10-fold CV. [d] 0.632 bootstrap estimate of prediction error.

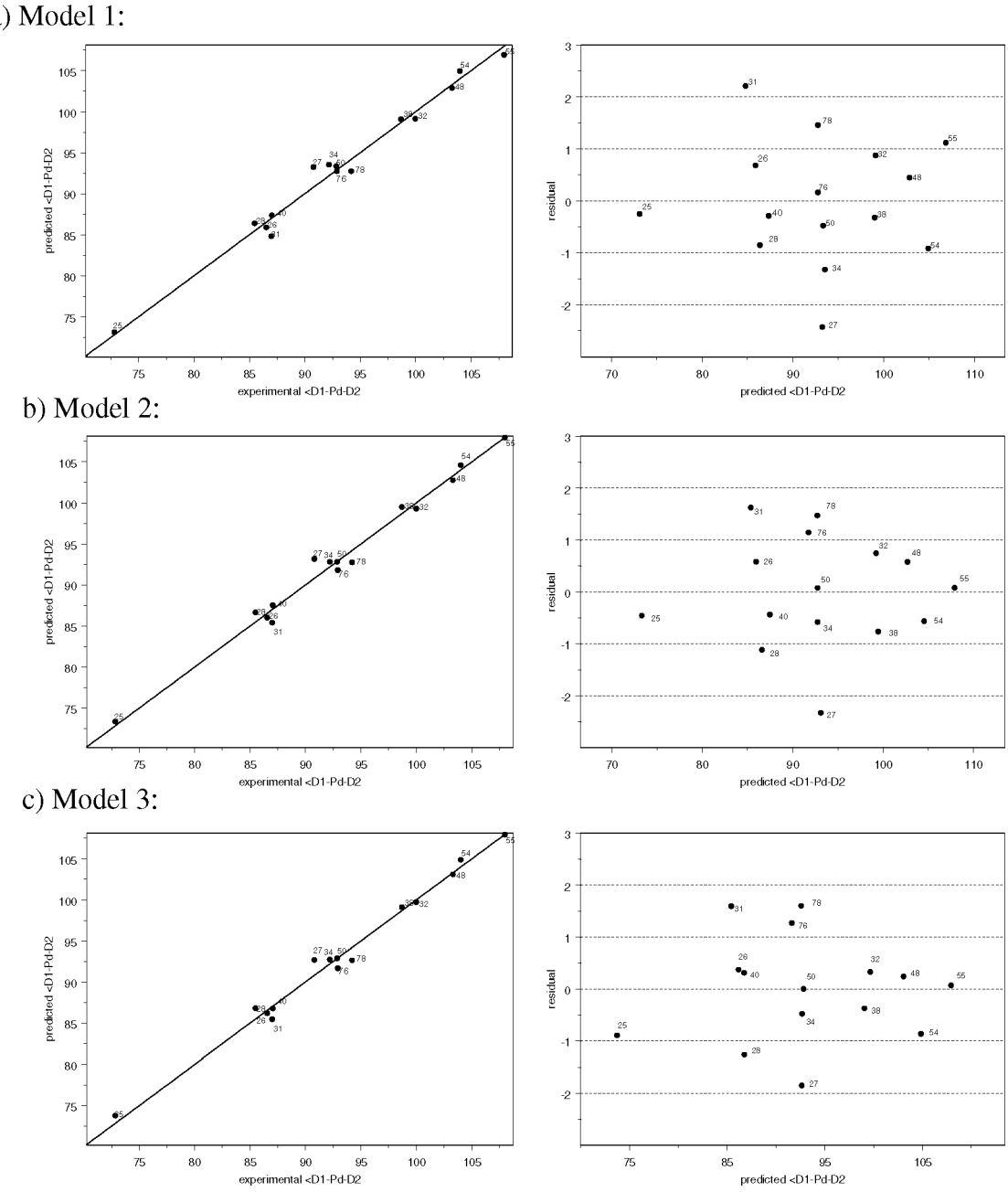


Figure 6. Diagnostic plots for multiple linear regression models 1–3 (Table 4; see Table S5 for experimental and predicted data).

this further. Some of the models derived are summarized in Table 4, with the relevant diagnostic plots for assessing the quality of fit displayed in Figure 6 and full model coefficients as well as experimental and predicted values listed in Tables S4 and S5 in the Supporting Information.

These models demonstrate that while better reproduction of the response data can be achieved by adding descriptors to the model (improving the regression coefficient R^2, which measures the quality of fit, on going from model 1 to model 3), their predictive performance will not necessarily be improved, resulting in larger prediction errors. In this case model 1, exploiting the high linear correlation between experimental and calculated bite angles in [$PdCl_2${LL}] complexes discussed above, gives acceptably good predictions, with residuals, i.e., the difference between experimental and predicted values, of less than 2.5°, while more complex models do not improve the quality of predictions, as indicated by higher prediction error estimates.

Table 5. Summary of Regression Models for BE(Cr) and Prediction Error Estimates (see Table S4 for model coefficients)

response	N^a	model	descriptors in model	R^2 (adj. R^2)^b	estimated prediction errors^c MLR	CV	boot^d
BE(Cr) (kcal mol^{-1})	96	4	PA_D1, BE(Zn), Zn−Cl, ΔD2−A(Zn), BE(Pd)	0.950 (0.947)	3.22	3.91	3.66
		5	PA_D1, BE(Zn), Zn−Cl, ΔD2−A(Zn), ΔA−D1−A(Zn), BE(Pd), Pd−Cl, ΔA−D1−A(Pd)	0.957 (0.953)	2.77	3.40	3.49
		6	E_{HOMO}_D2, PA_D2, He$_8$_wedge, BE(Zn), Zn−Cl, ΔD2−A(Zn), ΔA−D1−A(Zn), ΔZn−D2, BE(Pd), Pd−Cl, ΔPd−D1, ΔPd−D2	0.964 (0.959)	2.34	2.94	3.08

^{a} Number of ligands in sample. ^{b} Adjusted R^2 takes the number of variables in the model into account when computing the regression coefficient. ^{c} Mean squared errors. 10-fold CV. ^{d} 0.632 bootstrap estimate of prediction error.

With such a small data set a comprehensive exploration of models is difficult and sampling of ligand space remains limited. We have, however, also derived models for a calculated response variable, BE(Cr), i.e., the bond energy for complete dissociation of both donor atoms from a $Cr(CO)_4$ fragment. For this response, data have been calculated for 96 ligands, including most P,N-donor ligands. A range of models have been evaluated by assessing their complexity and comparing their prediction errors, and three models with good performance yet of different complexity are shown in Table 5, with the relevant diagnostic plots displayed in Figure 7 (see Figure S4 for plots with all ligand numbers). Model coefficients are listed in Table S4, and all experimental and predicted values are included as Table S6.

For this response variable, further descriptors in the model will actually improve the prediction errors, and given the large number of observations, models with additional variables are certainly acceptable from a statistical point of view. The diagnostic plots of residual versus predicted data allow the identification of ligands for which predictions are consistently difficult. These include ligand 82, where the donor atoms carry different and quite bulky substituents, as well as ligands 40 and 61, which are both fairly rigid with some steric hindrance. Other ligands giving high residuals are BISBI 35 and the related ligand 22 as well as ligand 76, where conformational issues may play a role, and rigid ligands such as 20 and 24, where the backbone rigidity may be important as the $Cr(CO)_4$ fragment is both sterically hindered and forces a small bite angle.

Comparing the descriptors in the models suggests that descriptors one would associate with σ-donation dominate in model 4, which already shows acceptable performance, while some improvement in terms of regression coefficients and prediction errors is realized in model 5, where the additional descriptors arise from the palladium fragment and ligand structural changes on coordination. These descriptors may actually be responsive to steric hindrance in the complex, but might also capture some π-bonding characteristics. Model 6 includes several descriptors specific to the second donor atom, which may help to capture the asymmetry of ligands. In addition, the He$_8$_wedge descriptor is now included, further highlighting the importance of representing the steric properties of ligands especially where hindered complexes are of interest. Standardized model coefficients are listed in Table S4 of the Supporting Information.

These examples have demonstrated that the relationship between different responses and these descriptors can be captured by simple linear models and hence that these descriptors are transferable to different chemical environments. Consideration of prediction errors suggests that such models could also be used to make predictions for novel or untested ligands based on these simple LKB-PP descriptors. However, the availability of suitable experimental response data remains the limiting factor in deriving predictive models based on LKB-PP descriptors.

Summary and Conclusions

This work has extended the ligand knowledge base approach to bidentate P,P- and P,N-donor ligands, developing suitable descriptors collated in a ligand knowledge base, LKB-PP. The design of both ligand set and descriptors has been discussed in the context of earlier LKB studies,[2] and descriptors have been related to data reported in the literature for chelating ligands, such as the popular bite angle parameter. New challenges, arising from the greater conformational flexibility of such ligands compared to monodentate ligands and the possibility for asymmetry of the donor atoms, have been addressed.

These descriptors have been projected to give the first map of chelating ligand space, and the observed clustering as well as potential applications of such maps in visualization and ligand design have been discussed. Simple linear regression models have been used to explore the transferability of descriptors to different chemical environments and the quality of predictions has been evaluated by estimating prediction errors using different resampling techniques. However, further exploration of this application remains hampered by a lack of suitable experimental data sets.

In many synthetic applications of phosphorus(III) donor ligands, both mono- and bidentate ligands are used interchangeably, with the choice of ligand determined by other factors such as catalyst stability, reaction conditions (temperature, solvent), and substrate of interest. While mapping and modeling of the separate ligand spaces is valuable in its own right, work is currently under way to explore how different knowledge bases might be combined to produce maps of wider ligand space.

Computational Details

All calculations used the Jaguar package[32] and the standard Becke−Perdew (BP86) density functional.[33] The Jaguar triple-ζ form of the standard Los Alamos ECP basis set (LACV3P) was used on the transition metal atoms, employing the 6-31G* basis for all other atoms. "Loose" convergence (5 times larger than default criteria) was used for all geometry optimizations. Test calculations using the more stringent default convergence criteria did not lead to significant changes in energies, bond lengths, or angles, but were much more time-consuming. Calculations were performed on isolated molecules, and NBO atomic charges were calculated.[34] Vibrational frequencies were not computed, and so the energetic data do not include a correction for zero-point energy, although we would expect this to be quite small. In the absence of frequency

(32) Schrödinger, L. *Jaguar, 6.0*; Schrödinger, LLC: New York, 2005.
(33) (a) Slater, J. C., *Quantum Theory of Molecules and Solids*, Vol. 4: *The Self-Consistent Field for Molecules and Solids*; McGraw-Hill: New York, 1974. (b) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100. (c) Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, *23*, 5048–5079. (d) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824. (e) Perdew, J. P. *Phys. Rev. B* **1986**, *34*, 7406.
(34) Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Weinhold, F. *NBO 5.0*; Theoretical Chemistry Institute, University of Wisconsin: Madison, 2001.

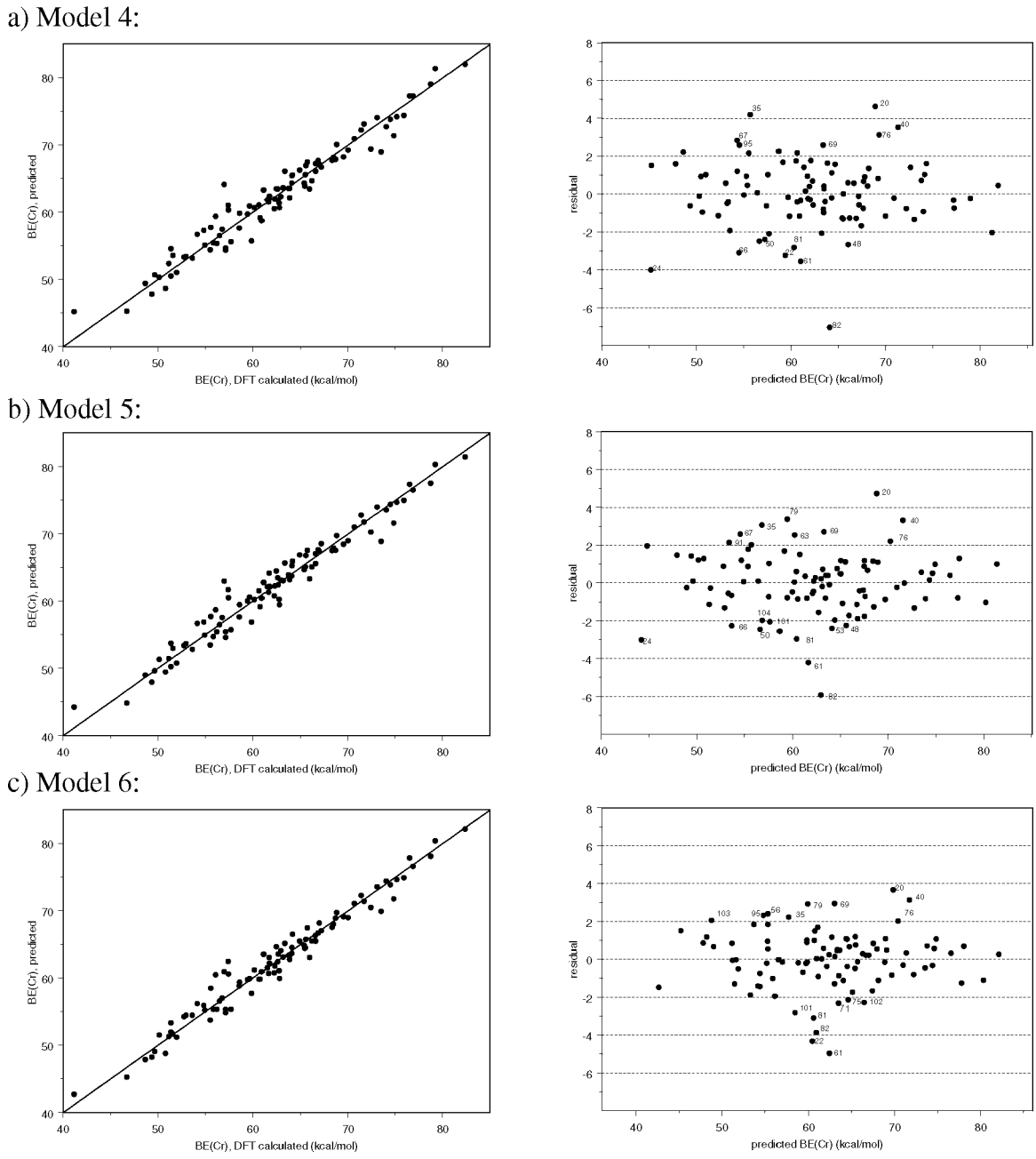


Figure 7. Diagnostic plots for multiple linear regression models 4–6 (Table 5; see Table S6 for experimental and predicted data and Figure S3 for all ligand numbers).

cacluations, stationary points have not been verified as minima. However, most ligands and complexes are large and optimization to transition states seems unlikely for these carefully built low symmetry starting geometries.

Multiple conformations are viable for some of the ligands and complexes, but a comprehensive exploration of conformational space with DFT remains computationally too demanding (vide supra). Instead, we have performed molecular mechanics (MM) conformational searches with the default MMX force field in PCModel.[35] The GMMX module was used for stochastic conformational searches with default settings unless stated otherwise. 500 iteration conformational searches were performed on free ligands and ligands bound to the zinc fragment (stop criteria defined as Emin found 10 times and duplicates found 50 times). Starting geometries for other metal complexes were based on the MM minimum energy conformer found for the zinc complex. The impact of potentially resulting "conformational noise" (i.e., variations in descriptor values between alternative conformers and due to discrepancies between MM and DFT) on the data is discussed above.

Statistical analyses were performed in SPSS for Windows[36] and S-Plus.[37] The leaps routine in S-Plus with adjusted R^2 as the model evaluation criterion was used to select regression models of different complexity (best subset). Multivariate linear regression models were evaluated by their regression coefficients (R^2 and adj. R^2) and by calculating the mean residual from the regression results. In addition, the crossValidation and bootstrapValidation functions as implemented in the S+ Resample library[38] were used to estimate mean squared-prediction errors for these linear regression models from 10-fold crossvalidation and 100 bootstrap replications (0.632 method[31]), respectively.

(35) Gilbert, K. *PCModel*; Serena Software: Bloomington, IN, 2004.

(36) *SPSS for Windows*, Release 14.0; SPSS Inc.: Chicago, IL, 2006.

Acknowledgment. We thank Gareth Owen-Smith, David Hollinshead, David Lathbury, and Dave Hose for many helpful discussions and AstraZeneca for generous funding of this project. J.N.H. thanks the EPSRC for an Advanced Research Fellowship.

Supporting Information Available: Lists of ligands and split ligand descriptors, principal component loadings, regression model coefficients, experimental and predicted response data, details of CSD bite angle survey, and principal component score plots. This material is available free of charge via the Internet at http://pubs.acs.org. A table of descriptor values for all ligands is available from the authors on request.

OM700840H

(37) *S-PLUS, 7.0*; Insightful Corporation: Seattle, WA, 2005.

(38) *S-PLUS Resample Library User's Manual*; Insightful Corporation: Seattle, WA, 2002.