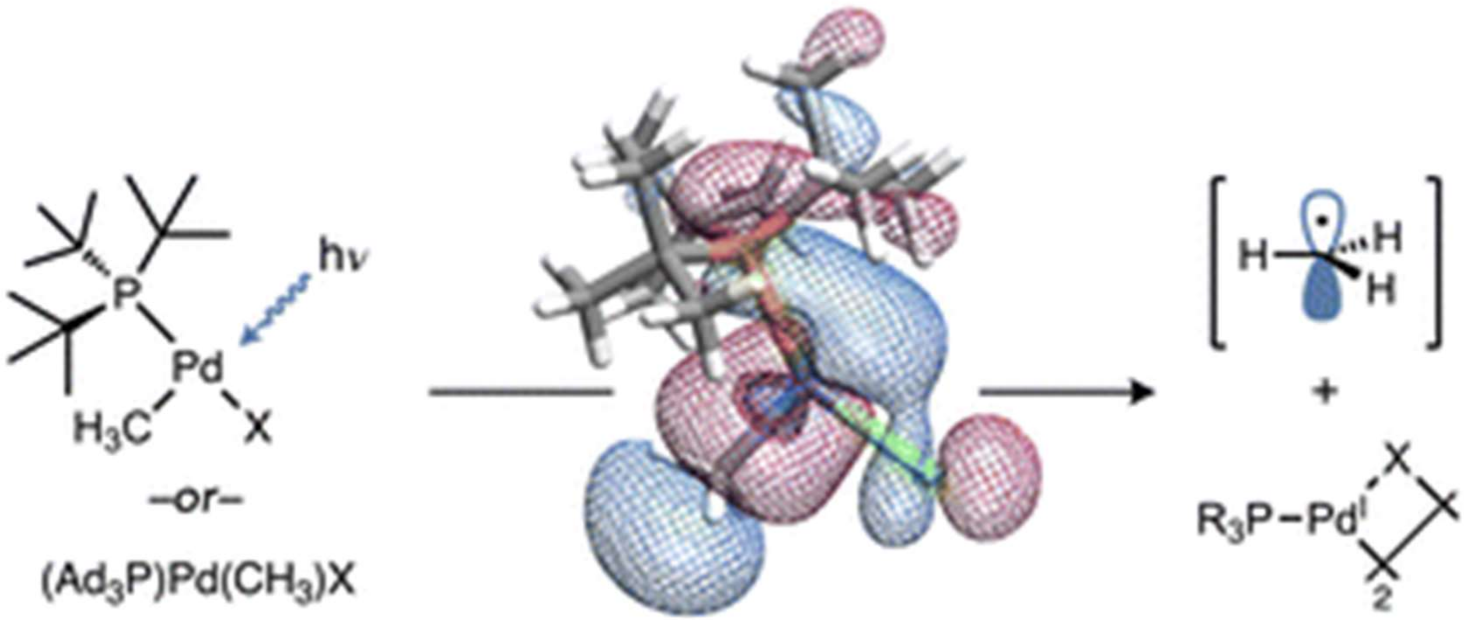
Computational and Machine Learning Exploration of Pd-C Photocleavage in T-shaped Organopalladium Complexes



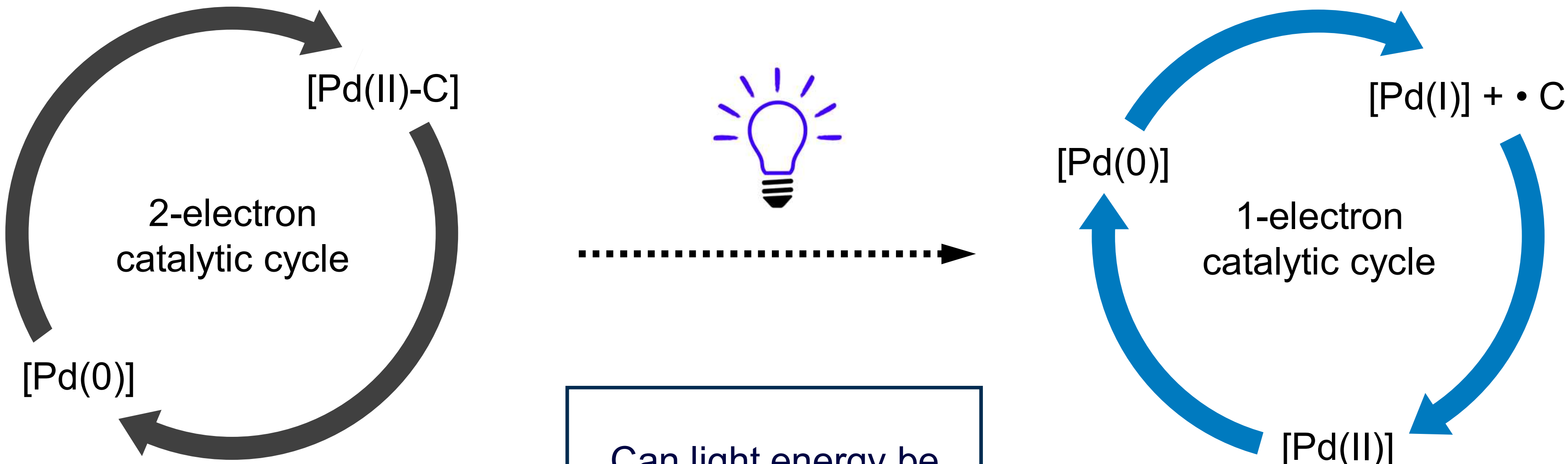
Peter Waddell, Ph.D.

Carrow *Chem. Sci.* 2023, 14, 14217.

ACCESSING 1-ELECTRON MANIFOLDS WITH Pd?

thermal reactivity

excited state reactivity

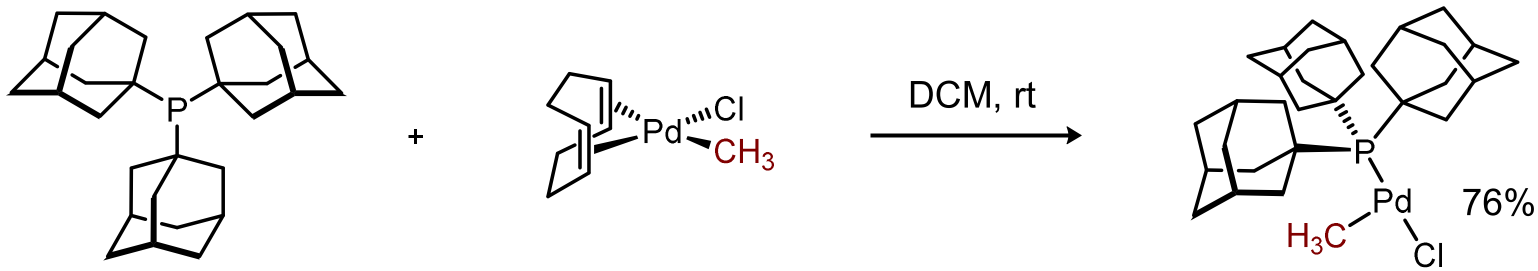


Pd is biased to this general manifold

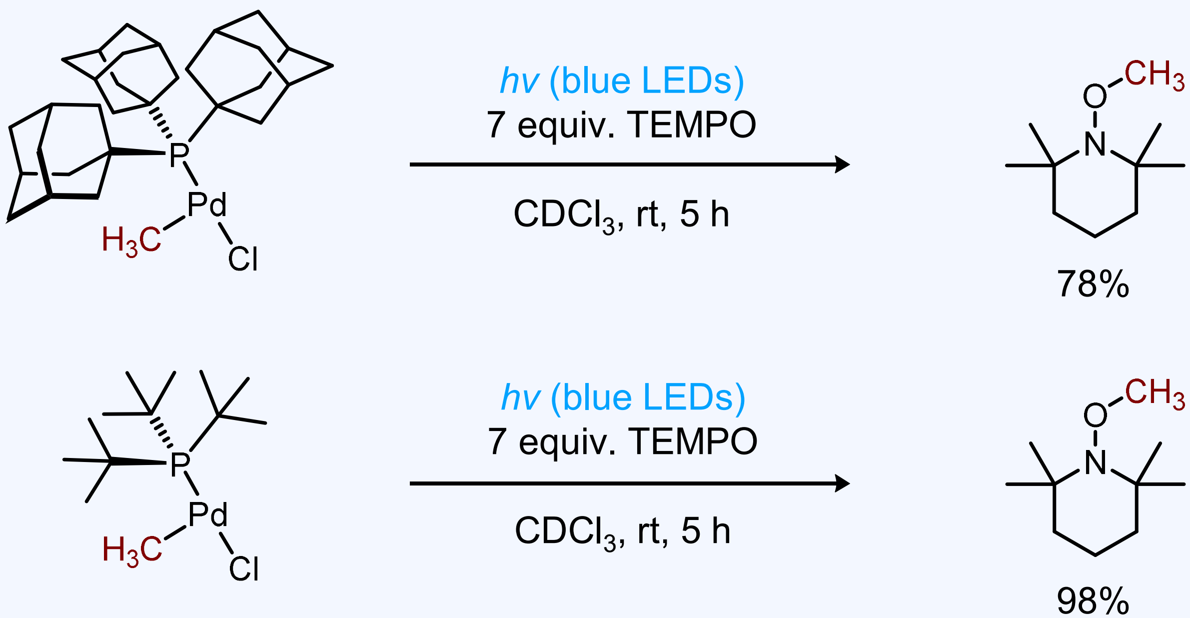
Can light energy be used to toggle between manifolds?

challenging to access

T-SHAPED COMPLEX Pd-C PHOTOCLEAVAGE WITH VISIBLE LIGHT



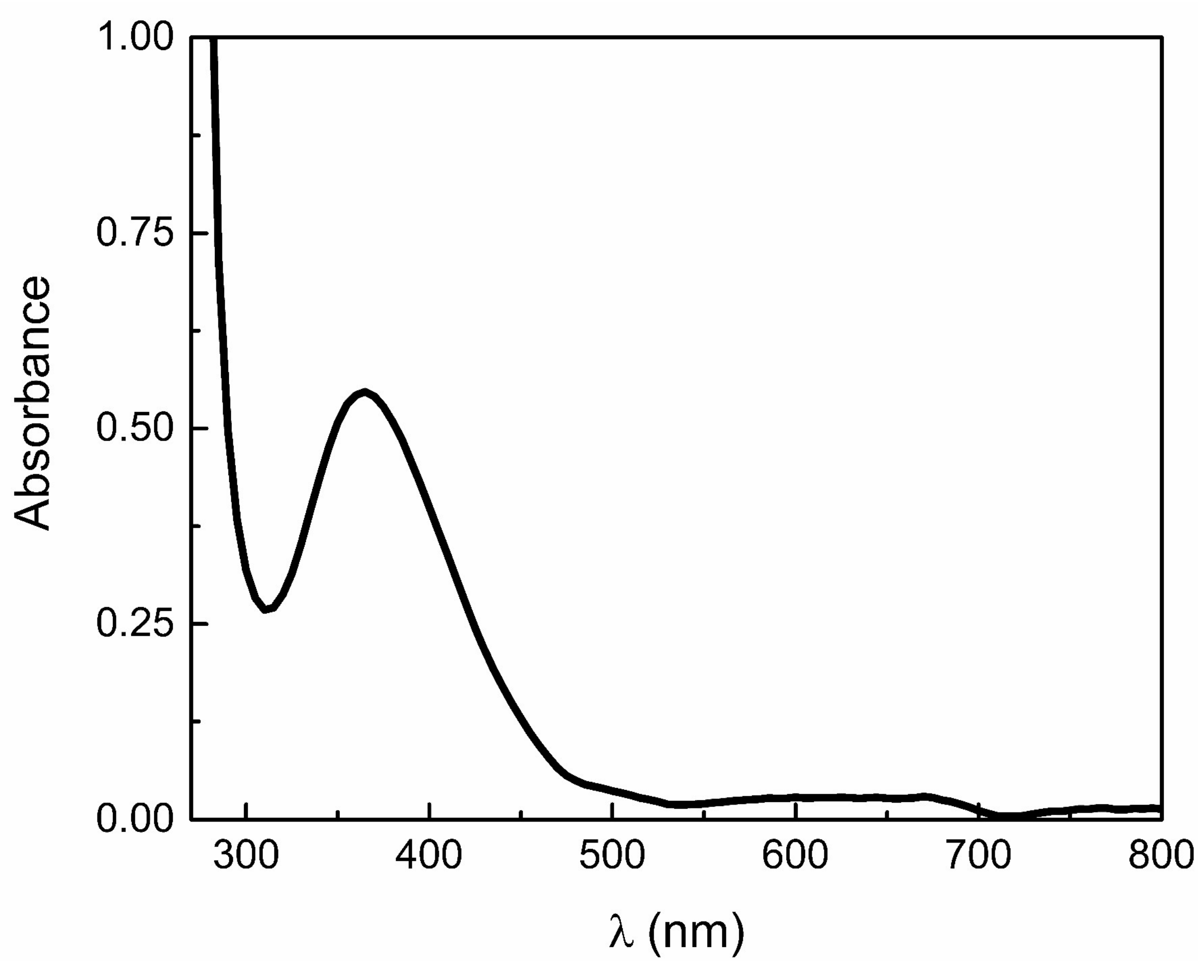
Do these T-shaped complexes exhibit any photoreactivity?



Visible light-induced bond weakening triggers SET chemistry in the most synthetically versatile metal and oxidation state (d^8, Pd^{II})!

Carrow *Chem. Sci.* 2023, *14*, 14217.

Pd-C PHOTOCLEAVAGE WITH VISIBLE LIGHT

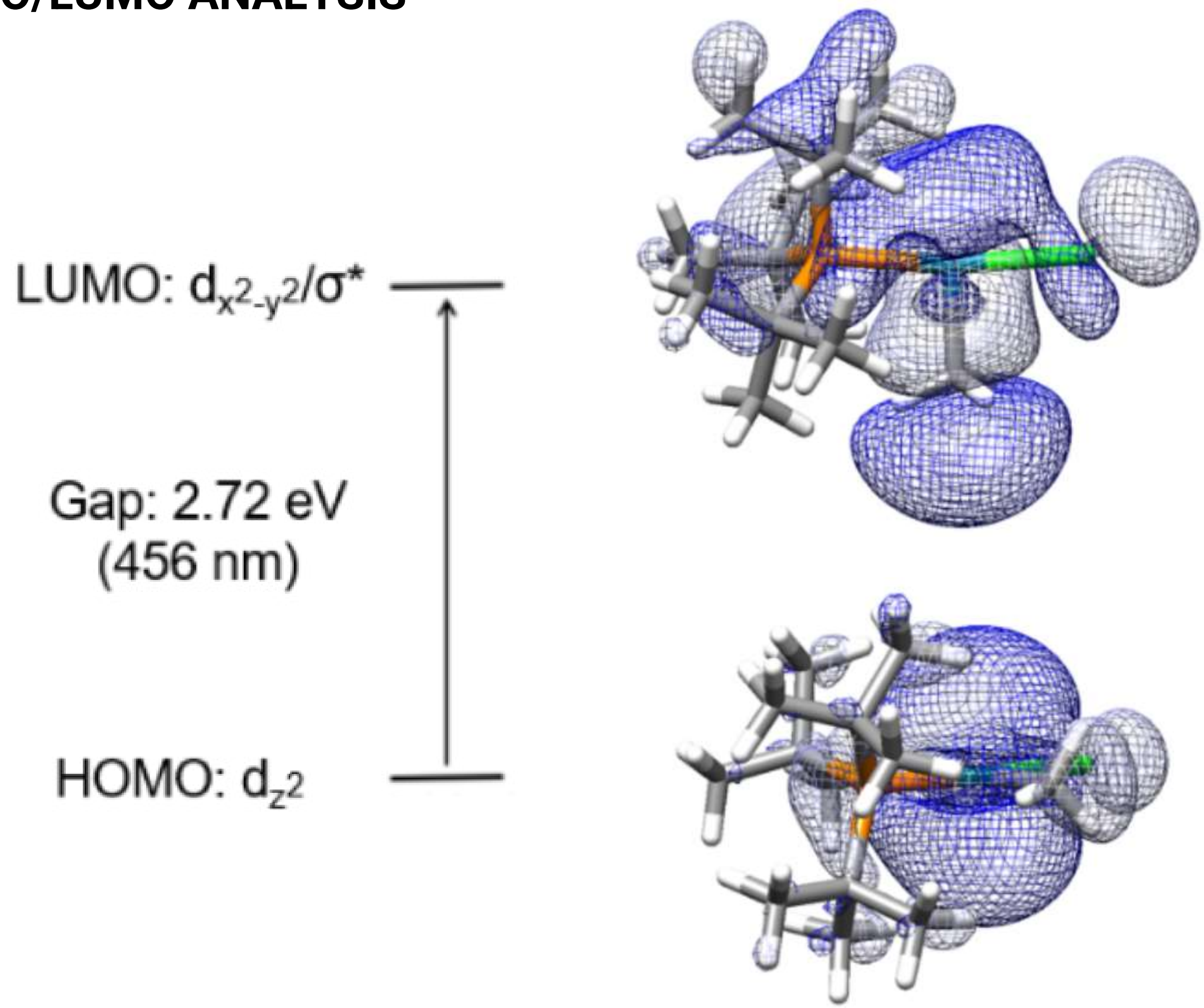


Absorption spectrum of the T-shaped complex shows the key peak for the transition that leads to Pd-C cleavage, with a λ_{max} just under 400 nm.

Carrow *Chem. Sci.* 2023, *14*, 14217.

HOMO/LUMO ANALYSIS

with Lei Tian



LUMO shows Pd-C σ* character, ***possibly accessible with visible light due to the absence of a ligand*** **trans** ***to C in the fourth site of the square plane.***

Carrow *Chem. Sci.* 2023, *14*, 14217.

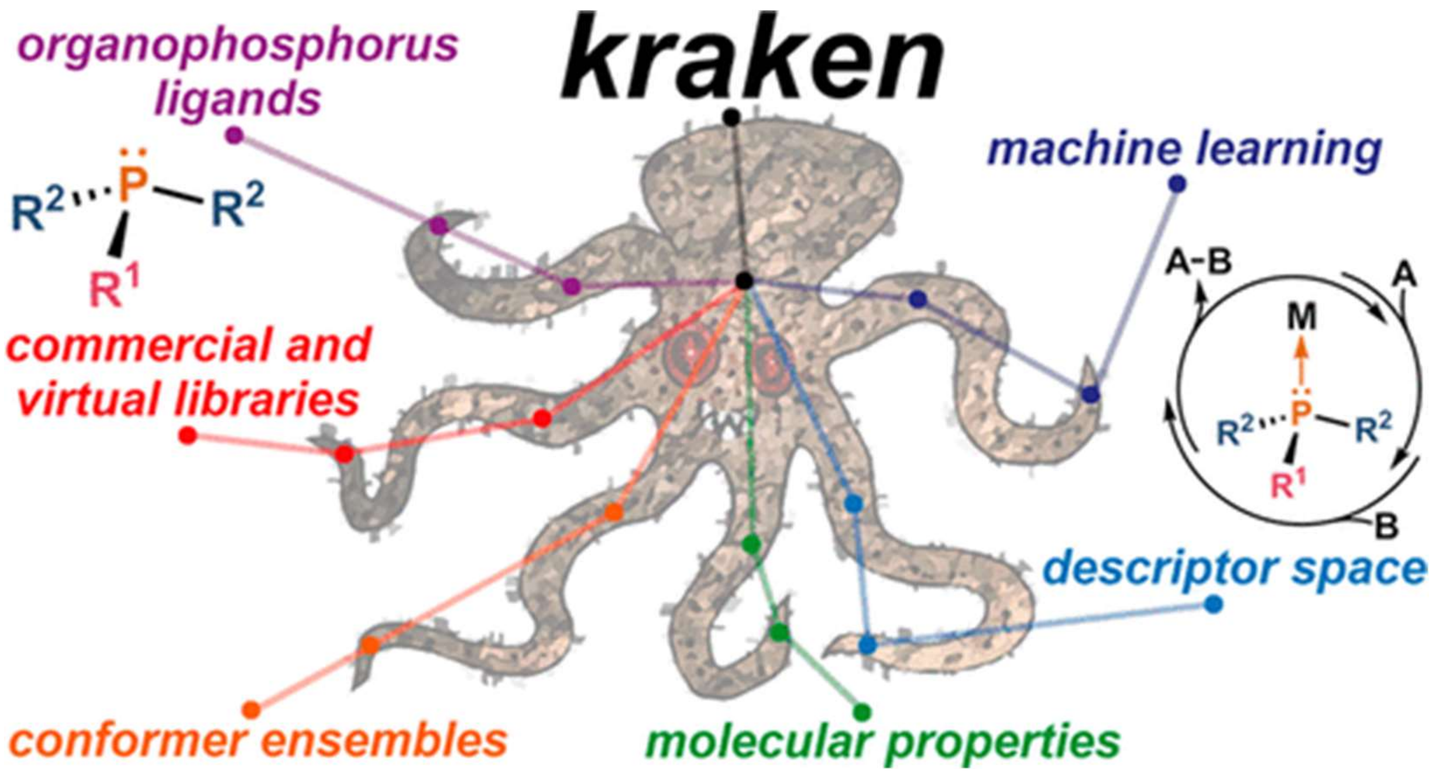
COMPUTATIONAL STUDY OBJECTIVES

- How do the steric and electronic properties of the ancillary ligand affect the UV-vis absorption spectrum of their organopalladium complexes? What characteristics cause red or blue shifts in the λ_{max} of the key peak? What is the lowest energy light that might be able to access this reactivity?

- Can we build a model with machine learning techniques to predict the spectral profile of a complex based only on information from its ancillary ligand?

- How conserved is the Pd-C photocleavage reactivity among a broad set of organophosphorus ligands?

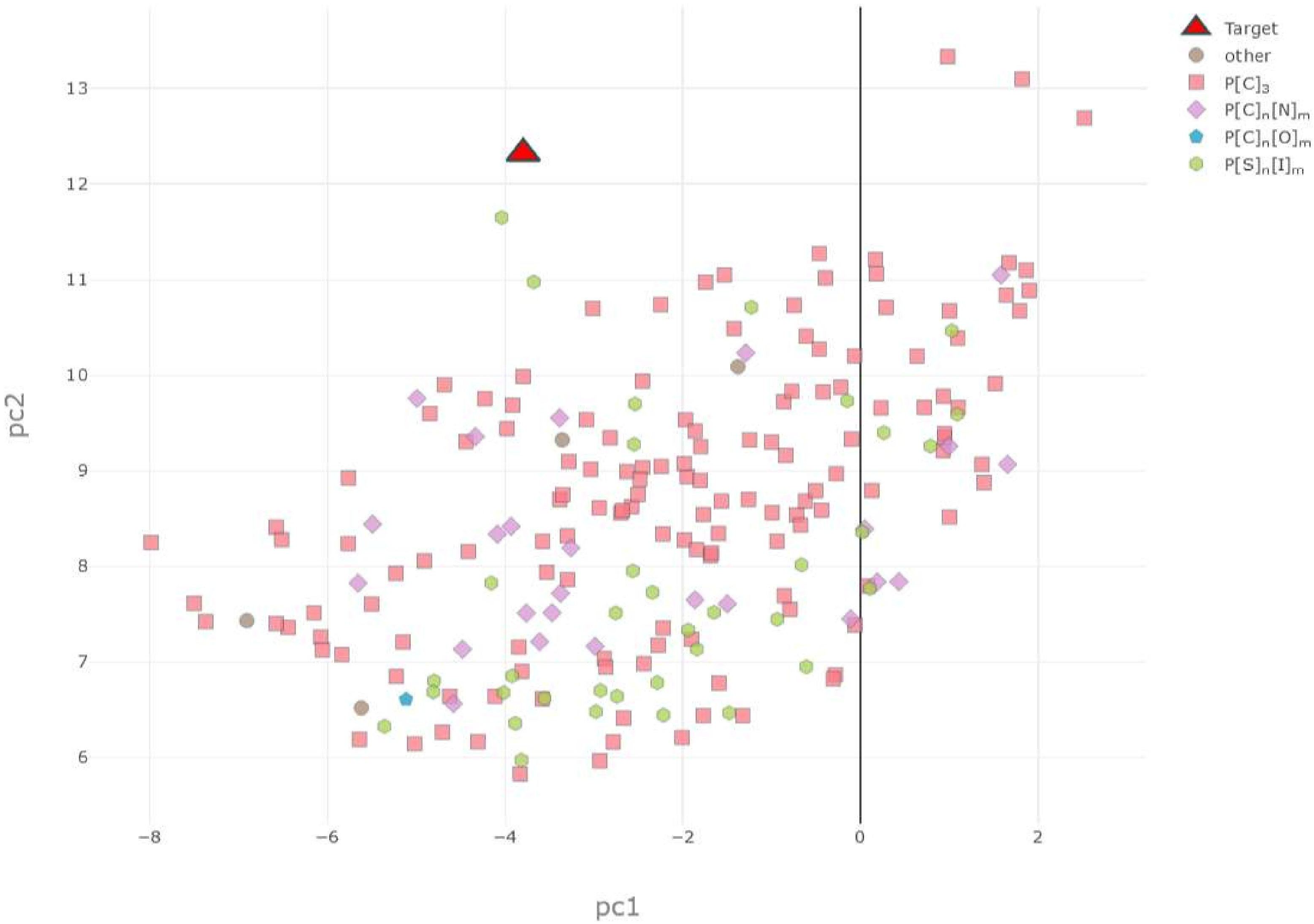
EXPLORING PHOSPHINE LIGAND SPACE WITH KRAKEN LIBRARY



The Kraken library of phosphine ligands is a perfect starting point to explore the chemical space of bulky organophosphorus ligands.

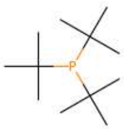
Gensch, Sigman, Aspuru-Guzik *J. Am. Chem. Soc.* **2022**, *144*, 1205.

EXPLORING PHOSPHINE LIGAND SPACE WITH KRAKEN LIBRARY

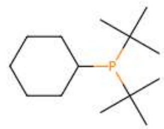


To start, I used the Kraken library to find the 210 closest (in principal component space) organophosphorus ligands to our target (tri-*t*-butylphosphine)

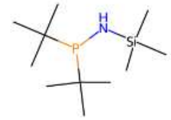
EXPLORING PHOSPHINE LIGAND SPACE WITH KRAKEN LIBRARY


8

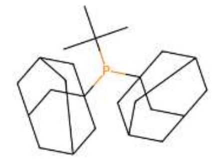

12

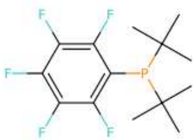

30

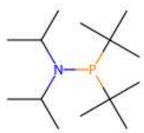

274

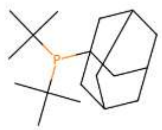

297

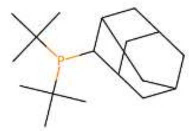

357

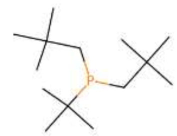

417


835

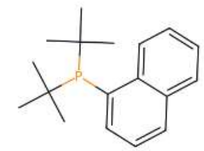

920


951


952

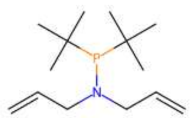

953


1136


1648


16670


16848


19864


19867


19869

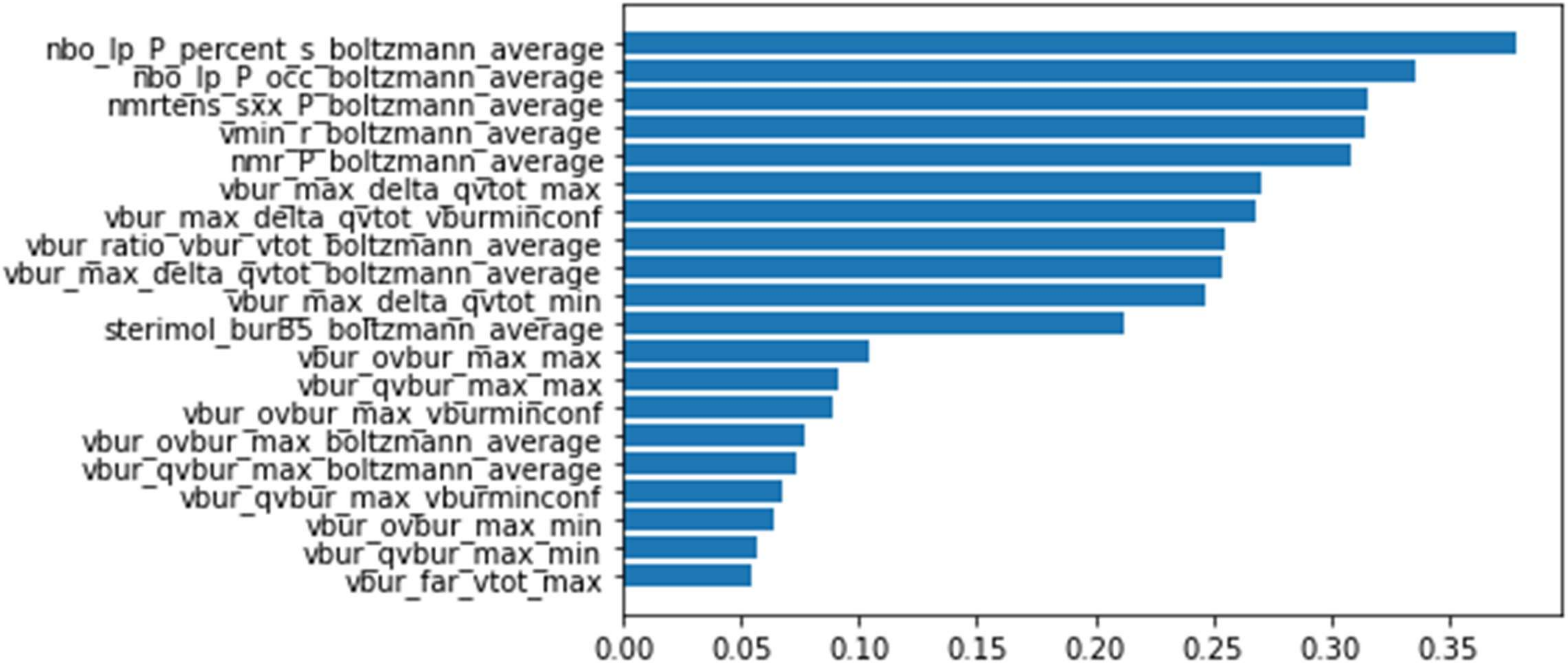

19872


19874

In order to ensure that the organophosphorus ligands would form T-shaped (monoligated) Pd complexes, I restricted this set to the ligands with Boltzmann-average $\%V_{bur}$ of 68% or more, very close to the value for tri-*t*-butylphosphine.

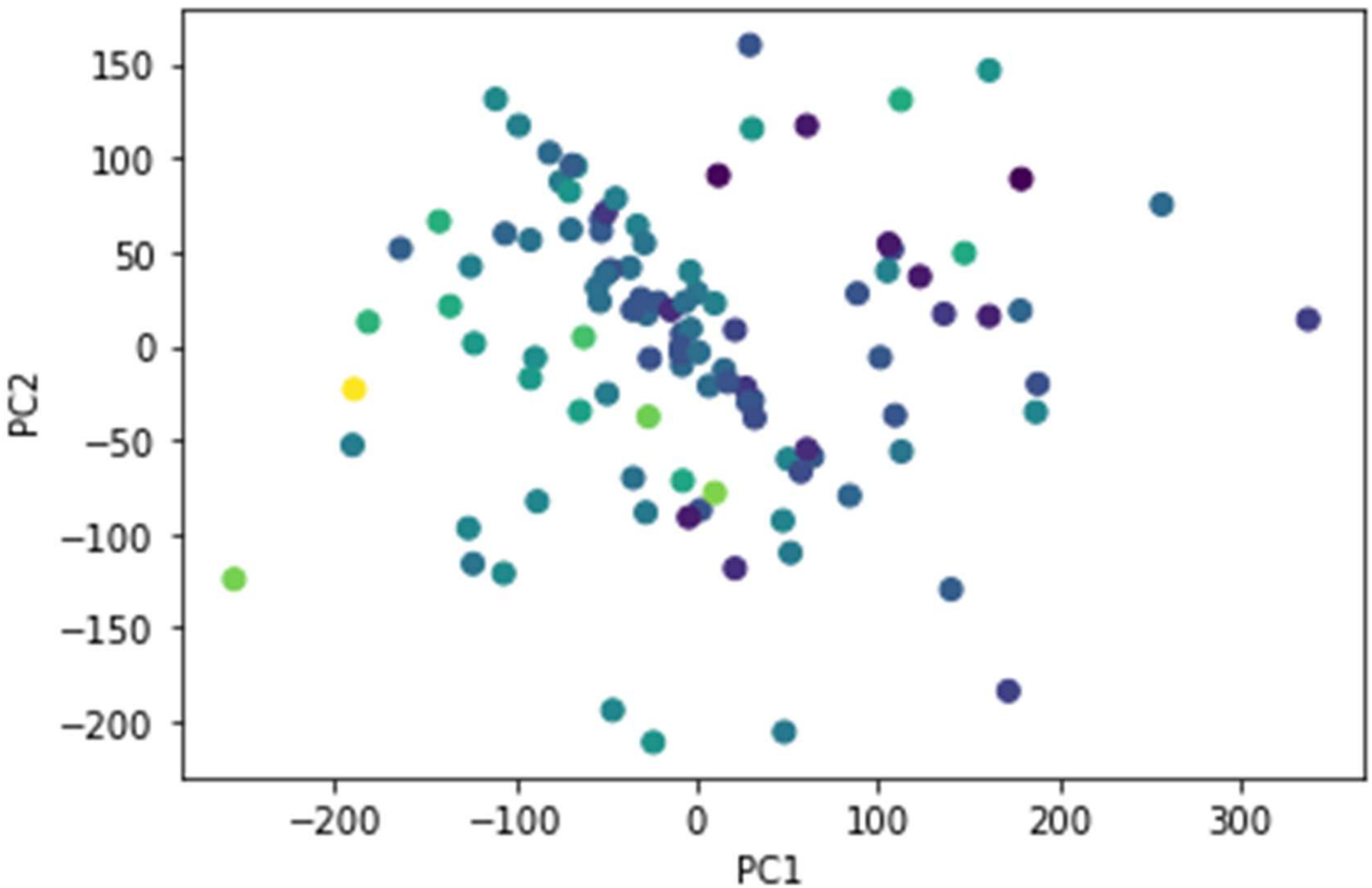
This gave a final set of **107 ligands** (small subset pictured).

LIGAND SET PRINCIPAL COMPONENT ANALYSIS



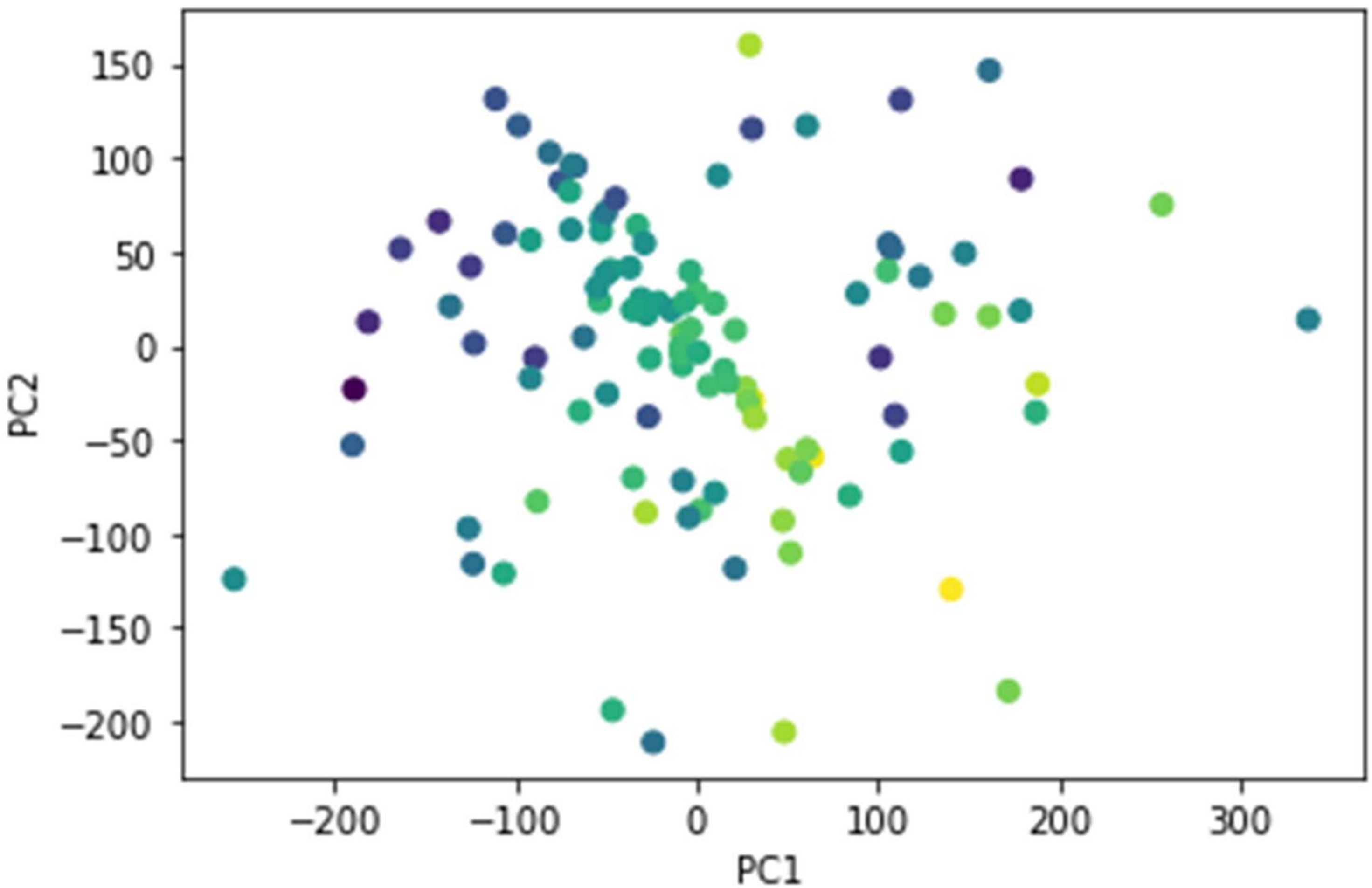
Here are the twenty most important features for the first principal component. Notably, electronic features such as Boltzmann-average NBO P lone pair %s and NMR values rank most highly.

LIGAND SET PRINCIPAL COMPONENT ANALYSIS



PCA on our ligand set, colored according to Boltzmann-average NBO P lone pair %s.

LIGAND SET PRINCIPAL COMPONENT ANALYSIS



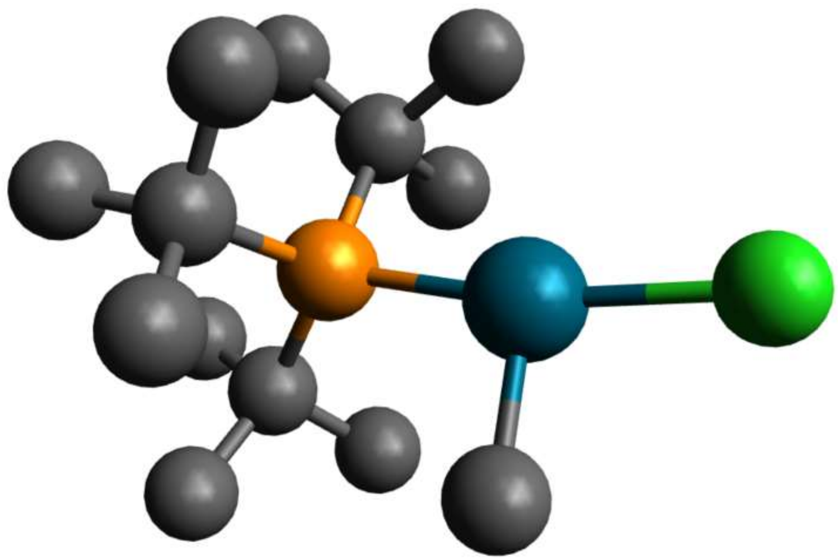
PCA on our ligand set, colored according to highest total volume difference in $\%V_{bur}$ between two neighboring quadrants.

COMPUTATIONAL WORKFLOW

- Gather and prune ligand set from Kraken database
- Generate SMILES strings for methylpalladium chloride T-shaped complexes for each ligand
- Use RDKit to run conformer searches for each complex
- Perform geometry optimizations based on the lowest energy conformer for each complex
- Perform TDDFT calculations and simulate the UV-vis spectra for each complex, determine the λ_{max} for the relevant transition
- Find trends in λ_{max} and ligand/complex properties
- Use machine learning to build and validate models to predict λ_{max} for other organophosphorus ligands, extract insights
- See the repo at: https://github.com/pmwaddell/pd-c-photochem-ML for the source code, all steps are automated with scripts

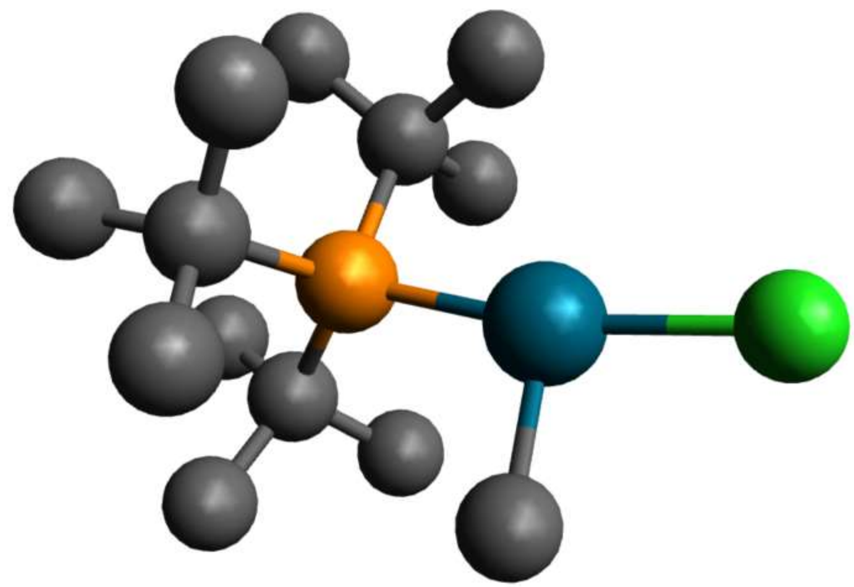
DFT: LEVEL OF THEORY, GEOMETRY OPTIMIZATION

Literature X-ray structure



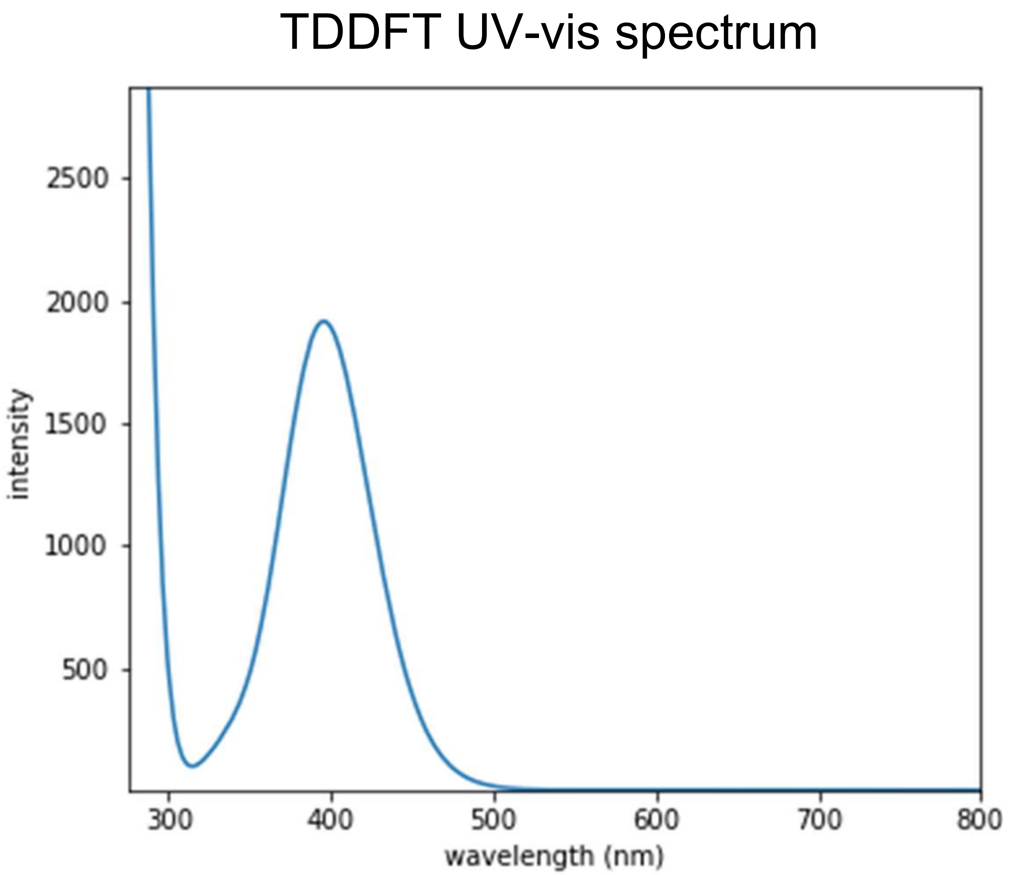
From Nozaki *Organometallics* **2006**, *25*, 4588.

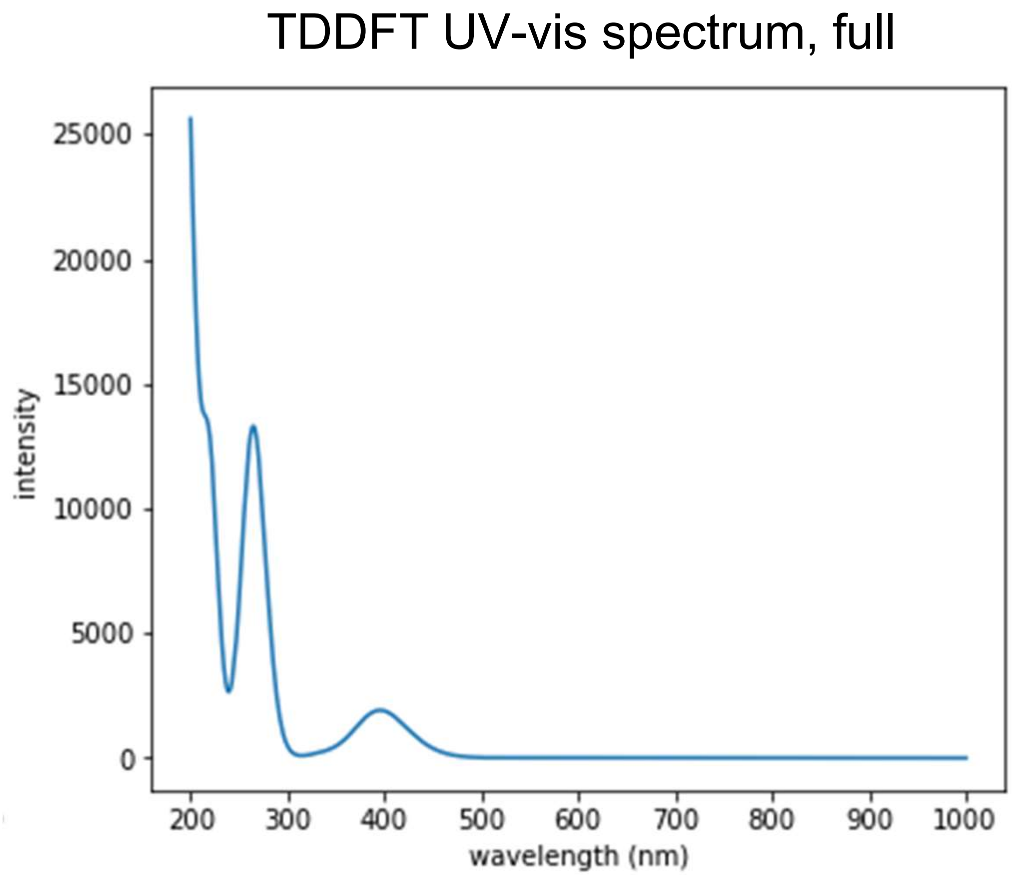
DFT geometry optimized structure



After screening various basis sets and functionals for geometry optimization, **B3LYP-D3/def2-TZVP with CPCM($CHCl_3$)** was found to give good agreement (key bond lengths within 0.05 Å, angles within 1°) with the reported X-ray structure while being relatively time-efficient.

DFT: LEVEL OF THEORY, TDDFT

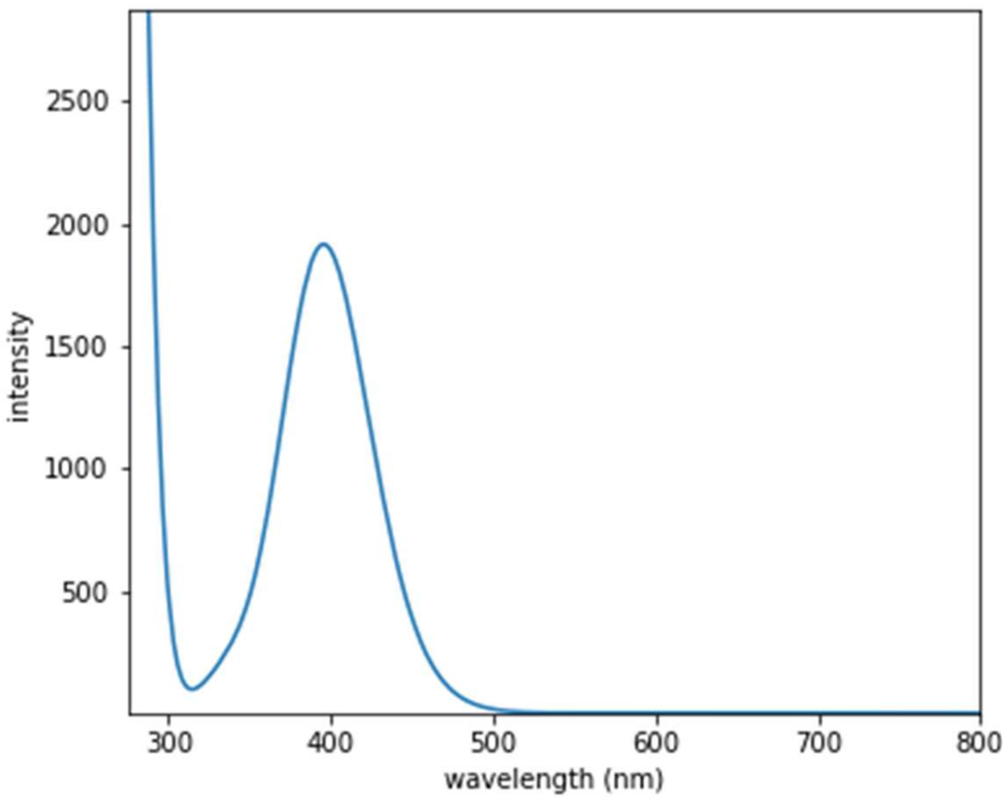




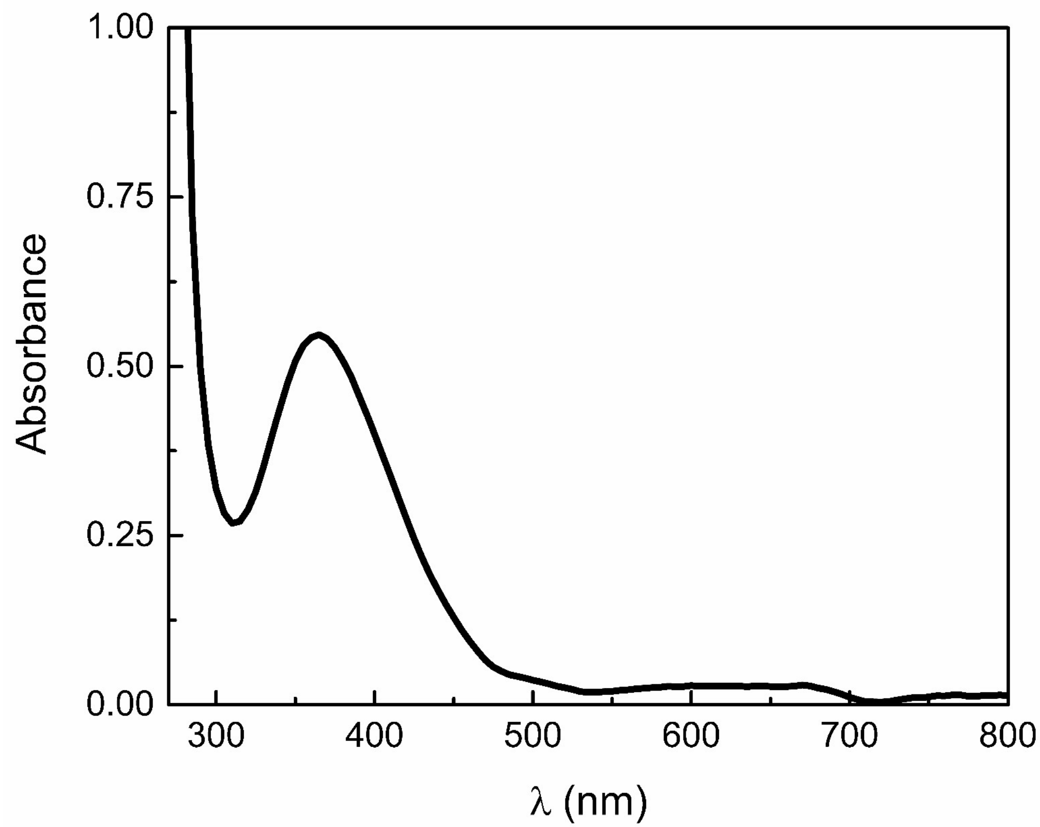
For TDDFT calculations, agreement with the experimental UV-vis absorption spectrum was used to benchmark performance. Similarly, **B3LYP/def2-TZVPP with CPCM($CHCl_3$)** was found to give the closest agreement.

DFT: LEVEL OF THEORY, TDDFT

TDDFT UV-vis spectrum

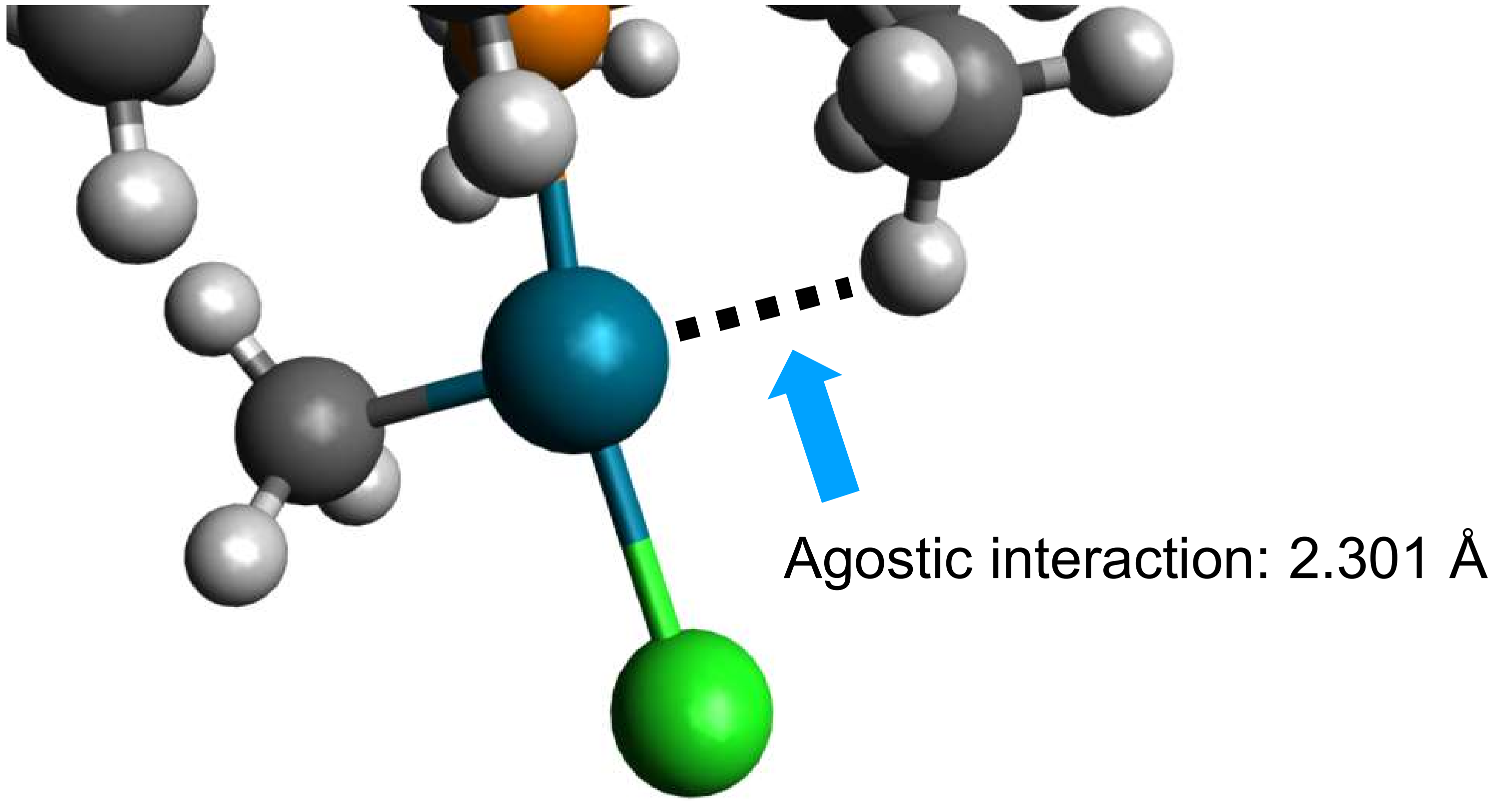


experimental UV-vis spectrum



For TDDFT calculations, agreement with the experimental UV-vis absorption spectrum was used to benchmark performance. Similarly, **B3LYP/def2-TZVPP with CPCM($CHCl_3$)** was found to give the closest agreement.

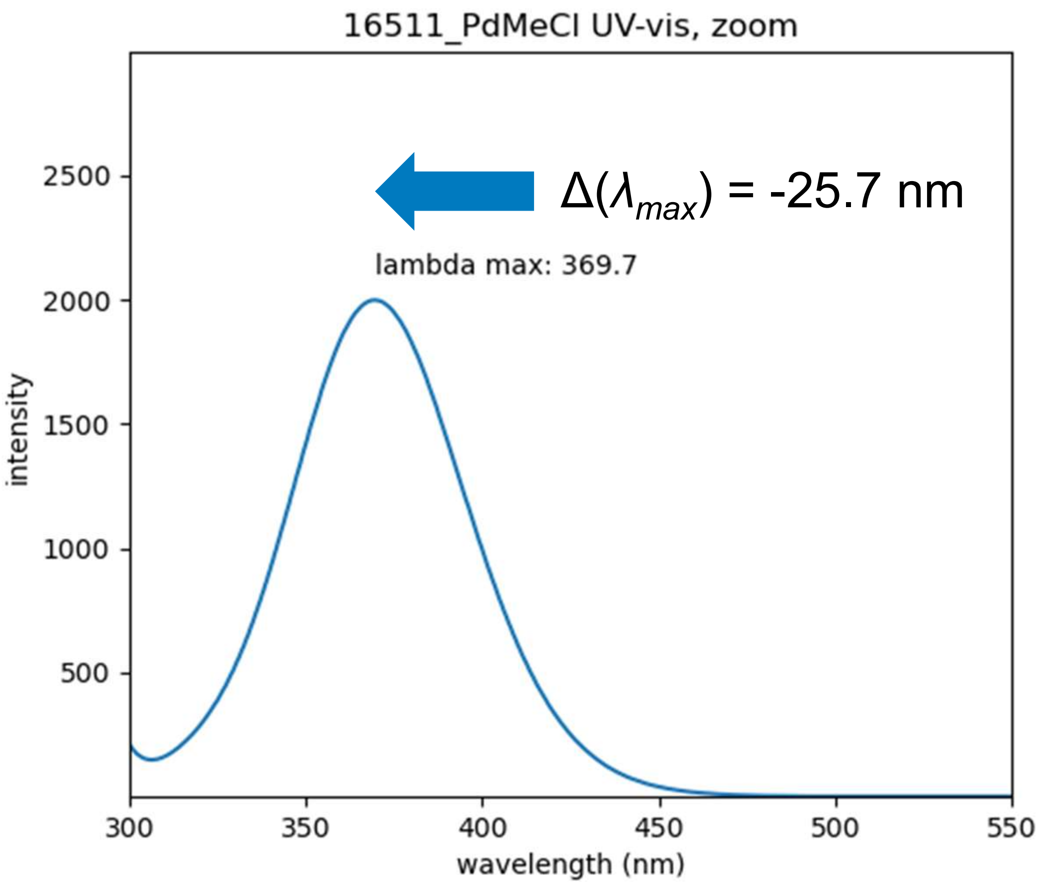
INITIAL OBSERVATIONS: TRENDS IN λMAX, AGOSTIC INTERACTIONS

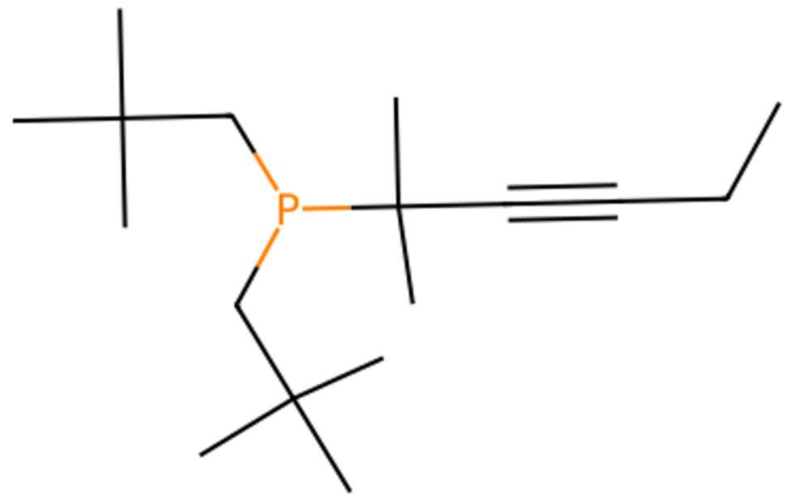


While we refer to these complexes as T-shaped, they often feature an agostic interaction in the fourth site. The distance between Pd and H correlates with the strength of the interaction.

My hypothesis is that the lack of a strong bonding interaction in this site lowers the LUMO (σ*) energy. Thus, λ_{max} should correlate with agnostic interaction strength.

INITIAL OBSERVATIONS: TRENDS IN λMAX, AGOSTIC INTERACTIONS

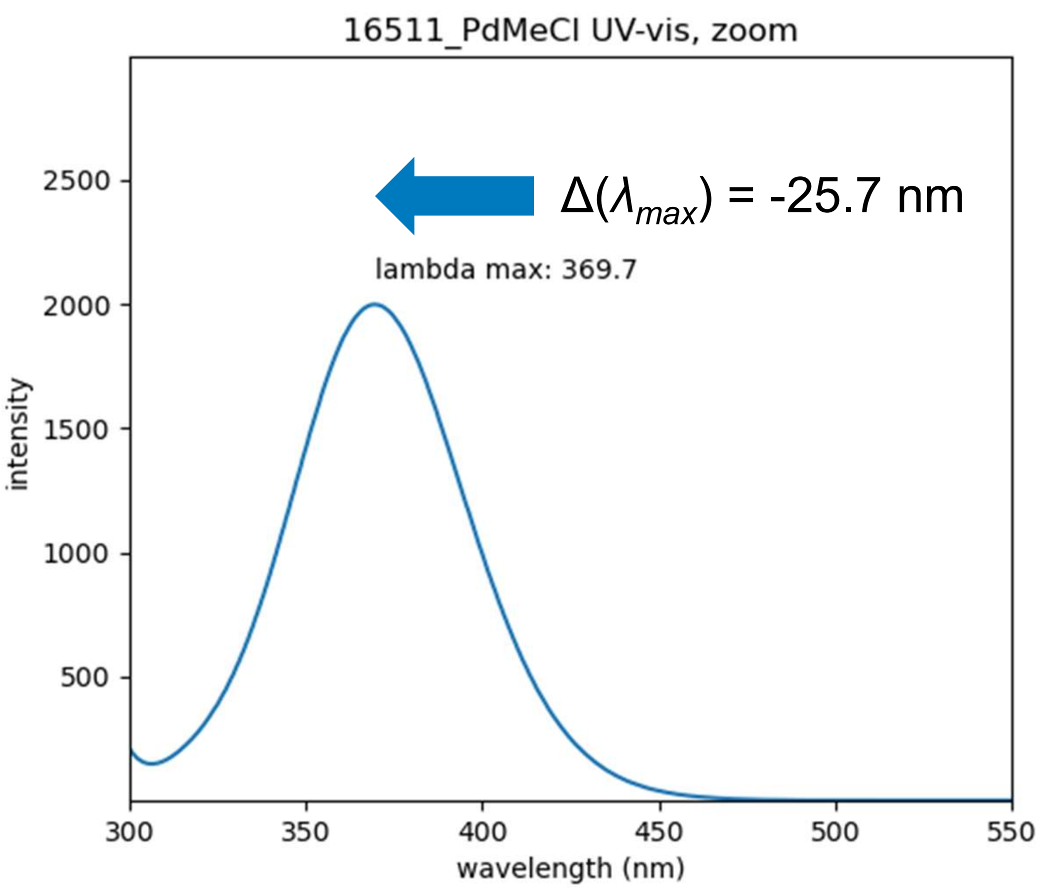


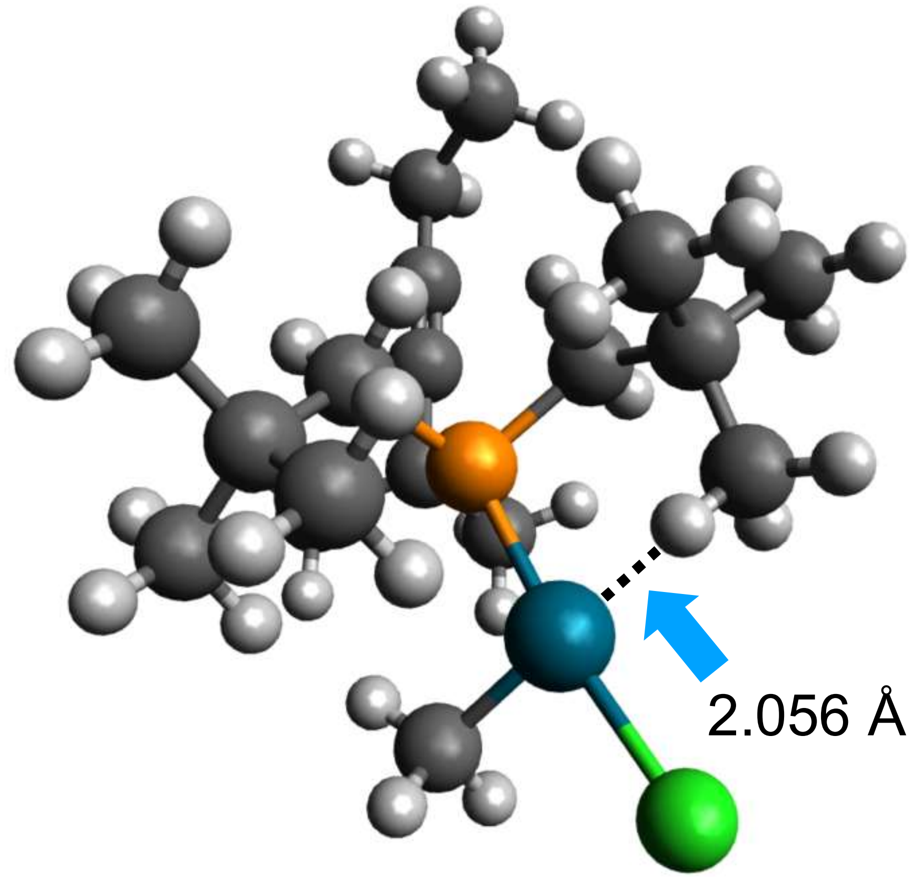


Ligand 16511 (Kraken designation)

As an example, this ligand features neopentyl groups which allow the H to come in closer proximity to Pd. As a result, the stronger agostic interaction leads to a λ_{max} blueshift of over 25 nm!

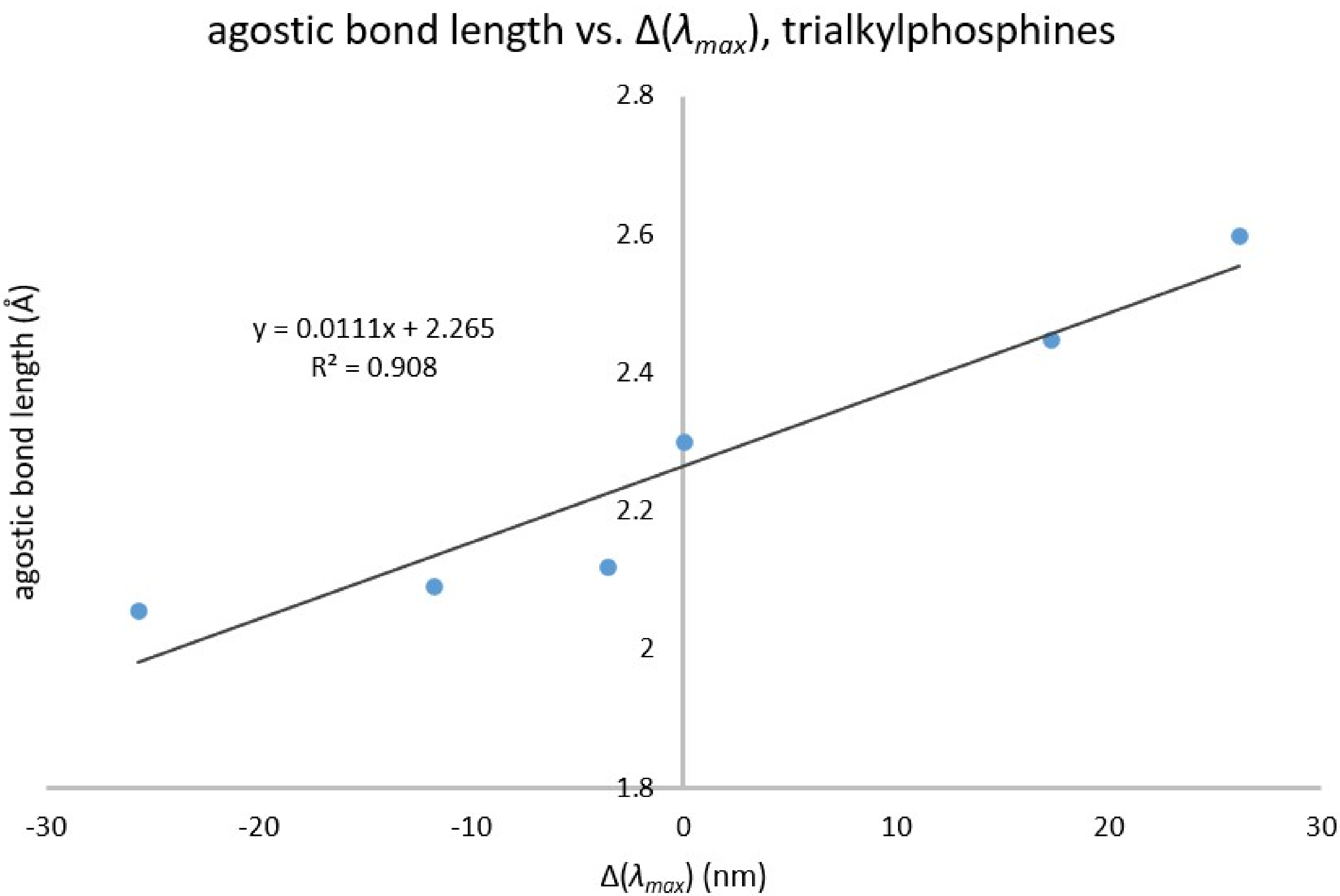
INITIAL OBSERVATIONS: TRENDS IN λMAX, AGOSTIC INTERACTIONS





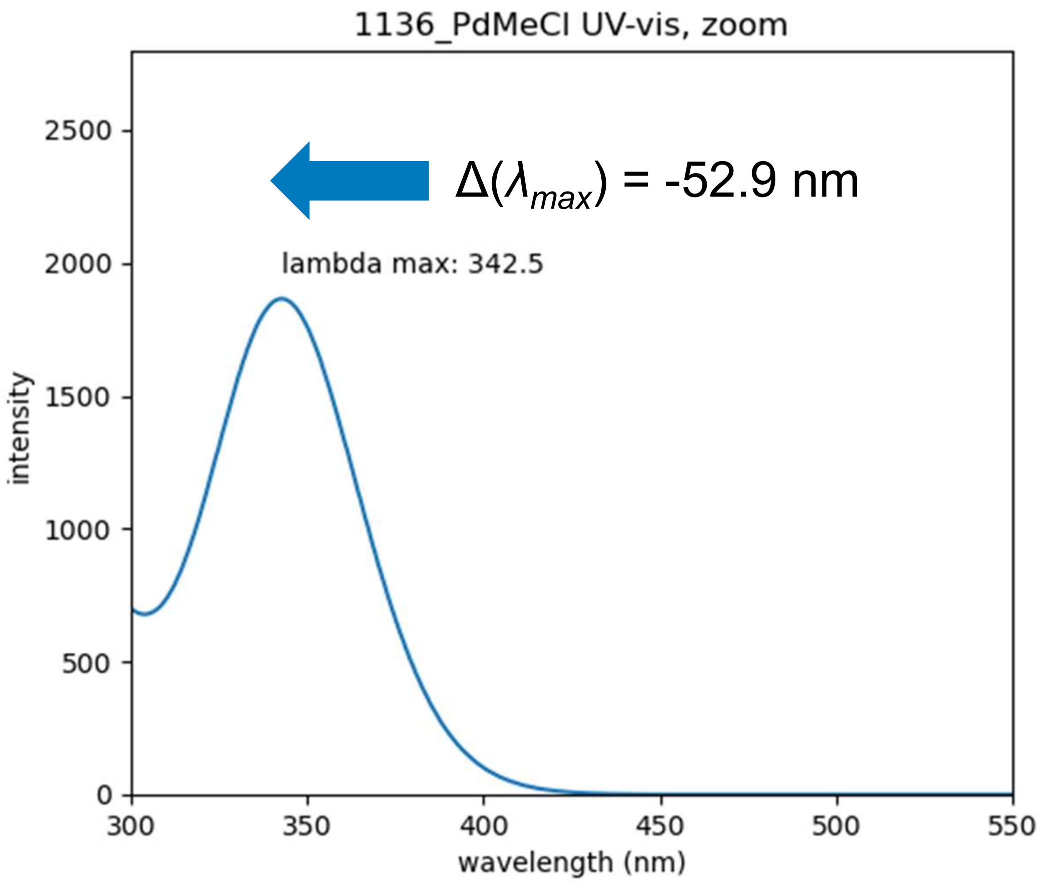
As an example, this ligand features neopentyl groups which allow the H to come in closer proximity to Pd. As a result, the stronger agostic interaction leads to a λ_{max} blueshift of over 25 nm!

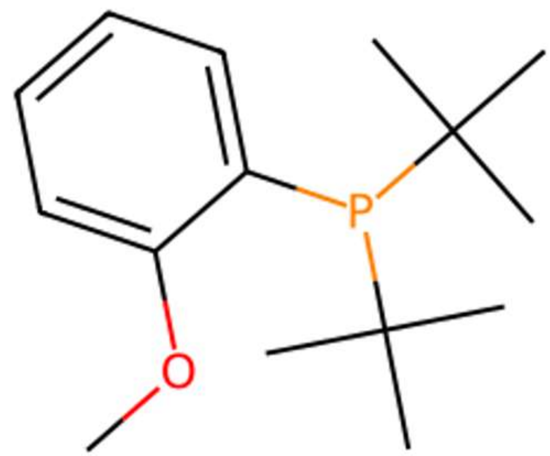
INITIAL OBSERVATIONS: TRENDS IN λMAX, AGOSTIC INTERACTIONS



Across a series of trialkylphosphines, the length of the agostic bond correlates well with the change in λ_{max} relative to tri-*t*-butylphosphine!

INITIAL OBSERVATIONS: TRENDS IN λMAX, CHELATION

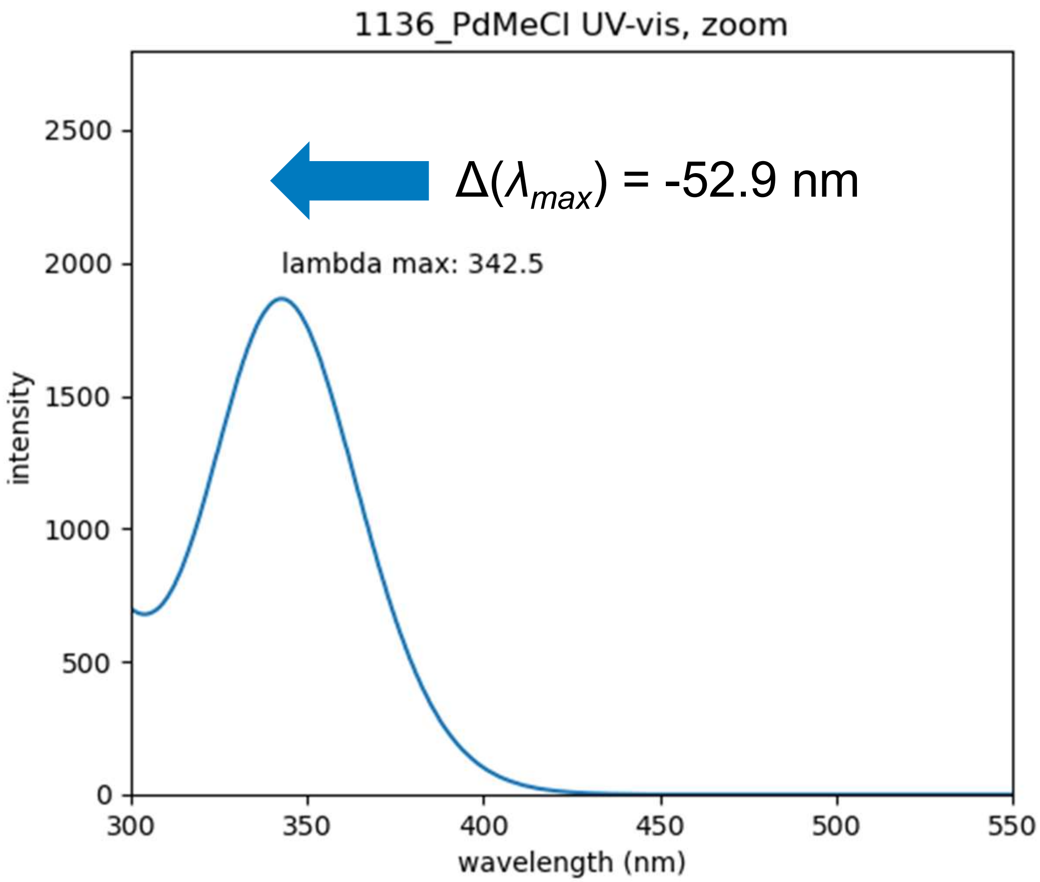


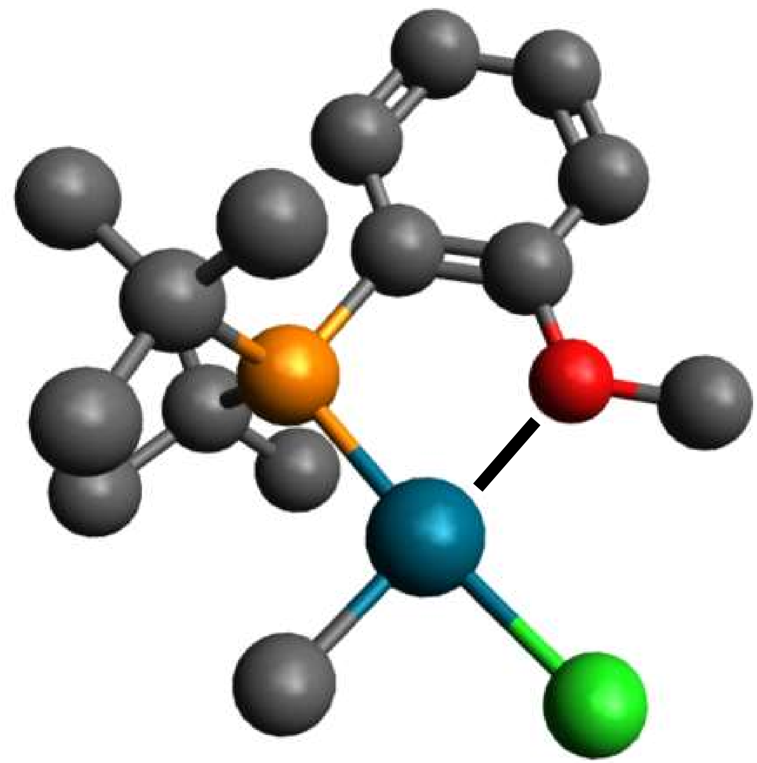


Ligand 1136

Furthermore, for ligands which have heteroatoms that allow chelation, very large λ_{max} blueshifts are observed, again consistent with my hypothesis. Indeed, these complexes are in fact square planar rather than T-shaped.

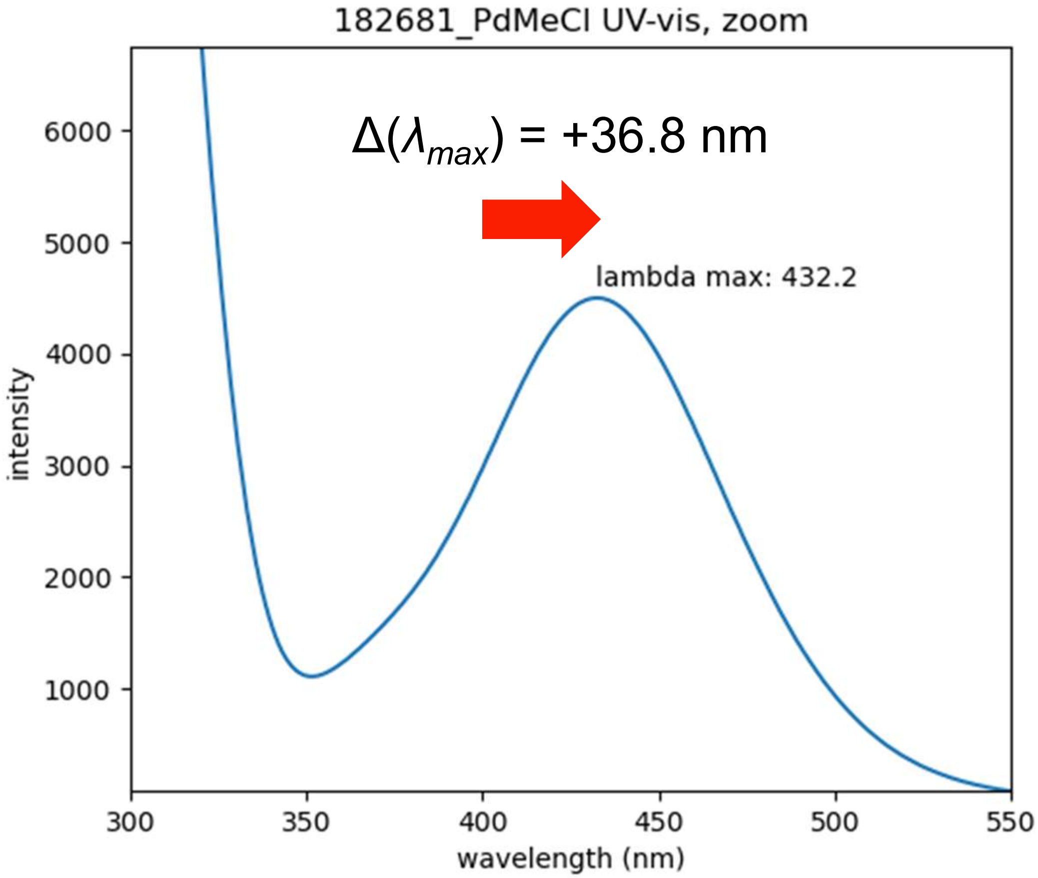
INITIAL OBSERVATIONS: TRENDS IN λMAX, CHELATION

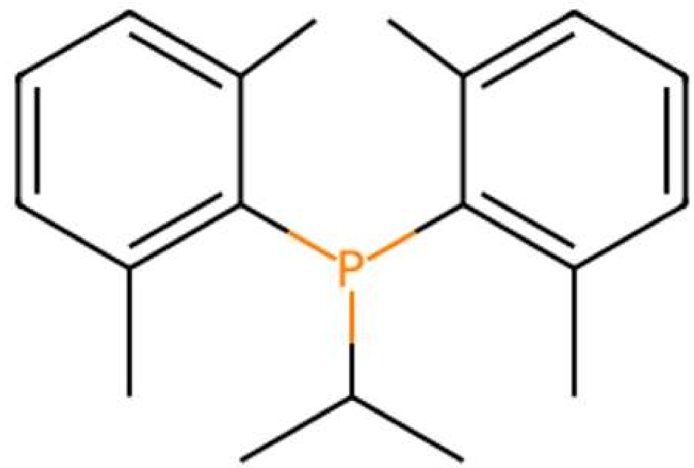




Furthermore, for ligands which have heteroatoms that allow chelation, very large λ_{max} blueshifts are observed, again consistent with my hypothesis. Indeed, these complexes are in fact square planar rather than T-shaped.

INITIAL OBSERVATIONS: TRENDS IN λMAX, OPEN SITE

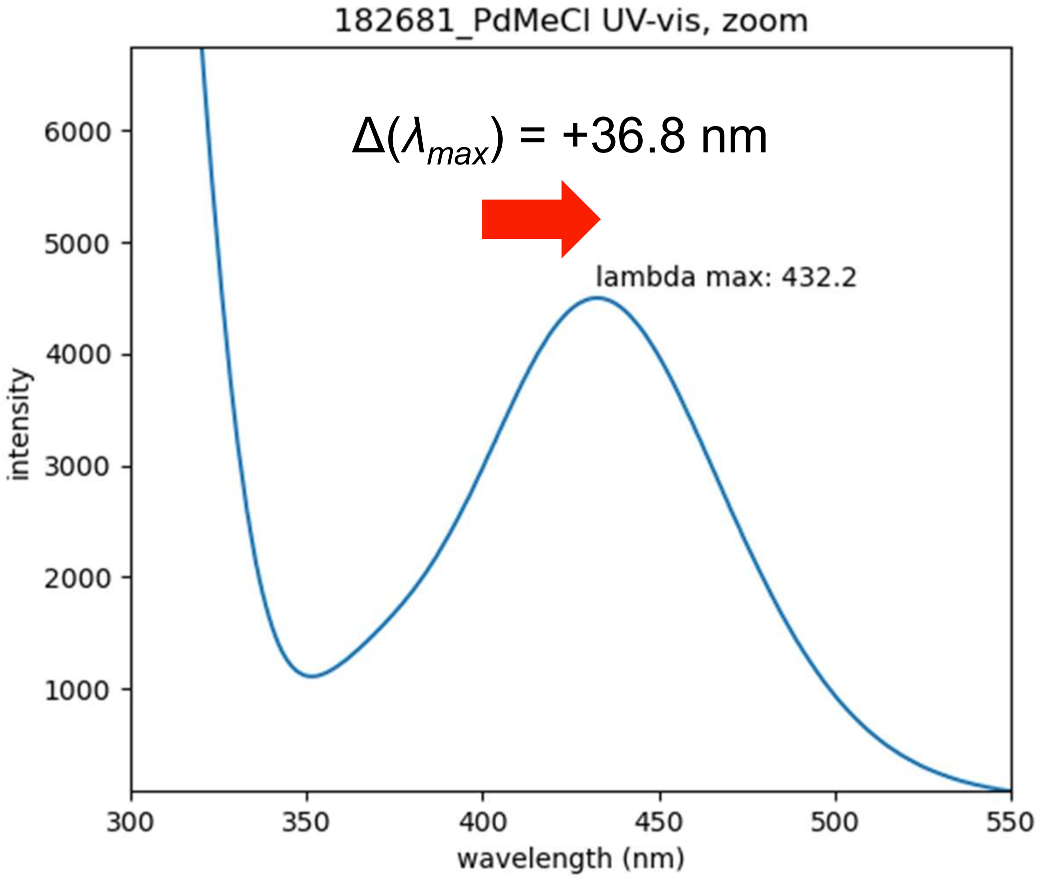


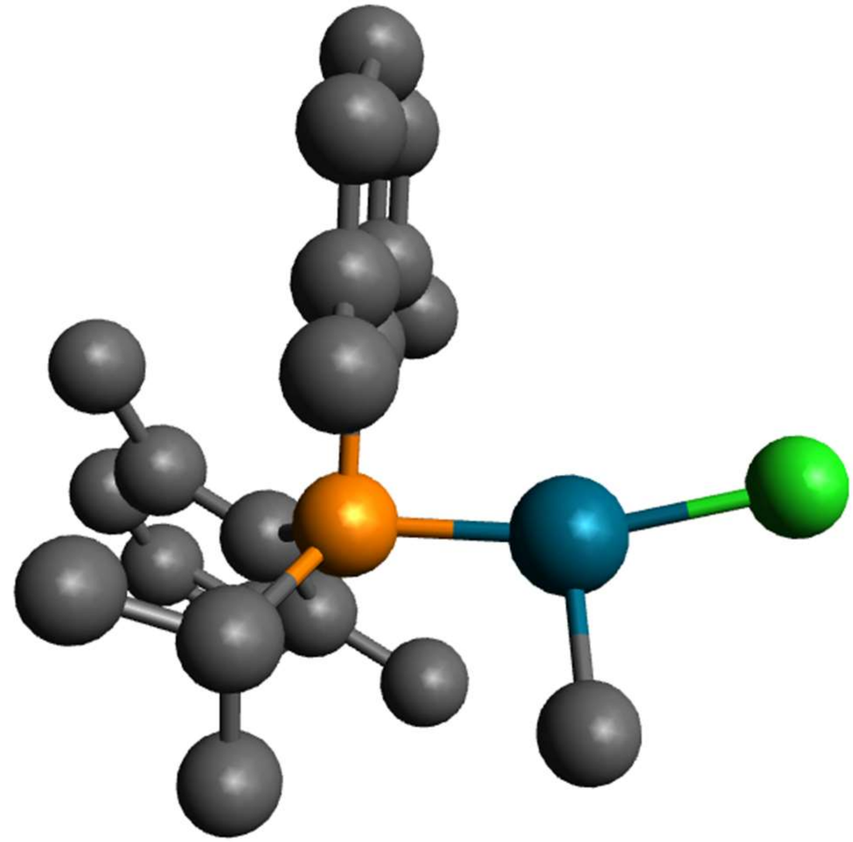


Ligand 182681

In some cases, the favored ligand conformation appears to leave the fourth site open entirely. In this limiting case, λ_{max} tends to be around 420-430 nm. However, in the condensed phase, this site would be open to occupation by solvent or e.g. bridging Cl ligand, so they may not be a good choice in practice for triggering this photochemistry with low energy light.

INITIAL OBSERVATIONS: TRENDS IN λMAX, OPEN SITE





In some cases, the favored ligand conformation appears to leave the fourth site open entirely. In this limiting case, λ_{max} tends to be around 420-430 nm. However, in the condensed phase, this site would be open to occupation by solvent or e.g. bridging Cl ligand, so they may not be a good choice in practice for triggering this photochemistry with low energy light.

INITIAL OBSERVATIONS: TRENDS IN λMAX, ELECTRONICS

Comments on electronic effects, P-C vs. P-N vs. P-O, aryl, etc.

MACHINE LEARNING MODELS TO PREDICT λMAX: ELASTIC NET

- Our case is highly dimensional: each ligand has 192 features, while we have at most around 100 data points.

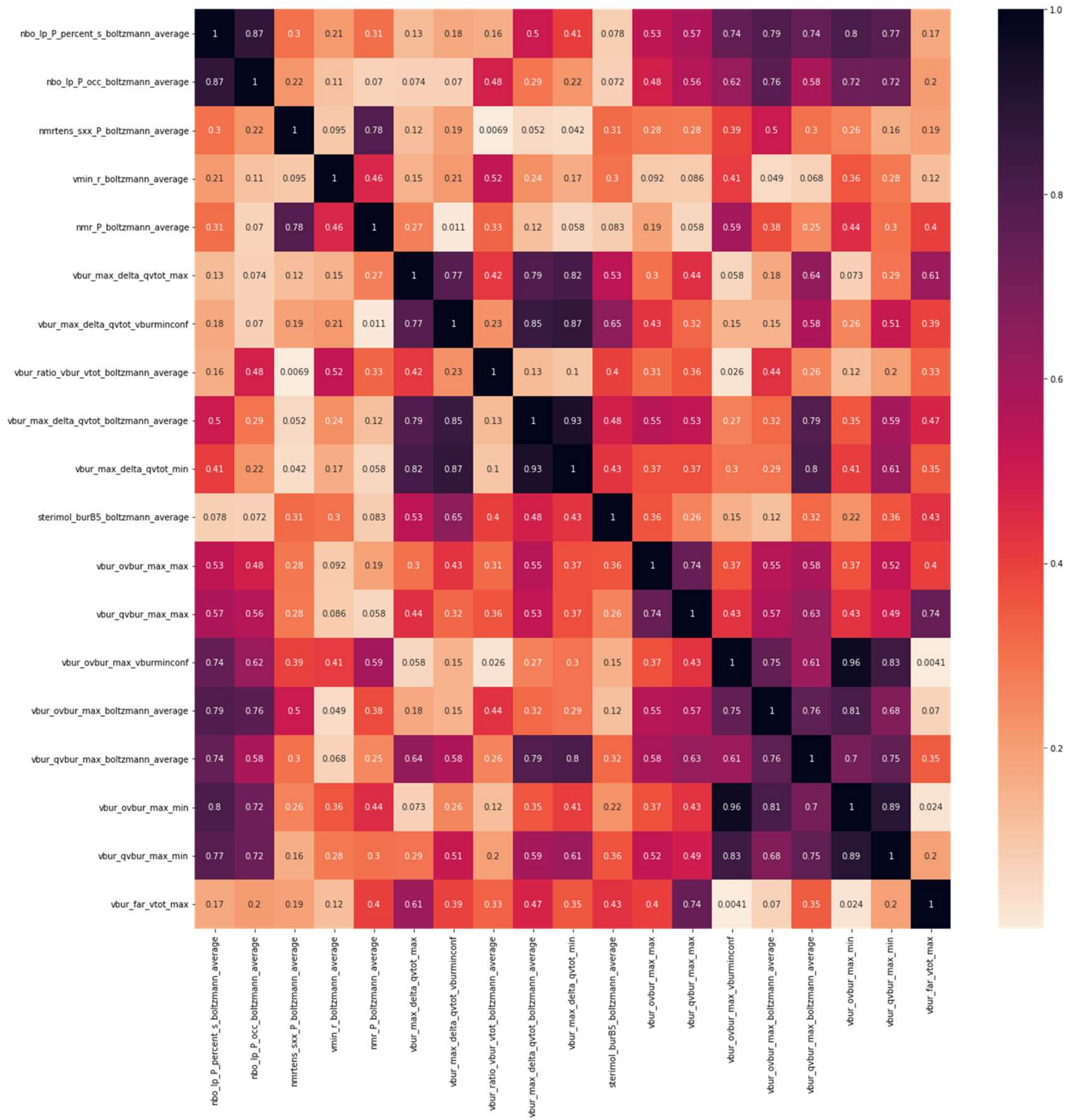
- To begin with, let us restrict ourselves to the top 20 features from our principal component analysis from earlier.

- Furthermore, there is significant multicollinearity in our feature set (see next slide)

- Therefore, a good model could be Elastic Net, which combines the L1 penalty of the Lasso and the L2 penalty of Ridge Regression. Thus, the model can do feature selection to get us down to a smaller number of the most relevant features, while the coefficients are also shrunk.

- **TODO: train model with CV, Grid Search for hyperparameter optimization (alpha, L1 ratio), use model against test set**

For simplicity, the absolute correlation is shown here, with darker meaning more highly correlated.



To be continued…