



## Supplementary Materials for

### **Accelerated dinuclear palladium catalyst identification through unsupervised machine learning**

Julian A. Hueffel *et al.*

Corresponding author: Franziska Schoenebeck, franziska.schoenebeck@rwth-aachen.de

*Science* **374**, 1134 (2021)  
DOI: 10.1126/science.abj0999

#### **The PDF file includes:**

Materials and Methods  
Supplementary Text  
Figs. S1 to S23  
Tables S1 to S6  
References

#### **Other Supplementary Material for this manuscript includes the following:**

Movie S1  
Data S1 to S3

# Table of Contents

<b>1. Methods and Terminology .....</b>	<b>4</b>
1.1. Principal component analysis (PCA).....	4
1.2. <i>k</i> -Means clustering.....	4
<b>2. Initial Clustering of the LKB-P.....</b>	<b>5</b>
2.1. Ligand Knowledge Base for Monodentate Phosphines (LKB-P) .....	5
2.2. Euclidean Distances between the Reference Ligands .....	11
2.2. <i>k</i> -Means Optimization .....	12
2.3. Results of the Initial Clustering .....	14
2.4. Statistical Evaluation .....	17
<b>3. Problem-specific Clustering .....</b>	<b>19</b>
3.1. Computational Details .....	19
3.1.1. DFT calculations.....	19
3.1.2. Descriptor extraction and data analysis.....	19
3.1.3. Calculation of Tolman cone angles.....	19
3.2. Problem-specific Descriptors .....	20
3.2.1. Pearson correlation analysis of problem-specific descriptors.....	22
3.2.2. Principal component analysis of problem-specific descriptors.....	24
3.3. <i>k</i> -Means Optimization .....	30
3.4. Results of the Problem-specific Clustering .....	32
3.5. Statistical Evaluation .....	37
3.6. General vs. Problem-Specific Descriptors.....	38
<b>4. Experimental Validation.....</b>	<b>41</b>
4.1. Experimental Details .....	41
4.1.1. Reagents, starting material and solvents.....	41
4.1.2. Characterization.....	41
4.2. Synthesis and Characterization of Phosphine Ligands.....	42
4.2.1. 1-Adamantyl acetate (1-Ad)OAc .....	42
4.2.2. 1-Adamantyl-di- <i>tert</i> -butylphosphine .....	42
4.2.3. Di-1-adamantyl- <i>tert</i> -butylphosphine.....	43
4.2.4. 1-( <i>tert</i> -Butyl)-2,2,6,6-tetramethylphosphinan-4-one .....	43
4.2.5. 2,2,6,6-Tetramethyl-1-phenylphosphinan-4-one .....	44
4.3. Synthesis and Characterization of Bis(phosphine)palladium(0) Complexes .....	45
4.3.1. General Procedure .....	45
4.3.2. Bis(di-1-adamantyl- <i>tert</i> -butylphosphine)palladium(0).....	45
4.3.3. Bis(di- <i>tert</i> -butyl(cyclohexyl)phosphine)palladium(0).....	45
4.3.4. Bis(di- <i>tert</i> -butyl(neopentyl)phosphine)palladium(0).....	46
4.3.5. Bis(di- <i>tert</i> -butyl(ferrocenyl)phosphine)palladium(0) .....	46
4.3.6. Bis(2,2,6,6-tetramethyl-1-( <i>tert</i> -butyl)-4-phosphorinanone)palladium(0) .....	46
4.3.7. Bis(2,2,6,6-tetramethyl-1-phenyl-4-phosphorinanone)palladium(0) .....	47

<b>4.4. Synthesis and Characterization of Iodide-bridged Pd(I) Dimers.....</b>	<b>48</b>
4.4.1. General Procedure A .....	48
4.4.2. General Procedure B. ....	48
4.4.3. Di- $\mu$ -iodobis(1-adamantyl-di- <i>tert</i> -butylphosphine)dipalladium(I).....	48
4.4.4. Di- $\mu$ -iodobis(di-1-adamantyl- <i>tert</i> -butylphosphine)dipalladium(I). ....	48
4.4.5. Di- $\mu$ -iodobis(di- <i>tert</i> -butyl(cyclohexyl)phosphine)dipalladium(I). ....	49
4.4.6. Di- $\mu$ -iodobis(di- <i>tert</i> -butyl(neopentyl)phosphine)dipalladium(I).....	49
4.4.7. Di- $\mu$ -iodobis(di- <i>tert</i> -butyl(ferrocenyl)phosphine)dipalladium(I) .....	49
4.4.8. Di- $\mu$ -iodobis(2,2,6,6-tetramethyl-1-( <i>tert</i> -butyl)-4-phosphorinanone)dipalladium(I)....	50
4.4.9. Di- $\mu$ -iodobis(2,2,6,6-tetramethyl-1-phenyl-4-phosphorinanone)dipalladium(I)....	50
<b>4.5. <math>^{31}\text{P}</math> NMR Studies .....</b>	<b>51</b>
4.5.1. General procedure employed for <i>in situ</i> testing of ligands. ....	51
4.5.2. Attempted comproportionation of $\text{Pd}(\text{PCy}_3)_2$ and $\text{PdI}_2$ .....	51
4.5.3. $[\text{Pd}(\mu\text{-I})(\text{QPhos})]_2$ .....	52
<b>4.6. Crystallographic data.....</b>	<b>53</b>
<b>4.7. NMR-spectra .....</b>	<b>55</b>
4.7.1. 1-( <i>tert</i> -Butyl)-2,2,6,6-tetramethylphosphinan-4-one .....	55
4.7.2. Bis(di-1-adamantyl- <i>tert</i> -butylphosphine)palladium(0).....	56
4.7.3. Bis(di- <i>tert</i> -butyl(cyclohexyl)phosphine)palladium(0).....	57
4.7.4. Bis(di- <i>tert</i> -butyl(neopentyl)phosphine)palladium(0) .....	58
4.7.5. Bis(2,2,6,6-tetramethyl-1-( <i>tert</i> -butyl)-4-phosphorinanone)palladium(0).....	60
4.7.6. Bis(2,2,6,6-tetramethyl-1-phenyl-4-phosphorinanone)palladium(0).....	61
4.7.7. Di- $\mu$ -iodobis(1-adamantyl-di- <i>tert</i> -butylphosphine)dipalladium(I).....	63
4.7.8. Di- $\mu$ -iodobis(di-1-adamantyl- <i>tert</i> -butylphosphine)dipalladium(I).....	64
4.7.9. Di- $\mu$ -iodobis(di- <i>tert</i> -butyl(cyclohexyl)phosphine)dipalladium(I):.....	65
4.7.10. Di- $\mu$ -iodobis(di- <i>tert</i> -butyl(neopentyl)phosphine)dipalladium(I).....	66
4.7.11. Di- $\mu$ -iodobis(di- <i>tert</i> -butyl(ferrocenyl)phosphine)dipalladium(I) .....	68
4.7.12. Di- $\mu$ -iodobis(2,2,6,6-tetramethyl-1-( <i>tert</i> -butyl)-4-phosphorinanone)dipalladium(I) ..	69
4.7.13. Di- $\mu$ -iodobis(2,2,6,6-tetramethyl-1-phenyl-4-phosphorinanone)dipalladium(I) ...	71
<b>5. Captions for Data Files .....</b>	<b>73</b>
<b>6. Captions for Movie Files .....</b>	<b>73</b>

# 1. Methods and Terminology

**1.1. Principal component analysis (PCA).** Principal component analysis (PCA) is a statistical multivariate technique that can be used to reduce the dimensionality of a dataset by emphasizing variation and bring out strong patterns in a dataset. Important information is extracted from the descriptor columns of the dataset and expressed as a small set of new orthogonal variables, so-called principal components (PCs). The high-dimensional data points are projected on these PCs in order to capture the most variation in the dataset (the distance between the different projections is maximum) in a way that the total distance from the original points to their corresponding projections is minimal. Hence, the representation is as close to the original data as possible. The best component (PC1) conveys maximum variation among data points and minimum error. PC2 describes the second most variation in the data and is orthogonal to the first PC etc. The principal components can be used as a new coordinate system to monitor similarity patterns by visualizing the observations as points in a 2D or 3D map. Furthermore, PCA offers the advantage to estimate the importance of the descriptors by analyzing the loadings of the PCs. The PC loadings correspond to the coefficients of the linear combination of descriptors from which the PC is constructed and thus reflect how much a given descriptor contributes to the variance that is captured by a PC, *i.e.* its importance.(41)

**1.2.  $k$ -Means clustering.**  $k$ -means, also known as Lloyd's algorithm,(42) is one of the most popular clustering algorithms to detect spherical clusters within a dataset. In  $k$ -means, a given dataset is partitioned into a predefined number  $k$  of subgroups (clusters). As a so-called hard clustering technique, each observation or point is assigned to only one cluster. Thereby, each cluster is associated with its (virtual) center, the so-called centroid.

In the original  $k$ -means algorithm, the centroids are initialized randomly by choosing  $k$  samples from the dataset. Since this can result in the algorithm being stuck in a local minimum during optimization more recent implementations, namely “ $k$ -means++” (43) (which was used in this study), initialize the centroids based on the weighted probability distribution of the squared distances between the clusters. This ensures that the centroids are initialized distant from each other and the algorithm then iteratively optimizes the location of the centroids, such that the sum of squared distances between the cluster members and the centroids becomes minimal, *i.e.* that the data points within a cluster become as similar as possible to each other, while differing as much as possible from points in other clusters. Since the algorithm depends on unweighted distances, all features are treated equally and thus, estimating the feature importance is not trivial.(44)

Optimizing the clustering parameters is a fundamental challenge for unsupervised machine learning algorithms and typically not trivial. Even though a quantitative optimization is not possible,  $k$ -means offers possibilities to estimate the optimum number of clusters  $k$  to be used. One approach is the so-called *elbow method*, which is based on plotting the inertia or distortion of a model as a function of the cluster number  $k$ . The point of inflection on the curve, *i.e.* the point after which the inertia or distortion starts to decrease in a linear fashion, then indicates the best fit of the model.(45, 46)

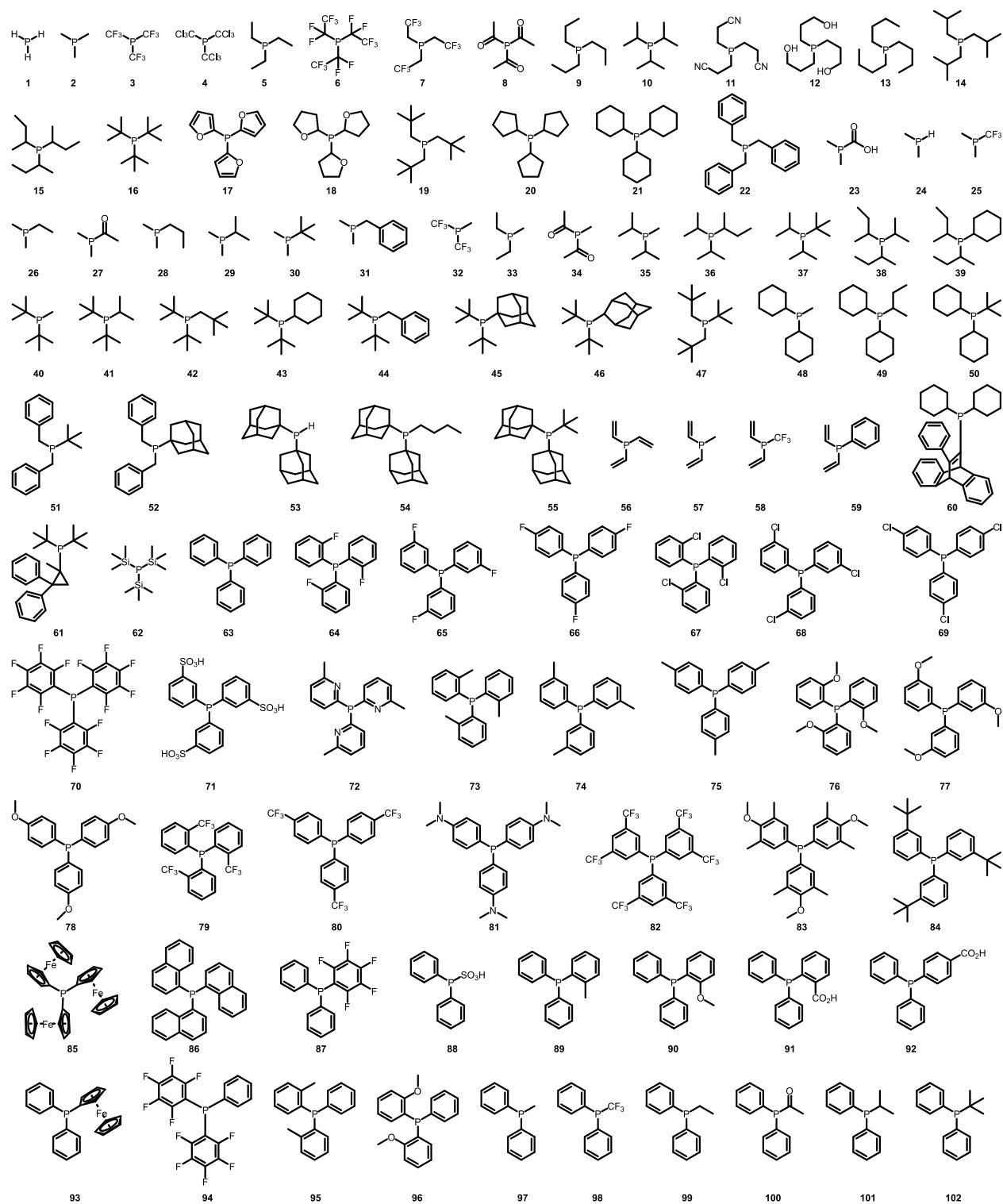
Another approach is the analysis of the *silhouette coefficient*, which is a measure of the distance between a point in one cluster to the points in the neighboring clusters. Silhouette coefficients range from 1 (very distant from the neighboring clusters) to -1 (part of neighboring cluster, *i.e.* misclassification), while a coefficient of 0 represents the decision boundary between two clusters. The best fit of cluster number  $k$  is indicated by a high average silhouette score, low fraction of negative silhouette scores and ideally an even cluster size.(47)

## 2. Initial Clustering of the LKB-P

### 2.1. Ligand Knowledge Base for Monodentate Phosphines (LKB-P)

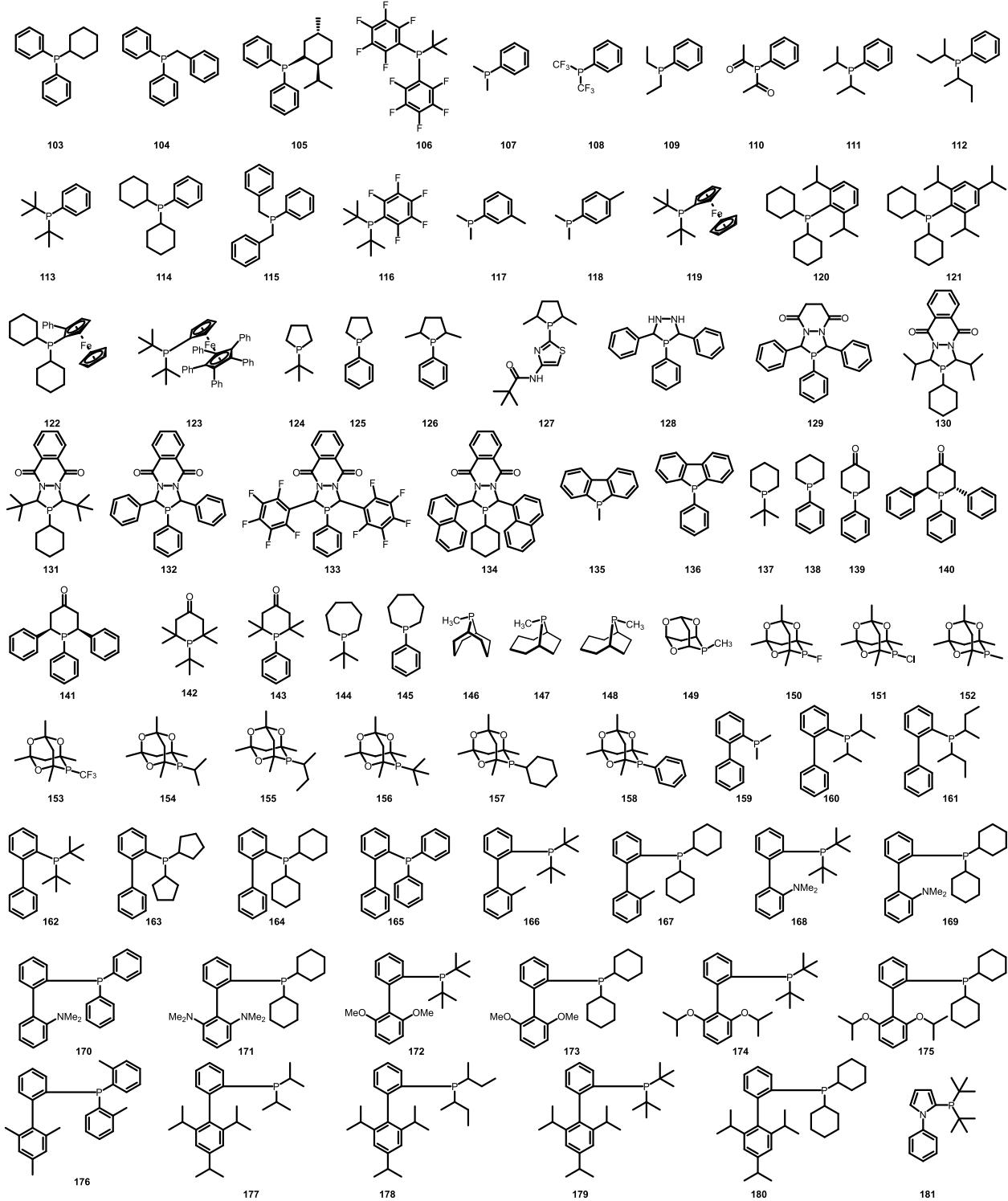
Starting point for our investigation was the ligand knowledge base for monodentate phosphine donors (LKB-P) that aims to describe the general properties of the included ligands. (8, 9) The database covers a total of 348 different phosphines (Fig. S1 - Fig. S4) which can conceptually be divided into 8 different ligand types. For a given ligand 28 different descriptors (see Table S1) are provided, ranging from certain ligand-specific data, such as proton affinity or HOMO/LUMO energies, to calculated data that describe the ligand's interaction in model complexes (coordination to Au, Pt, Pd, B). Principal component analysis (PCA) was employed to reduce the dimensionality of the descriptor space and visualize the results within 2D maps.

For our initial clustering we used the entire dataset as published. (11, 12) For visualization and comparison of the results we also employed the published first two principal components PC1 and PC2 which capture 65.1% of the variation within the data and reflect the electron-withdrawing character (PC1) and bulkiness of a ligand (PC2), respectively.



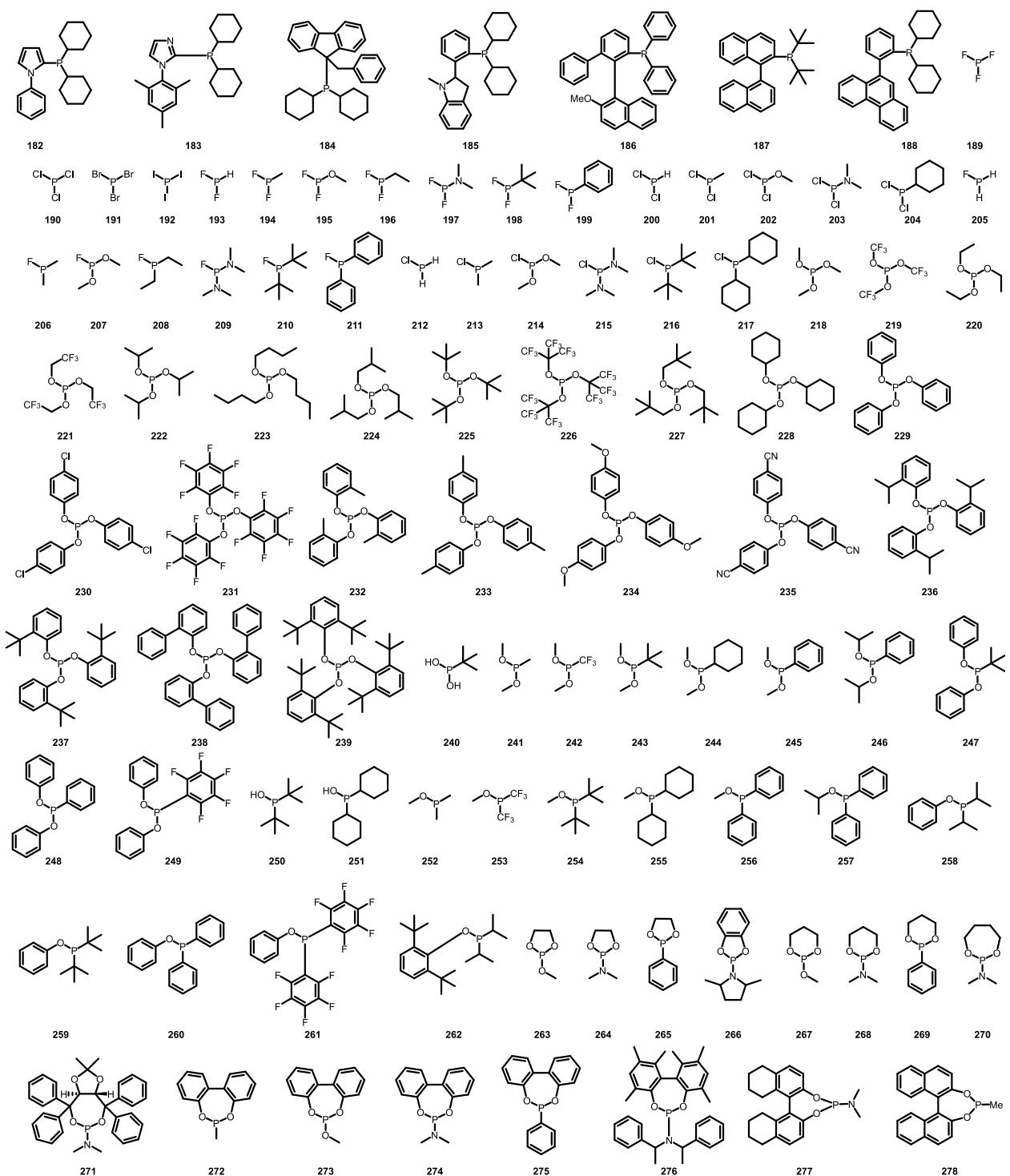
**Fig. S1.**

Structures of the ligands (IDs 1 – 102) included in the ligand knowledge base for monodentate phosphines (LKB-P).<sup>(12)</sup>



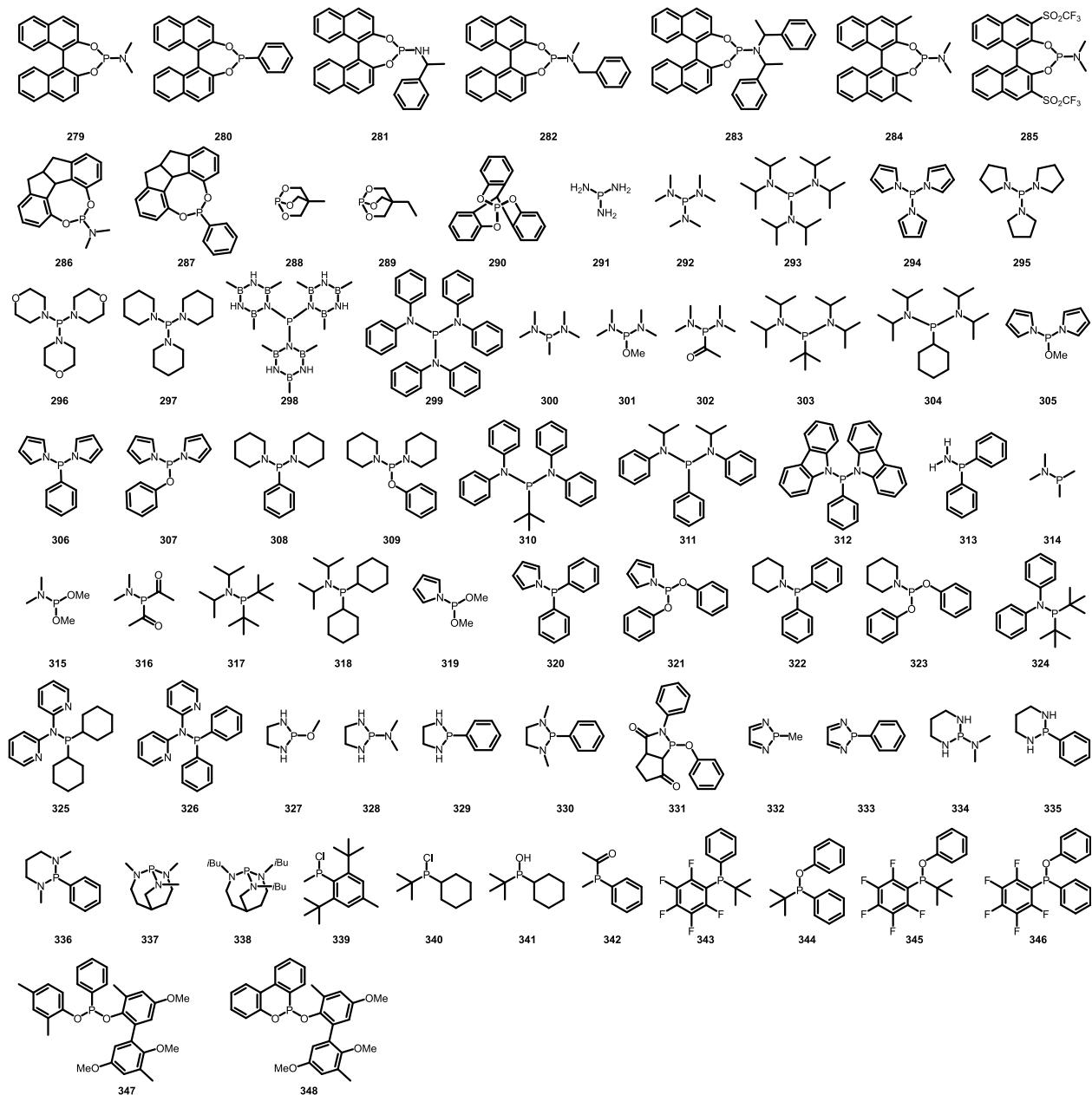
**Fig. S2.**

Structures of the ligands (IDs 103 – 181) included in the ligand knowledge base for monodentate phosphines (LKB-P).<sup>(12)</sup>



**Fig. S3.**

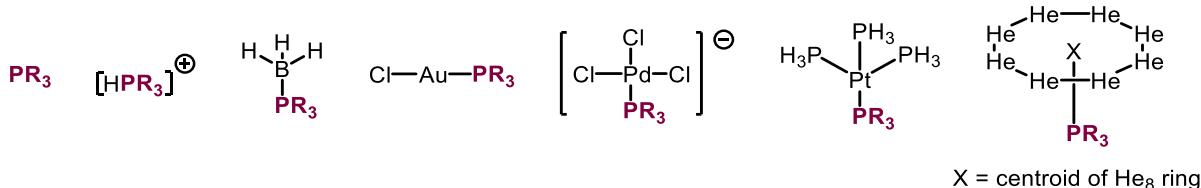
Structures of the ligands (IDs 182 – 278) included in the ligand knowledge base for monodentate phosphines (LKB-P).<sup>(12)</sup>



**Fig. S4.**

Structures of the ligands (IDs 279 – 348) included in the ligand knowledge base for monodentate phosphines (LKB-P).<sup>(12)</sup>

Model complexes used in LKB-P:



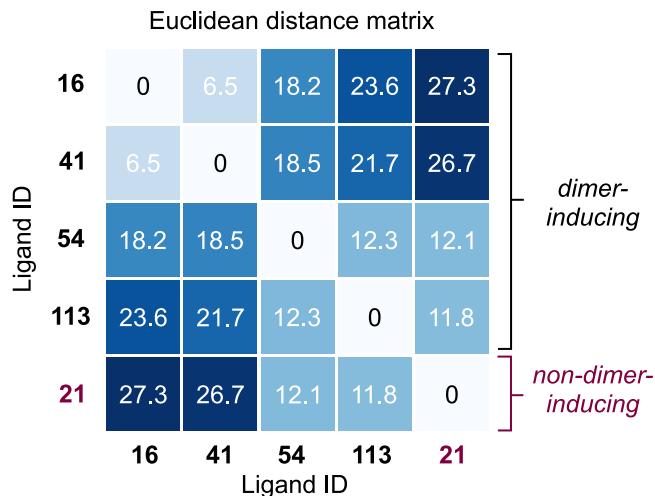
Descriptor	Unit	Derivation
E(HOMO)	Hartree	Energy of highest occupied molecular orbital of $\text{PR}_3$
E(LUMO)	Hartree	Energy of lowest unoccupied molecular orbital of $\text{PR}_3$
He8_steric	$\text{kcal mol}^{-1}$	Interaction energy between singlet $\text{PR}_3$ in ground-state conformation and ring of 8 helium atoms; $E_{\text{ster}} = E_{\text{tot}}(\text{system}) - [E_{\text{tot}}(\text{He}_8) + E_{\text{tot}}(\text{PR}_3)]$
PA	$\text{kcal mol}^{-1}$	Proton affinity of $\text{PR}_3$
Q(B)	-	NBO charge on $\text{BH}_3$ fragment
BE(B)	$\text{kcal mol}^{-1}$	Bond energy for dissociation of P-ligand from $\text{BH}_3$ fragment: $\text{BE} = [E_{\text{tot}}(\text{fragment}) + E_{\text{tot}}(\text{PR}_3)] - E_{\text{tot}}(\text{complex})$
P-B	$\text{\AA}$	P-B distance
$\Delta P\text{-A(B)}$	$\text{\AA}$	Change in average P-R bond length compared to free ligand
$\Delta A\text{-P\text{-}A(B)}$	deg	Change in average R-P-R angle compared to free ligand
Q(Au)	-	NBO charge on $\text{AuCl}$ fragment
BE(Au)	$\text{kcal mol}^{-1}$	Bond energy for dissociation of $\text{PR}_3$ from $[\text{AuCl}]$ fragment $\text{BE} = [E_{\text{tot}}(\text{fragment}) + E_{\text{tot}}(\text{PR}_3)] - E_{\text{tot}}(\text{complex})$
Au-Cl	$\text{\AA}$	Au-Cl distance
P-Au	$\text{\AA}$	Au-P distance
$\Delta P\text{-A(Au)}$	$\text{\AA}$	Change in average P-R bond length in complex compared to free ligand
$\Delta A\text{-P\text{-}A(Au)}$	deg	Change in average R-P-R angle compared to free ligand
Q(Pd)	-	NBO charge on $[\text{PdCl}_3]$ - fragment
BE(Pd)	$\text{kcal mol}^{-1}$	Bond energy for dissociation of $\text{PR}_3$ from $[\text{PdCl}_3]$ - fragment $\text{BE} = [E_{\text{tot}}(\text{fragment}) + E_{\text{tot}}(\text{PR}_3)] - E_{\text{tot}}(\text{complex})$
Pd-Cl trans	$\text{\AA}$	Pd-Cl distance, trans to ligand
P-Pd	$\text{\AA}$	Pd-P distance
$\Delta P\text{-A(Pd)}$	$\text{\AA}$	Change in average P-R bond length compared to free ligand
$\Delta A\text{-P\text{-}A(Pd)}$	deg	Change in average R-P-R angle compared to free ligand
Q(Pt)	-	NBO charge on $[(\text{Ph}_3\text{P})_3\text{Pt}]$ fragment
BE(Pt)	$\text{kcal mol}^{-1}$	Bond energy for dissociation of P-ligand from $[\text{Pt}(\text{Ph}_3)_3]$ fragment $\text{BE} = [E_{\text{tot}}(\text{fragment}) + E_{\text{tot}}(\text{PR}_3)] - E_{\text{tot}}(\text{complex})$
P-Pt	$\text{\AA}$	P-Pt distance
$\Delta P\text{-A(Pt)}$	$\text{\AA}$	Change in average P-R bond length compared to free ligand
$\Delta A\text{-P\text{-}A(Pt)}$	deg	Change in average R-P-R angle compared to free ligand
$\angle(\text{H}_3\text{P})\text{Pt}(\text{PH}_3)$	deg	Average $(\text{H}_3\text{P})\text{Pt}(\text{PH}_3)$ angle
S4' calc	deg	$(\Sigma \angle Z\text{PR} - \Sigma \angle R\text{PR})$ , where $Z = \text{BH}_3, [\text{PdCl}_3]\text{-}, [\text{Pt}(\text{Ph}_3)_3], [\text{AuCl}]$

**Table S1.**

Full list of descriptors used in the LKB-P and in our initial clustering. Adapted from reference (12).

## 2.2. Euclidean Distances between the Reference Ligands

Using all descriptors provided by the LKB-P, the Euclidean distance matrix between the reference ligands (positive: **16**, **41**, **54**, **113**; negative: **21**) was calculated (Fig. S5). Our analysis revealed that two of the dimer-inducing ligands (**54** and **113**) are closer to the non-dimer-inducing ligand **21** and thus considered more similar, than to the other dimer-inducing ligands (**16** and **41**). Clearly, this approach does not provide insights into the problem of interest in this study.



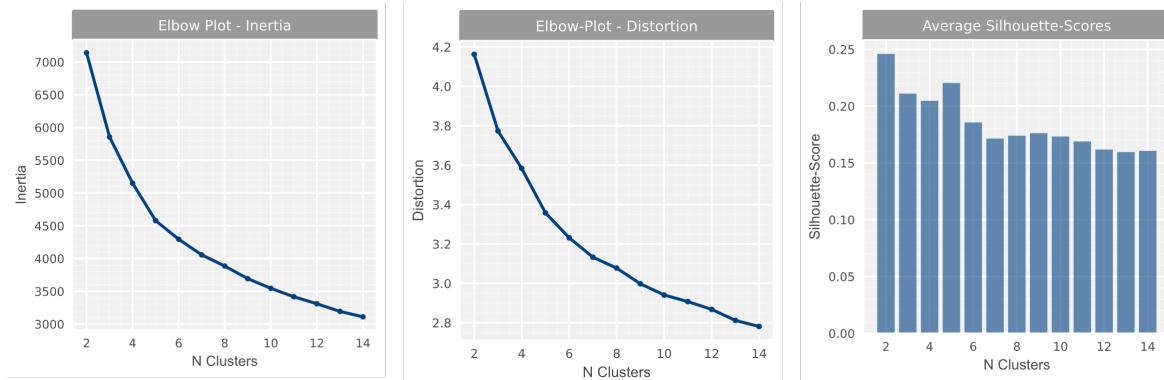
**Fig. S5.**

Euclidean distance matrix of the reference ligands in the feature space captured by the LKB-P descriptors.

## 2.2. $k$ -Means Optimization

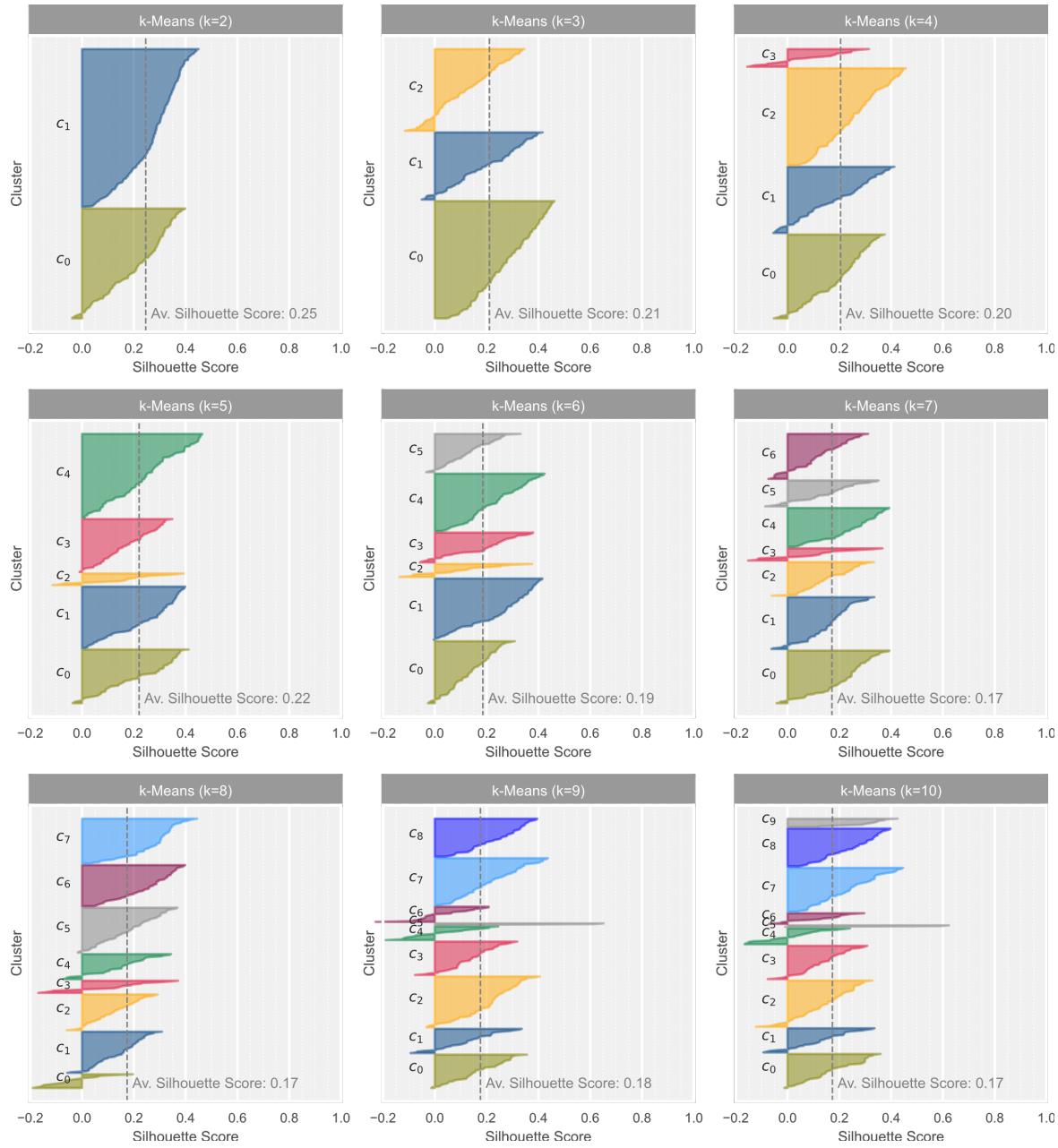
In order to find the optimum number of clusters  $k$ , we applied both the elbow method and analysis of silhouette scores (Fig. S6). For the elbow plots based on inertia (within-cluster sum-of-squares; Fig. S6, left) and distortion (sum of square errors; Fig. S6, middle), the point of inflection of the curve can be located between  $k = 6$  and  $k = 8$ , corresponding to the optimal cluster numbers according to the elbow method. Analysis of the average silhouette coefficients (measure of the distance between a point in one cluster to the points in the neighboring clusters; Fig. S6, right) indicates a similar performance for all values of  $k$ . When analyzing the silhouette coefficient per sample for each cluster number  $k$  (Fig. S7), we found that for all values of  $k$  small misclassifications (negative silhouette scores) occur for at least one of the clusters, which is likely a result of the high data density.

We therefore decided to choose a cluster number of  $k = 8$ , which is supported by the results from the elbow plot and corresponds to the number of different ligand types within the LKB-P. Hence it allowed us to evaluate whether the clustering algorithm would resemble this conceptual classification. The plot for the silhouette scores per sample for  $k = 8$  (Fig. S7, bottom-left) shows that all clusters contain ligands for which the silhouette coefficient is above the average silhouette score, which is generally desirable. While the cluster bearing our negative reference ligands (C6, see Fig. 2A in the main text) contains no misclassified samples (see section 1. for explanation), small misclassifications were found in the clusters bearing our positive references (C1 and C4, see Fig. 2A in the main text).



**Fig. S6.**

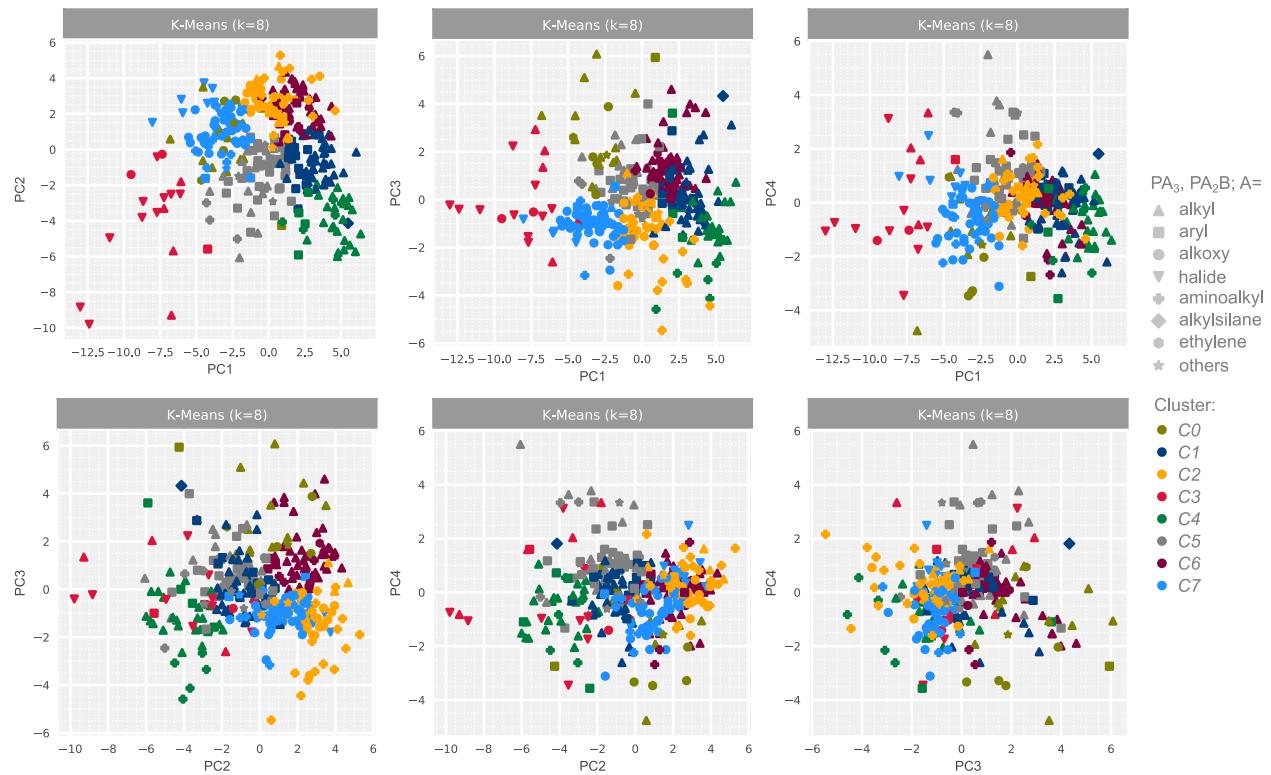
Optimization of the number of clusters  $k$  for the initial  $k$ -means clustering of the LKB-P using the *elbow-method* based on inertia (left) and distortion (middle) as well as the analysis of the average *silhouette score* (right).



**Fig. S7.**

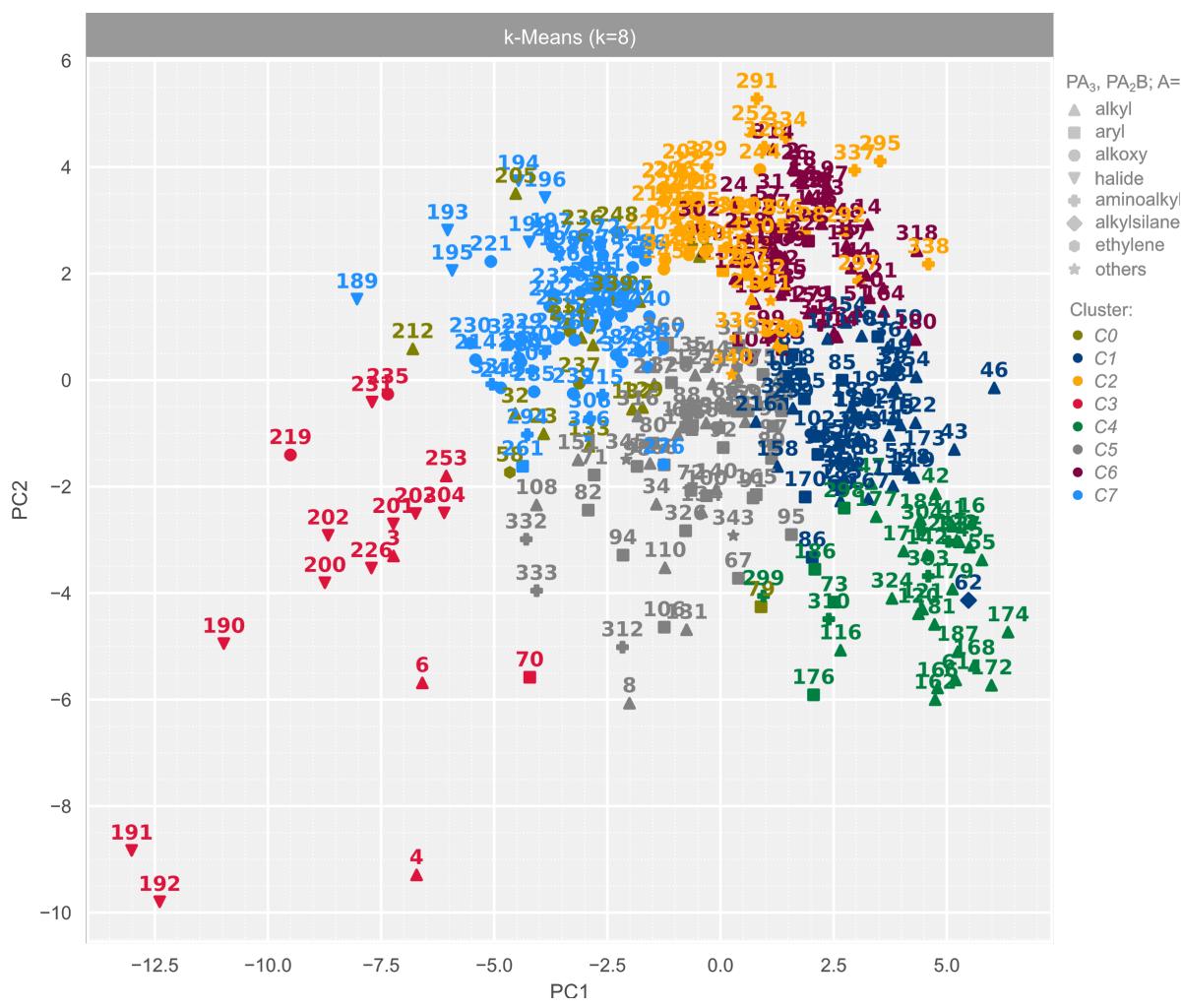
Silhouette scores per sample for different cluster numbers  $k$  in the initial clustering of the LKB-P.

### 2.3. Results of the Initial Clustering



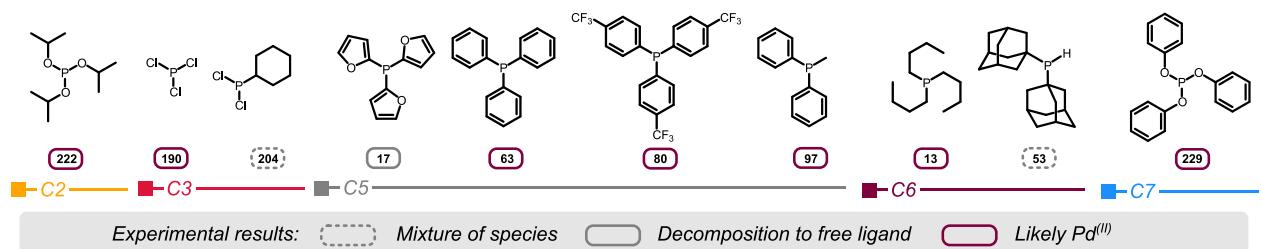
**Fig. S8.**

Results of the initial  $k$ -means clustering ( $k = 8$ ) of the LKB-P visualized using PC1 to PC4 which were adapted from reference (12).



**Fig. S9.**

Detailed overview of the clusters resulting from the initial  $k$ -means clustering ( $k = 8$ ) of the LKB-P. Principal components PC1 and PC2 as well as the ligand IDs are adapted from reference (12).

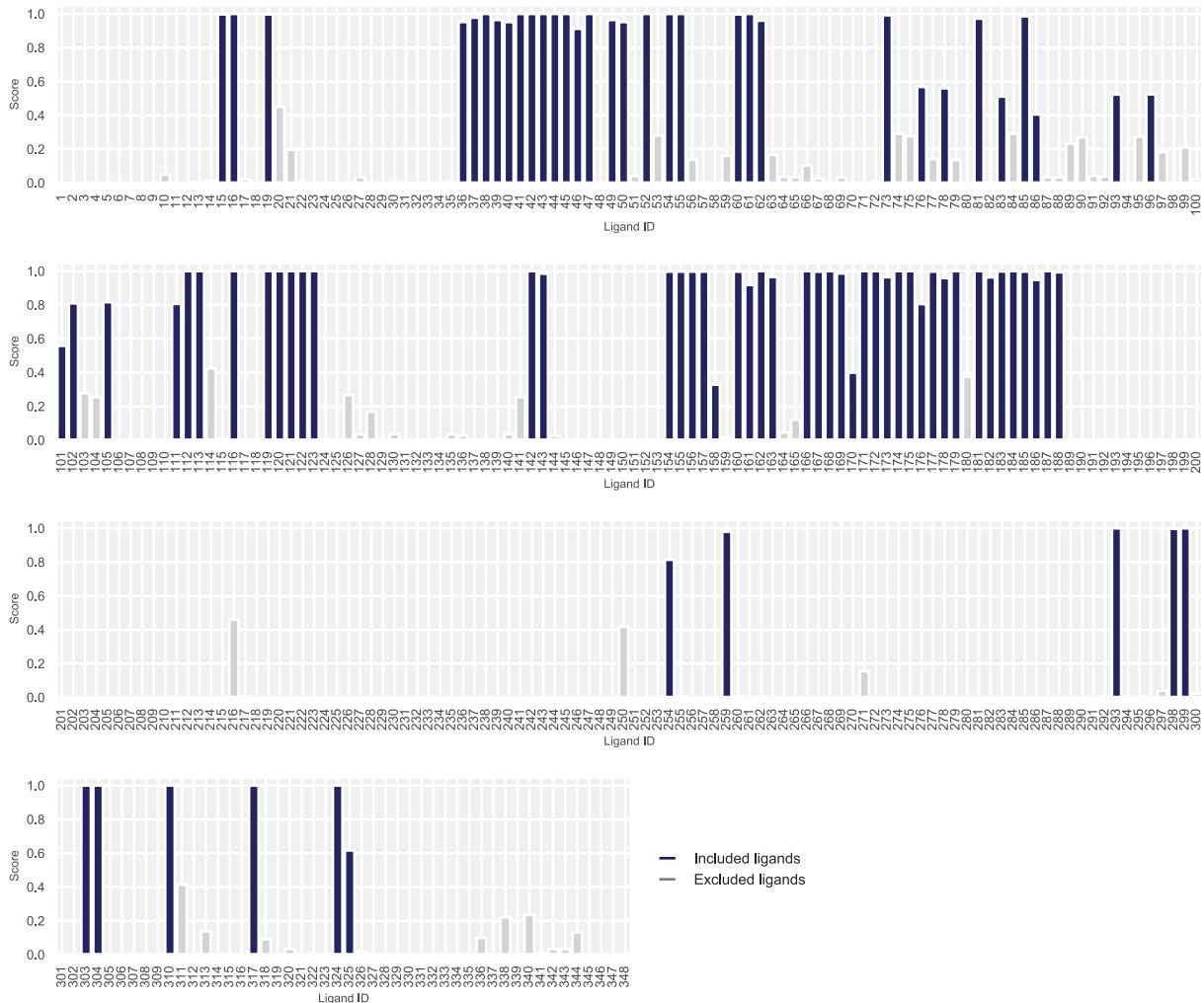


**Fig. S10.**

Experimental results for selected ligands that were not included in one of the positive reference-containing clusters yielded by the initial clustering of the LKB-P.

## 2.4. Statistical Evaluation

To ensure the reproducibility of our results, the clustering reported in the main text was performed using a fixed random state. Nevertheless, we set out to explore the relevance of our predictions in a statistical context. Therefore, we reperformed the clustering 1000 times with different random initializations and analyzed how many times  $N$  each ligand was clustered with our positive references (**16, 41, 54, 113**), *i.e.* how often each ligand was considered similar to our references. The results were summarized into a score  $S$  using the total number of models built ( $S = N/1000$ ) and are shown in Fig. S11. The vast majority of the ligands (78 of 89) predicted by the reported model exhibited a score  $S > 0.8$  (*i.e.* 80% of the 1000 models). The remaining 11 ligands (**76, 78, 83, 86, 93, 96, 101, 158, 170, 216, 325**) featured smaller scores ( $S \sim 0.5$ ). None of the ligands that were not in the initial prediction set exhibited a score  $S > 0.5$  (*i.e.* 50% of the 1000 models).



**Fig. S11.**

Statistical evaluation for the initial clustering. The clustering was performed 1000 times with different random initializations to evaluate how often each ligand was clustered with the dimer-inducing reference ligands (**16, 41, 54, 113**), which is reflected by the score  $S$ . Colored in blue are

the ligands that were predicted by the reported model, while the ligands which were not included in these predictions are colored in grey. Ligand IDs are adapted from reference (12).

### 3. Problem-specific Clustering

#### 3.1. Computational Details

**3.1.1. DFT calculations.** DFT calculations were employed for the calculation of problem-specific Pd-structures and to derive the corresponding problems-specific descriptors for further data analysis (as defined in Fig 2B of the main text).

Geometry optimization and frequency calculation were performed manually using Gaussian16 (48) with the B3LYP-D3(BJ) (49-52) functional employing SDD pseudo-potential and associated basis set for Pd, Fe and I in combination with the 6-31G(d) basis set for all other atoms. Minima were verified by ensuring that no imaginary frequency was present. Subsequently, single-point-energies were calculated using the CPCM solvent model for toluene while employing a combination of SDD pseudo-potentials for Pd, Fe and I and 6-311++G(d,p) basis set for all other atoms. In addition, Natural Bond Orbitals (NBOs) as well as the resulting Wiberg bond orders (WBOs) were calculated using the nbo 3.1 program package (53-60) included in Gaussian16. Based on previous, extensive conformational analysis of the dinuclear Pd-species  $[Pd^I(\mu-I)L]_2$  and  $[Pd^{II}(\mu-I)L]_2$  (structure D and G in Fig 2B of the main text, respectively) formed by the ligands 16, 21, 41, 43, 45, 50, 54 and 55 (IDs adapted from ref. (12)), an *anti*-conformation was kept for all ligands in the calculations of the dimers.

**3.1.2. Descriptor extraction and data analysis.** Descriptor extraction and data analysis was performed in Python 3.6.7. Energy related descriptors were extracted directly from the .log-files of the DFT calculations using regular expressions, while the openbabel library (61) was used for the extraction of structural descriptors on the basis of the optimized geometries. For the calculation of Sterimol descriptors, (31) the *sterimoltools.py*-script published by Paton and co-workers (62) was adapted and applied to the optimized geometries. The pandas library (63) was used to create the final .csv-file containing all descriptors and manage the data within the code. Data preprocessing and analysis was then carried out making heavy use of the scikit-learn library.(64)

**3.1.3. Calculation of Tolman cone angles.** Tolman cone angles  $\theta_T$  mentioned in the main text were calculated in Mathematica 12.1 (65) based on the optimized geometries obtained from the DFT calculations using the script published by Allen and co-workers.(66)

### 3.2. Problem-specific Descriptors

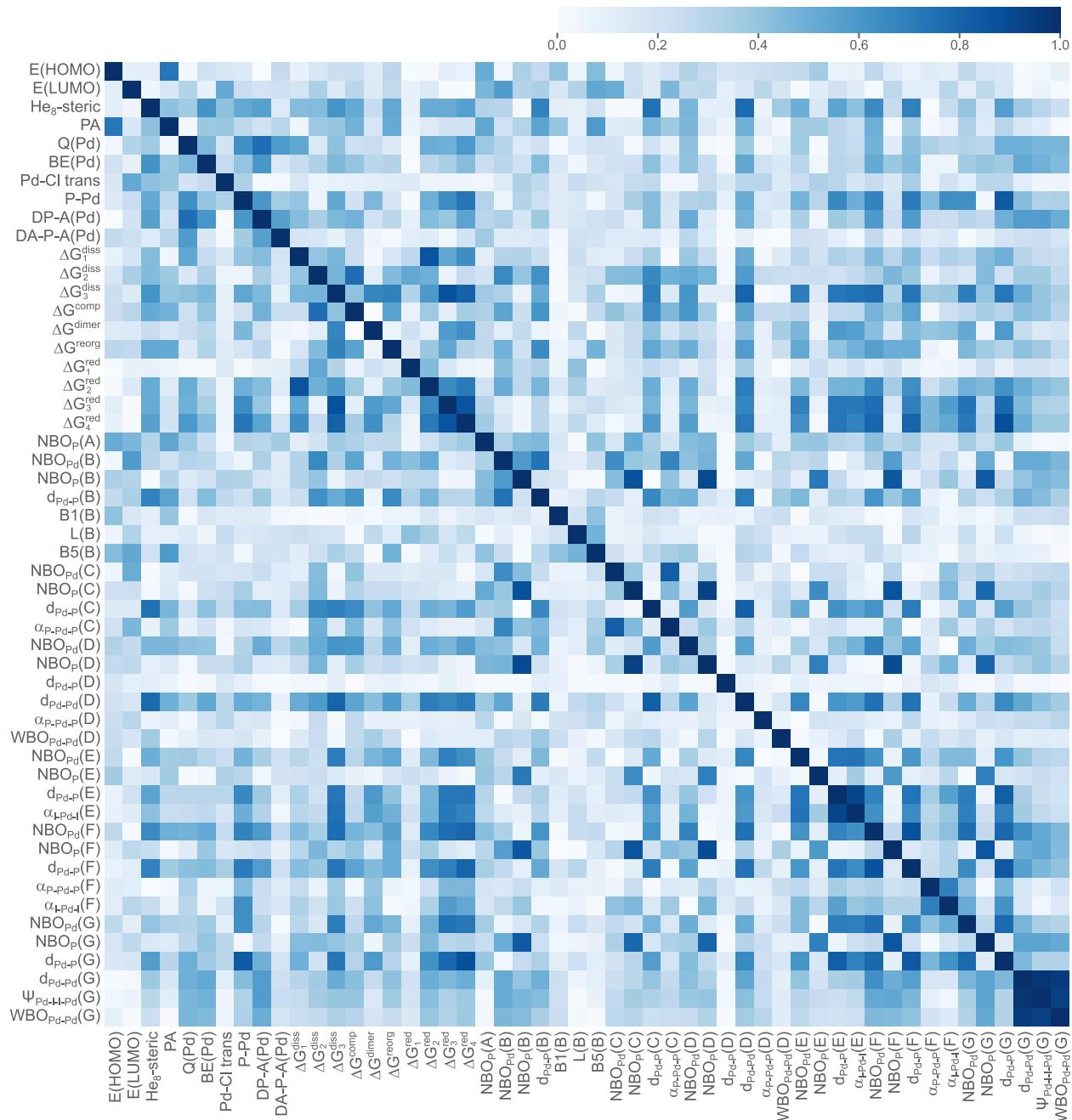
Descriptor	Unit	Derivation
E(HOMO)	Hartree	Energy of highest occupied molecular orbital (from LKB-P)
E(LUMO)	Hartree	Energy of lowest unoccupied molecular orbital (from LKB-P)
He8_steric	kcal mol <sup>-1</sup>	Interaction energy between singlet L in ground-state conformation and ring of 8 helium atoms; E <sub>ster</sub> = E <sub>tot</sub> (system) - [E <sub>tot</sub> (He8) + E <sub>tot</sub> (PR <sub>3</sub> )] (from LKB-P)
PA	kcal mol <sup>-1</sup>	Proton affinity (from LKB-P)
Q(Pd)	-	NBO charge of Pd in [PdCl <sub>3</sub> (PR <sub>3</sub> )] (from LKB-P)
BE(Pd)	kcal mol <sup>-1</sup>	Bond energy for dissociation of P-ligand from [PdCl <sub>3</sub> ]-fragment: BE = [E <sub>tot</sub> (PdCl <sub>3</sub> ) + E <sub>tot</sub> (PR <sub>3</sub> )] - E <sub>tot</sub> [[PdCl <sub>3</sub> (PR <sub>3</sub> ])] (from LKB-P)
Pd-Cl trans	Å	Pd-Cl distance in [PdCl <sub>3</sub> (PR <sub>3</sub> )] (trans to ligand) (from LKB-P)
P-Pd	Å	Pd-P distance in [PdCl <sub>3</sub> (PR <sub>3</sub> )] (from LKB-P)
DP-A(Pd)	Å	Change in average P-R bond length in [PdCl <sub>3</sub> (PR <sub>3</sub> )] compared to free ligand (from LKB-P)
DA-P-A(Pd)	deg	Change in average R-P-R angle in [PdCl <sub>3</sub> (PR <sub>3</sub> )] compared to free ligand (from LKB-P)
ΔG <sup>Diss</sup> (1)	kcal mol <sup>-1</sup>	Free energy for the dissociation of P-ligand from complex E
ΔG <sup>Diss</sup> (2)	kcal mol <sup>-1</sup>	Free energy for the dissociation of P-ligand from complex B
ΔG <sup>Diss</sup> (3)	kcal mol <sup>-1</sup>	Free energy for the dissociation of P-ligand from complex F
ΔG <sup>Comp</sup>	kcal mol <sup>-1</sup>	Free energy for the comproportionation of complex C with PdI <sub>2</sub> to form complex D
ΔG <sup>Dimer</sup>	kcal mol <sup>-1</sup>	Free energy for the dimerization of complex E to form complex G
ΔG <sup>Reorg</sup>	kcal mol <sup>-1</sup>	Free energy for the reorganization of complex F and PdI <sub>2</sub> to form complex G
ΔG <sup>Red</sup> (1)	kcal mol <sup>-1</sup>	Free energy for the reduction of complex D
ΔG <sup>Red</sup> (2)	kcal mol <sup>-1</sup>	Free energy for the reduction of complex E
ΔG <sup>Red</sup> (3)	kcal mol <sup>-1</sup>	Free energy for the reduction of complex F
ΔG <sup>Red</sup> (4)	kcal mol <sup>-1</sup>	Free energy for the reduction of complex G
NBO <sub>P</sub> (A)	a.u.	NBO charge of P in complex A
NBO <sub>Pd</sub> (B)	a.u.	NBO charge of Pd in complex B
NBO <sub>P</sub> (B)	a.u.	NBO charge of P in complex B
d <sub>Pd-P</sub> (B)	Å	Pd-P bond length in complex B
B1(B)	-	<i>Sterimol</i> descriptor B1: minimum ligand width in complex B
L(B)	-	<i>Sterimol</i> descriptor L: ligand width in complex B
B5(B)	-	<i>Sterimol</i> descriptor B5: maximum ligand width in complex B
NBO <sub>Pd</sub> (C)	a.u.	NBO charge of Pd in complex C
NBO <sub>P</sub> (C)	a.u.	NBO charge of P in complex C
d <sub>Pd-P</sub> (C)	Å	Averaged Pd-P bond length in complex C
α <sub>P-Pd-P</sub> (C)	deg	P-Pd-P bond angle in complex C
NBO <sub>Pd</sub> (D)	a.u.	Averaged NBO charge of Pd in complex D
NBO <sub>P</sub> (D)	a.u.	Averaged NBO charge of P in complex D
d <sub>Pd-P</sub> (D)	Å	Averaged Pd-P bond length in complex D
d <sub>Pd-Pd</sub> (D)	Å	Pd-Pd distance in complex D
α <sub>P-Pd-P</sub> (D)	Deg	Averaged P-Pd-P bond angle in complex D
WBO <sub>Pd-Pd</sub> (D)	-	Wiberg bond order between both Pd-centers in complex D
NBO <sub>Pd</sub> (E)	a.u.	NBO charge of Pd in complex E
NBO <sub>P</sub> (E)	a.u.	NBO charge of P in complex E
d <sub>Pd-P</sub> (E)	Å	Pd-P bond length in complex E
α <sub>I-Pd-I</sub> (E)	Deg	I-Pd-I bond angle in complex E
NBO <sub>Pd</sub> (F)	a.u.	NBO charge of Pd in complex F
NBO <sub>P</sub> (F)	a.u.	NBO charge of P in complex F
d <sub>Pd-P</sub> (F)	Å	Averaged Pd-P bond length in complex F

$\alpha_{P-Pd-P}(F)$	Deg	P-Pd-P bond angle in complex F
$\alpha_{I-Pd-I}(F)$	Deg	I-Pd-I bond angle in complex F
NBO <sub>Pd</sub> (G)	a.u.	Averaged NBO charge of Pd in complex G
NBO <sub>P</sub> (G)	a.u.	Averaged NBO charge of P in complex G
d <sub>Pd-P</sub> (G)	Å	Averaged Pd-P bond length in complex G
d <sub>Pd-Pd</sub> (G)	Å	Pd-Pd distance in complex G
$\psi_{Pd-I-I-Pd}(G)$	Deg	Pd-I-I-Pd dihedral angle in complex G
WBO <sub>Pd-Pd</sub> (G)	-	Wiberg bond order between both Pd centers in complex G

**Table S2.**

Full list of descriptors used for the problem-specific clustering. The first 9 descriptors are adapted from reference (12).

**3.2.1. Pearson correlation analysis of problem-specific descriptors.** The Pearson correlation coefficient (PCC) is a measure for the linear relationship between two variables and ranges from -1 to 1 (-1 or 1 = linear relationship; 0 = no linear correlation). In the heatmap shown in Fig. S12, the colors of the squares correspond the absolute PCC for each pair of descriptors. Typically, a small correlation between descriptors is desirable for the application of machine learning, since highly correlated descriptors could encode the same information and therefore alter its importance, leading to unintended biases of the model. In the present case, correlation among all descriptors is low, only for  $G-d_{Pd-Pd}$ ,  $G-\psi_{Pd-I-I-Pd}$  and  $G-WBO_{Pd-Pd}$  a correlation above a threshold of 0.95 was found. However, since their chemical meaning is still different, we decided to keep them for further analysis.



**Fig. S12.**

Absolute Pearson correlation coefficient (PCC) for all problem-specific descriptors.

**3.2.2. Principal component analysis of problem-specific descriptors.** In order to obtain an appropriate visualization of the new problem-specific ligand space (see Fig. 3A in the main text), we conducted principal component analysis (PCA, see section 1. for details). The first three principal components (PC1', PC2', PC3') capture over 60% of the variance in the data and were thus suited to represent our results (Table S3). The resulting values of the principal components (PCs) for each ligand are given in Table S4 while the loadings (*i.e.*, importance) for each descriptor are given in Table S5 for each PC. Additionally, the absolute loadings for each descriptor are visualized in Fig. S13, showing that all of them contribute to the PCs. Only for 2 descriptors,  $\Delta G^{\text{red}}_2$  and  $d_{\text{Pd-P}}(\text{D})$ , minor contributions (<0.05) are found.

PC	Variance [%]	Cumulative Variance [%]	Singular Value
PC1'	33.69	33.69	7.89
PC2'	17.78	51.47	5.73
PC3'	8.99	60.46	4.08
PC4'	5.99	66.45	3.33

**Table S3.**

Summary of the principal component analysis for the problem-specific descriptors.

No.	PC1'	PC2'	PC3'	PC4'
15	-0.99	0.41	0.02	-0.63
16	1.20	1.16	0.90	-0.06
19	-0.15	-0.32	0.88	-0.23
36	-1.14	0.28	0.18	-0.52
37	-0.90	0.45	0.11	-0.30
38	-1.04	0.32	0.13	-0.58
39	-1.28	0.44	0.07	-0.64
40	-0.36	0.18	0.82	-0.66
41	0.40	0.74	0.70	-0.61
42	0.66	1.03	0.32	0.00
43	-0.33	0.98	0.01	-0.51
44	-0.31	0.60	0.04	0.15
45	0.73	1.59	0.06	0.21
46	0.65	1.24	0.27	0.11
47	0.11	0.23	0.58	0.21
49	-1.07	0.42	0.06	-0.61
50	-0.94	0.64	-0.24	-0.37
52	-0.82	0.31	-0.30	0.11
54	-0.38	0.95	-0.27	-0.61
55	1.15	1.62	0.29	-0.10
60	0.22	-0.78	-0.47	-0.27
61	0.96	0.65	-0.19	-0.30
73	-0.52	-0.61	-0.13	0.02
76	-0.84	-0.25	-0.31	-0.19
78	-1.08	-0.29	-0.12	0.27
81	-1.11	-0.42	-0.75	-0.24
83	-1.04	-0.84	-0.36	0.45
85	-0.74	1.05	-0.78	0.84
86	-0.51	-0.88	-0.34	0.66
93	-1.13	-0.08	-0.31	0.49
96	-1.03	-0.56	-0.26	-0.07
101	-1.10	-0.47	0.25	0.32
102	-0.33	-0.33	0.67	0.29
105	-0.89	-0.48	0.14	0.30
111	-1.38	-0.22	0.07	-0.03
112	-0.97	-0.05	-0.06	0.06
113	-0.48	0.34	-0.09	0.04
116	-0.10	-0.24	-0.06	0.77
119	0.14	0.96	-0.36	0.54
120	-0.20	-0.07	-0.43	0.07
121	-0.29	0.08	-0.43	0.11
122	-0.14	0.02	-0.75	0.06
123	1.22	0.58	-0.43	0.49
142	0.73	1.31	0.07	0.71
143	-0.32	0.54	-0.02	1.00
154	-0.05	-0.72	1.23	0.19
155	-0.11	-0.69	1.13	0.05
156	0.49	-0.22	1.15	0.50
157	-0.12	-0.65	1.06	0.13
158	-0.02	-1.28	1.15	0.69
160	0.19	-1.12	-0.17	-0.23
161	-0.06	-0.89	-0.53	-0.17

<b>162</b>	1.86	-0.66	-0.43	-0.21
<b>163</b>	-0.41	-0.96	-0.60	-0.48
<b>166</b>	1.96	-0.60	-0.27	-0.10
<b>167</b>	0.10	-0.73	-0.01	-0.40
<b>168</b>	2.00	-0.44	-0.31	-0.11
<b>169</b>	0.13	-0.56	-0.18	-0.32
<b>170</b>	0.43	-0.88	0.14	0.53
<b>171</b>	0.75	-0.39	-0.25	-0.33
<b>172</b>	2.30	-0.27	-0.21	-0.14
<b>173</b>	0.31	-0.66	-0.03	-0.59
<b>174</b>	2.13	-0.30	-0.31	-0.29
<b>179</b>	2.46	-0.22	-0.12	-0.07
<b>184</b>	-0.63	0.77	-1.00	0.54
<b>187</b>	2.00	-0.79	-0.61	0.03

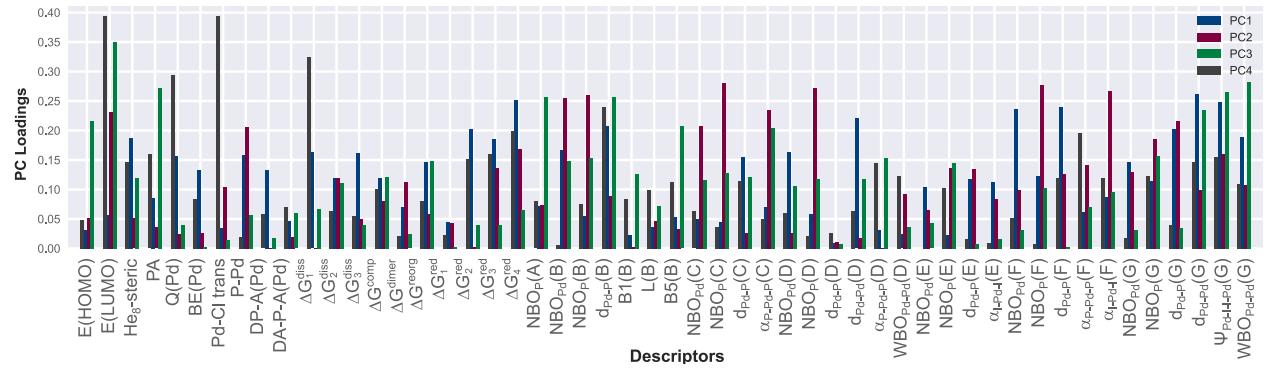
**Table S4.**

Resulting principal components for each ligand in the database. Ligand IDs are adapted from reference (12).

Descriptor	PC1'	PC2'	PC3'	PC4'
E(HOMO)	0.03	0.05	-0.22	-0.05
E(LUMO)	-0.06	0.23	0.35	-0.39
He8_steric	0.19	0.05	-0.12	-0.15
PA	0.09	0.04	-0.27	-0.16
Q(Pd)	0.16	0.02	0.04	0.29
BE(Pd)	-0.13	0.03	0.00	0.08
Pd-Cl trans	0.03	0.10	-0.01	-0.39
P-Pd	0.16	0.21	0.06	0.02
DP-A(Pd)	0.13	0.00	0.02	0.06
DA-P-A(Pd)	-0.05	-0.02	0.06	-0.07
$\Delta G^{\text{Diss}}(1)$	0.16	0.00	0.07	0.32
$\Delta G^{\text{Diss}}(2)$	0.12	-0.12	-0.11	-0.06
$\Delta G^{\text{Diss}}(3)$	0.16	0.05	-0.04	0.05
$\Delta G^{\text{Comp}}$	-0.12	0.08	0.12	0.10
$\Delta G^{\text{Dimer}}$	0.07	0.11	0.02	0.02
$\Delta G^{\text{Reorg}}$	-0.15	0.06	0.15	0.08
$\Delta G^{\text{Red}}(1)$	0.05	-0.04	0.00	-0.02
$\Delta G^{\text{Red}}(2)$	0.20	0.00	0.04	0.15
$\Delta G^{\text{Red}}(3)$	0.18	0.14	0.04	0.16
$\Delta G^{\text{Red}}(4)$	0.25	0.17	0.07	0.20
NBO <sub>P</sub> (A)	0.07	0.07	-0.26	0.08
NBO <sub>Pd</sub> (B)	0.17	-0.26	-0.15	0.01
NBO <sub>P</sub> (B)	-0.06	0.26	-0.15	-0.07
d <sub>Pd-P</sub> (B)	0.21	-0.09	-0.26	-0.24
B1(B)	-0.02	0.00	-0.13	0.08
L(B)	-0.04	-0.05	-0.07	0.10
B5(B)	0.05	-0.03	-0.21	0.11
NBO <sub>Pd</sub> (C)	0.05	-0.21	-0.12	0.06
NBO <sub>P</sub> (C)	-0.05	0.28	-0.13	-0.04
d <sub>Pd-P</sub> (C)	0.16	0.03	-0.12	-0.11
$\alpha_{P-Pd-P}$ (C)	-0.07	0.23	0.20	0.05
NBO <sub>Pd</sub> (D)	0.16	-0.03	-0.11	0.06
NBO <sub>P</sub> (D)	-0.06	0.27	-0.12	-0.02
d <sub>Pd-P</sub> (D)	-0.01	-0.01	0.01	0.03
d <sub>Pd-Pd</sub> (D)	0.22	0.02	-0.12	-0.06
$\alpha_{P-Pd-P}$ (D)	0.03	0.00	0.15	-0.14
WBO <sub>Pd-Pd</sub> (D)	0.02	-0.09	-0.04	-0.12
NBO <sub>Pd</sub> (E)	-0.10	-0.07	0.04	0.00
NBO <sub>P</sub> (E)	-0.02	0.14	-0.15	-0.10
d <sub>Pd-P</sub> (E)	0.12	0.13	0.01	-0.02
$\alpha_{I-Pd-I}$ (E)	-0.11	-0.08	-0.02	-0.01
NBO <sub>Pd</sub> (F)	0.24	0.10	-0.03	-0.05
NBO <sub>P</sub> (F)	-0.12	0.28	-0.10	0.01
d <sub>Pd-P</sub> (F)	0.24	0.13	0.00	-0.12
$\alpha_{P-Pd-P}$ (F)	-0.06	-0.14	-0.07	-0.20
$\alpha_{I-Pd-I}$ (F)	-0.09	-0.27	-0.10	-0.12
NBO <sub>Pd</sub> (G)	0.15	0.13	0.03	0.02
NBO <sub>P</sub> (G)	-0.11	0.19	-0.16	-0.12
d <sub>Pd-P</sub> (G)	0.20	0.22	0.03	0.04
d <sub>Pd-Pd</sub> (G)	0.26	-0.10	0.23	-0.15
$\psi_{Pd-I-I-Pd}$ (G)	0.25	-0.16	0.26	-0.15
WBO <sub>Pd-Pd</sub> (G)	-0.19	0.11	-0.28	0.11

**Table S5.**

Loadings on each principal component for the problem-specific descriptors.



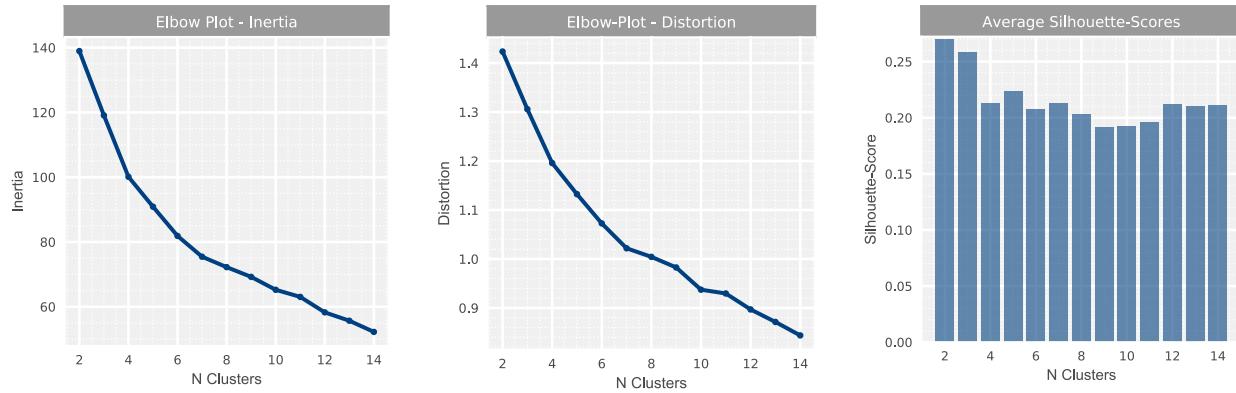
**Fig. S13.**

Absolute loadings on each of the new PCs (PC1' to PC4') for all problem-specific descriptors.

### 3.3. $k$ -Means Optimization

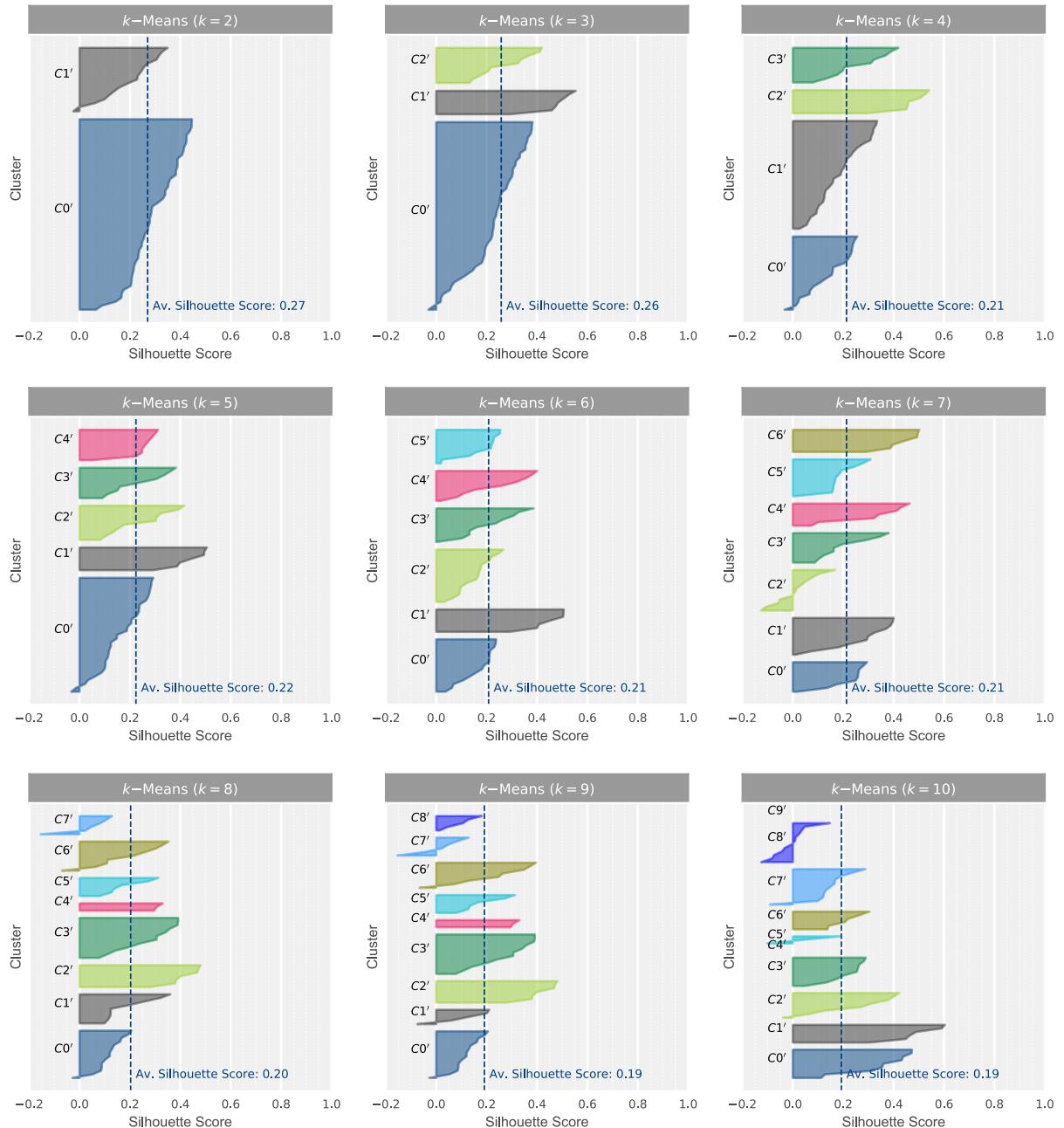
In analogy to the initial clustering (see Section 2.2), we analyzed elbow plots based on inertia (within-cluster sum-of-squares) and distortion (sum of square errors), as well as the silhouette coefficients to find the optimum number of clusters  $k$  (the analysis was done for  $k$  ranging from 2 to 14).

The point of inflection in both, the inertia- and distortion-based elbow plots occurs around  $k = 6$  (Fig. S14, left and middle). The average silhouette scores (Fig. S14, right) again indicate similar performance for all values of  $k$ . However, when analyzing the per-sample silhouette scores for each  $k$  (Fig. S15) we found that  $k = 6$  not only exhibited an even cluster size distribution but, more importantly, was the only value of  $k$ , for which no misclassification (see section 1. for explanation) occurred within any cluster. Thus, it was used to conduct all further analyses.



**Fig. S14.**

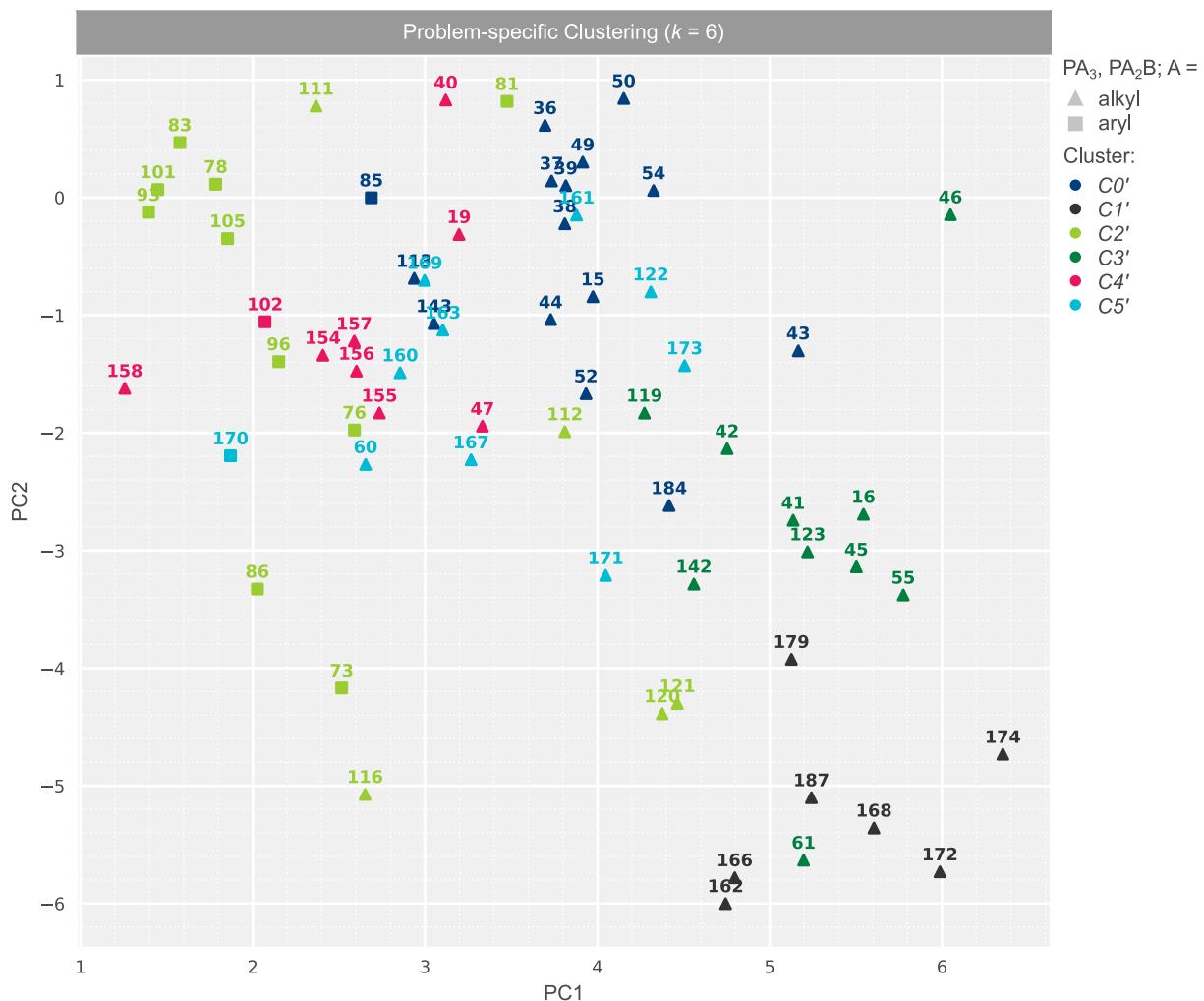
Optimization of the number of clusters  $k$  for the  $k$ -means clustering of problem-specific descriptors using the *elbow-method* and *silhouette score* analysis.



**Fig. S15.**

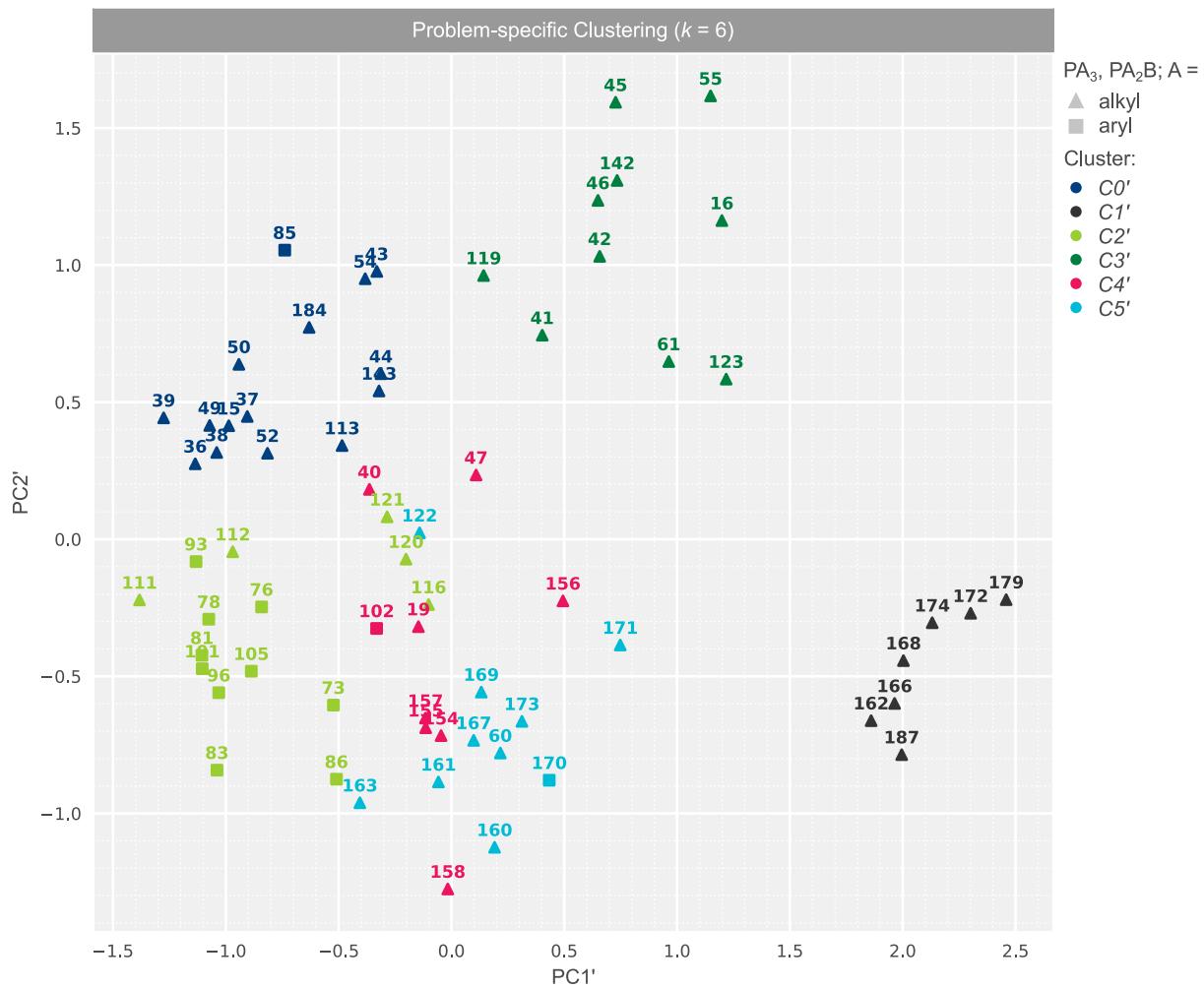
Silhouette scores per sample for different cluster numbers  $k$  in the  $k$ -means clustering of the problem-specific descriptors.

### 3.4. Results of the Problem-specific Clustering



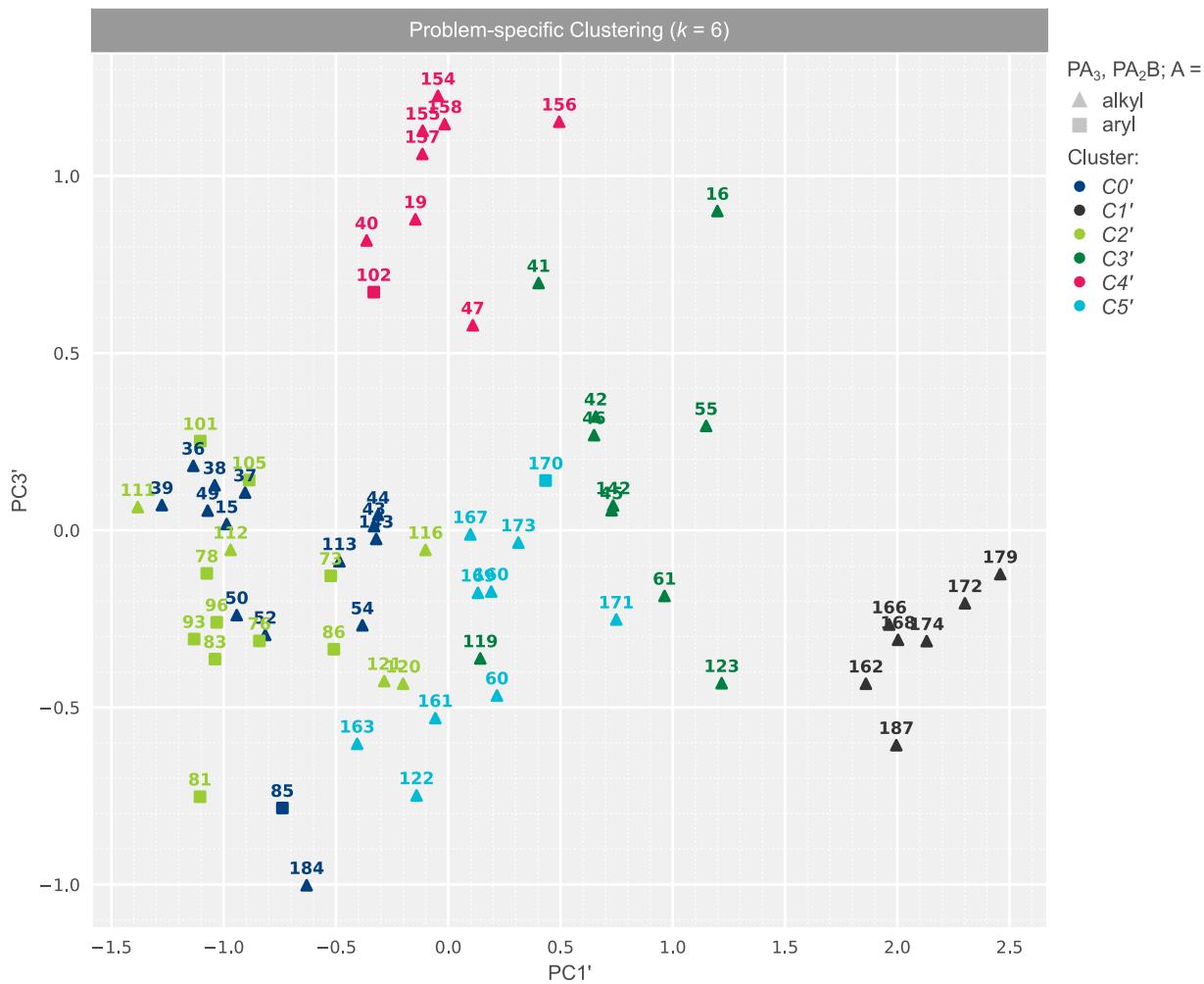
**Fig. S16.**

Results of the  $k$ -means clustering ( $k = 6$ ) of the problem-specific descriptors visualized within the space captured by PC1 and PC2 of the LKB-P. Principal components and ligand IDs are adapted from reference (12).



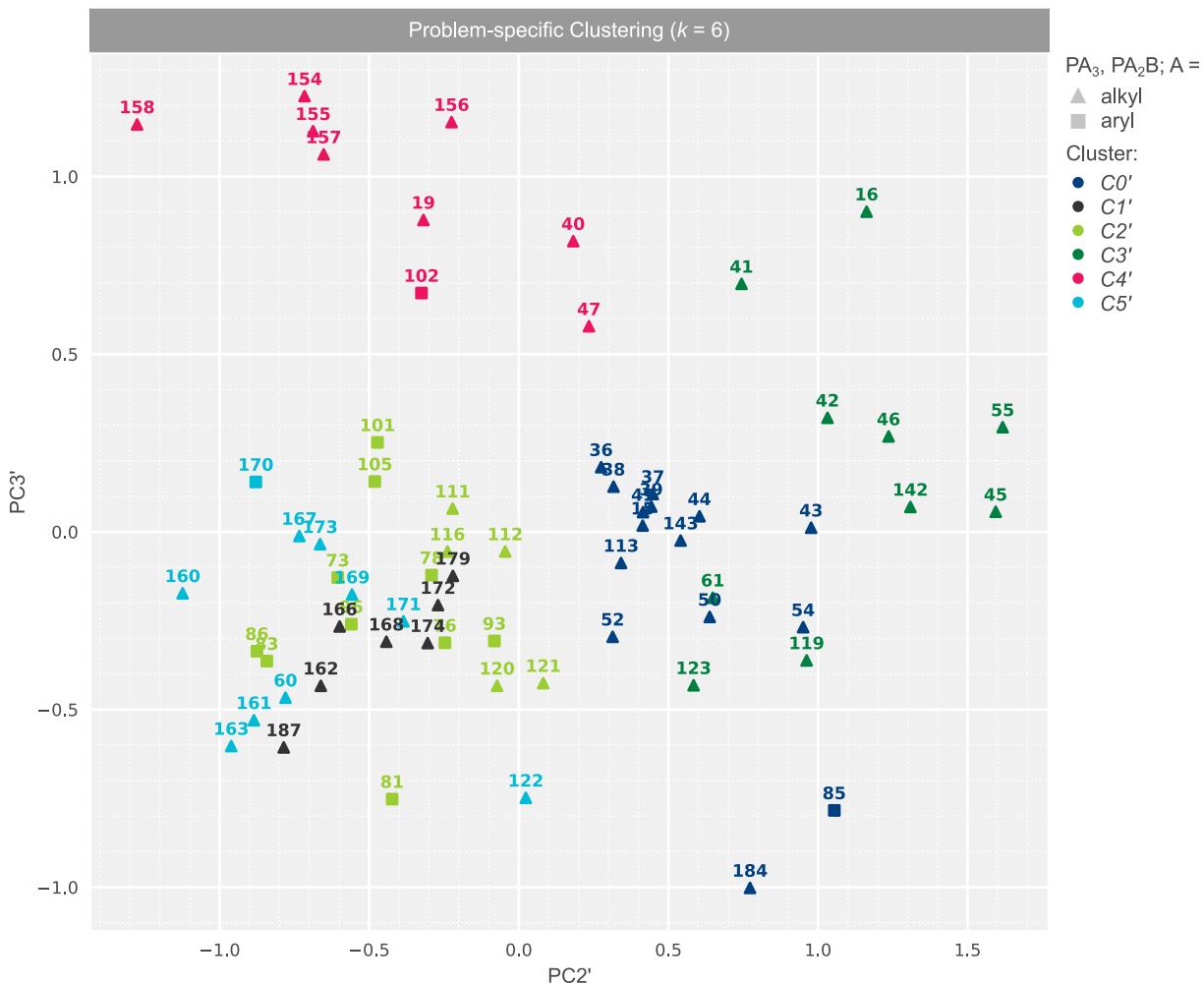
**Fig. S17.**

Results of the  $k$ -means clustering ( $k = 6$ ) of the problem-specific descriptors visualized within the space captured by the problem-specific principal components PC1' and PC2'. Ligand IDs are adapted from reference (12).



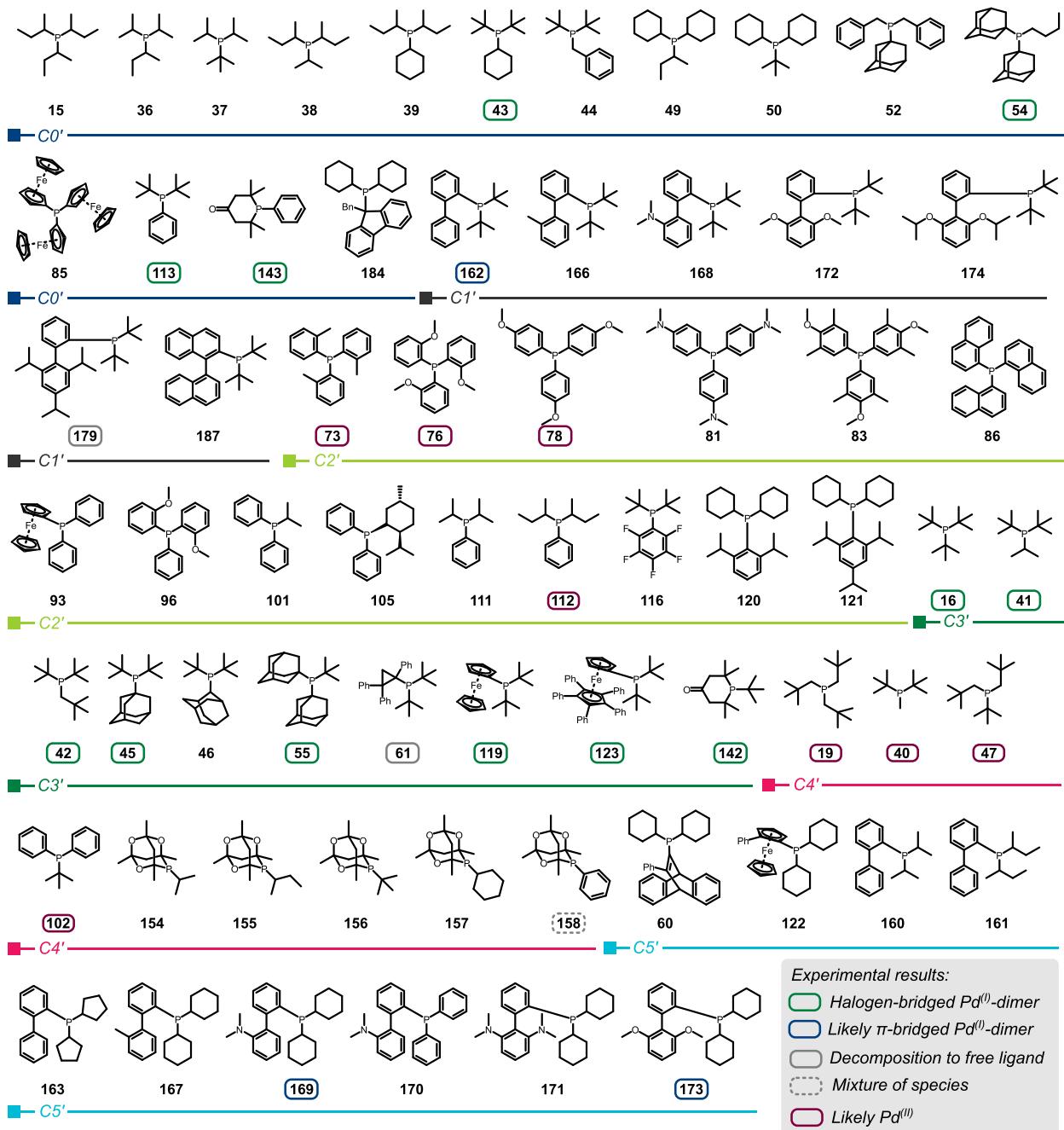
**Fig. S18.**

Results of the  $k$ -means clustering ( $k = 6$ ) of the problem-specific descriptors visualized within the space captured by the problem-specific principal components PC1' and PC3'. Ligand IDs are adapted from reference (12).



**Fig. S19.**

Results of the  $k$ -means clustering ( $k = 6$ ) of the problem-specific descriptors visualized within the space captured by the problem-specific principal components PC2' and PC3'. Ligand IDs are adapted from reference (12).

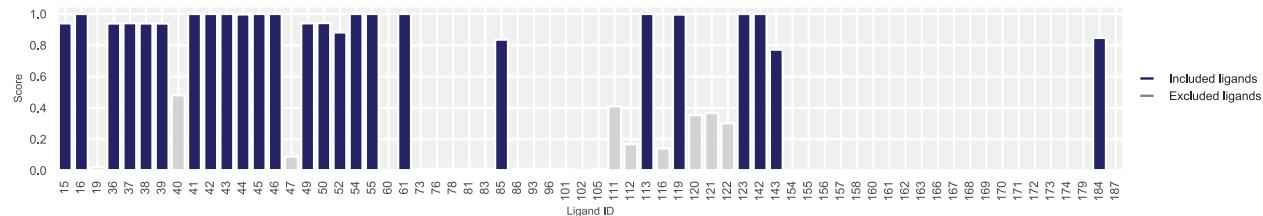


**Fig. S20.**

Structures of the ligands used for the problem-specific clustering, grouped by their resulting clusters. Experimentally tested ligands (for procedure see 4.5.1.) are highlighted accordingly. Ligand IDs are adapted from reference (12).

### 3.5. Statistical Evaluation

In analogy to section 2.4, we set out to explore the statistical relevance of our predictions by performing the clustering 1000 times with different random initializations and analyzed how many times each ligand was clustered with our positive references (**16, 41, 54, 113**). For the problem-specific clustering (Fig. S21) all ligands of the initial prediction set exhibit scores  $S > 0.8$  (i.e. 80% of the 1000 models) and are thus statistically meaningful predictions. Eight additional ligands bearing aryl-groups (**40, 111, 112, 116, 120, 121, 122**) were clustered with the dimer-inducing reference ligands only with low scores ( $S \sim 0.05$  to 0.4, *i.e.* 5% to 40% of the 1000 models)



**Fig. S21.**

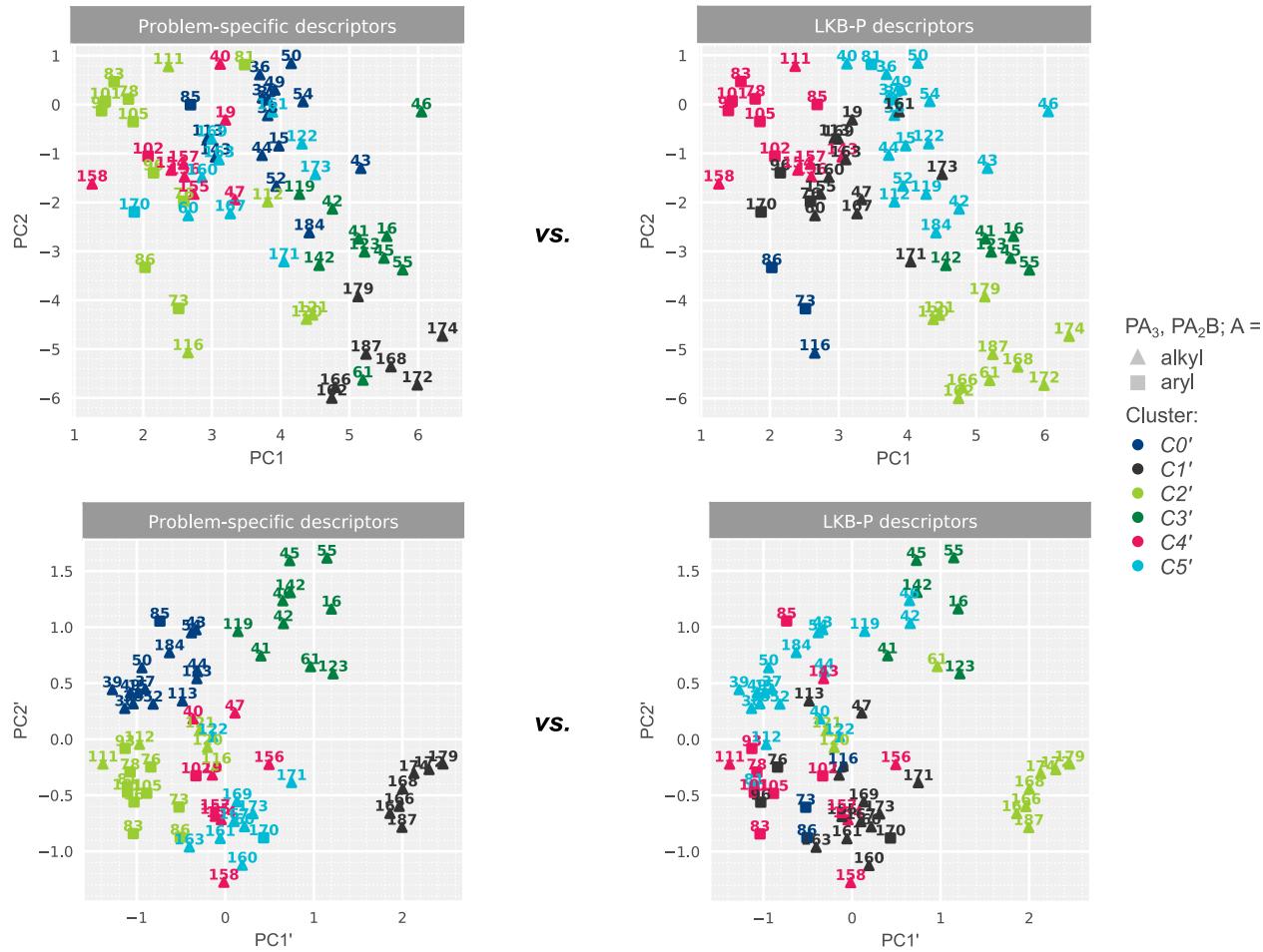
Statistical evaluation for the problem-specific clustering. The clustering was performed 1000 times with different random initializations to evaluate how often each ligand was clustered with the dimer-inducing reference ligands (**16, 41, 54, 113**), which is reflected by the score  $S$ . Colored in blue are the ligands that were predicted by the reported model, *i.e.* the final ligand candidates. Those ligands colored in grey were not predicted by the reported model, in agreement with their low score. Ligand IDs are adapted from reference (12).

### 3.6. General vs. Problem-Specific Descriptors

The need for the addition of problem-specific descriptors for further refinement after the initial clustering was investigated by performing the 2<sup>nd</sup> clustering using only the general LKB-P descriptors (instead of the additional problem-specific descriptors). Fig. S22 shows a comparison between the results for the clustering obtained from the problem-specific descriptors (left) and general LKB-P descriptors (right). Both approaches yield different results. Notably, in the clustering with general descriptors (right), the references are spread across 3 different clusters (black, green and turquoise) instead of 2. This results in a total of 39 ligands (>59% of the used data set) that would be considered as final candidates for experimental validation. The number of ligands obtained with the problem-specific descriptors is 25 ligands (in the green and blue cluster).

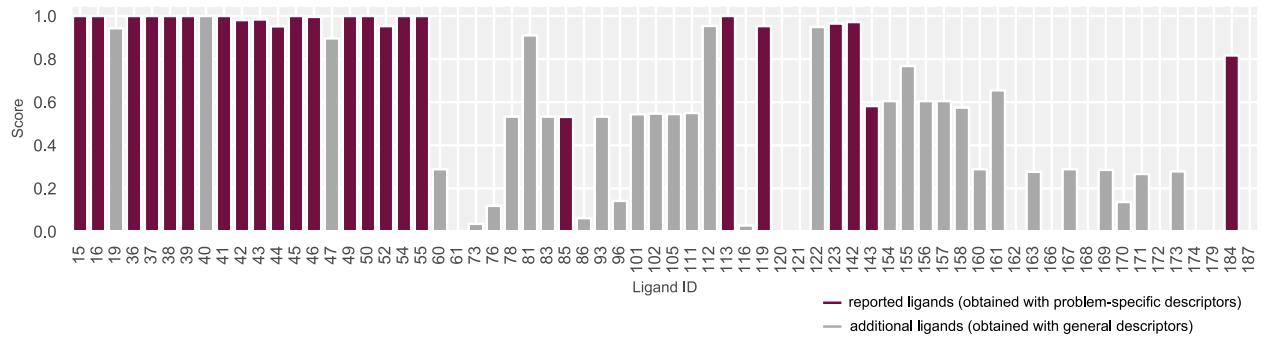
As discussed in section 2.4 and 3.5, the  $k$ -means algorithm is affected by random fluctuations in the initialization. Hence, we performed the clustering 1000 times with different random initialization seeds and subsequently analyzed how many times each ligand was clustered with our dimer-inducing reference ligands (**16, 41, 54, 113**). In Fig. S23, ligands that were predicted by the reported model (i.e. the model that used problem-specific descriptors) are shown in red, while the ones which were not included in these predictions are colored in dark grey.

For the clustering with general descriptors, 19 additional ligands were clustered with our dimer-inducing reference ligands in >50% of the cases. Four of these ligands (**19, 40, 47, 112**) that had particularly high scores ( $S > 0.95$ , *i.e.* >95% of the 1000 models) as well as five additional ligands (**78, 102, 158, 169, 173**) with slightly lower scores ( $0.2 < S < 0.6$ ) were experimentally verified to be false predictions (see Fig. S20, for procedure see 4.5.1.). This is in stark contrast to the results obtained with the problem-specific clustering, which resulted in a robust model and, based on the experimental tests we did, much higher quality of prediction.



**Fig. S22.**

Comparison of cluster results using problem-specific descriptors (left) and general LKB-P descriptors (right). Cluster numbers and colors are assigned arbitrary.



**Fig. S23.**

Statistical evaluation for the 2<sup>nd</sup> clustering using only the general LKB-P descriptors. The clustering was performed 1000 times with different random initializations to evaluate how often each ligand was clustered with the dimer-inducing reference ligands (**16, 41, 54, 113**) and the results are summarized in the score  $S$ . Colored in red are the ligands that were predicted by the reported problem-specific model, *i.e.* the final ligand candidates. Colored in dark grey are those ligands that were additionally predicted when using only the general LKB-P descriptors. Ligand IDs are adapted from reference (12).

## 4. Experimental Validation

### 4.1. Experimental Details

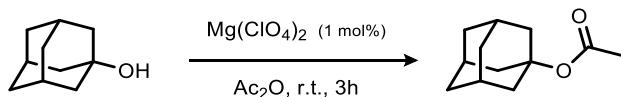
**4.1.1. Reagents, starting material and solvents.** All commercially available reagents and starting materials were used as received. Anhydrous toluene, THF, hexane, DCM and Et<sub>2</sub>O were dried using an Innovative Technology PS-MD-5 solvent purification system. All other anhydrous solvents were purchased (acetone and DMF from Acros) and degassed prior to use. Solvents used in work up and purification were distilled prior to use.

All synthesized Pd(0) complexes were stored at -30°C inside an Argon-filled glovebox and used as soon as practical.

**4.1.2. Characterization.** All <sup>1</sup>H, <sup>13</sup>C and <sup>31</sup>P NMR spectra were recorded on Varian VNMRS 600, Varian VNMRS 400, or Varian Mercury 300 spectrometers at ambient temperature. Chemical shifts ( $\delta$ ) are reported in parts per million (ppm) and were referenced either to residual solvent peak (CDCl<sub>3</sub>, CD<sub>2</sub>Cl<sub>2</sub>, C<sub>6</sub>D<sub>6</sub>, toluene-*d*<sub>8</sub> or THF-*d*<sub>8</sub>) or internal standards trimethyl phosphate O=P(OMe)<sub>3</sub> ( $\delta$  = 3.05 ppm) or O=PCl<sub>3</sub> ( $\delta$  = 2.79 ppm; within a closed glass capillary) for <sup>31</sup>P. Coupling patterns are reported as br (broad), s (singlet), d (doublet), t (triplet), q (quartet) or m (multiplet). Coupling constants ( $J$ ) are given in Hertz (Hz).

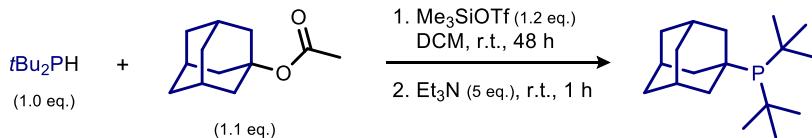
Gas chromatography coupled with mass spectrometry (GC-MS) was performed on an Agilent Technologies 5975 series MSD mass spectrometer under electrospray ionization (EI) mode coupled with an Agilent Technologies 7820A gas chromatograph employing an Agilent 19091s-433 HP-5MS column (30 m x 0.250  $\mu$ m x 0.250  $\mu$ m). GC-MS conditions: front inlet mode: split; temperature: 250°C; pressure: 718.4 mbar; total flow: 26.246 mL/min; split ratio: 20:1; split flow: 24 mL/min; run time: 25.5 min; oven programme: 60°C for 0.5 min then 10°C/min to 280°C then 280°C for 3 min; flow: 1.2 mL/min. Low-resolution masses (MS) of known compounds were extracted from their GC-MS chromatograms.

## 4.2. Synthesis and Characterization of Phosphine Ligands



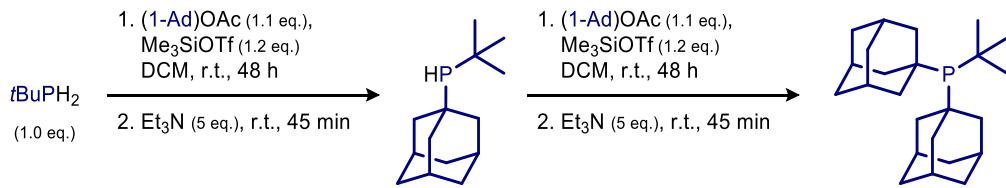
**4.2.1. 1-Adamantyl acetate (1-Ad)OAc.**(67) The title compound was prepared using a slightly modified procedure of Chakraborti *et al.*(67) To a mixture of 1-adamantol (11.03 g, 72.4 mmol, 1.0 eq.) and  $\text{Mg}(\text{ClO}_4)_2$  (162 mg, 0.72 mmol, 1 mol%) was added acetic anhydride (6.8 mL, 72.4 mmol, 1.0 eq.). The reaction mixture was stirred at ambient temperature using a big olive-shaped stir bar. The initially heterogeneous mixture liquified within 1 h and was quenched by the addition of water after 2 h. The mixture was extracted using DCM, dried over  $\text{MgSO}_4$  and the solvent was removed under reduced pressure. To remove excess acetic acid, the obtained crude was redissolved in  $\text{CHCl}_3$  and all volatiles removed *in vacuo*. This process was repeated 3 times (until the pungent smell of acetic acid was gone). The obtained oily solid was dried under high vacuum to yield the product (13.41 g, 69.0 mmol, 95%).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  2.06 (s, 3H), 2.01 (2 overlapping s, 6H), 1.86 (s, 3H), 1.57 (s, 6H).  $^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  170.0, 80.0, 41.2, 36.1, 30.7, 22.5. MS (70eV, EI): *m/z* (%): 194 (2) [ $\text{M}^+$ ], 136 (3), 135 (27), 134 (100), 119 (12), 105 (8), 95 (38), 94 (5), 93 (28), 92 (72), 91 (18), 79 (22), 77 (11), 67 (9), 56 (7), 55 (6). This is consistent with previously reported spectroscopic data.(68)



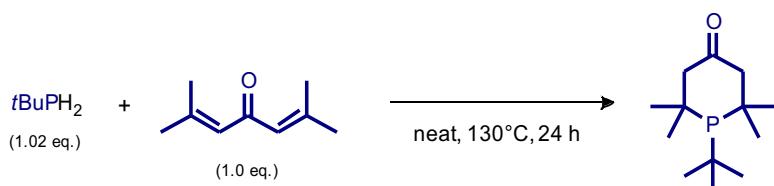
**4.2.2. 1-Adamantyl-di-*tert*-butylphosphine.** The title compound was prepared using a procedure analogous to that reported for tri-1-adamantylphosphine.(69) Di-*tert*-butylphosphine ( $t\text{Bu}_2\text{PH}$ , 4.00 g, 27.36 mmol, 1.0 eq.) and 1-adamantyl acetate ((1-Ad)OAc, 5.85 g, 30.09 mmol, 1.1 eq.) were dissolved in anhydrous DCM (135 mL) inside the glovebox. The flask was closed with a rubber septum and taken out of the glovebox.  $\text{Me}_3\text{SiOTf}$  (5.9 mL, 32.83 mmol, 1.2 eq.) was added with a syringe and the reaction mixture was stirred at ambient temperature for 48 h. Then, triethylamine (19.1 mL, 136.8 mmol, 5 eq.) was added and the obtained mixture was stirred for an additional 1 h. All volatiles were then removed under vacuum and the crude was redissolved in pentane inside the glovebox. The mixture was cooled at -30°C for 12 h, filtered through a sintered funnel and the solvent removed to yield the product as a white solid (7.07 g, 25.2 mmol, 92%).

$^1\text{H}$  NMR (600 MHz, benzene- $d_6$ )  $\delta$  2.15 – 2.08 (m, 6H), 1.86 (s, 3H), 1.66 (q (br),  $J = 12.4$  Hz, 6H), 1.31 (d,  $J = 10.0$  Hz, 18H).  $^{31}\text{P}$  NMR (243 MHz, benzene- $d_6$ )  $\delta$  62.2. This is consistent with previously reported spectroscopic data.(70)



**4.2.3. Di-1-adamantyl-*tert*-butylphosphine.** The title compound was prepared in two steps using a procedure analogous to that reported for tri-1-adamantylphosphine.(69) *tert*-Butylphosphine (*t*BuPH<sub>2</sub>, 1.00 g, 11.1 mmol, 1.0 eq.) and 1-adamantyl acetate ((1-Ad)OAc, 2.37 g, 12.2 mmol, 1.1 eq.) were dissolved in anhydrous DCM (50 mL) inside the glovebox. The flask was closed with a rubber septum and taken out of the glovebox. Me<sub>3</sub>SiOTf (2.4 mL, 13.3 mmol, 1.2 eq.) was added with a syringe and the reaction mixture was stirred at ambient temperature for 48 h. Then, triethylamine (7.7 mL, 111 mmol, 10 eq.) was added and the obtained mixture was stirred for additional 45 minutes. All volatiles were then removed under vacuum and the crude was suspended in hexane inside the glovebox. The liquid was decanted and the solvent evaporated. The crude was redissolved in DCM (50 mL) and (1-Ad)OAc (2.37 g, 12.2 mmol, 1.1 eq.) was added. The flask was again taken out of the glovebox and Me<sub>3</sub>SiOTf (2.4 mL, 13.3 mmol, 1.2 eq.) was added via syringe through a rubber septum. After 48 hours of stirring at ambient temperature, triethylamine (7.7 mL, 111 mmol, 10 eq.) was added and the obtained mixture stirred for additional 45 minutes. Inside the glovebox all volatiles were removed in *vacuo*, the crude was redissolved in hexane (250 mL, until all white powder had dissolved) and filtered. The filtrate was evaporated to a minimum amount of hexane and the formed precipitate was collected by filtration (washing with small amounts of hexane) through a sintered funnel to yield the product as a white solid (1.77 g, 4.94 mmol, 45% over 2 steps).

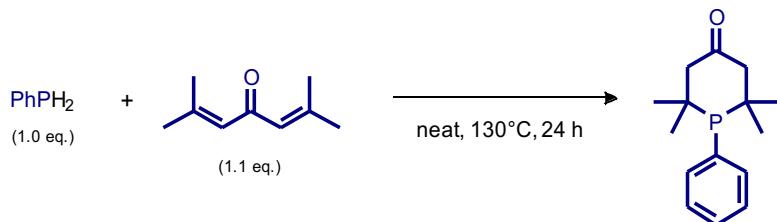
<sup>1</sup>H NMR (600 MHz, benzene-*d*<sub>6</sub>) δ 2.25 (d, *J* = 3.1 Hz, 12H), 1.93 – 1.86 (m, 6H), 1.72 (d (br), *J* = 11.6 Hz, 6H), 1.66 (d (br), *J* = 12.0 Hz, 6H), 1.40 (d, *J* = 9.4 Hz, 9H). <sup>31</sup>P NMR (243 MHz, benzene-*d*<sub>6</sub>) δ 61.6. This is consistent with previously reported spectroscopic data.(70)



**4.2.4. 1-(*tert*-Butyl)-2,2,6,6-tetramethylphosphinan-4-one.** *tert*-Butylphosphine (1.0 g, 11.1 mmol, 1.02 eq.) and 2,6-dimethylhepta-2,5-dien-4-one (1.5 g, 10.9 mmol, 1.0 eq.) were weighed into a pressure tube. The tube was sealed with a screw cap and Teflon tape and was heated at 130°C using an aluminum heating block and aluminum foil for isolation. After 24 hours the reaction mixture was allowed to cool to ambient temperature. A small amount of methanol (1 mL) was added and the resulting yellow solution was cooled at -30°C. Colorless needles formed within 24 h. The yellow solution was quickly separated before the crystals melted at ambient temperature. The separated molten product was dried *in vacuo* to give an analytically pure sample for spectroscopic analyses. The remaining methanol solution was directly employed in synthesis without further purification.

<sup>1</sup>H NMR (600 MHz, Benzene-*d*<sub>6</sub>) δ 2.48 (dd, *J* = 12.7, 4.3 Hz, 2H), 1.99 (dd, *J* = 22.3, 12.8 Hz, 2H), 1.19 (d, *J* = 2.7 Hz, 6H), 1.12 (two overlapping doublets; *J* = 16.1, 10.8 Hz, 15H, overlap). <sup>13</sup>C NMR (151 MHz, Benzene-*d*<sub>6</sub>) δ 207.4, 58.9 (d, *J* = 11.2 Hz), 37.3 (d, *J* = 29.9 Hz), 34.7 (d, *J*

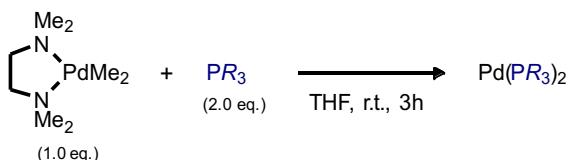
$= 26.4$  Hz), 33.9 (d,  $J = 33.4$  Hz), 32.3 (d,  $J = 14.1$  Hz), 26.1 (d,  $J = 4.6$  Hz).  $^{31}\text{P}$  NMR (243 MHz, Benzene- $d_6$ )  $\delta$  48.6.



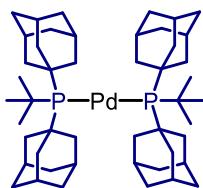
**4.2.5. 2,2,6,6-Tetramethyl-1-phenylphosphinan-4-one.** Phenylphosphine (1.09 g, 9.9 mmol, 1.0 eq.) and 2,6-dimethylhepta-2,5-dien-4-one (1.5 g, 10.9 mmol, 1.1 eq.) were weighed into a pressure tube. The tube was sealed with a screw cap and Teflon tape and was heated at 130°C using an aluminum heating block and aluminum foil for isolation. After 24 hours the reaction mixture was allowed to cool to ambient temperature. A small amount of methanol (5 mL) was added and the resulting yellow solution was cooled at -30°C. Colorless solid formed within 24 h. The yellow solution was separated and the remaining crude was dried *in vacuo* to give the product as a white powder (2.15 g, 8.65 mmol, 87%).

$^1\text{H}$  NMR (600 MHz, Benzene- $d_6$ )  $\delta$  7.57 (td,  $J = 7.9, 1.4$  Hz, 2H), 7.15 – 7.07 (m, 3H), 2.54 (d,  $J = 13.0$  Hz, 2H), 2.17 (dd,  $J = 12.8, 5.1$  Hz, 2H), 1.24 (d,  $J = 17.6$  Hz, 6H), 0.73 (d,  $J = 11.1$  Hz, 6H).  $^{13}\text{C}$  NMR (151 MHz, Benzene- $d_6$ )  $\delta$  209.1, 136.9 (d,  $J = 26.4$  Hz), 136.2 (d,  $J = 23.7$  Hz), 129.7 (d,  $J = 1.3$  Hz), 128.5 (d,  $J = 8.5$  Hz), 53.0 (d,  $J = 3.0$  Hz), 35.1 (d,  $J = 19.0$  Hz), 31.1 (d,  $J = 31.7$  Hz), 30.2 (d,  $J = 9.6$  Hz).  $^{31}\text{P}$  NMR (243 MHz, Benzene- $d_6$ )  $\delta$  16.7. This is consistent with previously reported spectroscopic data.(71)

### 4.3. Synthesis and Characterization of Bis(phosphine)palladium(0) Complexes

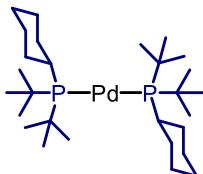


**4.3.1. General Procedure.** Inside a glovebox  $\text{Pd}(\text{tmeda})(\text{Me})_2$  (1.0 equiv.) (18) and free phosphine (2.0 equiv.) were dissolved in THF (0.015m) and stirred at ambient temperature. The conversion was monitored by  $^{31}\text{P}$  NMR. Upon full conversion (mostly 3h) of the free phosphine the solvent was evaporated *in vacuo*. The remainder was then triturated with DMF, filtered over a sintered funnel and washed with a small amount of cold hexane to remove traces of DMF. The obtained white solid was dried *in vacuo* to give the product.



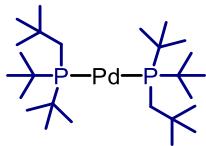
**4.3.2. Bis(di-1-adamantyl-tert-butylphosphine)palladium(0).** The title compound was synthesized from  $[\text{Pd}(\text{tmeda})(\text{Me})_2]$  (98 mg, 0.39 mmol, 1.0 eq.) and phosphine (279 mg, 0.78 mmol, 2.0 eq.) in toluene (2.5 mL). After a reaction time of 2 h, the reaction mixture was layered with *n*-hexane and stored at -35°C for 24 h. The formed precipitate was collected by filtration (washing with *n*-hexane) to yield the product as a slightly off-white powder (313 mg, 0.38 mmol, 98%).

*Note:* The product showed very low solubility that prevented its further characterization.  $^{31}\text{P}$  NMR (243 MHz, benzene- $d_6$ )  $\delta$  82.2.



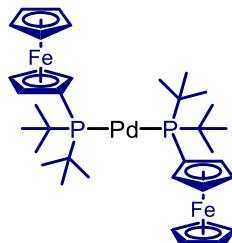
**4.3.3. Bis(di-tert-butyl(cyclohexyl)phosphine)palladium(0).** The title compound was synthesized according to the general procedure from  $[\text{Pd}(\text{tmeda})(\text{Me})_2]$  (553 mg, 2.19 mmol, 1.0 eq.) and phosphine (1.0 g, 4.38 mmol, 2.0 eq.) in THF (7 mL). After a reaction time of 2 h, the product was obtained as a white powder (452 mg, 0.803 mmol, 37%).

$^1\text{H}$  NMR (400 MHz, Benzene- $d_6$ )  $\delta$  2.70 (d,  $J = 12.5$  Hz, 4H), 1.76 (d,  $J = 12.2$  Hz, 6H), 1.69 – 1.42 (m, 40H), 1.34 – 1.02 (m, 8H).  $^{13}\text{C}$  NMR (101 MHz, Benzene- $d_6$ )  $\delta$  39.7, 35.9, 35.2 (t,  $J = 4.0$  Hz), 32.1 (t,  $J = 5.6$  Hz), 29.2 (t,  $J = 5.2$  Hz), 27.2.  $^{31}\text{P}$  NMR (121 MHz, Benzene- $d_6$ )  $\delta$  72.06.



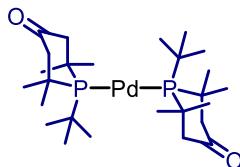
**4.3.4. Bis(di-*tert*-butyl(neopentyl)phosphine)palladium(0).** The title compound was synthesized according to the general procedure from  $[\text{Pd}(\text{tmEDA})(\text{Me})_2]$  (800 mg, 3.17 mmol, 1.0 eq.) and phosphine (1.37 g, 6.33 mmol, 2.0 eq.) in THF (20 mL). After a reaction time of 24 h, the product was obtained as a white powder (1.12 g, 2.08 mmol, 66%).

$^1\text{H}$  NMR (600 MHz, Benzene- $d_6$ )  $\delta$  1.46 (br, 22H), 1.36 (t,  $J = 5.7$  Hz, 36H).  $^{13}\text{C}$  NMR (151 MHz, Benzene- $d_6$ )  $\delta$  35.9 (t,  $J = 1.9$  Hz), 34.4 (t,  $J = 4.3$  Hz), 32.9 (t,  $J = 3.7$  Hz), 31.5 (t,  $J = 4.6$  Hz), 31.2 (t,  $J = 5.6$  Hz).  $^{31}\text{P}$  NMR (243 MHz, Benzene- $d_6$ )  $\delta$  44.1.



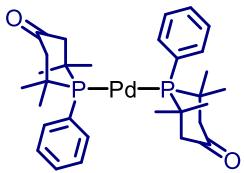
**4.3.5. Bis(di-*tert*-butyl(ferrocenyl)phosphine)palladium(0).** The title compound was synthesized according to the general procedure from  $[\text{Pd}(\text{tmEDA})(\text{Me})_2]$  (306 mg, 1.21 mmol, 1.0 eq.) and phosphine (800 mg, 2.42 mmol, 2.0 eq.) in THF (8 mL). After a reaction time of 2.5 h, the product was obtained as a light orange powder (556 mg, 0.725 mmol, 60%).

$^1\text{H}$  NMR (600 MHz, THF- $d_8$ )  $\delta$  4.54 (br, 4H), 4.36 (br, 4H), 4.31 (br, 10H), 1.44 (t,  $J = 5.7$  Hz, 36H).  $^{13}\text{C}$  NMR (151 MHz, THF- $d_8$ )  $\delta$  80.1 (t,  $J = 6.7$  Hz), 74.6 (t,  $J = 6.1$  Hz), 71.0, 69.8 (t,  $J = 2.5$  Hz), 36.3 (t,  $J = 4.6$  Hz), 32.2 (t,  $J = 5.6$  Hz).  $^{31}\text{P}$  NMR (243 MHz, THF- $d_8$ )  $\delta$  55.8. This is consistent with previously reported spectroscopic data.(72)



**4.3.6. Bis(2,2,6,6-tetramethyl-1-(tert-butyl)-4-phosphorinanone)palladium(0).** The title compound was synthesized according to the general procedure from  $[\text{Pd}(\text{tmEDA})(\text{Me})_2]$  (252.7 mg, 1.0 mmol, 1.0 eq.) and phosphine (2.0 mmol, 2.0 eq., 1M solution in MeOH) in THF (5 mL). After 48 h reaction time, the product was obtained as an off-white powder (163.8 mg, 0.291 mmol, 29%).

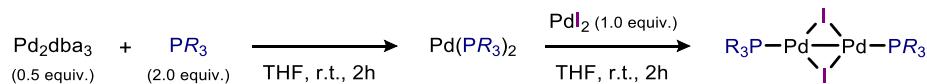
$^1\text{H}$  NMR (600 MHz, Benzene- $d_6$ )  $\delta$  3.58 (d,  $J = 11.9$  Hz, 4H), 2.10 – 1.97 (m, 4H), 1.59 – 1.50 (m, 12H), 1.34 – 1.29 (m, 18H), 1.25 (t,  $J = 2.6$  Hz, 12H).  $^{13}\text{C}$  NMR (151 MHz, Benzene- $d_6$ )  $\delta$  206.8, 60.4 (t,  $J = 4.9$  Hz), 40.7, 36.3, 35.8 (t,  $J = 8.5$  Hz), 32.6 (t,  $J = 5.5$  Hz), 24.6.  $^{31}\text{P}$  NMR (243 MHz, Benzene- $d_6$ )  $\delta$  74.3.



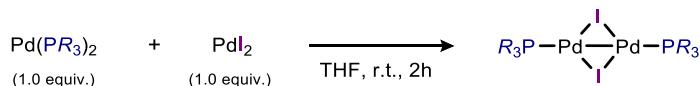
**4.3.7. Bis(2,2,6,6-tetramethyl-1-phenyl-4-phosphorinanone)palladium(0).** The title compound was synthesized according to the general procedure from [Pd(tmeda)(Me)<sub>2</sub>] (252.7 mg, 1.0 mmol, 1.0 eq.) and phosphine (496.6 mg, 2.0 mmol, 2.0 eq.) in THF (5 mL). After 1 h reaction time, the product was obtained as a pale yellow powder (179.8 mg, 0.298 mmol, 30%).

<sup>1</sup>H NMR (600 MHz, Benzene-*d*<sub>6</sub>) δ 8.44 – 8.12 (m, 2H), 7.29 – 6.83 (m, 3H), 2.58 (d, *J* = 13.0 Hz, 2H), 2.62 – 2.41 (m, 2H), 1.63 (t, *J* = 8.6 Hz, 6H), 0.96 (t, *J* = 6.5 Hz, 6H). <sup>13</sup>C NMR (151 MHz, Benzene-*d*<sub>6</sub>) δ 207.9, 137.6 (t, *J* = 10.5 Hz), 136.2 (t, *J* = 7.9 Hz), 130.2, 53.0, 37.0 (t, *J* = 5.2 Hz), 33.0 (t, *J* = 11.7 Hz), 29.9 (t, *J* = 3.4 Hz). <sup>31</sup>P NMR (243 MHz, Benzene-*d*<sub>6</sub>) δ 50.2.

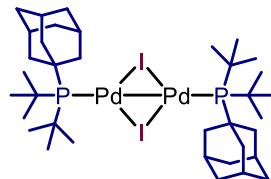
#### 4.4. Synthesis and Characterization of Iodide-bridged Pd(I) Dimers



**4.4.1. General Procedure A.** Inside a glovebox free phosphine ligand (2.0 equiv.) and tris(dibenzylideneacetone)dipalladium(0) (0.5 equiv.) were dissolved in THF (0.07M). The reaction mixture was stirred at ambient temperature for 2 h before palladium(II) iodide (1.0 equiv.) was added. The reaction was monitored by  $^{31}\text{P}$  NMR. Upon full conversion of the starting material (2-6h), acetone (10-fold amount compared to THF) was added and the resulting mixture was left to crystallize at -30°C overnight. Pd(I) dimers were collected by filtration and washed with acetone (3x 10 mL).

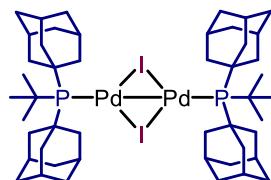


**4.4.2. General Procedure B.** Inside a glovebox bis(phosphine)palladium(0) [ $\text{Pd}(\text{PR}_3)_2$ ] (1.0 equiv.) and palladium(II) iodide (1.0 equiv.) were dissolved in THF (0.05M) and the reaction mixture stirred at ambient temperature. The reaction was monitored by  $^{31}\text{P}$  NMR. Upon full conversion of the starting material (monitored by  $^{31}\text{P}$  NMR, usually 3h), acetone (10-fold amount compared to THF) was added and the resulting mixture was left to crystallize at -30°C overnight. Pd(I) dimers were collected by filtration and washed with acetone (3x 10 mL).



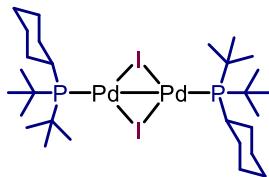
**4.4.3. Di- $\mu$ -iodobis(1-adamantyl-di-*tert*-butylphosphine)dipalladium(I).** Prepared according to general procedure A (reaction time: 6h). The title compound was obtained as a very dark purple (almost black) powder (2.63 g, 2.56 mmol, 96%). Crystals suitable for X-ray diffraction analysis were grown by slow evaporation from a solution of THF.

*Note:* A small amount of *n*-hexane was added to the NMR sample to increase solubility.  $^1\text{H}$  NMR (600 MHz, benzene- $d_6$ )  $\delta$  2.23 (br, 12H), 1.72 (br, 6H), 1.53 (d (br),  $J = 12.0$  Hz, 6H), 1.44 (d (br),  $J = 12.0$  Hz, 6H), 1.32 (dd,  $J = 6.0, 6.0$  Hz, 36H).  $^{13}\text{C}$  NMR (151 MHz, benzene- $d_6$ )  $\delta$  41.9, 41.7 (br), 36.8, 36.5, 33.4 (br), 29.4 (t,  $J = 3.8$  Hz).  $^{31}\text{P}$  NMR (243 MHz, benzene- $d_6$ )  $\delta$  104.6.



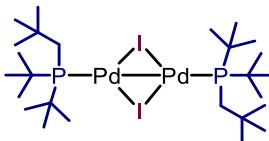
**4.4.4. Di- $\mu$ -iodobis(di-1-adamantyl-*tert*-butylphosphine)dipalladium(I).** The title compound was prepared according to general procedure B (reaction time: 3h) and was obtained as a dark purple powder (654 mg, 0.552 mmol, 91%).

*Note:* The product showed very low solubility that prevented its further characterization.  $^{31}\text{P}$  NMR (243 MHz, benzene- $d_6$ )  $\delta$  106.4.



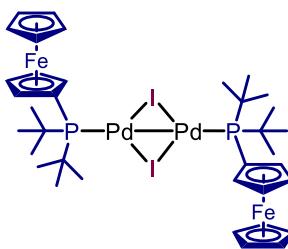
**4.4.5. Di- $\mu$ -iodobis(di-*tert*-butyl(cyclohexyl)phosphine)dipalladium(I).** The title compound was prepared according to general procedure B and was obtained as a dark purple crystalline solid (340 mg, 0.368 mmol, 69%). Crystals suitable for X-ray diffraction analysis were grown by slow evaporation from a solution of THF.

$^1\text{H}$  NMR (600 MHz, Benzene- $d_6$ )  $\delta$  2.36 (d,  $J$  = 11.7 Hz, 4H), 1.89 (t,  $J$  = 11.5 Hz, 2H), 1.62 – 1.45 (m, 7H), 1.44 – 1.30 (m, 6H), 1.25 (t,  $J$  = 6.2 Hz, 36H), 1.16 – 0.93 (m, 6H).  $^{13}\text{C}$  NMR (151 MHz, Benzene- $d_6$ )  $\delta$  42.7 (t,  $J$  = 3.5 Hz), 34.5 (t,  $J$  = 3.9 Hz), 33.7, 31.5 (t,  $J$  = 3.3 Hz), 28.6 (t,  $J$  = 4.9 Hz), 26.7.  $^{31}\text{P}$  NMR (243 MHz, Benzene- $d_6$ )  $\delta$  88.0.



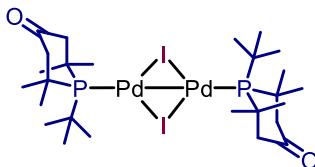
**4.4.6. Di- $\mu$ -iodobis(di-*tert*-butyl(neopentyl)phosphine)dipalladium(I).** Inside an argon-filled glovebox bis(di-*tert*-butyl(neopentyl)phosphine)palladium(0) (431 mg, 0.8 mmol, 1.0 eq.) and palladium(II)iodide (288 mg, 0.8 mmol, 1.0 eq.) were dissolved in THF (16 mL). After 2 hours of stirring at ambient temperature, the solvent was evaporated. The obtained residue was redissolved in toluene (25 mL) and filtered through a sintered funnel. The solvent was evaporated and the obtained residue was washed with MeOH (3x 5 mL) and a small amount of hexane (1 mL). The product was obtained as a dark greenish crystalline powder (138.5 mg, 0.154 mmol, 19%). Crystals suitable for X-ray diffraction analysis were grown by slow evaporation from a solution of THF.

$^1\text{H}$  NMR (600 MHz, Benzene- $d_6$ )  $\delta$  1.30 – 1.25 (m, 4H), 1.24 – 1.03 (m, 54H).  $^{13}\text{C}$  NMR (151 MHz, Benzene- $d_6$ )  $\delta$  36.2 (t,  $J$  = 3.8 Hz), 33.1 (t,  $J$  = 6.3 Hz), 32.8 (t,  $J$  = 2.9 Hz), 30.7 (t,  $J$  = 2.6 Hz), 30.6 (t,  $J$  = 3.8 Hz).  $^{31}\text{P}$  NMR (243 MHz, Benzene- $d_6$ )  $\delta$  60.5.



**4.4.7. Di- $\mu$ -iodobis(di-*tert*-butyl(ferrocenyl)phosphine)dipalladium(I).** The title compound was prepared according to general procedure B and was obtained as a dark purple crystalline solid (540.1 mg, 0.479 mmol, 82%). Crystals suitable for X-ray diffraction analysis were grown by slow evaporation from a solution of THF.

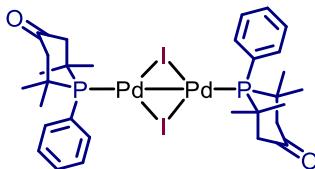
<sup>1</sup>H NMR (600 MHz, Benzene-*d*<sub>6</sub>) δ 4.50 (s, 4H), 4.29 (s, 10H), 4.05 (s, 4H), 1.28 (t, *J* = 6.9 Hz, 36H). <sup>13</sup>C NMR (151 MHz, Benzene-*d*<sub>6</sub>) δ 75.4 (t, *J* = 5.5 Hz), 74.8 (t, *J* = 12.4 Hz), 71.0, 70.1 (t, *J* = 3.2 Hz), 34.0 (t, *J* = 6.7 Hz), 31.2 (t, *J* = 3.9 Hz). <sup>31</sup>P NMR (243 MHz, Benzene-*d*<sub>6</sub>) δ 74.2.



#### 4.4.8. Di-μ-iodobis(2,2,6,6-tetramethyl-1-(*tert*-butyl)-4-phosphorinanone)dipalladium(I).

The title compound was prepared according to general procedure B and was obtained as a greenish dark crystalline solid (137.3 mg, 0.149 mmol, 71%). Crystals suitable for X-ray diffraction analysis were grown by slow evaporation from a solution of THF.

<sup>1</sup>H NMR (600 MHz, Benzene-*d*<sub>6</sub>) δ 3.41 (d, *J* = 12.5 Hz, 4H), 1.90 – 1.76 (m, 4H), 1.32 – 1.25 (m, 12H), 1.21 – 1.14 (m, 18H), 1.10 – 0.99 (m, 12H). <sup>13</sup>C NMR (151 MHz, Benzene-*d*<sub>6</sub>) δ 205.7, 59.3 (t, *J* = 2.7 Hz), 39.6 (t, *J* = 2.5 Hz), 35.7, 35.4 (t, *J* = 6.4 Hz), 32.3 (t, *J* = 4.0 Hz), 25.1. <sup>31</sup>P NMR (243 MHz, Benzene-*d*<sub>6</sub>) δ 89.6.

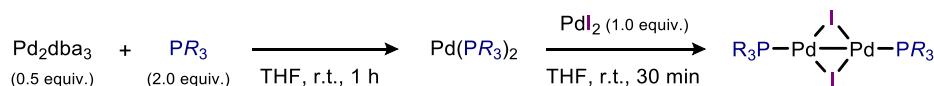


#### 4.4.9. Di-μ-iodobis(2,2,6,6-tetramethyl-1-phenyl-4-phosphorinanone)dipalladium(I).

The title compound was prepared according to general procedure B and was obtained as a dark green powder (122.1 mg, 0.127 mmol, 49%). Crystals suitable for X-ray diffraction analysis were grown by slow evaporation from a solution of THF.

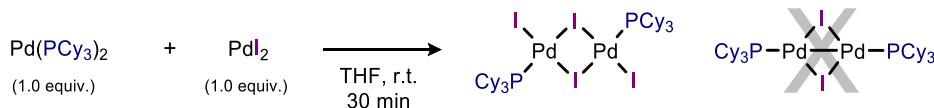
<sup>1</sup>H NMR (600 MHz, Methylene Chloride-*d*<sub>2</sub>) δ 8.05 – 7.96 (m, 2H), 7.55 – 7.39 (m, 3H), 3.09 (d, *J* = 13.3 Hz, 2H), 2.44 (dt, *J* = 13.8, 7.2 Hz, 2H), 1.44 (t, *J* = 8.8 Hz, 6H), 1.05 (t, *J* = 7.4 Hz, 6H). <sup>13</sup>C NMR (151 MHz, Methylene Chloride-*d*<sub>2</sub>) δ 208.9, 136.3 (t, *J* = 7.3 Hz), 132.8 (m), 131.5, 129.0 (t, *J* = 4.9 Hz), 52.3, 36.1 (t, *J* = 6.6 Hz), 32.6 (t, *J* = 9.1 Hz), 29.4 (t, *J* = 2.8 Hz). <sup>31</sup>P NMR (243 MHz, Methylene Chloride-*d*<sub>2</sub>) δ 67.5.

#### 4.5. $^{31}\text{P}$ NMR Studies

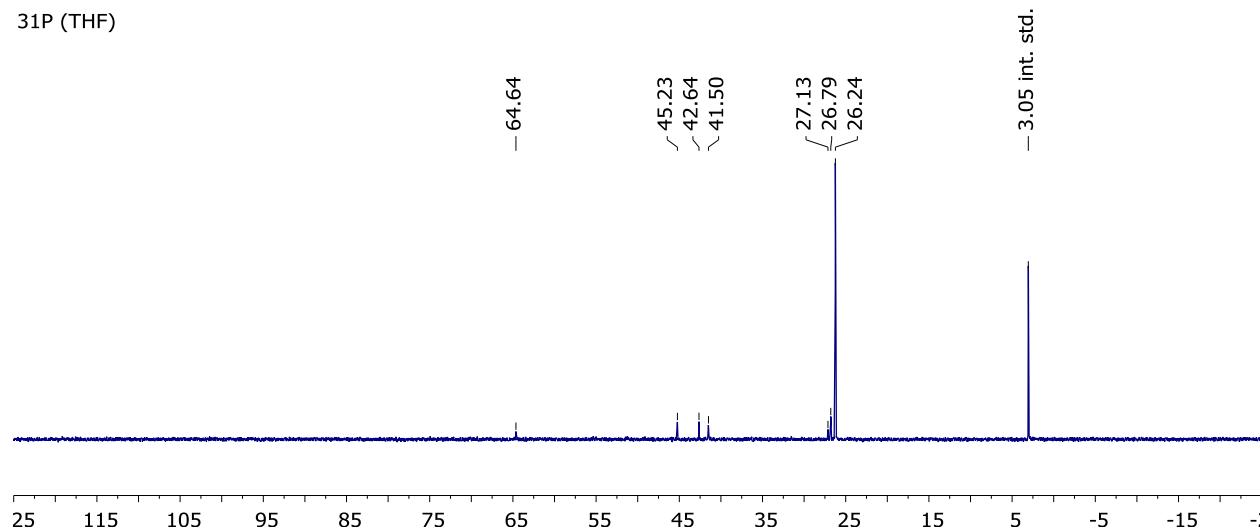


**4.5.1. General procedure employed for *in situ* testing of ligands.** Inside a glovebox free phosphine ligand (0.1 mmol, 2.0 equiv.) and tris(dibenzylideneacetone)dipalladium(0) (22.9 mg, 0.025 mmol, 0.5 equiv.) were dissolved in THF (1.5 mL). The reaction mixture was stirred at ambient temperature for 1 h. Half of the obtained mixture was then analyzed by  $^{31}\text{P}$  NMR. To the remaining half palladium(II) iodide (9.0 mg, 0.025 mmol, 1.0 equiv.) was added. After 30 minutes of further stirring at ambient temperature the resulting mixture was analyzed by  $^{31}\text{P}$  NMR.

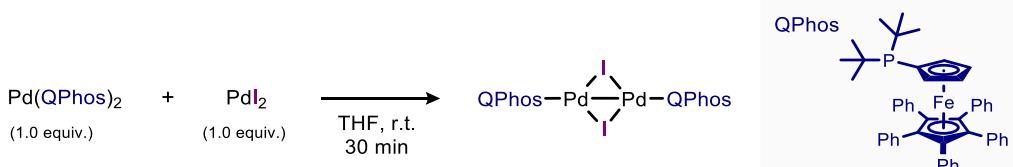
Obtained  $^{31}\text{P}$  NMR shifts were compared with reported Pd(0), and Pd(II) iodo complexes and/or close analogues (*e.g.* chloro or bromo complexes) in order to identify the potential formation of dinuclear Pd(I) complexes.



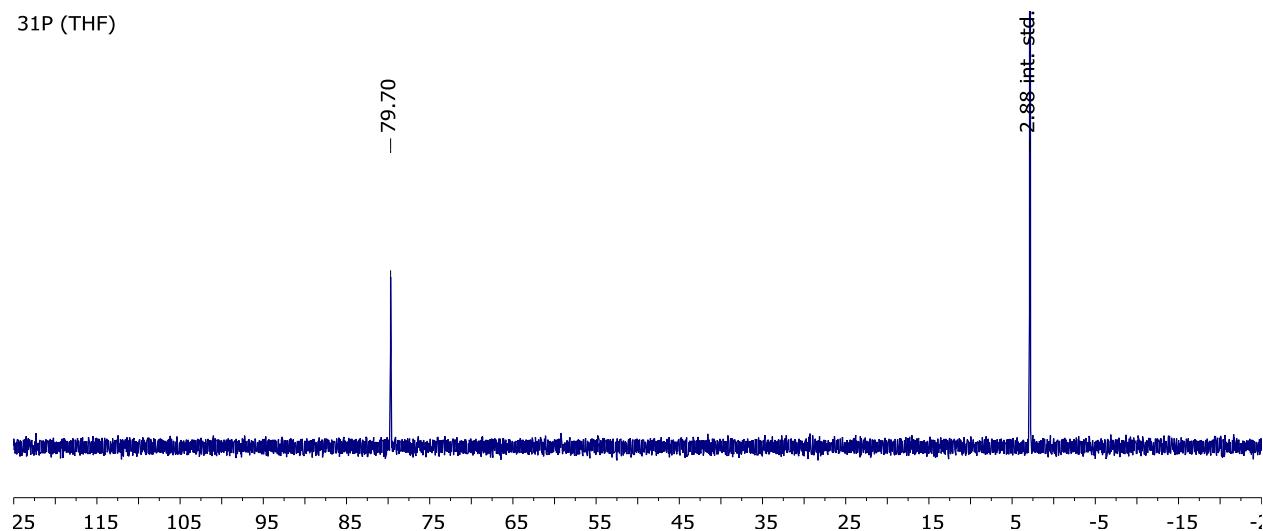
**4.5.2. Attempted comproportionation of  $\text{Pd}(\text{PCy}_3)_2$  and  $\text{PdI}_2$ .** Inside a glovebox [ $\text{Pd}(\text{PCy}_3)_2$ ] (33.4 mg, 0.05 mmol, 1.0 equiv.;  $^{31}\text{P}$  signal at 39.0 ppm) and palladium(II) iodide (18.0 mg, 0.05 mmol, 1.0 equiv.) were dissolved in THF (1.0 mL) and the reaction mixture stirred at ambient temperature. After 30 min an aliquot was taken and analyzed by  $^{31}\text{P}$  NMR after filtration through a syringe filter to remove the formed black precipitate. No changes in  $^{31}\text{P}$  NMR spectrum occurred within 90 min:



*Note:* The same  $^{31}\text{P}$  shift (at 26 ppm) was obtained after 30 minutes of exposure when  $\text{PdI}_2$  was added to a THF solution containing 1 equivalent of  $\text{PCy}_3$  (*i.e.* Pd:P ratio of 1:1).



**4.5.3.  $[\text{Pd}(\mu-\text{I})(\text{QPhos})]_2$ .** Inside a glovebox  $[\text{Pd}(\text{QPhos})_2]$  (76.4 mg, 0.05 mmol, 1.0 equiv.;  $^{31}\text{P}$  signal at 60.8 ppm) and palladium(II) iodide (18.0 mg, 0.05 mmol, 1.0 equiv.) were dissolved in THF (1.0 mL) and the reaction mixture stirred at ambient temperature. After 30 min the mixture was analyzed by  $^{31}\text{P}$  NMR:



*Note:* The observed signal at 79.7 ppm is similar to that of the Pd(I) dimer with the related di-*tert*-butyl(ferrocenyl)phosphine ligand (74.2 ppm; corresponding Pd(0) complex at 55.8 ppm) and hence suggests the formation of a dinuclear Pd(I) complex.

#### 4.6. Crystallographic data

Single crystal X-ray data were collected on a Bruker-Nonius Kappa CCD diffractometer with an APEX-II detector with graphite-monochromatized Mo- $K\alpha$  ( $\lambda = 0.71073 \text{ \AA}$ ) radiation at 170 K (for **D3**, **D8**, and **D10**), 273 K (**D11**) or 293 K (**D23**). Data collection and reduction for **D3**, **D8**, and **D10** were performed using the programs *COLLECT* (73) and *HKL DENZO AND SCALEPACK*,(74) and for **D11** and **D23** using the program *SAINT* (75), and in all structures the intensities were corrected for absorption using *SADABS*.(76) Single-crystal X-ray data of **D4** were measured using a Rigaku SuperNova dual-source Oxford diffractometer equipped with an Atlas detector using mirror-monochromated Cu- $K\alpha$  ( $\lambda = 1.54184 \text{ \AA}$ ) radiation. The data collection and reduction were performed using the program *CrysAlisPro* (77) and Gaussian face index absorption correction method was applied.(77) The structures were solved with intrinsic phasing (*SHELXT*) (78) and refined by full-matrix least squares on  $F^2$  using the *OLEX2* software (79), which utilises the *SHELXL-2015* module.(80) The main details of crystal data collection and refinement parameters are presented in Table S6.

CCDC 2055171-2055174, 2064562 and 2064863 contain the supplementary crystallographic data for this paper. These data can be obtained free of charge via <http://www.ccdc.cam.ac.uk/conts/retrieving.html>.

**Comments.** Anisotropic displacement parameters were assigned to non-H atoms. All hydrogen atoms were refined using riding models with  $U_{\text{eq}}(\text{H})$  of  $1.5U_{\text{eq}}(\text{C})$  for terminal methyl groups and  $U_{\text{eq}}(\text{H})$  of  $1.2U_{\text{eq}}(\text{C})$  for all other C-H groups (methylene, cyclohexyl, ferrocenyl).

**D11** was refined as a twinned dataset. One of the *t*-Bu groups was found to be disordered over two positions in a 57:43% ratio and treated by gently restraining the anisotropic displacement parameters of the disorder using the RIGU restraint.

Parameter	D3	D4	D8	D10	D11	D23
CCDC Number	2055171	2064562	2055172	2055173	2055174	2064863
Formula	C <sub>26</sub> H <sub>58</sub> I <sub>2</sub> P <sub>2</sub> Pd <sub>2</sub>	C <sub>36</sub> H <sub>66</sub> I <sub>2</sub> P <sub>2</sub> Pd <sub>2</sub>	C <sub>36</sub> H <sub>54</sub> Fe <sub>2</sub> I <sub>2</sub> P <sub>2</sub> Pd <sub>2</sub>	C <sub>26</sub> H <sub>50</sub> I <sub>2</sub> O <sub>2</sub> P <sub>2</sub> Pd <sub>2</sub>	C <sub>28</sub> H <sub>58</sub> I <sub>2</sub> P <sub>2</sub> Pd <sub>2</sub>	C <sub>30</sub> H <sub>42</sub> I <sub>2</sub> O <sub>2</sub> P <sub>2</sub> Pd <sub>2</sub>
Formula Weight [g/mol]	899.26	1027.42	1127.03	923.20	923.28	963.17
Colour	green	black	brown	black	violet	Green
Shape	plate	plate	block	plate	plate	needle
Crystal System	monoclinic	triclinic	monoclinic	triclinic	triclinic	monoclinic
Space Group	<i>P</i> 2 <sub>1</sub> / <i>c</i>	<i>P</i> -1	<i>P</i> 2 <sub>1</sub> / <i>n</i>	<i>P</i> -1	<i>P</i> -1	<i>P</i> 2 <sub>1</sub> / <i>n</i>
<i>a</i> [Å]	10.4724(3)	8.0167(4)	10.1692(4)	8.2264(4)	8.0429(10)	8.1159(3)
<i>b</i> [Å]	13.6976(5)	10.0881(6)	16.0064(4)	8.4748(3)	9.3961(12)	26.1938(11)
<i>c</i> [Å]	12.0688(4)	13.2295(7)	12.6486(4)	13.4422(6)	12.4272(15)	8.2221(3)
$\alpha$ [Å]	90	108.675(5)	90	103.001(2)	74.787(4)	90
$\beta$ [Å]	102.168(2)	95.314(4)	109.763(2)	94.597(2)	73.950(4)	108.355(2)
$\gamma$ [Å]	90	108.217(5)	90	116.065(2)	76.133(5)	90
V [Å <sup>3</sup> ]	1692.34(10)	940.44(10)	1937.57(11)	802.92(6)	856.65(19)	1658.98(11)
Z	2	1	2	1	1	2
$\rho_{\text{calc}}$ [g/cm <sup>3</sup> ]	1.765	1.814	1.932	1.909	1.790	1.928
<i>F</i> (000)	884	510	1100	450	454	932
$\mu$ [mm <sup>-1</sup> ]	2.99	21.57	3.35	3.16	2.96	3.07
Temperature [K]	170.0(1)	120.0(1)	170.0(1)	170.0(1)	273.0(1)	293.0
$\theta_{\text{max}}$ [°]	28.7	76.1	28.7	27.9	25.2	25.0
Total Refl.	12761	6131	14431	6341	118240	100462
Unique Refl.	3737	3668	4262	3519	49373	2926
Reflections with $I_0 > 2\sigma(I_0)$	2691	3460	3515	3089	43906	2817
$R_{\text{int}}$	0.048	0.028	0.049	0.030	0.074	0.027
Parameters	154	196	205	161	193	176
Restraints	0	0	0	0	36	0
GooF on F <sup>2</sup>	1.08	1.15	1.08	1.12	1.19	1.35
$R_I$ ( $I_0 > 2\sigma(I_0)$ )	0.046	0.034	0.046	0.031	0.082	0.022
$wR_2$ ( $I_0 > 2\sigma(I_0)$ )	0.091	0.086	0.090	0.092	0.226	0.048
$R_I$ (all reflections)	0.070	0.035	0.060	0.037	0.089	0.023
$wR_2$ (all reflections)	0.098	0.087	0.095	0.095	0.231	0.049
Largest Peak [e/Å <sup>3</sup> ]	1.08	0.91	1.75	0.67	2.43	0.50
Largest Hole [e/Å <sup>3</sup> ]	-0.83	-0.97	-1.40	-1.32	-1.51	-0.44

**Table S6.**

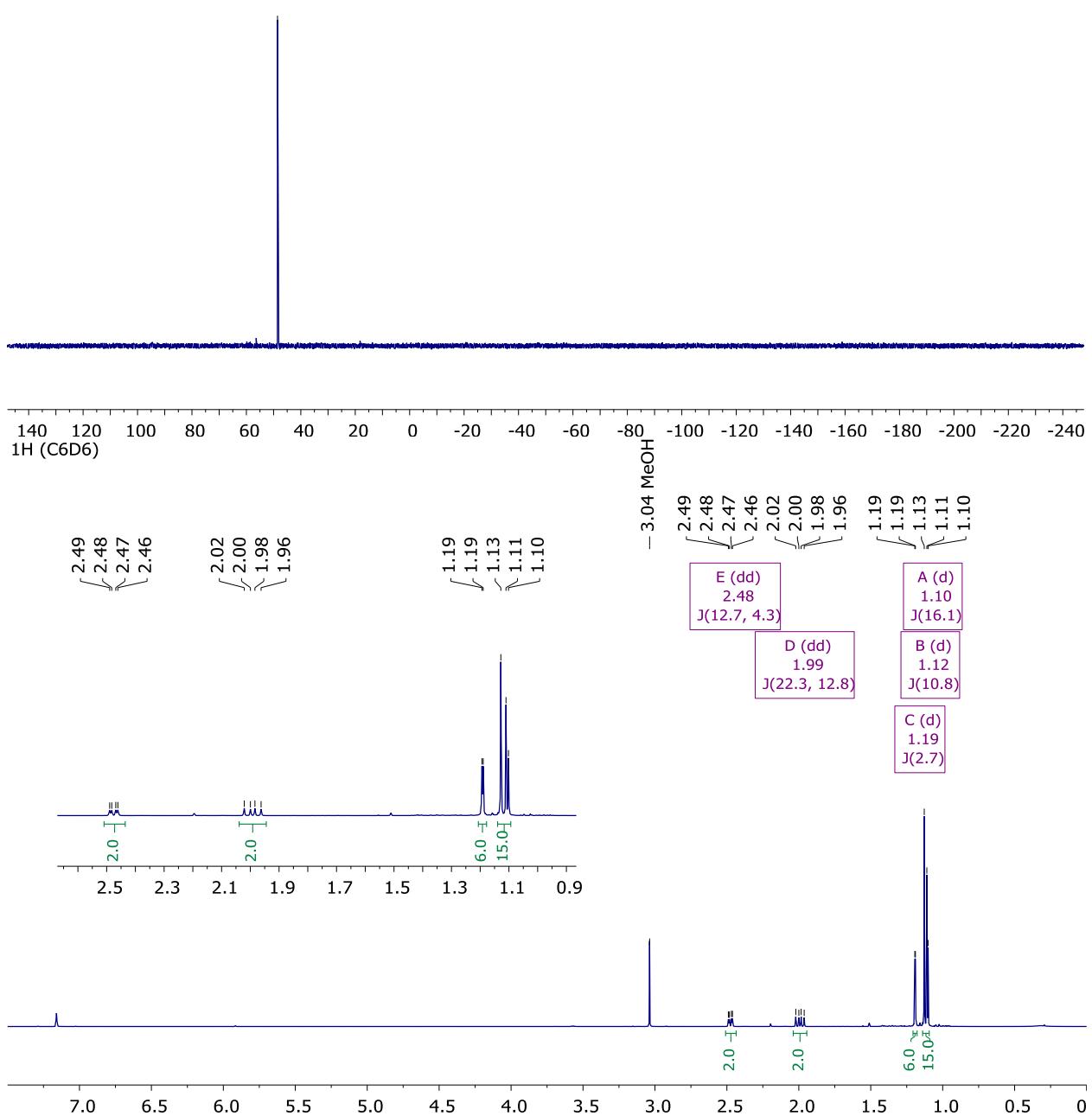
Crystallographic data of compounds **D3**, **D4**, **D8**, **D10**, **D11** and **D23**.

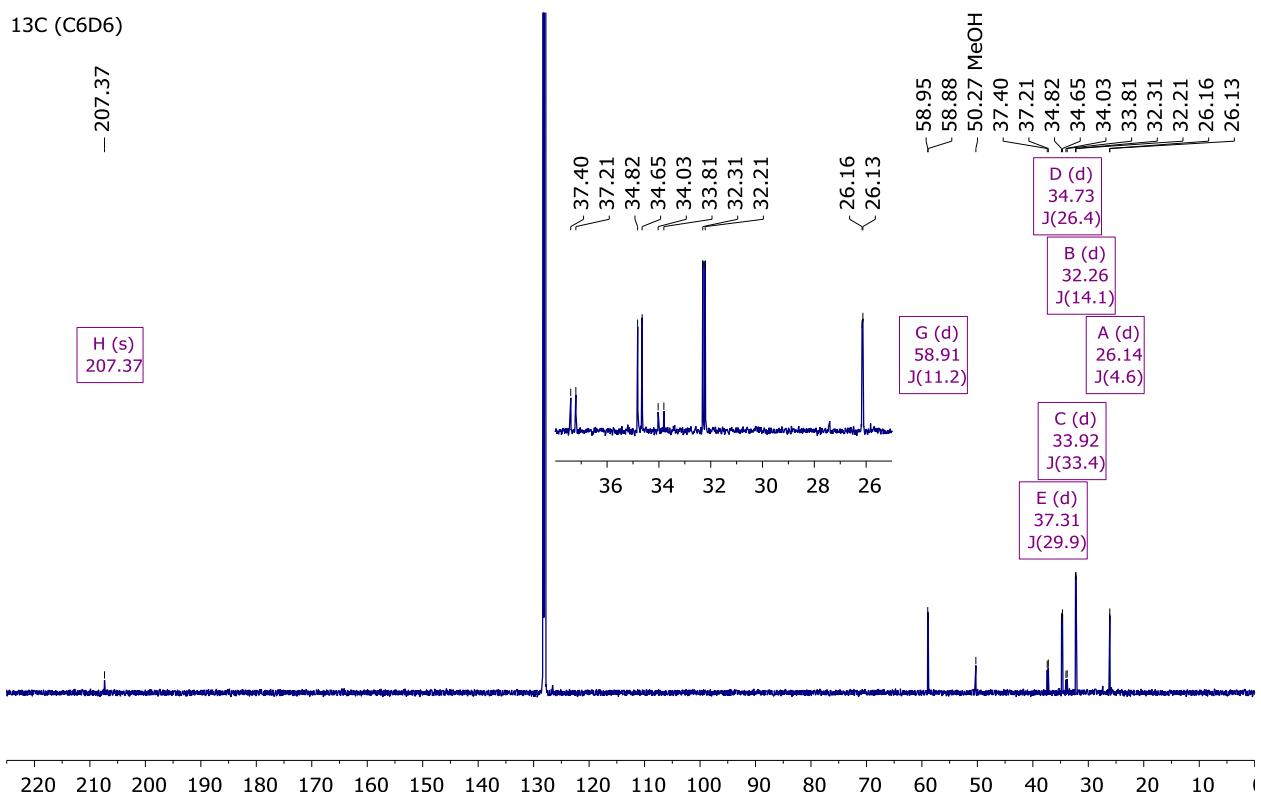
## 4.7. NMR-spectra

### **4.7.1. 1-(*tert*-Butyl)-2,2,6,6-tetramethylphosphinan-4-one**

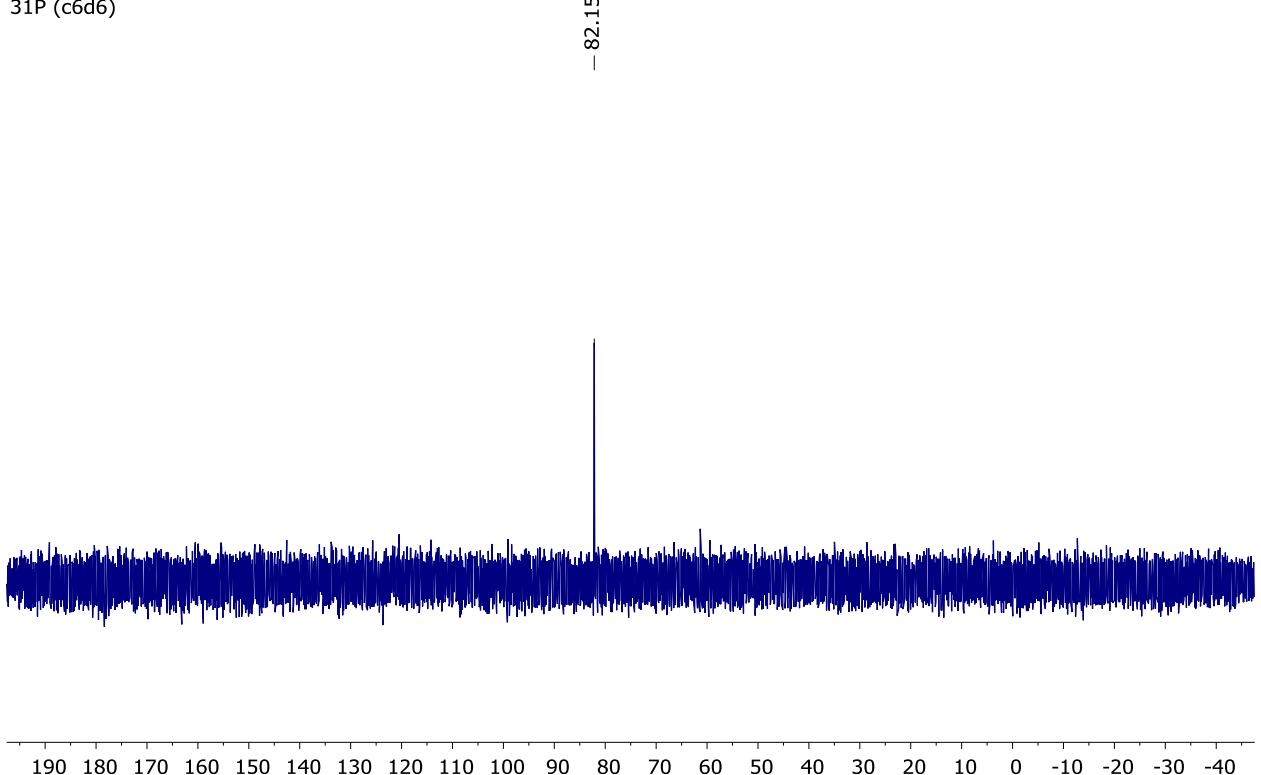
<sup>31</sup>P (C6D6)

— 48.56

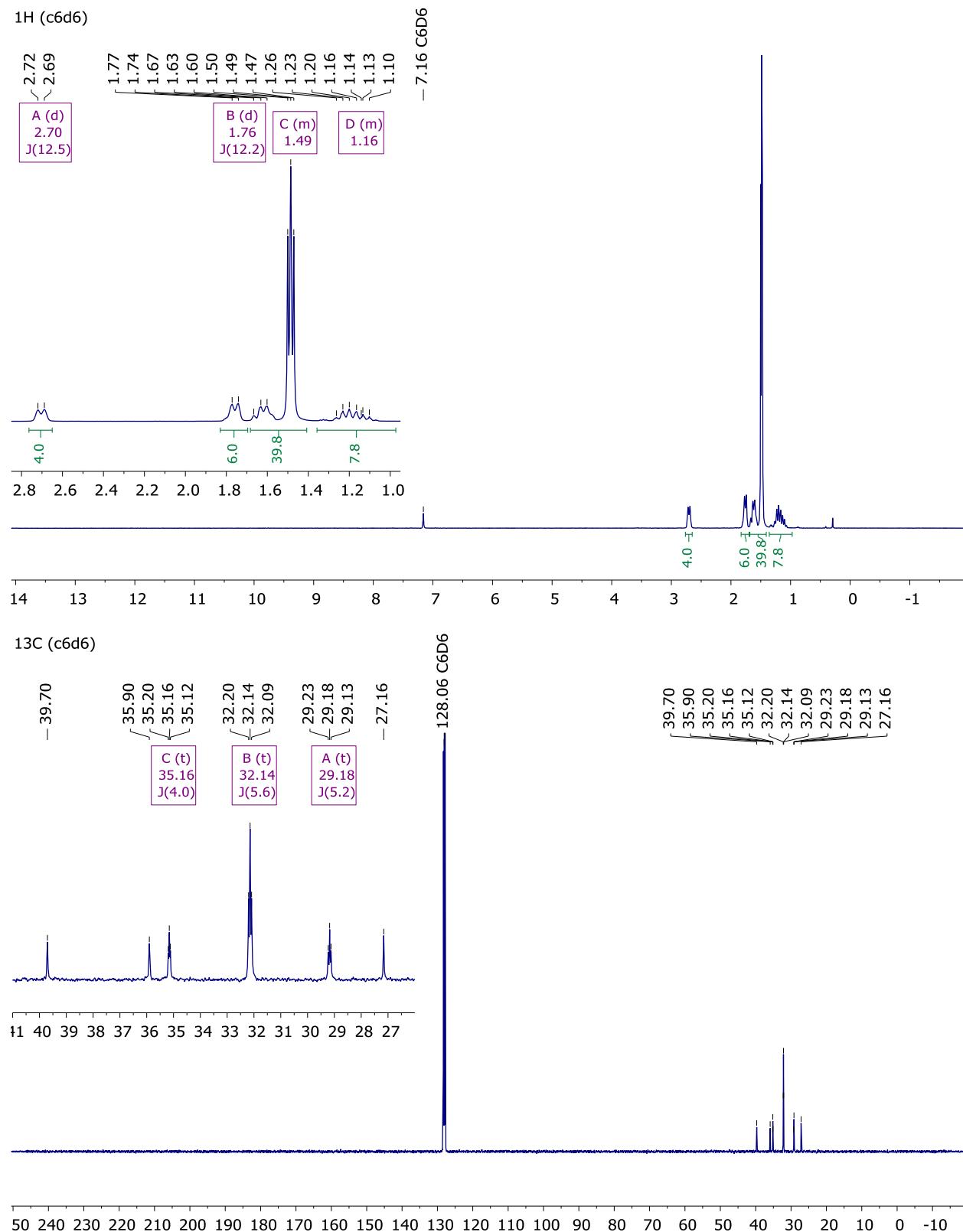


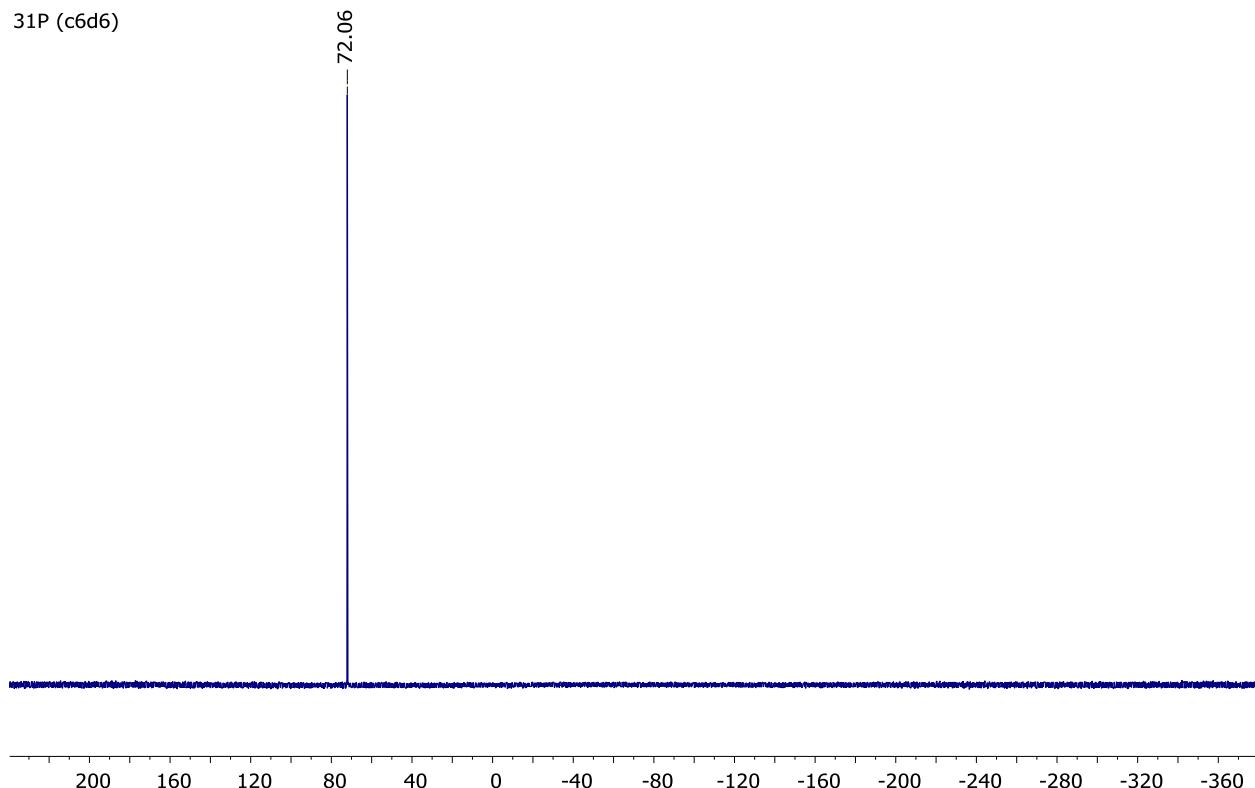


#### 4.7.2. Bis(di-1-adamantyl-*tert*-butylphosphine)palladium(0)

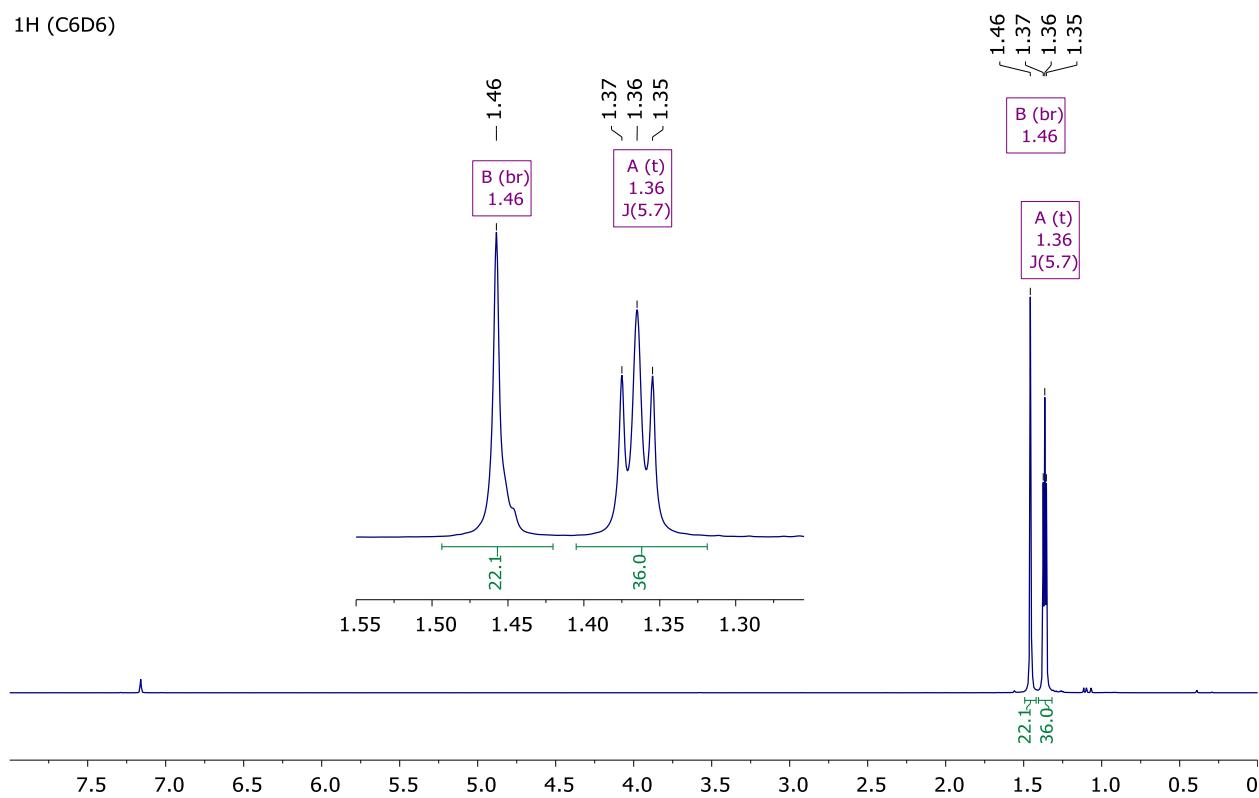


#### 4.7.3. Bis(di-*tert*-butyl(cyclohexyl)phosphine)palladium(0)

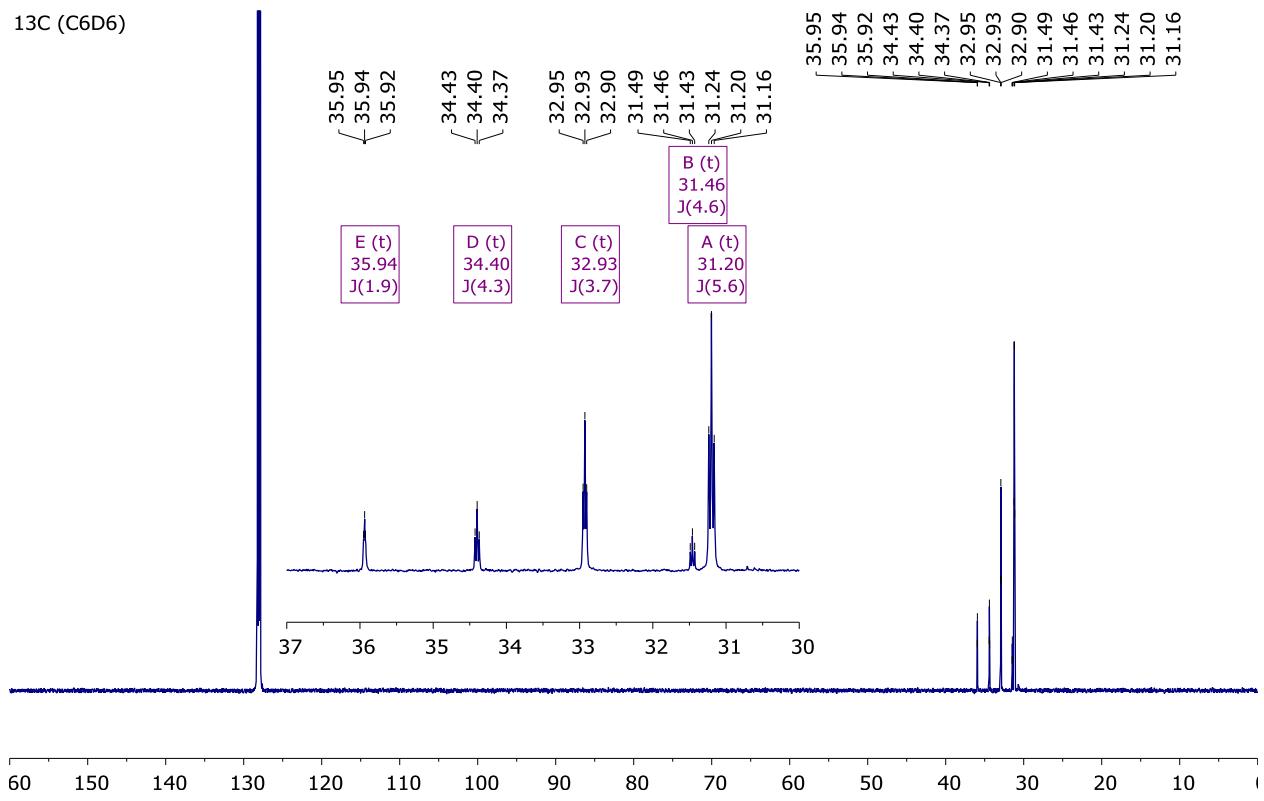




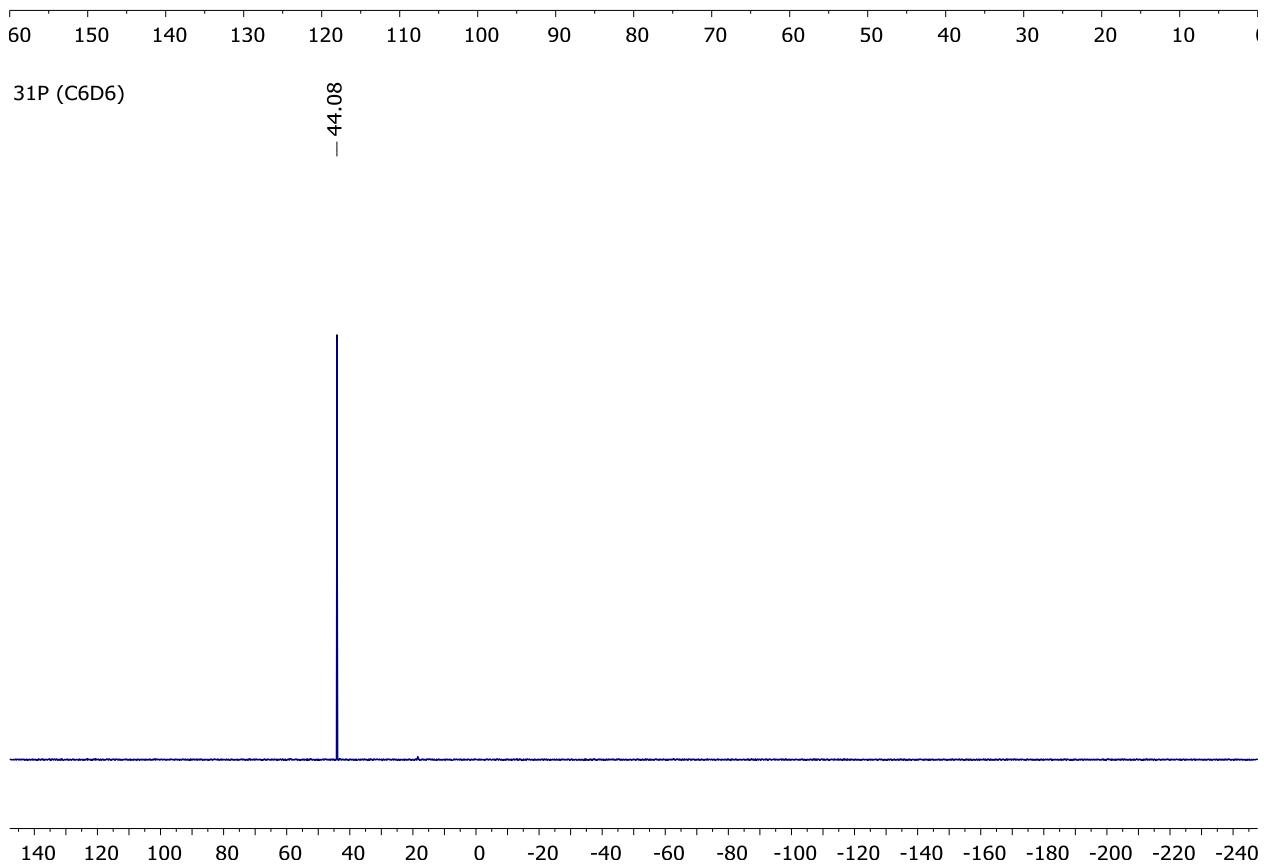
#### 4.7.4. Bis(di-*tert*-butyl(neopentyl)phosphine)palladium(0)



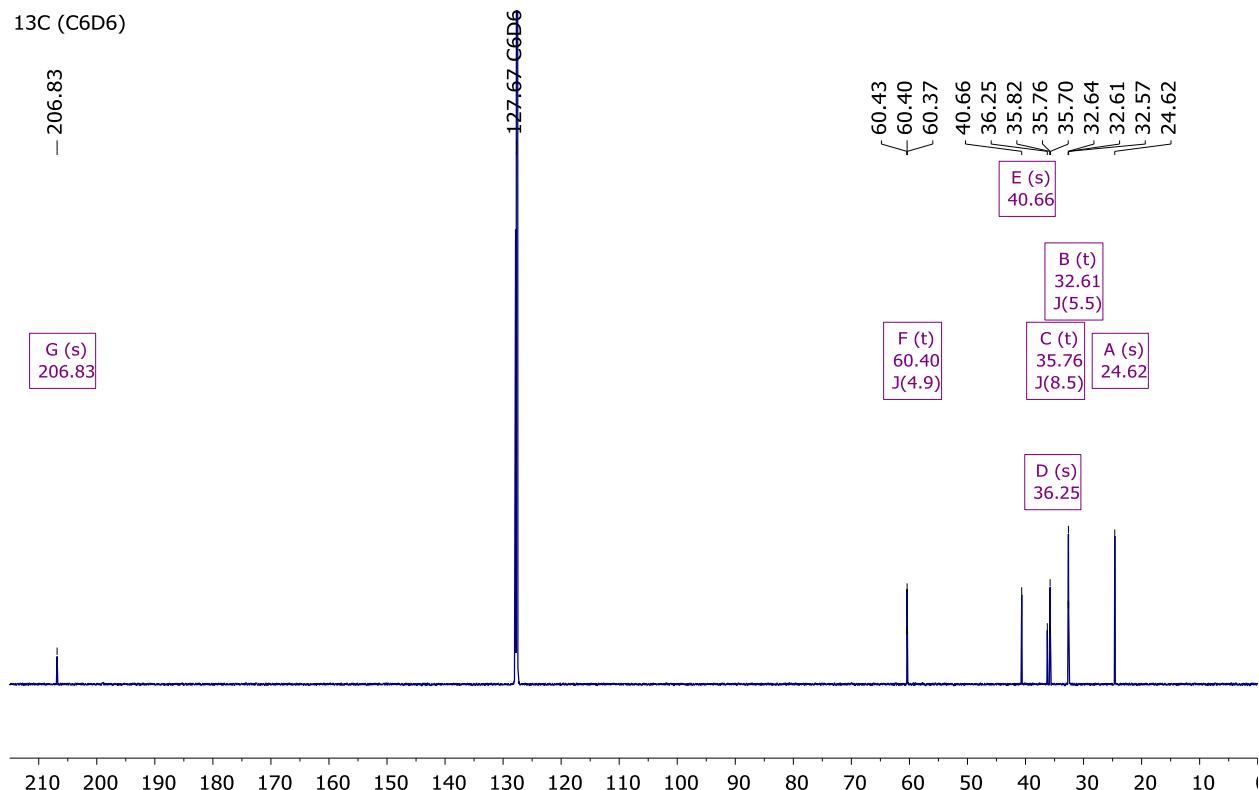
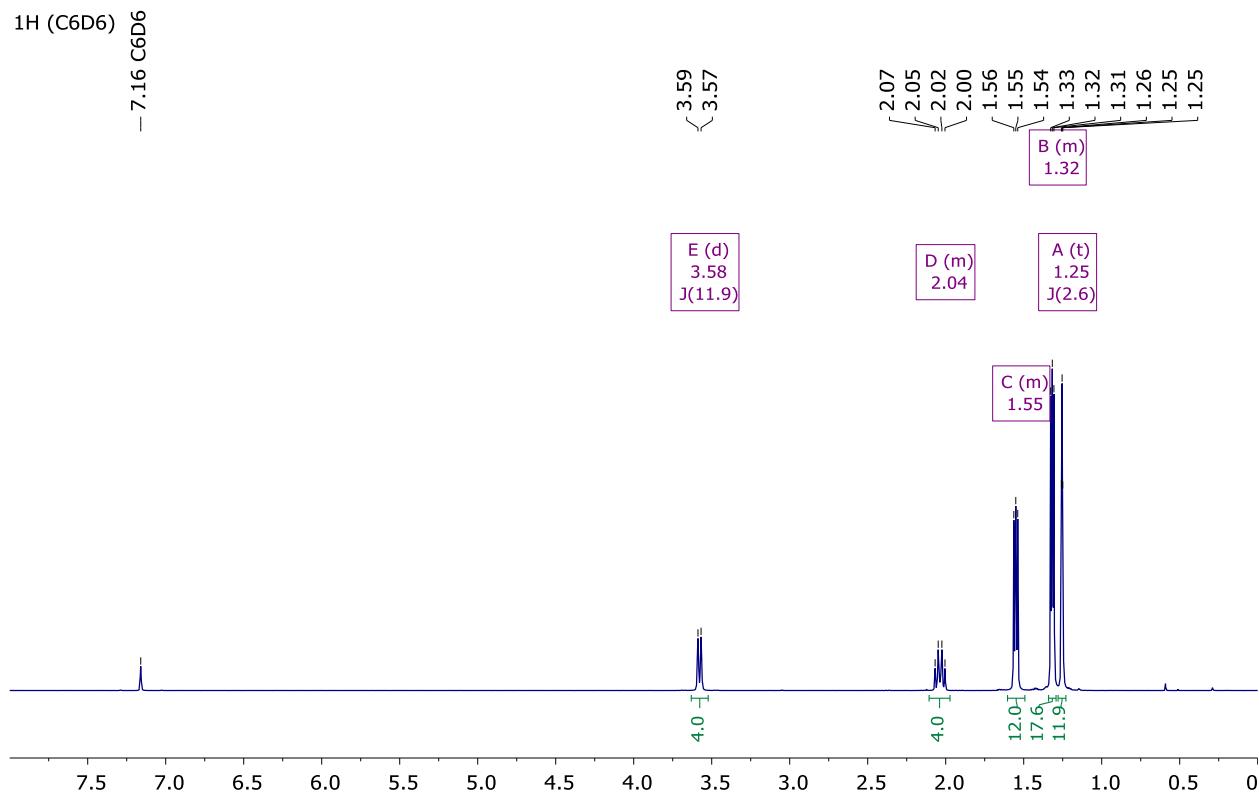
<sup>13</sup>C (C6D6)



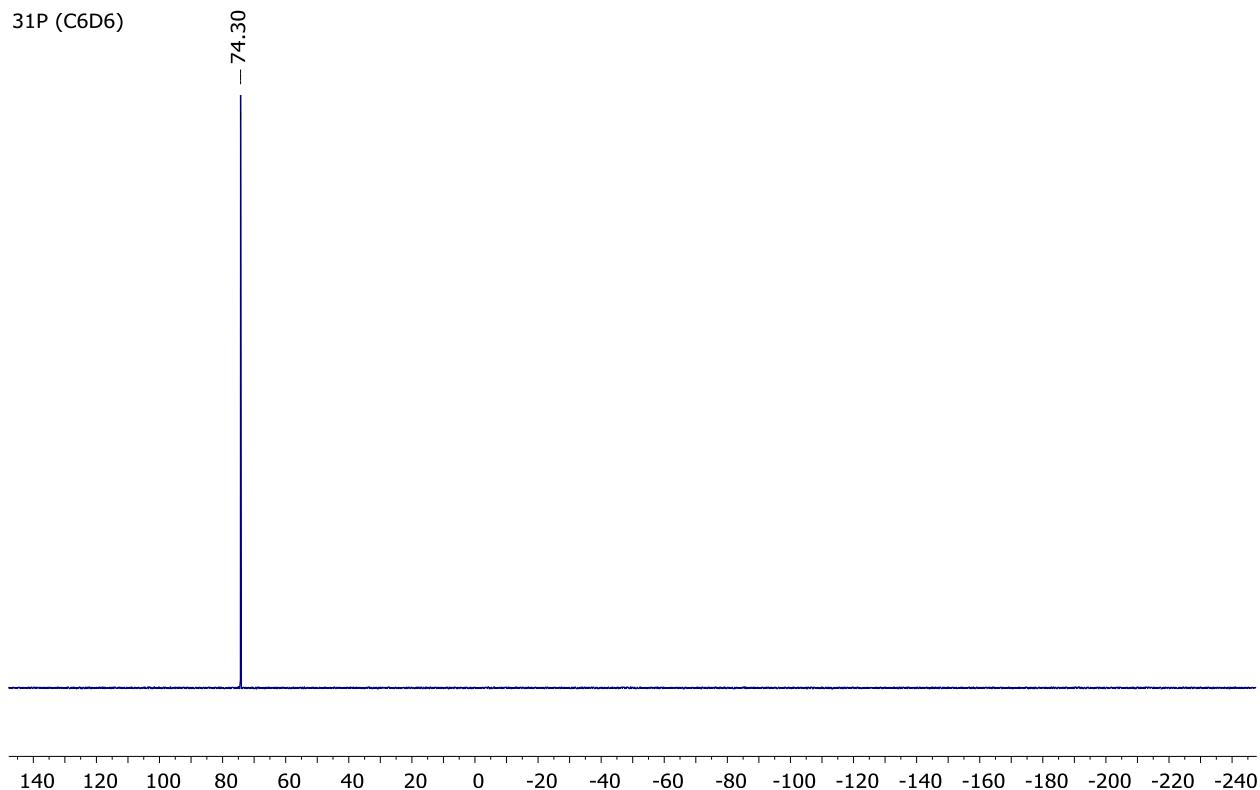
<sup>31</sup>P (C6D6)



#### 4.7.5. Bis(2,2,6,6-tetramethyl-1-(*tert*-butyl)-4-phosphorinanone)palladium(0)

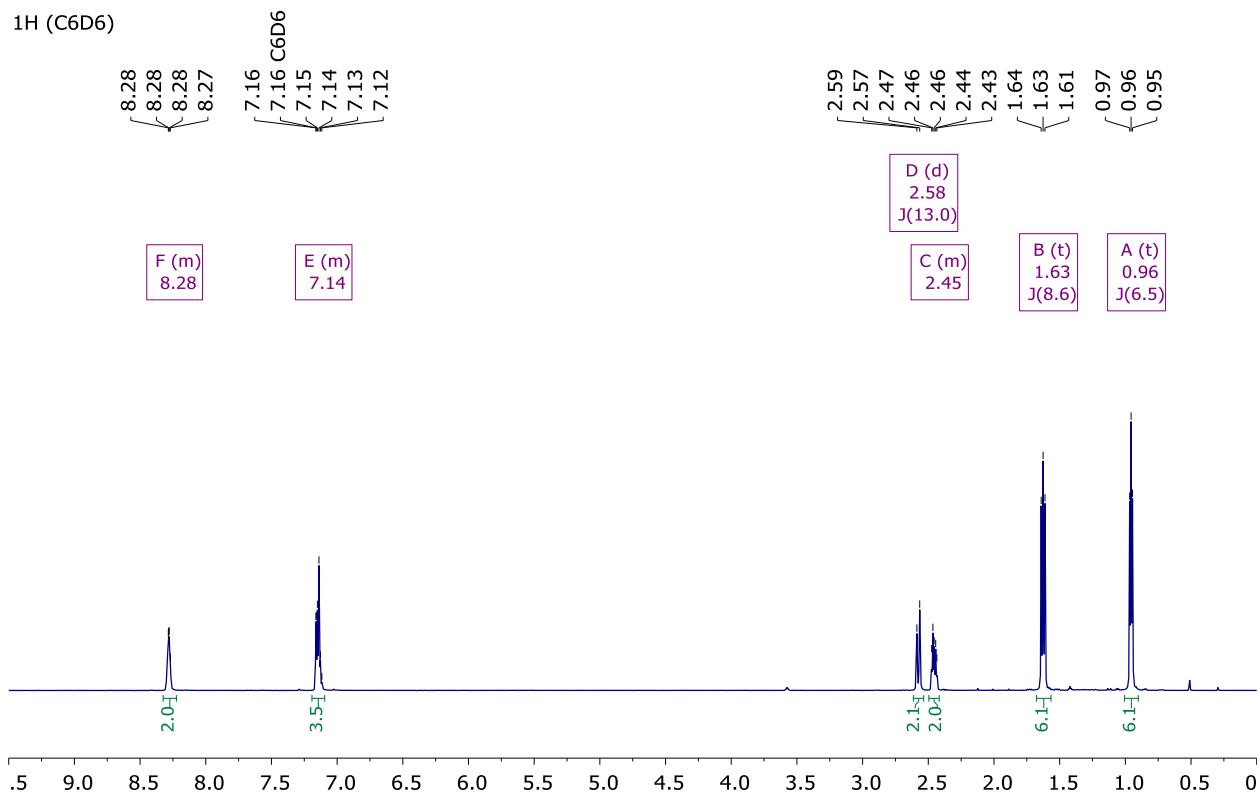


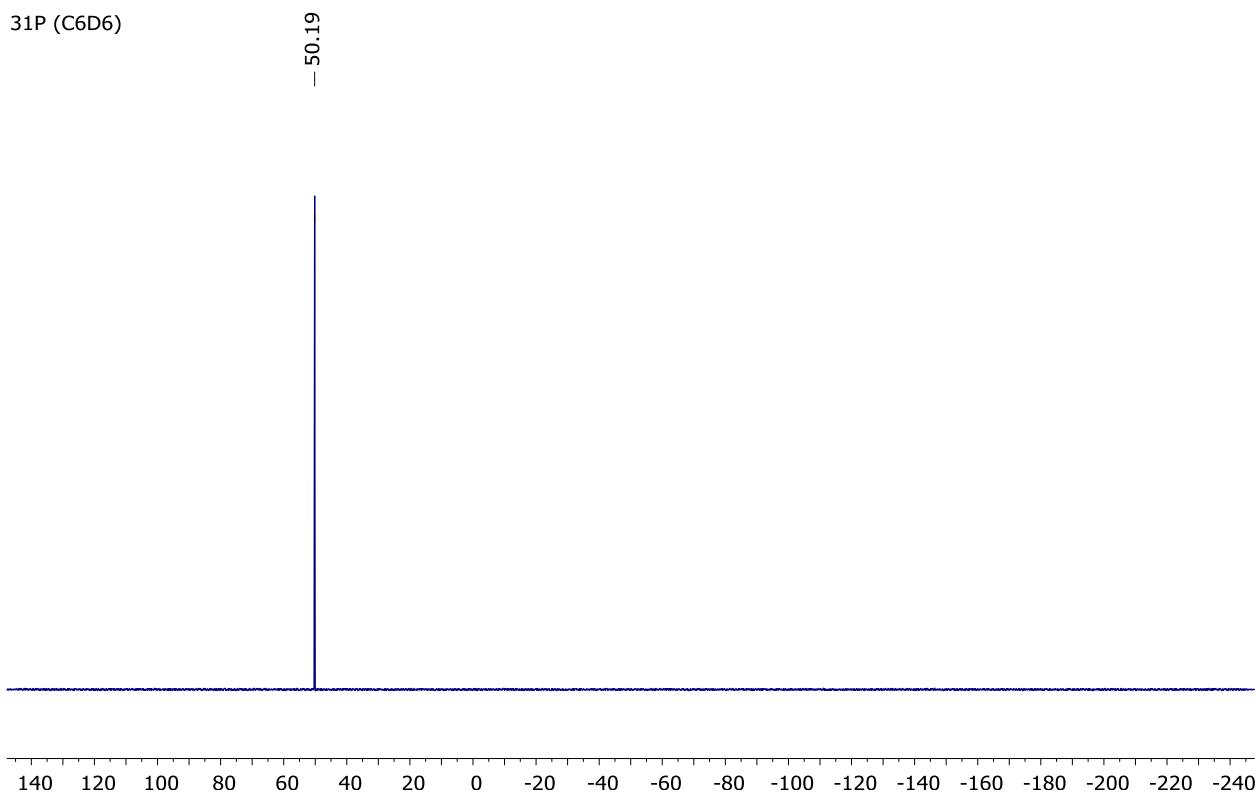
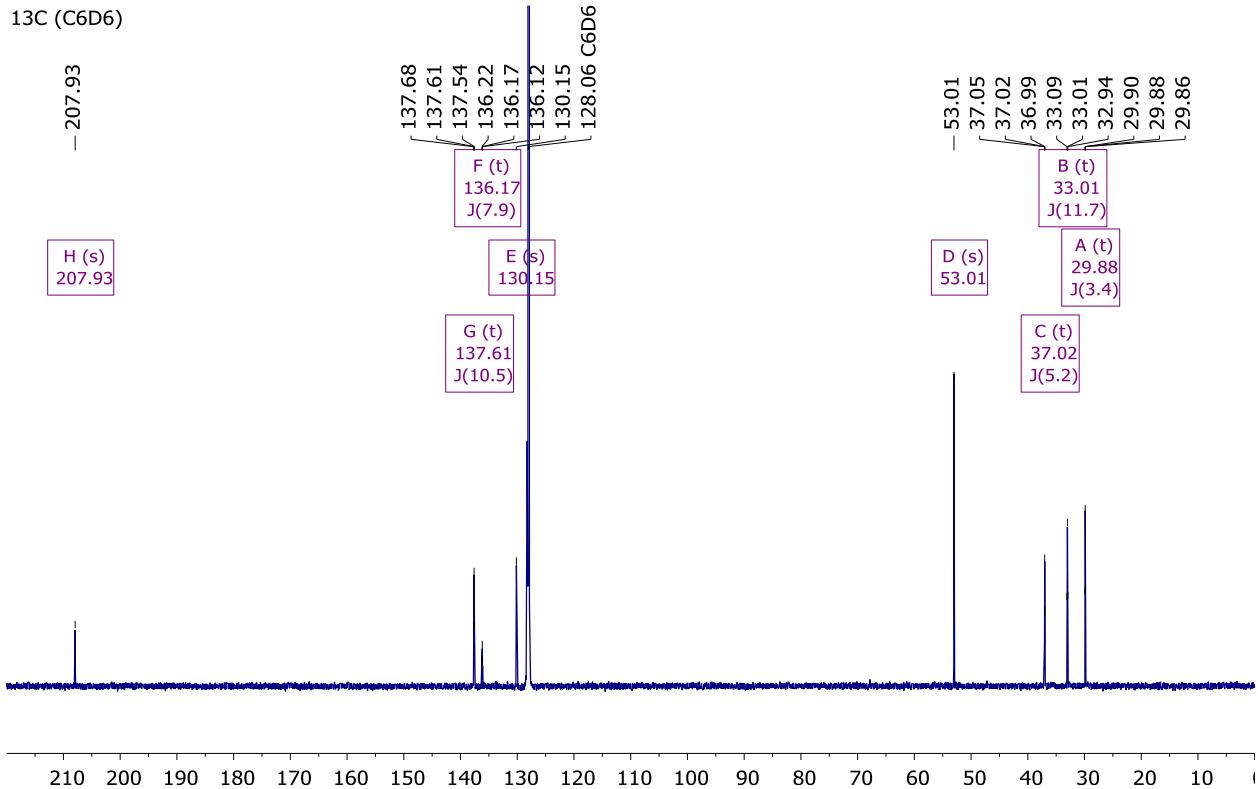
<sup>31</sup>P (C6D6)



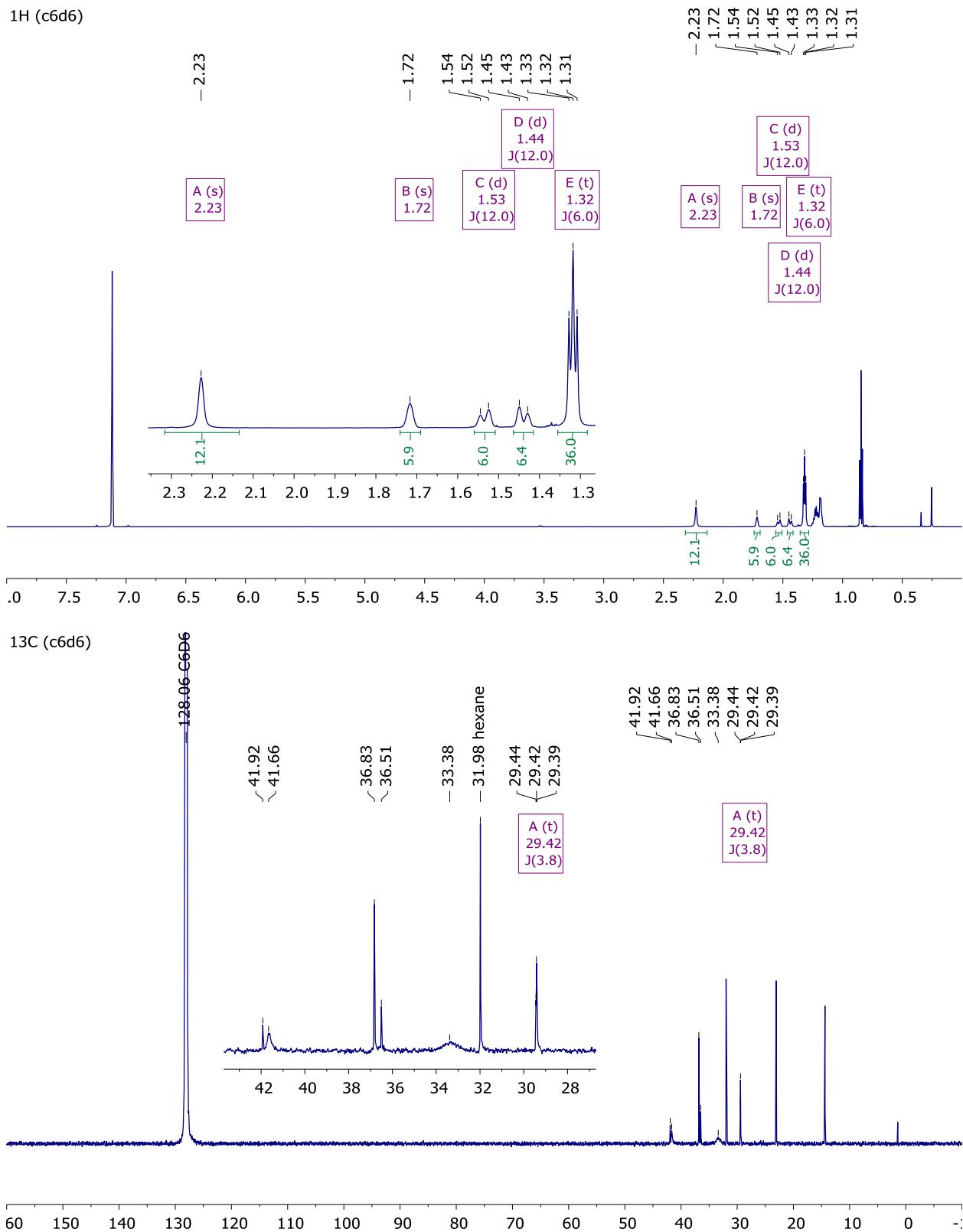
#### 4.7.6. Bis(2,2,6,6-tetramethyl-1-phenyl-4-phosphorinanone)palladium(0)

<sup>1</sup>H (C6D6)



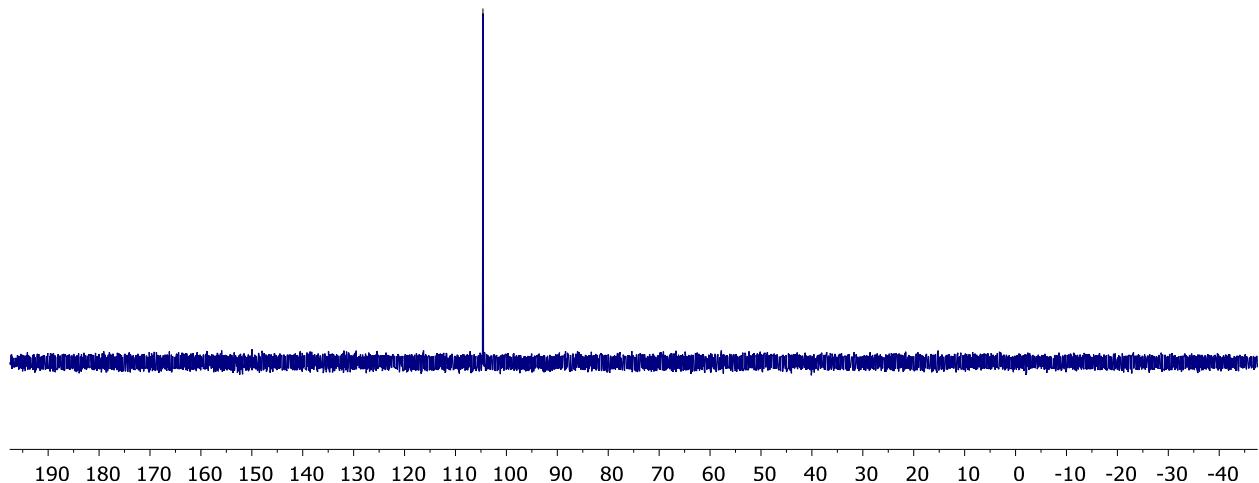


#### 4.7.7. Di- $\mu$ -iodobis(1-adamantyl-di-*tert*-butylphosphine)dipalladium(I)



31P (c6d6)

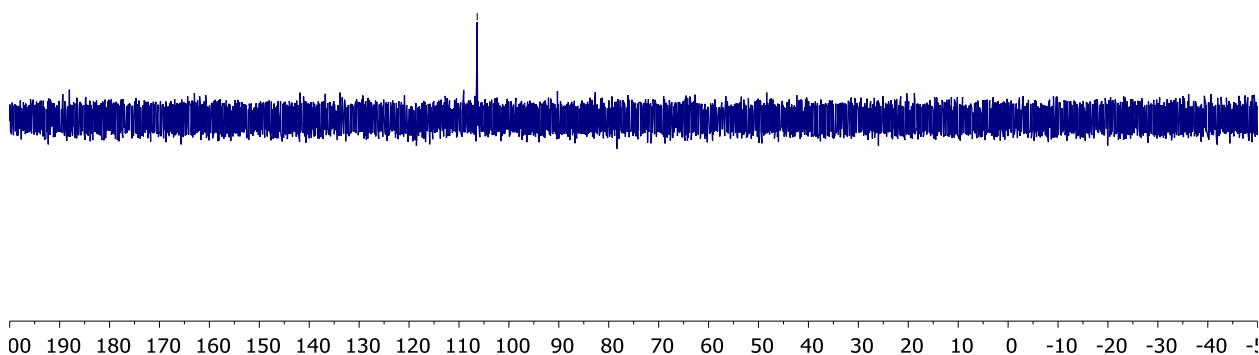
- 104.60



#### 4.7.8. Di- $\mu$ -iodobis(di-1-adamantyl-*tert*-butylphosphine)dipalladium(I)

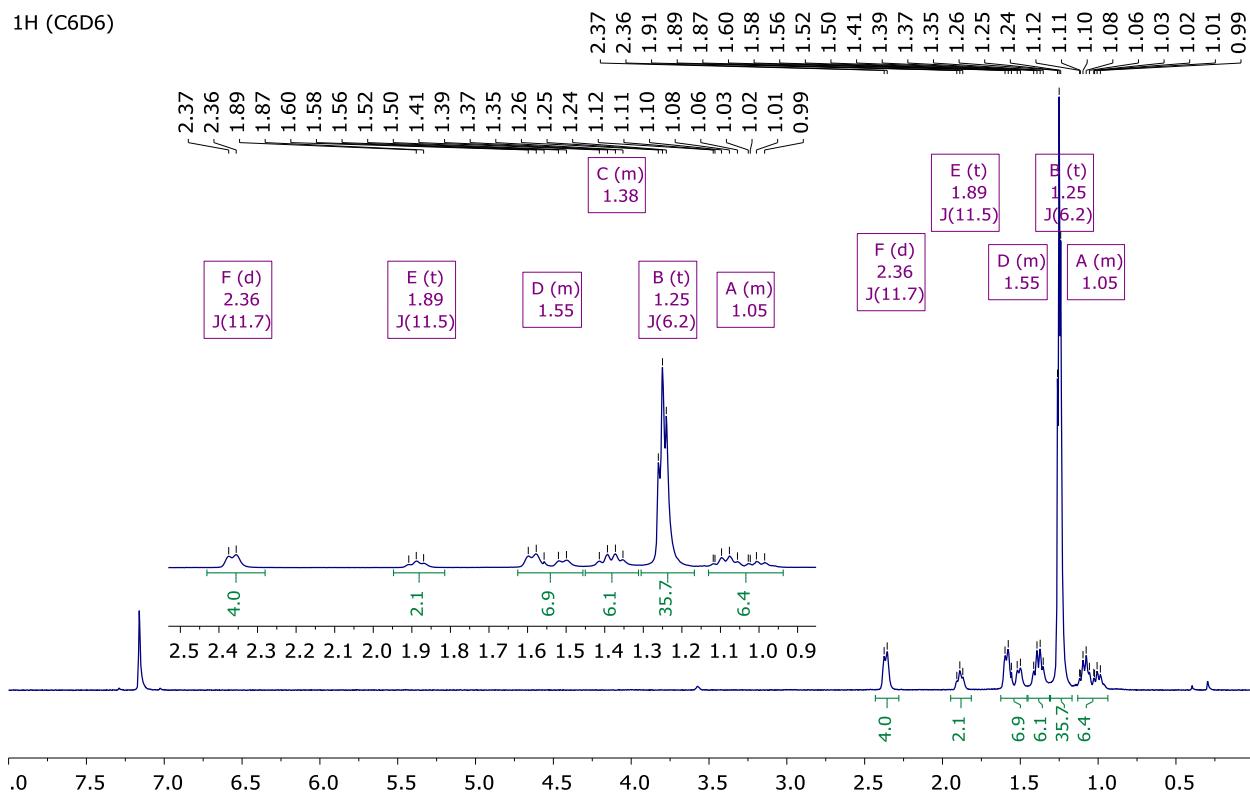
31P (c6d6)

- 106.36

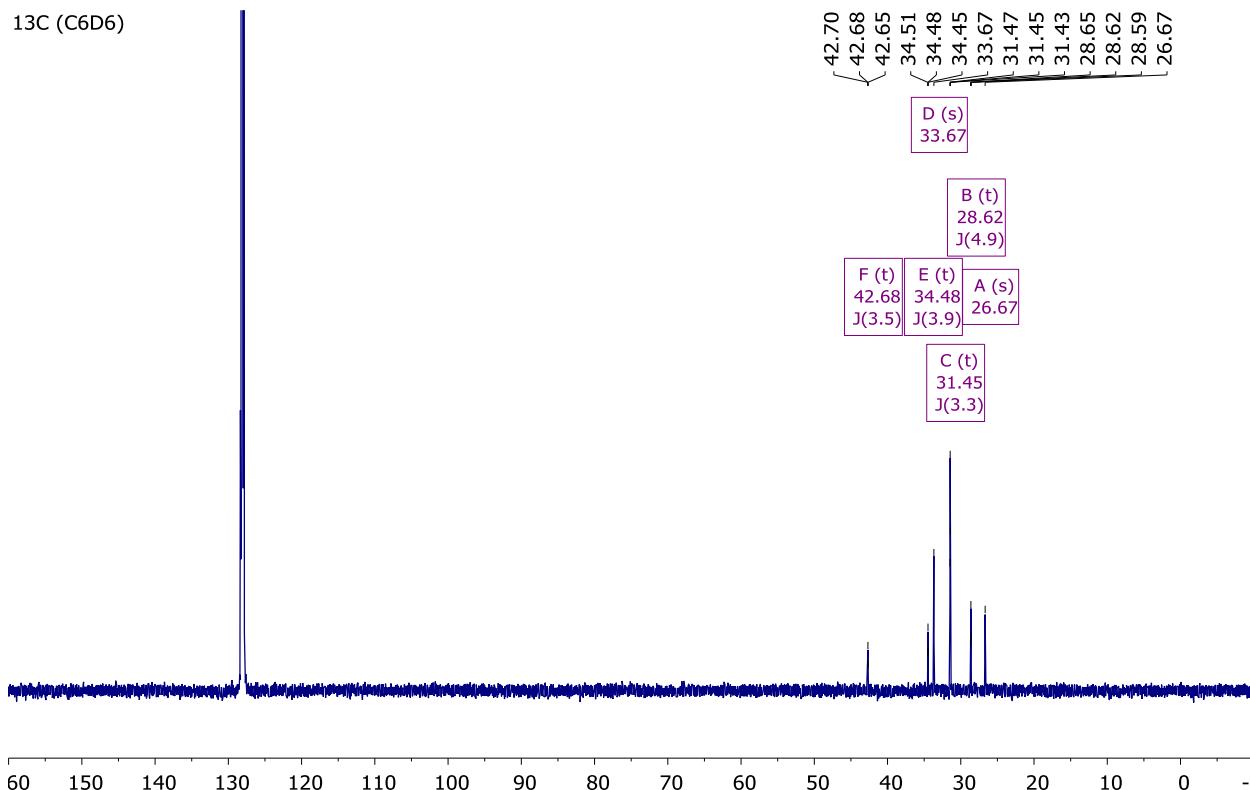


#### 4.7.9. Di- $\mu$ -iodobis(di-*tert*-butyl(cyclohexyl)phosphine)dipalladium(I):

<sup>1</sup>H (C6D6)

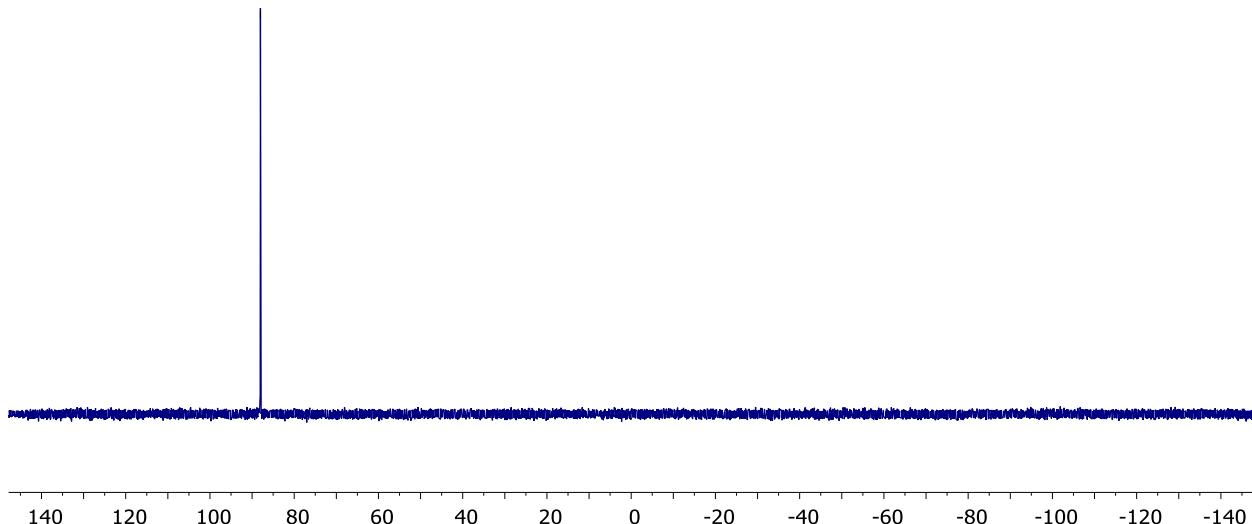


<sup>13</sup>C (C6D6)



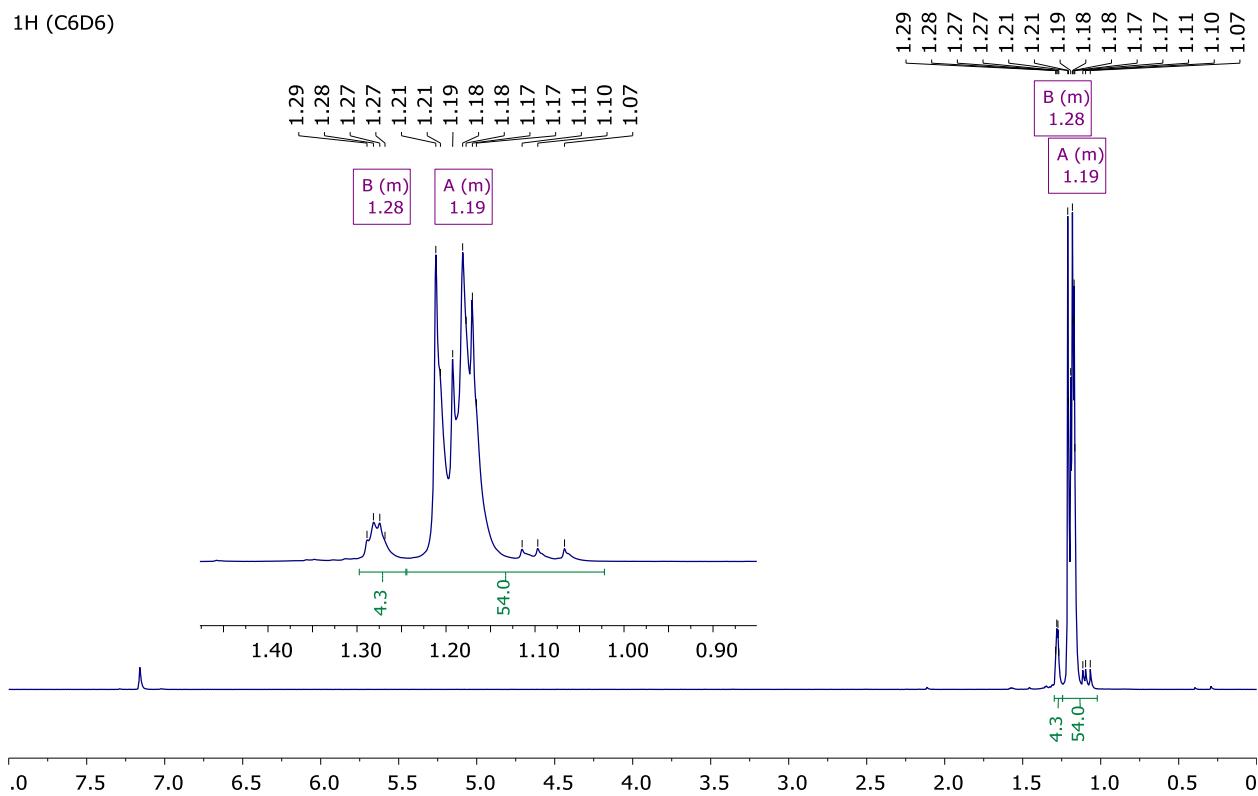
<sup>31</sup>P (C6D6)

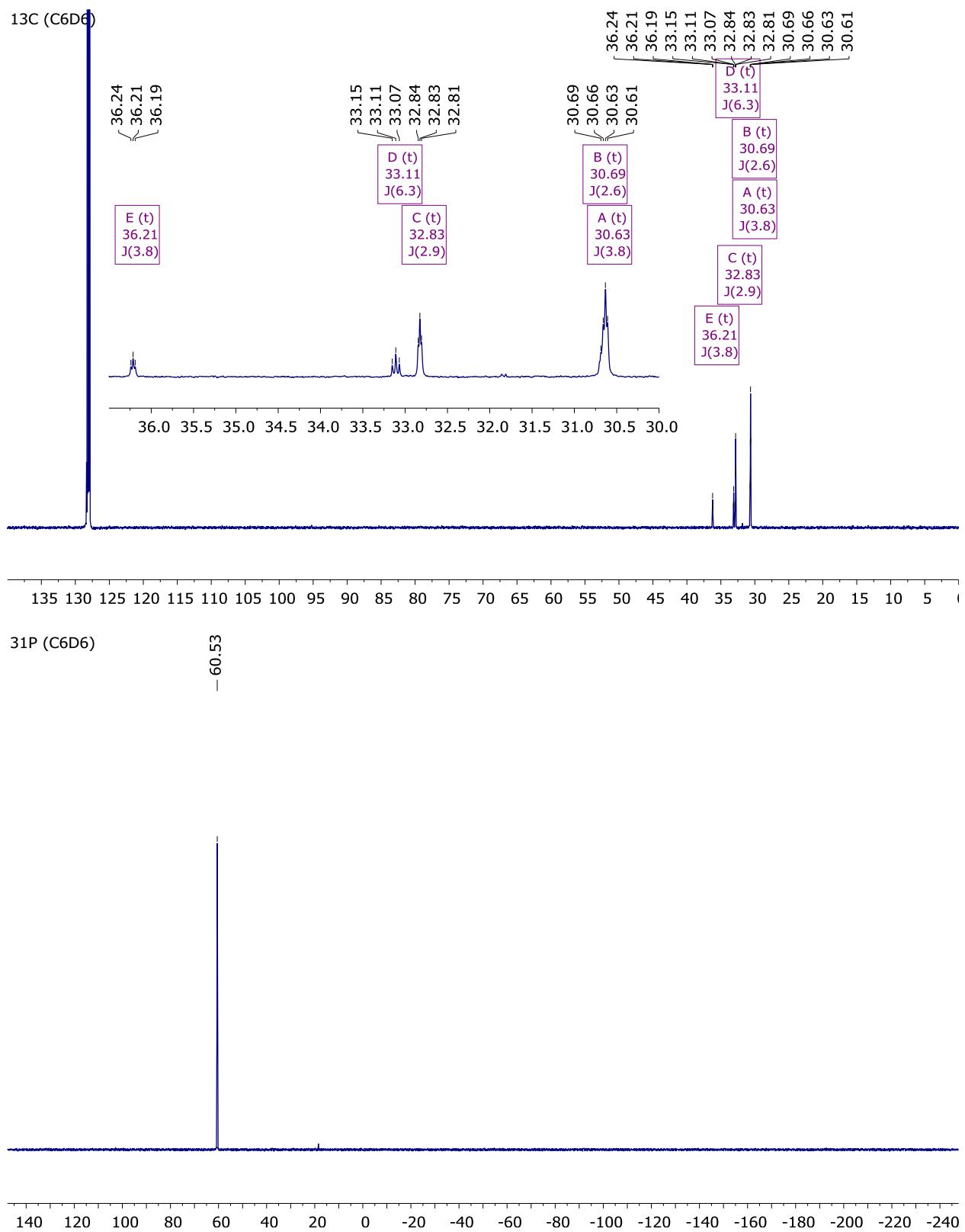
- 88.02



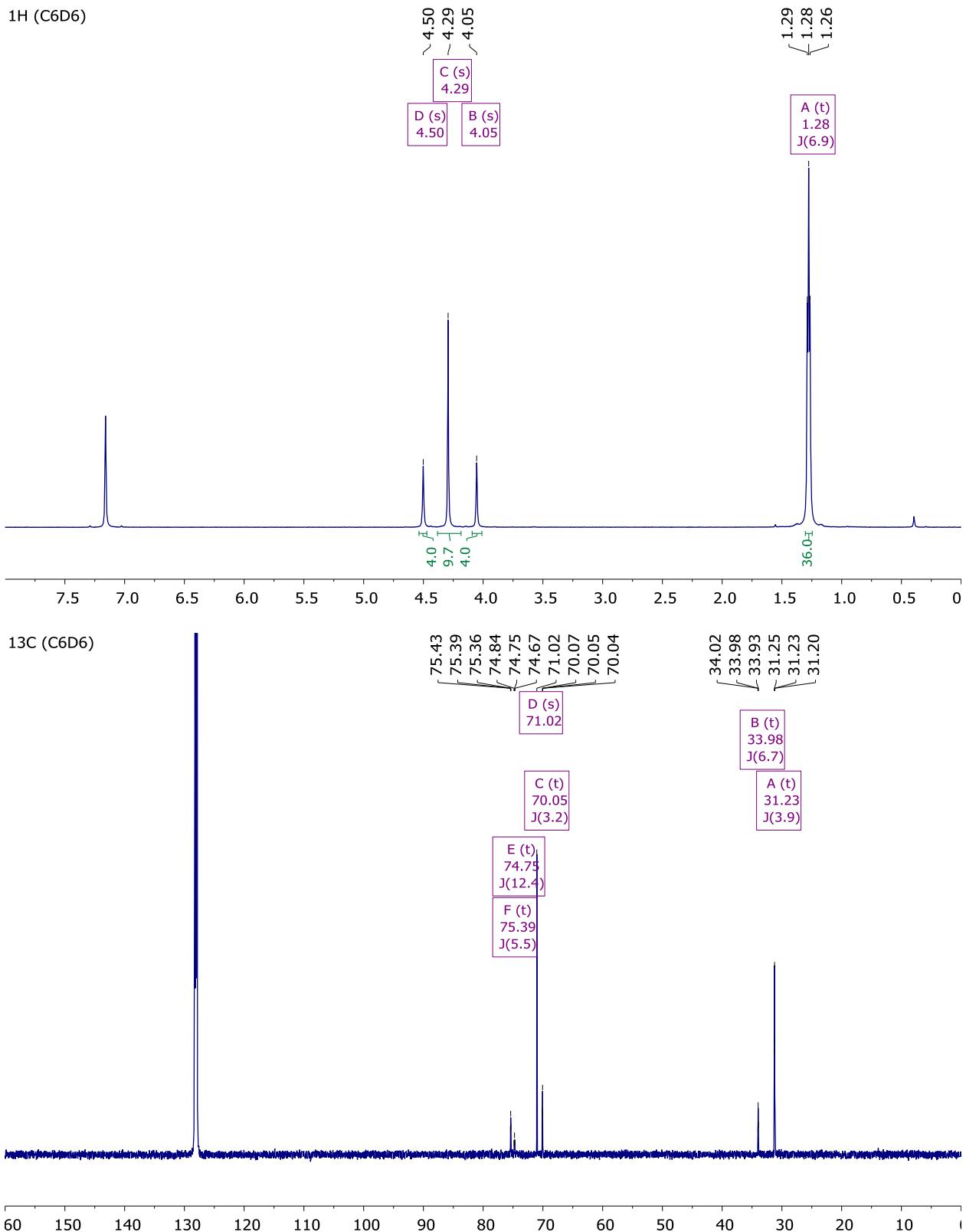
#### 4.7.10. Di- $\mu$ -iodobis(di-*tert*-butyl(neopentyl)phosphine)dipalladium(I)

<sup>1</sup>H (C6D6)

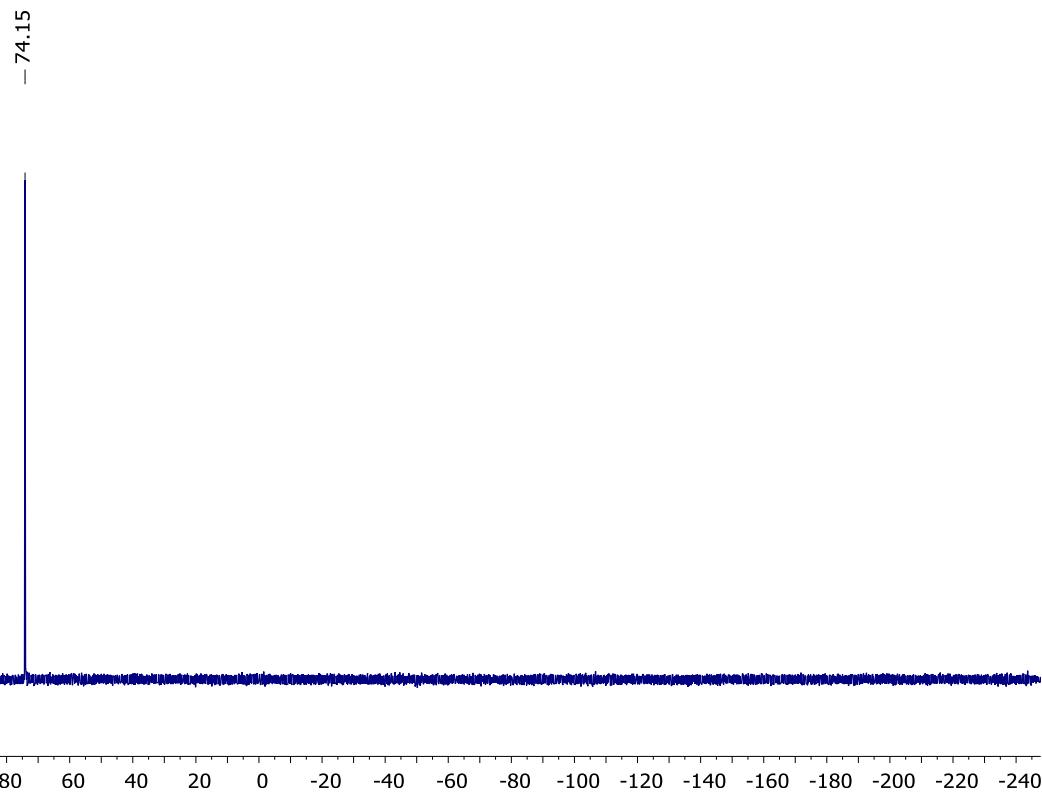




#### 4.7.11. Di- $\mu$ -iodobis(di-*tert*-butyl(ferrocenyl)phosphine)dipalladium(I)

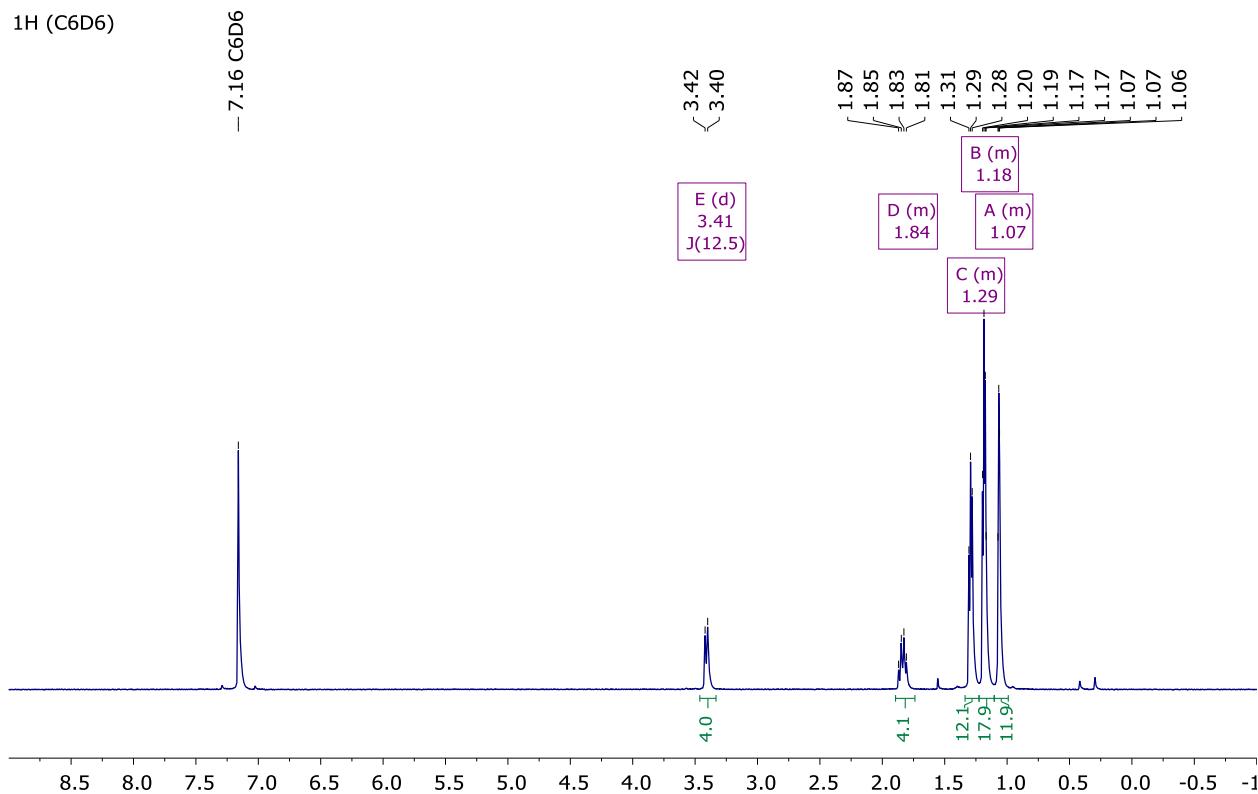


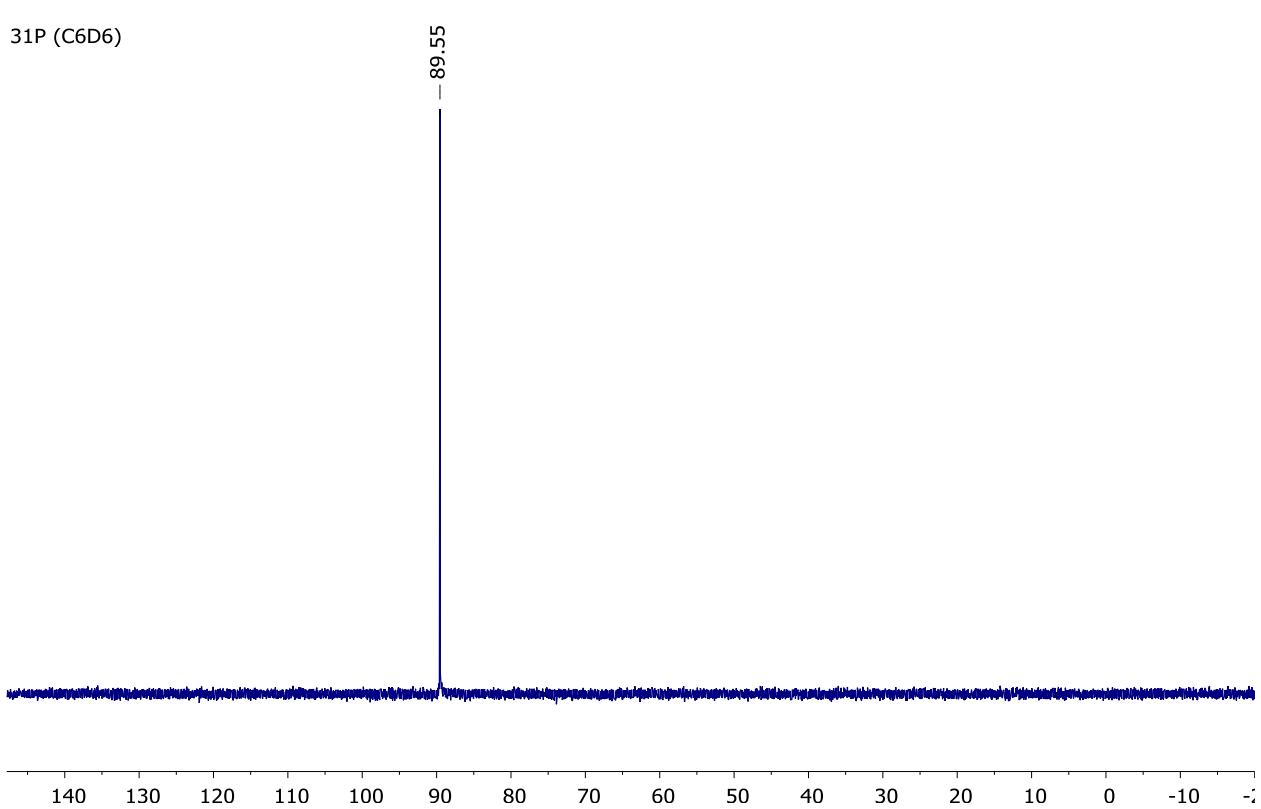
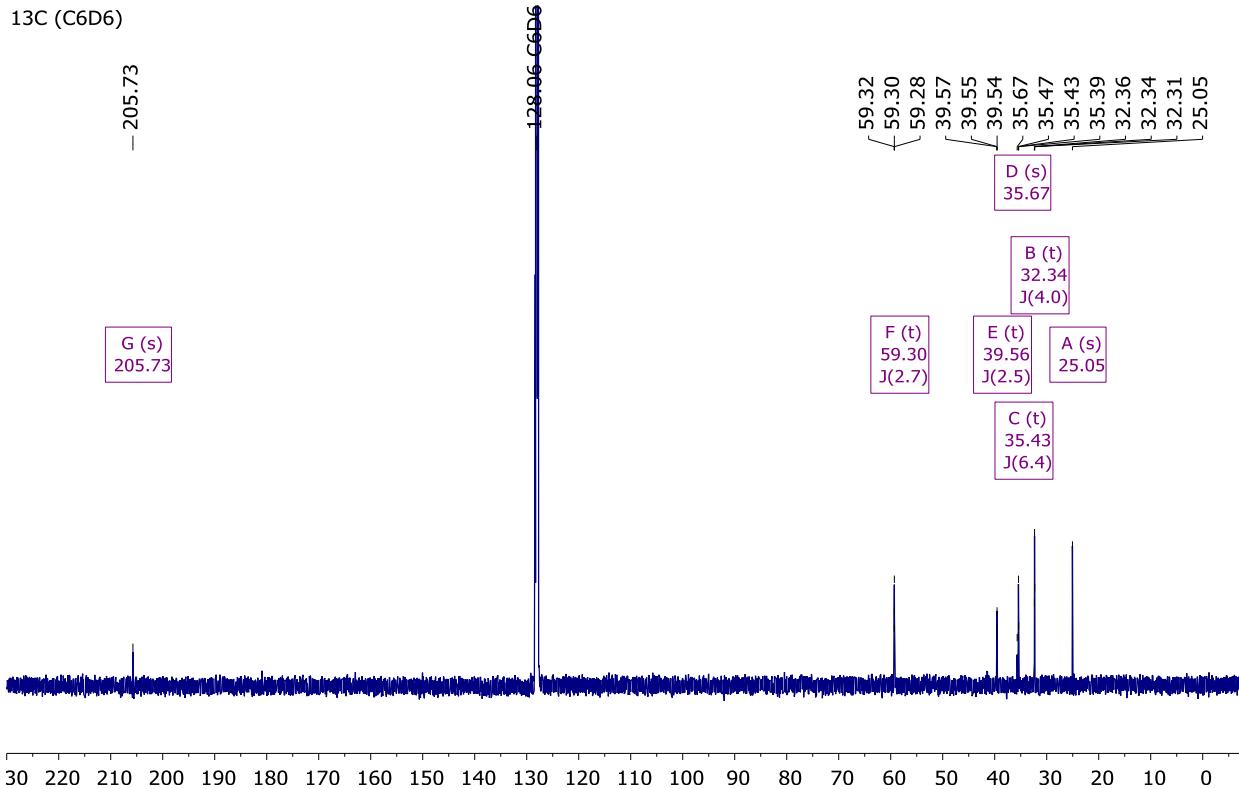
<sup>31</sup>P (C6D6)



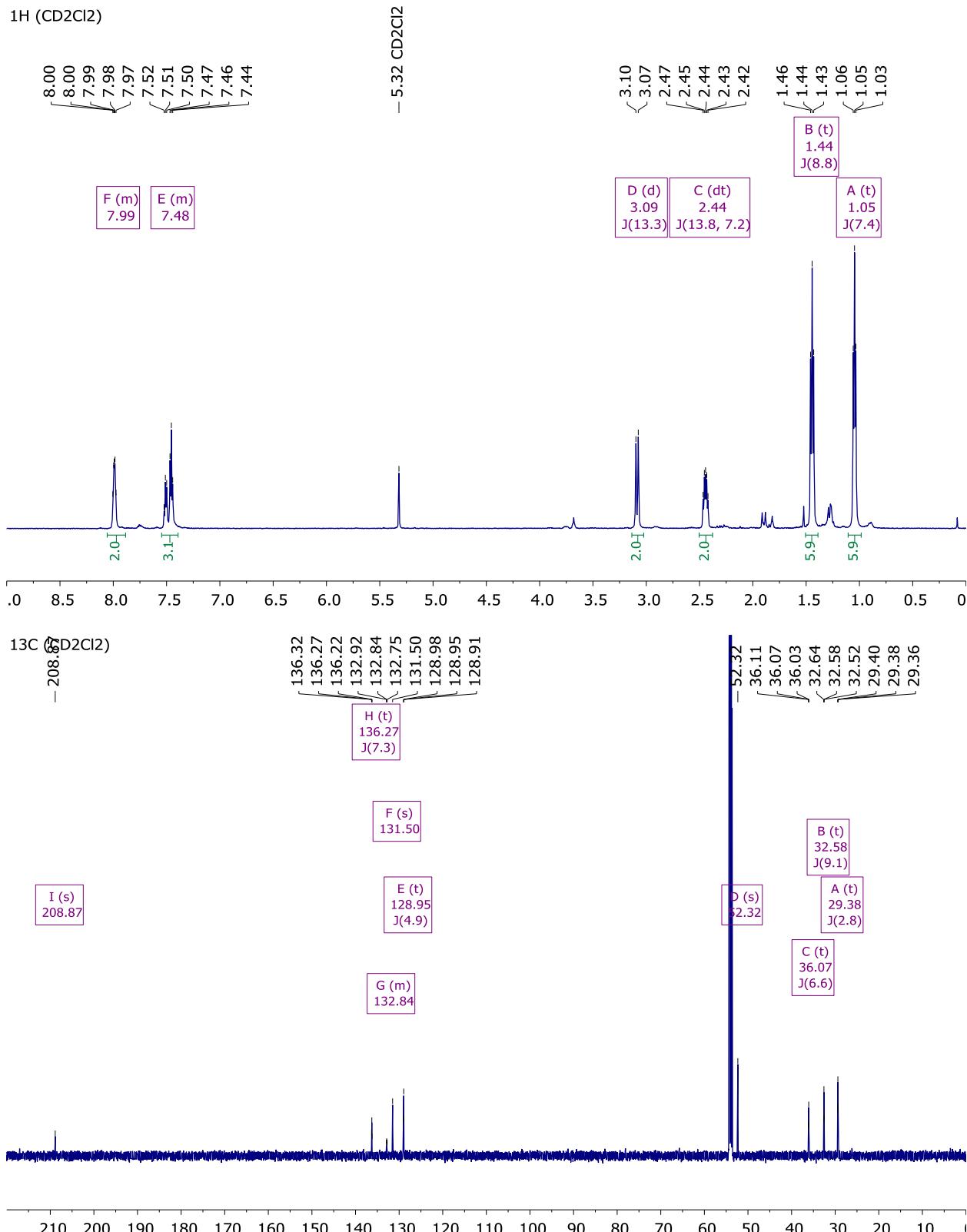
#### 4.7.12. Di- $\mu$ -iodobis(2,2,6,6-tetramethyl-1-(*tert*-butyl)-4-phosphorinanone)dipalladium(I)

<sup>1</sup>H (C6D6)



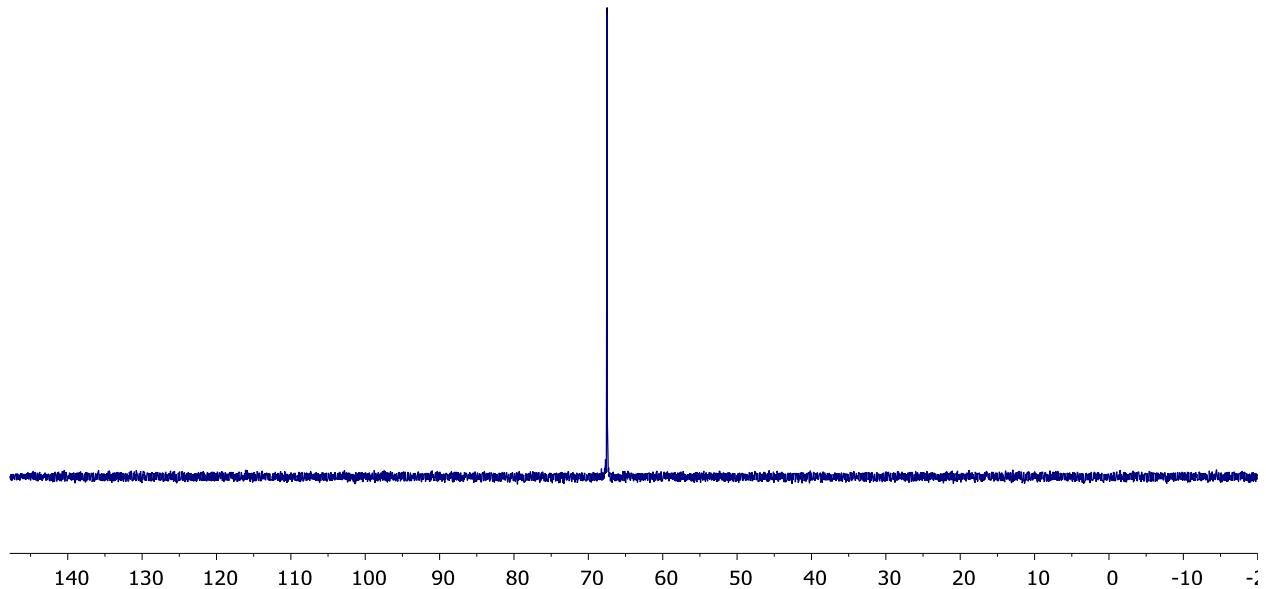


#### 4.7.13. Di- $\mu$ -iodobis(2,2,6,6-tetramethyl-1-phenyl-4-phosphorinanone)dipalladium(I)



<sup>31</sup>P (CD<sub>2</sub>Cl<sub>2</sub>)

-67.48



## **5. Captions for Data Files**

### **Data S1. (LKB-P\_clusters.csv)**

Dataset for the initial clustering. Contains all descriptors of the LKB-P database by Fey and the eight resulting clusters of the initial clustering.

### **Data S2. (Dimer-LKB-P\_clusters.csv)**

Dataset for the problem-specific clustering. Contains the descriptors taken from the LKB-P, the newly generated descriptors as well as the six resulting clusters of the problem-specific clustering.

### **Data S3. (XYZ\_coordinates.pdf)**

Cartesian coordinates for all species calculated during the problem-specific data acquisition, as well as the resulting energies and thermochemistry data.

## **6. Captions for Movie Files**

### **Movie S1. (cluster\_animation.mp4)**

3D animation of the problem-specific clusters within the spaces spanned by the problem-specific principal components (PC1', PC2' and PC3').

## References and Notes

1. J. F. Hartwig, *Organotransition Metal Chemistry: From Bonding to Catalysis* (University Science Books, New York, 2010).
2. F. Barrios-Landeros, B. P. Carrow, J. F. Hartwig, Autocatalytic oxidative addition of PhBr to Pd(P'Bu<sub>3</sub>)<sub>2</sub> via Pd(P'Bu<sub>3</sub>)<sub>2</sub>(H)(Br). *J. Am. Chem. Soc.* **130**, 5842–5843 (2008).  
[doi:10.1021/ja711159y](https://doi.org/10.1021/ja711159y) [Medline](#)
3. C. J. Diehl, T. Scattolin, U. Englert, F. Schoenebeck, C-I-Selective Cross-Coupling Enabled by a Cationic Palladium Trimer. *Angew. Chem. Int. Ed.* **58**, 211–215 (2019).  
[doi:10.1002/anie.201811380](https://doi.org/10.1002/anie.201811380) [Medline](#)
4. M.-E. Moret, R. J. M. Gebbink, Non-Noble Metal Catalysis: Molecular Approaches and Reactions (Wiley-VCH, Weinheim, Germany, 2019).
5. C. A. Tolman, Steric effects of phosphorus ligands in organometallic chemistry and homogeneous catalysis. *Chem. Rev.* **77**, 313–348 (1977). [doi:10.1021/cr60307a002](https://doi.org/10.1021/cr60307a002)
6. W. Strohmeier, F.-J. Müller, Klassifizierung phosphorhaltiger Liganden in Metallcarbonyl-Derivaten nach der  $\pi$ -Acceptorstärke. *Chem. Ber.* **100**, 2812–2821 (1967).  
[doi:10.1002/cber.19671000907](https://doi.org/10.1002/cber.19671000907)
7. C. A. Tolman, Electron donor–acceptor properties of phosphorus ligands. Substituent additivity. *J. Am. Chem. Soc.* **92**, 2953–2956 (1970). [doi:10.1021/ja00713a006](https://doi.org/10.1021/ja00713a006)
8. K. A. Bunten, L. Chen, A. L. Fernandez, A. J. Poë, Cone angles: Tolman's and Plato's. *Coord. Chem. Rev.* **233-234**, 41–51 (2002). [doi:10.1016/S0010-8545\(02\)00099-1](https://doi.org/10.1016/S0010-8545(02)00099-1)
9. Z. L. Niemeyer, A. Milo, D. P. Hickey, M. S. Sigman, Parameterization of phosphine ligands reveals mechanistic pathways and predicts reaction outcomes. *Nat. Chem.* **8**, 610–617 (2016). [doi:10.1038/nchem.2501](https://doi.org/10.1038/nchem.2501) [Medline](#)
10. D. J. Durand, N. Fey, Computational Ligand Descriptors for Catalyst Design. *Chem. Rev.* **119**, 6561–6594 (2019). [doi:10.1021/acs.chemrev.8b00588](https://doi.org/10.1021/acs.chemrev.8b00588) [Medline](#)
11. N. Fey, A. C. Tsipis, S. E. Harris, J. N. Harvey, A. G. Orpen, R. A. Mansson, Development of a ligand knowledge base, part 1: Computational descriptors for phosphorus donor ligands. *Chemistry* **12**, 291–302 (2005). [doi:10.1002/chem.200500891](https://doi.org/10.1002/chem.200500891) [Medline](#)
12. J. Jover, N. Fey, J. N. Harvey, G. C. Lloyd-Jones, A. G. Orpen, G. J. J. Owen-Smith, P. Murray, D. R. J. Hose, R. Osborne, M. Purdie, Expansion of the Ligand Knowledge Base for Monodentate P-Donor Ligands (LKB-P). *Organometallics* **29**, 6245–6258 (2010).  
[doi:10.1021/om100648v](https://doi.org/10.1021/om100648v)
13. See supplementary materials for additional details and definition of terminology.

14. C. C. C. Johansson Seechurn, M. O. Kitching, T. J. Colacot, V. Snieckus, Palladium-catalyzed cross-coupling: A historical contextual perspective to the 2010 Nobel Prize. *Angew. Chem. Int. Ed.* **51**, 5062–5085 (2012). [doi:10.1002/anie.201107017](https://doi.org/10.1002/anie.201107017) [Medline](#)
15. C. Fricke, T. Sperger, M. Mendel, F. Schoenebeck, Catalysis with Palladium(I) Dimers. *Angew. Chem. Int. Ed.* **60**, 3355–3366 (2021). [doi:10.1002/anie.202011825](https://doi.org/10.1002/anie.202011825) [Medline](#)
16. J. P. Stambuli, R. Kuwano, J. F. Hartwig, Unparalleled rates for the activation of aryl chlorides and bromides: Coupling with amines and boronic acids in minutes at room temperature. *Angew. Chem. Int. Ed.* **41**, 4746–4748 (2002). [doi:10.1002/anie.200290036](https://doi.org/10.1002/anie.200290036) [Medline](#)
17. D. M. Ohlmann, N. Tschauder, J.-P. Stockis, K. Goossen, M. Dierker, L. J. Goossen, Isomerizing olefin metathesis as a strategy to access defined distributions of unsaturated compounds from fatty acids. *J. Am. Chem. Soc.* **134**, 13716–13729 (2012). [doi:10.1021/ja303822c](https://doi.org/10.1021/ja303822c) [Medline](#)
18. X.-Y. Chen, M. Pu, H.-G. Cheng, T. Sperger, F. Schoenebeck, Arylation of Axially Chiral Phosphorothioate Salts by Dinuclear Pd<sup>I</sup> Catalysis. *Angew. Chem. Int. Ed.* **58**, 11395–11399 (2019). [doi:10.1002/anie.201906063](https://doi.org/10.1002/anie.201906063) [Medline](#)
19. S. T. Keaveney, G. Kundu, F. Schoenebeck, Modular Functionalization of Arenes in a Triply Selective Sequence: Rapid C(sp<sup>2</sup>) and C(sp<sup>3</sup>) Coupling of C-Br, C-OTf, and C-Cl Bonds Enabled by a Single Palladium(I) Dimer. *Angew. Chem. Int. Ed.* **57**, 12573–12577 (2018). [doi:10.1002/anie.201808386](https://doi.org/10.1002/anie.201808386) [Medline](#)
20. Calculated at the CPCM (toluene) B3LYP-D3(BJ)/6-311++G(d,p)(SDD for Pd, Fe, I) //B3LYP-D3(BJ)/6-31G(d) (SDD for Pd, Fe, I) level of theory.
21. F. Proutiere, E. Lyngvi, M. Aufiero, I. A. Sanhueza, F. Schoenebeck, Combining the Reactivity Properties of PCy<sub>3</sub> and PtBu<sub>3</sub> into a Single Ligand, P(iPr)(tBu)<sub>2</sub>. Reaction via Mono- or Bisphosphine Palladium(0) Centers and Palladium(I) Dimer Formation. *Organometallics* **33**, 6879–6884 (2014). [doi:10.1021/om5009605](https://doi.org/10.1021/om5009605)
22. J. B. O. Mitchell, Machine learning methods in chemoinformatics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 468–481 (2014). [doi:10.1002/wcms.1183](https://doi.org/10.1002/wcms.1183) [Medline](#)
23. J. P. Reid, M. S. Sigman, Holistic prediction of enantioselectivity in asymmetric catalysis. *Nature* **571**, 343–348 (2019). [doi:10.1038/s41586-019-1384-z](https://doi.org/10.1038/s41586-019-1384-z) [Medline](#)
24. A. R. Rosales, J. Wahlers, E. Limé, R. E. Meadows, K. W. Leslie, R. Savin, F. Bell, E. Hansen, P. Helquist, R. H. Munday, O. Wiest, P.-O. Norrby, Rapid virtual screening of enantioselective catalysts using CatVS. *Nat. Catal.* **2**, 41–45 (2019). [doi:10.1038/s41929-018-0193-3](https://doi.org/10.1038/s41929-018-0193-3)
25. A. Milo, A. J. Neel, F. D. Toste, M. S. Sigman, Organic chemistry. A data-intensive approach to mechanistic elucidation applied to chiral anion catalysis. *Science* **347**, 737–743 (2015). [doi:10.1126/science.1261043](https://doi.org/10.1126/science.1261043) [Medline](#)
26. A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow, S. E. Denmark, Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science* **363**, eaau5631 (2019). [doi:10.1126/science.aau5631](https://doi.org/10.1126/science.aau5631) [Medline](#)

27. Y. Amar, A. M. Schweidtmann, P. Deutsch, L. Cao, A. Lapkin, Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis. *Chem. Sci.* **10**, 6697–6706 (2019). [doi:10.1039/C9SC01844A](https://doi.org/10.1039/C9SC01844A) [Medline](#)
28. H. Gao, T. J. Struble, C. W. Coley, Y. Wang, W. H. Green, K. F. Jensen, Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **4**, 1465–1476 (2018). [doi:10.1021/acscentsci.8b00357](https://doi.org/10.1021/acscentsci.8b00357) [Medline](#)
29. D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, A. G. Doyle, Predicting reaction performance in C-N cross-coupling using machine learning. *Science* **360**, 186–190 (2018). [doi:10.1126/science.aar5169](https://doi.org/10.1126/science.aar5169) [Medline](#)
30. C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, K. F. Jensen, Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **3**, 434–443 (2017). [doi:10.1021/acscentsci.7b00064](https://doi.org/10.1021/acscentsci.7b00064) [Medline](#)
31. M. Foscato, V. R. Jensen, Automated in Silico Design of Homogeneous Catalysts. *ACS Catal.* **10**, 2354–2377 (2020). [doi:10.1021/acscatal.9b04952](https://doi.org/10.1021/acscatal.9b04952)
32. T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K.-i. Shimizu, Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **10**, 2260–2297 (2020). [doi:10.1021/acscatal.9b04186](https://doi.org/10.1021/acscatal.9b04186)
33. I. Funes-Ardoiz, F. Schoenebeck, Established and Emerging Computational Tools to Study Homogeneous Catalysis - From Quantum Mechanics to Machine Learning. *Chem* **6**, 1904–1913 (2020). [doi:10.1016/j.chempr.2020.07.008](https://doi.org/10.1016/j.chempr.2020.07.008)
34. G. dos Passos Gomes, R. Pollice, A. Aspuru-Guzik, Navigating through the Maze of Homogeneous Catalyst Design with Machine Learning. *Trends Chem.* **3**, 96–110 (2021). [doi:10.1016/j.trechm.2020.12.006](https://doi.org/10.1016/j.trechm.2020.12.006)
35. A. V. Brethomé, S. P. Fletcher, R. S. Paton, Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* **9**, 2313–2323 (2019). [doi:10.1021/acscatal.8b04043](https://doi.org/10.1021/acscatal.8b04043)
36. S. G. Newman, M. Lautens, Palladium-catalyzed carboiodination of alkenes: Carbon-carbon bond formation with retention of reactive functionality. *J. Am. Chem. Soc.* **133**, 1778–1780 (2011). [doi:10.1021/ja110377q](https://doi.org/10.1021/ja110377q) [Medline](#)
37. K. O. Kirlikovali, E. Cho, T. J. Downard, L. Grigoryan, Z. Han, S. Hong, D. Jung, J. C. Quintana, V. Reynoso, S. Ro, Y. Shen, K. Swartz, E. Ter Sahakyan, A. I. Wixtrom, B. Yoshida, A. L. Rheingold, A. M. Spokoyny, Buchwald-Hartwig amination using Pd(i) dimer precatalysts supported by biaryl phosphine ligands. *Dalton Trans.* **47**, 3684–3688 (2018). [doi:10.1039/C8DT00119G](https://doi.org/10.1039/C8DT00119G) [Medline](#)
38. Ligand 20 that was identified in our study gave a successful dimer also which we meanwhile applied in olefin isomerization; see (39).
39. G. Kundu, T. Sperger, K. Rissanen, F. Schoenebeck, A Next-Generation Air-Stable Palladium(I) Dimer Enables Olefin Migration and Selective C-C Coupling in Air. *Angew. Chem. Int. Ed.* **59**, 21930–21934 (2020). [doi:10.1002/anie.202009115](https://doi.org/10.1002/anie.202009115) [Medline](#)
40. J. A. Hueffel, J-Hueffel/PdDimer: PdDimer v1.0, Zenodo (2021); <https://doi.org/10.5281/zenodo.5541842>.

41. H. Abdi, L. J. Williams, Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2**, 433–459 (2010). [doi:10.1002/wics.101](https://doi.org/10.1002/wics.101)
42. S. Lloyd, Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982). [doi:10.1109/TIT.1982.1056489](https://doi.org/10.1109/TIT.1982.1056489)
43. D. Arthur, S. Vassilvitskii, k-means++: the advantages of careful seeding, in *Proceedings of the Eighteenth Annual ACM-SIAM symposium on Discrete Algorithms* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2007), pp. 1027–1035.
44. A. Smellie, Accelerated K-means clustering in metric spaces. *J. Chem. Inf. Comput. Sci.* **44**, 1929–1935 (2004). [doi:10.1021/ci0499222](https://doi.org/10.1021/ci0499222) [Medline](#)
45. A. P. P. Bholowalia, EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *Int. J. Comput. Appl.* **105**, 17–24 (2014).
46. C. Yuan, H. Yang, Research on K-Value Selection Method of K-Means Clustering Algorithm. *J. Multidiscip. Res.* **2**, 226–235 (2019). [doi:10.3390/j2020016](https://doi.org/10.3390/j2020016)
47. P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987). [doi:10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
48. M. J. Frisch *et al.*, *Gaussian 16, Revision A.03* (Gaussian, Inc., Wallingford CT, 2016).
49. A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993). [doi:10.1063/1.464913](https://doi.org/10.1063/1.464913)
50. C. Lee, W. Yang, R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter* **37**, 785–789 (1988). [doi:10.1103/PhysRevB.37.785](https://doi.org/10.1103/PhysRevB.37.785) [Medline](#)
51. P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **98**, 11623–11627 (1994). [doi:10.1021/j100096a001](https://doi.org/10.1021/j100096a001)
52. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104–154121 (2010). [doi:10.1063/1.3382344](https://doi.org/10.1063/1.3382344) [Medline](#)
53. A. E. Reed, R. B. Weinstock, F. Weinhold, Natural population analysis. *J. Chem. Phys.* **83**, 735–746 (1985). [doi:10.1063/1.449486](https://doi.org/10.1063/1.449486)
54. A. E. Reed, F. Weinhold, Natural localized molecular orbitals. *J. Chem. Phys.* **83**, 1736–1740 (1985). [doi:10.1063/1.449360](https://doi.org/10.1063/1.449360)
55. A. E. Reed, F. Weinhold, Natural bond orbital analysis of near-Hartree–Fock water dimer. *J. Chem. Phys.* **78**, 4066–4073 (1983). [doi:10.1063/1.445134](https://doi.org/10.1063/1.445134)
56. A. E. Reed, L. A. Curtiss, F. Weinhold, Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **88**, 899–926 (1988). [doi:10.1021/cr00088a005](https://doi.org/10.1021/cr00088a005)
57. F. Weinhold, J. E. Carpenter, "The Natural Bond Orbital Lewis Structure Concept for Molecules, Radicals, and Radical Ions" in *The Structure of Small Molecules and Ions*, R. Naaman, Z. Vager, Eds. (Springer, Boston, MA, 1988), pp. 227–236.

58. J. E. Carpenter, "Extension of Lewis structure concepts to open-shell and excited-state molecular species," thesis, University of Wisconsin, Madison, WI (1987).
59. J. P. Foster, F. Weinhold, Natural hybrid orbitals. *J. Am. Chem. Soc.* **102**, 7211–7218 (1980). [doi:10.1021/ja00544a007](https://doi.org/10.1021/ja00544a007)
60. J. E. Carpenter, F. Weinhold, Analysis of the geometry of the hydroxymethyl radical by the “different hybrids for different spins” natural bond orbital procedure. *J. Mol. Struct. THEOCHEM* **169**, 41–62 (1988). [doi:10.1016/0166-1280\(88\)80248-3](https://doi.org/10.1016/0166-1280(88)80248-3)
61. N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open Babel: An open chemical toolbox. *J. Cheminform.* **3**, 33 (2011). [doi:10.1186/1758-2946-3-33](https://doi.org/10.1186/1758-2946-3-33) [Medline](#)
62. Paton Lab, K. Jackson, Bobbypaton/Sterimol: Sterimol Version 1.0, Zenodo (2018); <https://doi.org/10.5281/zenodo.1320768>.
63. The pandas development team, pandas-dev/pandas: Pandas 1.0.3, Zenodo (2020); <https://doi.org/10.5281/zenodo.3715232>.
64. F. Pedregosa *et al.*, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
65. Wolfram Research Inc, *Mathematica, Version 12.2* (Champaign, IL, 2020).
66. J. A. Bilbrey, A. H. Kazeez, J. Locklin, W. D. Allen, Exact ligand cone angles. *J. Comput. Chem.* **34**, 1189–1197 (2013). [doi:10.1002/jcc.23217](https://doi.org/10.1002/jcc.23217) [Medline](#)
67. A. K. Chakraborti, L. Sharma, R. Gulhane, Shivani, Shivani, Electrostatic catalysis by ionic aggregates: Scope and limitations of Mg(ClO<sub>4</sub>)<sub>2</sub> as acylation catalyst. *Tetrahedron* **59**, 7661–7668 (2003). [doi:10.1016/j.tet.2003.08.007](https://doi.org/10.1016/j.tet.2003.08.007)
68. Z. Liu, Q. Ma, Y. Liu, Q. Wang, 4-(*N,N*-Dimethylamino)pyridine hydrochloride as a recyclable catalyst for acylation of inert alcohols: Substrate scope and reaction mechanism. *Org. Lett.* **16**, 236–239 (2014). [doi:10.1021/o14030875](https://doi.org/10.1021/o14030875) [Medline](#)
69. L. Chen, P. Ren, B. P. Carrow, Tri(1-adamantyl)phosphine: Expanding the Boundary of Electron-Releasing Character Available to Organophosphorus Compounds. *J. Am. Chem. Soc.* **138**, 6392–6395 (2016). [doi:10.1021/jacs.6b03215](https://doi.org/10.1021/jacs.6b03215) [Medline](#)
70. J. P. Stambuli, S. R. Stauffer, K. H. Shaughnessy, J. F. Hartwig, Screening of homogeneous catalysts by fluorescence resonance energy transfer. Identification of catalysts for room-temperature Heck reactions. *J. Am. Chem. Soc.* **123**, 2677–2678 (2001). [doi:10.1021/ja0058435](https://doi.org/10.1021/ja0058435) [Medline](#)
71. R. Doherty, M. F. Haddow, Z. A. Harrison, A. G. Orpen, P. G. Pringle, A. Turner, R. L. Wingad, Bulky 4-phosphacyclohexanones: Diastereoselective complexations, orthometallations and unprecedented [3.1.1]metallabicycles. *Dalton Trans.* **0**, 4310–4320 (2006). [doi:10.1039/b607490a](https://doi.org/10.1039/b607490a) [Medline](#)
72. G. Mann, Q. Shelby, A. H. Roy, J. F. Hartwig, Electronic and Steric Effects on the Reductive Elimination of Diaryl Ethers from Palladium(II). *Organometallics* **22**, 2775–2789 (2003). [doi:10.1021/om030230x](https://doi.org/10.1021/om030230x)
73. R. W. W. Hooft, *COLLECT* (Nonius BV, 1998).

74. Z. Otwinowski, W. Minor, “Processing of X-ray diffraction data collected in oscillation mode” in *Methods in Enzymology, Macromolecular Crystallography, Part A*, C. W. Carter, R. M. Sweet, Eds. (Academic Press, 1997), vol. 276, pp. 307–326.
75. A. X. S. Bruker, Inc., *SAINT* (Madison, WI, 2017).
76. G. M. Sheldrick, *SADABS, Version 2008/2* (University of Göttingen, Germany, 1996).
77. Rigaku Oxford Diffraction, *CrysAlisPro, Version 38.46* (Rigaku Corp., Oxford, UK, 2018).
78. G. M. Sheldrick, SHELXT - integrated space-group and crystal-structure determination. *Acta Crystallogr. A Found. Adv.* **71**, 3–8 (2015). [doi:10.1107/S2053273314026370](https://doi.org/10.1107/S2053273314026370) [Medline](#)
79. O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard, H. Puschmann, OLEX2: A complete structure solution, refinement and analysis program. *J. Appl. Cryst.* **42**, 339–341 (2009). [doi:10.1107/S0021889808042726](https://doi.org/10.1107/S0021889808042726)
80. G. M. Sheldrick, Crystal structure refinement with SHELXL. *Acta Crystallogr. C Struct. Chem.* **71**, 3–8 (2015). [doi:10.1107/S2053229614024218](https://doi.org/10.1107/S2053229614024218) [Medline](#)