Quantitative Models for Accurate Reactivity Predictions and Mechanistic Elucidation

by

Jingru Lu
M.Sc., Carnegie Mellon University, 2014
B.Sc., University of Minnesota, Twin Cities, 2012

A Dissertation Submitted for Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Chemistry

Supervisory Committee

Quantitative Models for Accurate Reactivity Predictions and Mechanistic Elucidation

by

Jingru Lu
M.Sc., Carnegie Mellon University, 2014
B.Sc., University of Minnesota, Twin Cities, 2012

Supervisory Committee

Dr. David C. Leitch, Supervisor
Department of Chemistry

Dr. Irina Paci, Departmental Member
Department of Chemistry

Dr. Fraser Hof, Departmental Member
Department of Chemistry

Dr. Nishant Mehta, Outside Member
Department of Computer Science

Abstract

Accurate prediction of reaction outcomes is among the most important goals in chemical and pharmaceutical synthesis. In recent years, the ultrafast growth in computing power and the advancement of high-throughput experimental (HTE) technology have paved new ways to apply data-rich approaches in chemistry research. In organic synthesis, data-driven methods have found many successful applications in accelerating reaction condition optimization and developing machine learning models for reaction prediction. Despite all the impressive progress made in this area, accurate prediction of chemical reactivity remains challenging.

This thesis describes development of quantitative reactivity models for accurate reaction prediction and mechanistic elucidation in organic synthesis. Starting with a minireview/perspective in Chapter 1, recent progress is discussed in data-rich approaches to reaction development and quantitative predictions for palladium-catalyzed reaction systems. In Chapter 2 and Chapter 3, quantitative predictive models are developed for two pharmaceutically important reaction systems: nucleophilic aromatic substitution (S_NAr) and oxidative addition to palladium(0), a fundamental and usually the rate/selectivity determining step in palladium-catalyzed cross-coupling reactions. Both models focus on structure-reactivity relationships of the electrophiles. Diverse and reliable reaction rate data for training set was collected using high-throughput competition experimentation. These were used to construct multivariate linear regression models by quantitatively mapping a group of ground state molecular descriptors to the experimental reaction rates. Predictive accuracy is validated via a series of random train-test splits, as well as predicting outcomes for a wide variety of external reaction data. Following the procedures described above, generally applicable models for quantitative predictions on both the reaction rates and site-selectivity for both reaction systems have been realized.

In addition to making quantitative reaction predictions, a structure-reactivity model constructed using high-quality data and mechanistically meaningful descriptors is also very useful in gaining mechanistic insights. This is demonstrated by the solvent effect study in Chapter 4 and the reaction mechanistic study in

Chapter 5. From the quantitative reactivity scales constructed for oxidative addition to palladium(0) in different solvents, specific electrophiles were identified that exhibit significant solvent effects; the role of solvent was investigated case by case. These include the importance of solvent hydrogen-bond basicity as well as solvent polarity. Finally, the underlying mechanistic causes behind a series of systematic prediction outliers from our oxidative addition model were investigated. These reveal that the frontier orbital symmetry also plays an important role in determining reaction outcomes. Insights into these mechanistic aspects, which have a significant impact on both the reaction rate and site-selectivity in oxidative addition to palladium(0), enabled a refined quantitative model that incorporates frontier orbital descriptors.

Table of Contents

Supervisory Committee ..... **ii**

Abstract ..... **iii**

Table of Contents ..... **v**

List of Figures ..... **x**

List of Tables ..... **xvii**

List of Abbreviations ..... **xviii**

Acknowledgements ..... **xx**

Chapter 1 – Organopalladium Catalysis as a Proving Ground for Data-Rich Approaches to Reaction Development and Quantitative Predictions **1**

1.1 Abstract **1**

1.2 Introduction **2**

1.3 Datasets From Existing Sources **4**

1.3.1 The Power (and Pitfalls) of Training on the Literature 4

1.3.2 Perspectives on the Available Data. 10

1.4 Data-Rich Computational Approaches **12**

1.4.1 High-Throughput Computation of Reaction Energy Profiles 12

1.4.2 Mapping Catalyst Chemical Space 16

1.5 Datasets from High-Throughput Experimentation **21**

1.5.1 Big Data on Demand via Miniaturized Array-Based Experiments. 21

1.5.2 Combining High-Throughput Experimental and Computational Approaches. 23

1.5.3 Quantitative Structure-Rate Relationships (QSRRs). 28

1.6 Conclusions and Outlook **36**

1.7 References **38**

Chapter 2 – A Broadly Applicable Quantitative Relative Reactivity Model for Nucleophilic Aromatic Substitution (S_NAr) Using Simple Descriptors **50**

2.1 Abstract **50**

2.2 Introduction **51**

2.3 Results and Discussion **56**

2.3.1 Creating the Training/Test Set 56
2.3.2 Model Generation and Performance. 59
2.3.3 Predicting Site Selectivity in Multihalogenated Heterocycles 63
2.3.4 External Case Study #1: S_NAr Rate Correlations 67
2.3.5 External Case Study #2: Site Selectivity Predictions 70
2.3.6 External Case Study #3: Complex Molecule Synthetic Planning 76
2.4 Conclusions 79
2.5 References 80
Chapter 3 – A Reactivity Model for Oxidative Addition to Palladium Enables Quantitative Predictions for Catalytic Cross-Coupling Reactions 93
3.1 Abstract 93
3.2 Introduction 94
3.3 Results and Discussion 97
3.3.1 Development of the Oxidative Addition Reactivity Model 97
3.3.2 Mechanistic Aspects of Oxidative Addition Linked to Molecular Descriptors 102
3.3.3 Case Study #1: Predicting Reaction Rates 104
3.3.4 Case Study #2: Predicting Site Selectivity 107
3.3.5 Case Study #3: Retrospective Synthesis Planning 111
3.4 Recent Progress on Pd Oxidative Addition Prediction 113
3.4.1 Reactivity Prediction of Aryl Chlorides in Pd Oxidative Addition 113
3.4.2 A Four-descriptor Model for Halobenzene Oxidative Addition 115
3.4.3 Reactivity of 2-Pyridyl Triflate in Oxidative Addition to Pd(0) 117
3.4.4 Reactivity of 2-Pyridyl Tosylate in Oxidative Addition to Pd(0) 118
3.4.5 Reactivity of (Hetero)aryl Iodide in Oxidative Addition to Pd(0) 120
3.5 Conclusions 123
3.6 References 124
Chapter 4 – Solvent Effects on Oxidative Addition to $Pd(PCy_3)_2$ 132
4.1 Abstract 132
4.2 Introduction 133
4.3 Results and Discussion 135
4.3.1 Oxidative Addition Reactivity Models in Different Solvents 135
4.3.2 Solvent Effect via Solvent Hydrogen Bond Basicity (pK_{HB}) 138

4.3.3 Influence of Solvent Polarity on Site-Selectivity 144
4.3.4 Mechanistic Study of Oxidative Addition of 2-Pyridyl Triflate and Solvent Effects on Triflate versus Chloride/Bromide Selectivity 145
4.4 Conclusions 154
4.5 References 155
Chapter 5 – Mechanism and Site-Selectivity of Halopyridine Oxidative Addition to $L_2Pd(0)$ Are Controlled by Frontier Molecular Orbital Symmetry 157
5.1 Abstract 157
5.2 Introduction 158
5.3 Results and Discussion 161
5.3.1 Mechanism of 2-Halopyridine Oxidative Addition Is Controlled by Frontier Orbital Symmetry 161
5.3.2 Investigating the Oxidative Addition Mechanism of 2-Chloro-6-EDG-Pyridines 169
5.3.3 Incorporating Frontier Orbital Based Descriptors into a Predictive Model 170
5.4 Conclusions 177
5.5 References 178
Chapter 6 – Conclusions and Future Work 182
6.1 Summary and Conclusions 182
6.2 Future Work 184
6.2.1 Quantitative Models for Amine Reactivity in Amide Coupling 184
6.2.2 Quantitative Model for Amine Reactivity in Buchwald-Hartwig Amination 185
6.2.3 Ligand Effects in Palladium-Catalyzed Reactions 187
6.3 References 189
Appendices 191
Appendix A: Supporting Information for Chapter 2 191
General Considerations 191
Experimental details for rate measurement by competition experimentation 192
Experimental Details of kinetic analysis for the touchstone reactions 197
Hammett Analyses of para- and meta-Substituted Substrates 199
Preparative Scale Synthesis of S_NAr products 202
Computational Determination of Molecular Descriptors 209
Construction of the Multivariate Linear Regression Model 210

Cross Validation and Out-of-Sample Prediction 214
LUMO energy as the Electron Affinity alternatives 216
Transition State Calculations of S_NAr by DFT 219
Experimental details for assessing outliers from external case studies (Fig. 2.9C) 225
Reference 225
Appendix B: Supporting Information for Chapter 3 227
General Considerations 227
Experimental Details for Oxidative Addition Competition Studies 228
Preparative Scale Synthesis of Oxidative Addition Complexes 237
Computational Determination of Molecular Descriptors 256
Hammett Analyses of para-Substituted Substrates 258
π-Complex Intermediates and Transition States for Oxidative Addition 263
Construction of the Multivariate Linear Regression Model 273
Cross Validation and Out-of-Sample Prediction 277
Including Ar-OTf substrates into the predictive model 281
Sonogashira Initial Rate Prediction Modelling 284
Site Selectivity Predictions 291
Experimental and computational data for the additional substrates in section 3.4 . 291
Cross Validation for the four-descriptor predictive model from Fig. 3.10B, main text. 291
Reference 293
Appendix C: Supporting Information for Chapter 4 295
General Considerations 295
Experimental Details for Intermolecular Oxidative Addition Competition Reactions 296
Experimental Details of reaction progress kinetic analysis for the touchstone reactions 296
Synthesis of Dichloropyridine 300
Experimental Details for Oxidative Addition Intramolecular Site-selectivity Determination 301
Structural Characterization of the Oxidative Addition Complexes from 2,6-dichloro-pyridin-3-ol 306
Computational Determination of Molecular Descriptors 314
Transition State Calculations for Oxidative Addition by DFT 317

Cross Validation for the three reactivity models in different solvents (Fig. 4.3) ... 319
Reference ... 323
Appendix D: Supporting Information for Chapter 5 ... 324
General Considerations ... 324
Experimental Details for Intermolecular Oxidative Addition Competition Reactions ... 324
Experimental Details of reaction progress kinetic analysis for the touchstone reactions ... 325
Experimental Details for Oxidative Addition Intramolecular Site-selectivity Determination ... 325
Synthesis of chlorinated pyridines ... 326
Preparative Scale Synthesis and structural elucidation of the Oxidative Addition Complexes from the substrates in Fig. 5.6 ... 337
Structural elucidation of the Oxidative Addition Complexes from the substrates in Fig. 5.7 ... 368
Computational Determination of Molecular Descriptors including the %lumo coefficients ... 383
Transition State Calculations for Oxidative Addition by DFT ... 384
Cross Validation for the six-descriptor predictive model from Fig. 5.9B (main text) ... 389
References ... 391

List of Figures

Figure 1.1. Data foundations for building quantitative models in catalyst reactivity and selectivity predictions. 4

Figure 1.2. Comparison of frequency of reported yield values for Buchwald-Hartwig coupling reactions taken from the literature (Reaxys database) or from an HTE-based dataset. Values obtained from ref. 27. 6

Figure 1.3. Diversity analysis of reported reaction conditions for C–N couplings from a combination of three databases (CAS, Reaxys, and USPTO). Plots represent % of reported conditions (y-axis) that contain the top N settings (x-axis). For example, ~80% of reactions in the complete database use just the top two solvents (top N = 2). Reproduced with permission from ref. 29. Copyright 2023, American Chemical Society. 9

Figure 1.4. Molecular volcano plot of the calculated thermodynamics of oxidative addition (Rxn A), transmetallation (Rxn B) and reductive elimination (Rxn C) for a series of Ni, Pd, Pt, Cu, Ag, and Au catalysts Optimal performance is achieved by Pd-based catalysts, which have intermediate exergonicities of oxidative addition as compared to other transition metals Reproduced with permission from ref. 46. Copyright 2021, American Chemical Society. 14

Figure 1.5. A comparison of two monodentate phosphine databases: A) LKB-P; and B) Kraken, with representative phosphines shown. Each database is represented by two-dimensional principal component analysis (PCA) plot. In the LKB-P, PC1 is predominantly electronic in nature, while PC2 is predominantly steric. In the Kraken visualization, PC1-PC4 have the indicated contributors. Figure 1.5A reproduced with permission from ref. 64, copyright 2010, American Chemical Society. Figure 1.5B reproduced with permission from ref. 66, copyright 2022, American Chemical Society. 18

Figure 1.6. A) Hayashi-Heck coupling required in high yield, enantioselectivity, and regioselectivity. B) Step-wise multi-objective optimization via a data-rich computational / HTE approach, with univariate threshold analysis for activity, followed by MLR analysis for regioselectivity, and finally virtual screening for extrapolation to superior catalyst. C) Improved outcome with new catalyst system. Adapted with permission from ref. 108. Copyright 2023, American Chemical Society. 27

Figure 1.7. A) Generic catalytic cycle for Pd-catalyzed cross-coupling, highlighting the importance of oxidative addition. B) Experimental approach to rapid relative rate collection via competition experiments, with analysis of L2Pd(Ar)(X) complexes by 31P qNMR spectroscopy. C) Reactivity scale for representative substrates. Figure adapted with permission from ref. 47. Copyright 2022, Royal Society of Chemistry. 31

Figure 1.8. Multivariate QSRR model for Ar–X oxidative addition to Pd(PCy3)2. Figure adapted with permission from ref. 47. Copyright 2022, Royal Society of Chemistry. 33

Figure 1.9. Applying predicted $\Delta G^{\ddagger}_{OA}$ values to external catalytic reaction datasets: unified QSRR of Plenio and co-workers' Sonogashira initial rates for all substrate/catalyst combinations. Figure adapted with permission from ref. 47. Copyright 2022, Royal Society of Chemistry. 35

Figure 2.1. Examples of APIs prepared by multi-step synthesis that involves at least one S_NAr reaction (nucleophilic atom and bond formed by S_NAr highlighted in red). **Bottom**: Top 10 most used reactions in medicinal chemistry in 2014 ranked by frequency.[58] 52

Figure 2.2. Approaches to developing quantitative structure-reactivity relationships (QSRR) for S_NAr reactions. **A)** Empirical parameters derived from experimental data. **B)** Calculated descriptors from DFT analysis (FMO = frontier molecular orbital theory; TS = transition state). **C)** Recent hybrid DFT/ML approach. **D)** Bottom-up approach combining new experimental data with simple calculated descriptors. 53

Figure 2.3. Experimental approach to collecting free energies of activation for 74 S_NAr reactions; Bn = benzyl. **A)** Touchstone reaction progress analysis under *pseudo* first order conditions. **B)** Competition experiments to establish relative rates across electrophile library. **C)** Representative primary data for determining $\Delta\Delta G^{\ddagger}_{SNAr}$ from competition experiments. **D)** Quantitative reactivity scale for representative electrophiles. 58

Figure 2.4. Quantitative model generation and performance. **A)** Molecular descriptors used in multivariate regression analysis, with percent contribution determined by min/max normalization. **B)** All data linear regression analysis for experimental versus predicted $\Delta G^{\ddagger}_{SNAr}$ with accompanying statistics (MAE = mean absolute error); linear correlation uses non-normalized descriptors. **C)** One of five 60/40 training/test validations, with accompanying statistics. **D)** Predicted versus residuals plot for the 74 data points, with accompanying box plot (right); one outlier is identified ($|R| > 5$ kJ mol^{-1} red point with accompanying structure)..... 60

Figure 2.5. Importance of ESP_2 descriptor in predicting $\Delta G^{\ddagger}_{SNAr}$ for multiple substrate classes. **A)** Bivariate model incorporating only *EA* and ESP_1 descriptors, with two sets of outliers highlighted. **B)** Comparison of substrate pairs with very similar *EA* and ESP_1 values but significantly different $\Delta G^{\ddagger}_{SNAr}$ values, revealing the importance of ESP_2 in differentiating reactivity. *ESP* maps for each substrate structure are shown, with colour gradient indicating local *ESP* (red = maximum positive; green = 0; blue = maximum negative)..... 63

Figure 2.6. Site selectivity in multihalogenated heterocycles that are part of the training set. LUMO+1 energies are approximated by subtracting the LUMO/LUMO+1 energy gap from the *EA* value for the substrate. 64

Figure 2.7. A) Calculated energy diagram for S_NAr of 2,4-dichloropyridine and the anion of benzyl alcohol in DMSO. The blue pathway is the S_NAr coordinate at C_4, the major reactive site; and the red pathway is the S_NAr coordinate at C_2, the minor reactive site. **B)** FMO analysis of S_NAr selectivity with 2,4-dichloropyridine, revealing orbital symmetry effects in the substrate (LUMO versus LUMO+1) and transition states (HOMO contributions from *ortho* and *para* sites). 65

Figure 2.8. Calculated S_NAr energy diagrams of **A)** 2-chloro-3-fluoropyridine and **B)** 2-chloro-4-fluoropyridine at both reactive sites. The blue pathway is the S_NAr at the major site (Py-F); and the red pathway is the S_NAr at the minor site (Py-Cl). The concerted mechanism is found in all cases. 67

Figure 2.9. Model validation through assessing correlations between experimental $\Delta G^{\ddagger}$ values and predicted $\Delta G^{\ddagger}_{SNAr}$ for three external data sets (A, B, and C)..... 69

Figure 2.10. Site selectivity predictions and rate correlation for S_NAr between fluorinated arenes and ammonia. Experimental data from ref [112]. 70

Figure 2.11. Site selectivity predictions and rate correlation for S_NAr between fluorinated arenes and methoxide. Experimental data from ref[113]. 71

Figure 2.12. Site selectivity predictions and rate correlation for SNAr between fluorinated heterocycles and ammonia. Experimental data from refs [112–114]. 72

Figure 2.13. Qualitative site selectivity predictions for combinations of (hetero)aryl halides with anionic (A and C) and neutral (B and D) nucleophiles, and for mixed halide aromatics (E). 73

Figure 2.14. Site selectivity predictions using examples from reference [80]. The original experimental data was obtained from the public USPTO database. 75

Figure 2.15. Example applications of S_NAr predictions to route development for investigational API synthesis, including regioselectivity for specific substrates, and comparison of potential substrate regioselectivity/reactivity. 78

Figure 3.1. (A) Simplified cross-coupling mechanism, highlighting oxidative addition as the rate and/or selectivity determining step. **(B)** Competition experiment approach to determining relative rates of oxidative addition by quantifying ratios of Pd(II) products via ^{31}P NMR spectroscopy. **(C)** Relative reactivity scale for oxidative addition to $Pd(PCy_3)_2$ for selected substrates; observed $\Delta G^{\ddagger}_{OA}$ for 2-bromo-5-nitropyridine set to 0 kJ mol^{-1}. **(D)** Utility of reactivity model in predictions for cross-coupling in synthesis. 96

Figure 3.2. Design and performance of a quantitative reactivity model for oxidative addition to Pd(0). **(A)** General mechanism for oxidative addition to $L_nPd(0)$, with π-complex intermediate preceding either Pd insertion into C–X bond, or an S_NAr-like displacement of X. **(B)** Molecular descriptors used to model oxidative addition reactivity as a function of substrate structure. **(C)** Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 70 Ar–Cl and Ar–Br substrates in THF, including all data points in regression analysis. **(D)** Representative multivariate linear regression model generated using a 60/40 training/test split. **(E)** Univariate plot of $IBSI_{C-X}$ versus $\Delta G^{\ddagger}_{OA}$ for Ar–Cl, Ar–Br, and Ar–OTf, revealing that bond strength is poorly correlated to $\Delta G^{\ddagger}_{OA}$ within each leaving group set. **(F)** Unified linear regression model of $\Delta G^{\ddagger}_{OA}$ for Ar–Cl, Ar–Br, and Ar–OTf substrates in THF, including all data points in regression analysis. MAE = mean absolute error. Colour-coding on R^2, Q^2, and MAE values corresponds to the matching data subset, values in black are for all data. 99

Figure 3.3. Electronic and steric features of oxidative addition. **(A)** ESP_{Pd} for calculated π-complex intermediate structures correlates with oxidative addition rates; structures for 7 of 11 examples shown; electrostatic potential maps for each intermediate are overlaid onto the line structures. **(B)** Calculated structures of π-complex intermediates reveal how steric strain induced by R_1 and R_2 (here, $–CF_3$ groups) in 2-halopyridines affect oxidative addition reactivity in equal proportions. 103

Figure 3.4. Translating oxidative addition predictions to quantitative models of catalytic reactivity. **(A)** General reaction scheme and chemical space explored for 410 Sonogashira reactions, with two distinct substrate sets; initial rates determined previously.[56,57] **(B)** Univariate linear correlations between predicted $\Delta G^{\ddagger}_{OA}$ for oxidative addition to $Pd(PCy_3)_2$ and ln *k* for Sonogashira coupling with three phosphines; out-of-model substrates are Ar–Br molecules not included in $\Delta G^{\ddagger}_{OA}$ training set. **(C)** Unified three-descriptor model for predicting ln *k* for the

entire set of 410 reactions (29 substrates, 17 ligands), with data partitioned into training (60% of set #1), test (40% of set #1), and external validation (set #2); two external outlier points (red) are not included in the external validation statistics. **(D)** Subset of the model with 13 "small" phosphines (%V_{bur} < 75). (E) Subset of the model with 4 "large" phosphines (%V_{bur} > 75); two external outlier points (red) are not included in the external validation statistics. MAE = mean absolute error. Colour-coding on R^2, Q^2, and MAE values corresponds to the matching data subset, values in black are for all data. 105

Figure 3.5. Predicted and reported selectivities for multihalogenated heterocycles in Suzuki-Miyaura and Buchwald-Hartwig cross-coupling reactions (examples of Buchwald-Hartwig substrates denoted with "BH"). Coloured labels on the heterocycles correspond to predicted major site (blue), predicted minor site (red, along with percentage of exceptions as reported in ref. [51]), and observed site (purple sphere). The magnitude of $\Delta\Delta G^{\ddagger}_{OA}$ between the two sites is given in green. Contents circled in dashed blue line are the predicted and observed oxidative addition rates from my experimental data. 108

Figure 3.6. (A) Quantitative selectivity predictions for dihalogenated heterocycles with small-to-medium $\Delta\Delta G^{\ddagger}_{OA}$ between two sites, and observed product ratios. **(B)** Predictions for substrates with observed tunable selectivity, demonstrating that "simple" catalysts are quantitatively consistent with predicted selectivities; overriding predicted reactivity requires targeted screening and/or catalyst design. Coloured labels on the heterocycles correspond to predicted major site (blue), predicted minor site (red, along with percentage of exceptions as reported in ref. [51]), and observed site (purple sphere). The magnitude of $\Delta\Delta G^{\ddagger}_{OA}$ between the two sites is given in green. 110

Figure 3.7. Retrospective analysis of applying $\Delta G^{\ddagger}_{OA}$ predictions to synthesis design: dragmacidin D. **(A)** Retrosynthesis of core structure, involving selective fragment coupling to a dihalogenated pyrazine. **(B)** Approach involving differential halogenation, tosylate protecting group on indole 1, and switch from 5-chloro to 5-bromopyrazine intermediates for selective coupling.[74] **(C)** Approach involving regioselective coupling to dibromopyrazine, TBS protecting group on indole 1, and regioselective Stille coupling.[75] 112

Figure 3.8. Performance of the quantitative model for oxidative addition of aryl chlorides to Pd(0). **(A)** Structures of the 10 4-substituted aryl chlorides in my reaction dataset. **(B)** Multivariate linear regression model using all data points except for the 10 aryl chlorides for training set, and the 10 aryl chlorides for external test set. **(C)** Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 89 Ar–Cl, Ar–Br and Ar–OTf substrates in THF, including all data points in regression analysis. MAE = mean absolute error. Colour-coding on R^2, Q^2, and MAE values corresponds to the matching data subset, values in black are for all data. 115

Figure 3.9. Performance of the four-descriptor (*ESP_1, A_1, IBSI,* pK_a) model for oxidative addition of the halobenzene subset. **(A)** Structural features included in the 43-member halo-benzene subset. **(B)** Multivariate linear regression model using all the 43 data points training set. **(C)** Residual plot for the predicted $\Delta G^{\ddagger}_{OA}$ using the equation from Fig. 3.9B. **(D)** Multivariate linear regression model generated using a 70/30 training/test split. MAE = mean absolute error. Colour-coding on R^2, Q^2, and MAE values corresponds to the matching data subset, values in black are for all data. 116

Figure 3.10. Prediction of the relative oxidative addition rate of 2-pyridyl-triflate to Pd(0) and the DFT calculated oxidative addition mechanism. **(A)** Relative rate prediction using the five-

descriptor model from Fig. 3.8C. **(B)** Relative rate prediction using the four-descriptor model from Fig. 3.9B. **(C)** DFT calculated reaction mechanism of 2-pyridyl-triflate oxidative addition to $Pd(0)(PCy_3)_2$ in THF. TS and INT search, Gibbs energy correction at RI BP86/def2-SVP, def2/J, D3BJ/def2-TZVP(Pd)/CPCM; Single point energy at RI-B2PLYP/def2-TZVP, def2-TZVP/C, D3/CPCM. **(D)** Calculated transition state with bond length (Pd-N, Pd-C and Pd-O) and bond angle (Pd-C-O) measurements. 118

Figure 3.11. Relative rate prediction for oxidative addition of 2-pyridyl tosylate to $Pd(PCy_3)_2$ using the five-descriptor model from Fig. 3.8C. 119

Figure 4.1. Summary of the solvent effect cases discussed in this Chapter. **A)** Multivariate linear regression models of $\Delta G^{\ddagger}_{OA}$ in THF, 1:1 THF/DMF and toluene. **B)** Influence of solvent hydrogen bonding on oxidative addition outcomes, here shows the site-selectivity of 2,6-dichloro-3-aminopyridine as a representative example. **C)** Solvent effects on site-selectivity via one type of mechanism being favoured over another, here shows 2,6-dichloro-3-nitropyridine as an example. **D)** Mechanistic study on oxidative addition of 2-pyridyl triflate and solvent effect on site-selectivity between a triflate and a chloride/bromide site, here shows 2-chloropyridyl-3-triflate as an example. 134

Figure 4.2. Touchstone reaction progress analysis under pseudo first order conditions in toluene. The $\Delta G^{\ddagger}_{OA}$ of 2-chloro-6-methylpyridine **(left)** was used as the calibration point to obtain the absolute rates of the entire array of substrates, and the $\Delta G^{\ddagger}_{OA}$ of 2-chloro-4-methylpyridine **(right)** was used to confirm the validity of the competition determined $\Delta G^{\ddagger}_{OA}$ for this substrate. 136

Figure 4.3. A) Molecular descriptors used to model oxidative addition reactivity as a function of substrate structure. **B)** Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 97 substrates in THF. **C)** Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 49 substrates in 1:1 THF/DMF. **D)** Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 50 substrates in toluene. **E)** relative contribution of each descriptor to the predictive accuracy of the 3 multivariate linear regression models. 137

Figure 4.4. Experimental rates in three solvents for oxidative addition of the 11 overlapping Ar-Br substrates, including five 4-substituted bromobenzenes and five substituted 2-bromopyridines. 138

Figure 4.5. Experimental rates of the 12 overlapping substituted 2-chloropyridine derivatives oxidative addition to $Pd(PCy_3)_2$ in toluene, THF and 1:1 THF/DMF. 139

Figure 4.6. Oxidative addition of 2-chloro-3-aminopyridine to $Pd(PCy_3)_2$ via the nucleophilic displacement mechanism. 139

Figure 4.7. A) Experimental $\Delta G^{\ddagger}_{OA}$ for oxidative addition of 2-chloro-3-aminopyridine, 2-chloro-3-hydroxypyridine and 2-chloropyridine to $Pd(PCy_3)_2$ in toluene, THF and 1:1 THF/DMF. **B)** Intermolecular H-bond between DMF and the $–NH_2$ group, and its deactivating effect on the C_2 site. **C)** Computational analysis of the solvent hydrogen bond basicity (SHBB) effect on the *ESP* at the C_2 site; ΔESP_{C2} values are the difference between ESP_{C2} of the substrate-solvent H-bond complex in the CPCM solvation model and ESP_{C2} of the substrate itself in the CPCM solvation model. **D)** *ESP* isosurfaces at 0.002 au of 2-chloro-3-aminopyridine and the 2-chloro-3-aminopyridine–DMF H-bond complex; blue represents negative *ESP*, and red represents positive *ESP*. 141

Figure 4.8. A) Experimental C_2/C_6 selectivity for oxidative addition of 2,6-dichloro-3-EDG-pyridines (EDG = NH_2, NHCOMe, OH) to $Pd(PCy_3)_2$ in different solvents (>> indicates only C_2 is observed). **B)** Computational analysis of the solvent hydrogen bond basicity (SHBB) effect on the electrostatic potential (*ESP*) at the C_2 and C_6 sites; the Δ*ESP* values at C_2 and C_6 are the difference between the *ESP* values of the substrate- solvent H-bonding complex in the CPCM solvation model and the *ESP* values of the substrate in the CPCM solvation model. **C)** *ESP* isosurfaces at 0.002 au of 2,6-dichloro-3-hydroxypyridine, the substrate–THF complex, and the substrate–DMF complex; blue represents negative *ESP*, and red represents positive *ESP*. 143

Figure 4.9. A) Observed C_6/C_2 selectivity of 2,6-dichloro-3-EWG-pyridines (EWG = CN, NO_2, CHO) in different solvents. **B)** A proposed hypothesis that the increased C_6/C_2 selectivity in polar solvents is caused by change in mechanism at the two sites; this will be further discussed and proved in Chapter 5. 145

Figure 4.10. Experimental rates of the 8 overlapping para-substituted aryl triflates oxidative addition to $Pd(PCy_3)_2$ in toluene, THF and 1:1 THF/DMF 146

Figure 4.11. Solvent effect on C-OTf vs. C-Cl selectivity reported in literature. 148

Figure 4.12. Free energies of the oxidative addition product of 2-pyridyl triflate in DMF (CPCM) in the form of cationic Py-$Pd(PCy_3)_2^+$, uncharged trans-Py-$Pd(OTf)(PCy_3)_2$ and solvent-coordinated trans-Py-$Pd(DMF)(PCy_3)_2^+$. 149

Figure 4.13. A) Calculated energy diagram for oxidative addition of 2-chloropyridine to $Pd(PCy_3)_2$ in DMF, THF and toluene. **B)** Calculated energy diagram for oxidative addition of 2-pyridyl triflate to $Pd(PCy_3)_2$ in DMF, THF and toluene. The blue/red/green pathways are the calculated reaction coordinate in DMF, THF and toluene, respectively. Transition state geometries were optimized in implicit solvation model CPCM at RI BP86 def2-SVP def2/J level with D3BJ dispersion for all atoms except for Pd, for which a def2-TZVP basis set was used; and the single point energy in implicit solvation model CPCM at RI-B2PLYP D3 def2-TZVP def2-TZVP/C level. 151

Figure 5.1. Representative literature examples of directing group pre-coordination in palladium-catalyzed cross-coupling reactions. 159

Figure 5.2. Systematic discrepancy between predicted and experimental oxidative addition relative rates (given as $\Delta\Delta G^{\ddagger}_{OA}$, where $\Delta\Delta G^{\ddagger}_{OA} = \Delta G^{\ddagger}_{OA}$ (2-Cl-5-EDG-py) – $\Delta G^{\ddagger}_{OA}$ (2-Cl-3-EDG-py)) between three related substrate pairs: 2-chloro-3/5-EDG-pyridine (EDG = NH_2, OMe, F). 162

Figure 5.3. Calculated energy diagram for oxidative addition of 2-chloro-3/4/5-NH_2-pyridine to $Pd(PCy_3)_2$ in THF. The blue/red/green pathways are the oxidative addition coordinates of 3-NH_2, 4-NH_2 and 5-NH_2, respectively. Transition state geometries were optimized with CPCM(THF) at RI BP86 def2-SVP def2/J level with D3BJ dispersion for all atoms except for Pd, for which a def2-TZVP basis set was used; and the single point energy at RI-B2PLYP D3 def2-TZVP def2-TZVP/C level with CPCM(THF). 164

Figure 5.4. Key frontier orbitals of the $Pd(PCy_3)_2$ and the -NH_2 group that are involved in oxidative addition. 166

Figure 5.5. Experimental $\Delta G^{\ddagger}_{OA}$ values for 3-EDG / 5-EDG pairs in oxidative addition to $Pd(PCy_3)_2$. 166

Figure 5.6. A) *Top:* LUMO features and observed site-selectivity for 2,6-dichloro-3-EDG-pyridines/pyrazine. *Bottom:* Solid-state molecular structures (via X-ray crystallography) of the major oxidative addition products. **B)** LUMO features and observed site-selectivity for 2,6-dichloro-3-EWG-pyridines/pyrazine 168

Figure 5.7. LUMO features and observed site-selectivity of 2-chloro-5-bromo-3-EDG-pyridines. 169

Figure 5.8. A) Molecular descriptors used to model oxidative addition reactivity as a function of substrate structure; examples of the LUMO-based descriptors for bonding (2-chloro-4-aminopyridine) and antibonding (2-chloro-3-aminopyridine) symmetries through the C=N bond. 172

Figure 5.9. Quantitative selectivity predictions for dihalogenated heterocycles reported in literature. Coloured labels on the heterocycles correspond to predicted major site (blue), predicted minor site (red). The magnitude of the predicted $\Delta\Delta G^{\ddagger}_{OA}$ between the two sites is given in green. 175

Figure 6.1. left: competition experiments for amide coupling to establish relative rates across amine library. right: representative amine structures from amine library. 185

Figure 6.2. top: limited substrate scope for S_NAr: only the extremely electron deficient heteroarenes form C-N bond with amines via S_NAr. bottom left: synthesis of the oxidative addition product from the $Pd^{(II)}$ precursor DAB-Pd-$(CH_2TMS)_2$. Bottom right: competition experiments for Buchwald-Hartwig coupling to establish relative rates across amine library. ... 186

Figure 6.3. ^{31}P NMR spectra for oxidative addition of 2-chloro-5-bromopyridine (green), 2,6-dichloro-3-fluoro-5-cyanopyridine (red), 2-chloro-5-bromo-3-fluoropyridine (blue) to Pd(tBu$_3$)$_2$ in THF at room temperature. 188

List of Tables

Table 3.1. Identified outliers from the regression model in Fig. 3.2F. 101

Table 3.2. Linear regression coefficients and % contribution to predicted $\Delta G^{\ddagger OA}$ for min/max scaled descriptors. 102

Table 3.3. $\Delta G^{\ddagger}_{OA}$ prediction for the seven iodide electrophiles using the equation from Fig. 3.8C. 120

Table 3.4. $\Delta G^{\ddagger}_{OA}$ prediction for the seven iodide electrophiles using the equation from Fig. 3.9B. 121

Table 3.5. $\Delta G^{\ddagger}_{OA}$ prediction after including iodide electrophiles in the training set. 121

Table 3.6. Linear regression coefficients %contribution (min/max normalization). 122

Table 4.1. Experimental and predicted site-selectivity for oxidative addition of halopyridyl triflates to $Pd(PCy_3)_2$ in three solvent systems. 153

Table 5.1. LUMO features of the 2-chloro-6-EDG-pyridines; experimental $\Delta G^{\ddagger}_{OA}$ and computed $\Delta E^{\ddagger}_{OA}$ of the 3,4,5,6-EDG substrates in each group (EDG = NH_2, OMe and F). 170

Table 5.2. Linear regression coefficients and % contribution to predicted $\Delta G^{\ddagger}_{OA}$ for min/max scaled descriptors. 172

Table 5.3. Comparison of the prediction accuracy on relative rates ($\Delta\Delta G^{\ddagger}_{OA} = \Delta G^{\ddagger}_{OA}$ (5-EDG) - $\Delta G^{\ddagger}_{OA}$ (3-EDG)) using the model in Fig. 5.2 with the model in Fig. 5.8B. 173

Table 5.4. Comparison of the prediction accuracy on site-selectivity using the models before and after the LUMO features are included. 173

Table 5.5. Comparison of the prediction accuracy on site-selectivity using the models before and after the LUMO features are included. 176

List of Abbreviations

$\%V_{bur}$ ..... Percent buried volume
Ad ..... Adamantyl group
API ..... Active pharmaceutical ingredient
Ar ..... Aromatic ring
BDE ..... Bond-dissociation energy
Bn ..... Benzyl
CC BY-NC-ND ..... Attribution-NonCommercial-NoDerivs
CPCM ..... Conductor-like polarizable continuum model
Cy ..... Cyclohexyl group
DAB ..... diazabutadiene
DFT ..... Density functional theory
DMF ..... Dimethylformamide
DMSO ..... Dimethylsulfoxide
DPPF ..... 1,1′-Bis(diphenylphosphino)ferrocene
EA ..... Electron affinity
EDG ..... Electron donating group
ESP ..... Electrostatic potential
EWG ..... Electron withdrawing group
FMO ..... Frontier molecular orbital theory
GC ..... Gas chromatography
HMBC ..... Heteronuclear Multiple Bond Correlation
HOMO ..... Highest occupied molecular orbital
HPLC ..... High-performance liquid chromatography
HRMS ..... High-resolution mass spectrometry
HSQC ..... Heteronuclear Single Quantum Coherence
HTE ..... High-throughput experimentation
IBSI ..... Intrinsic bond strength index
L ..... Ligand
LCMS ..... Liquid chromatography–mass spectrometry
LUMO ..... Lowest unoccupied molecular orbital

MAE ..... Mean absolute error
ML ..... Machine learning
MLR ..... Multivariate linear regression
NHC ..... N-Heterocyclic Carbene
NMP N-Methyl-2-pyrrolidone
NMR ..... Nuclear magnetic resonance
OTf ..... Trifluoromethanesulfonate
OTs ..... Toluenesulfonate
PCA ..... Principal component analysis
Ph ..... Phenyl group
QSAR ..... Quantitative structure–activity relationship
QSRR ..... Quantitative structure-rate relationship
QSSR ..... Quantitative Structure-Selectivity Relationships
R ..... Functional group
R&D ..... Research and Development
S_NAr ..... Nucleophilic aromatic substitution
*t*Bu ..... tert-butyl
THF ..... Tetrahydrofuran
TMS ..... Trimethylsilyl group
TST ..... Transition state theory
UPLC ..... Ultra-performance liquid chromatography
USPTO ..... United States Patent and Trademark Office
X ..... Halogen
XRD ..... X-ray diffraction
$\Delta G^{\ddagger}$ ..... Free energies of activation
$\Delta\Delta G^{\ddagger}$ ..... Relative free energies of activation

Acknowledgements

I started my PhD in 2019 after a few years of working in industry, at that time I was both excited for the upcoming new journey yet a bit uncertain for making transition from workplace back to the academic environment. Looking back, I've had an incredible time during the past four years at UVic and there are a number of great things worth to mention here; off the top of my list, the best part of doing PhD is being able to spend the entire four years working for myself – to learn what I need, to do research I'm interested in and to watch myself making improvements.

First and foremost, I would like to thank my supervisor Dr. David Leitch for his constant support and encouragement throughout my entire PhD career. It has been such a great experience working with him, I have not only grown academically but also became a lot more confident than before. I'm truly thankful for his guidance on my research projects and the time he spent on refining my writing.

I would like to thank my co-supervisor Dr. Irina Paci for guiding me through the theory and applications of computational chemistry, a branch of chemistry that intimidated me previously. I highly appreciate her patience and guidance in helping with my calculations during our weekly meetings.

I would also like to thank my committee members Dr. Fraser Hof and Dr. Nishant Mehta; it was great meeting you during the online and in-person meetings. Thanks to all my committee members for taking time reading this nearly 400 pages thesis.

A huge thank you goes to everyone from the Leitch group. I feel fortunate being part of a group where everyone is so nice and supportive; thanks for everyone for your support and help, and all the good times we spent in and outside the lab throughout the four years. I'm also grateful to Christopher Barr, Ori Granot and Tyler Trefz for NMR and LCMS support, as well as everyone in the UVic chemistry department who assisted me with my research.

I want to thank my parents for everything, their support gives me the freedom to do what I want. Thanks to all my friends, especially two of my best friends: Ma Jie, we became friends since primary school and Lin Chen, who always shares interests with me; there's never been a stupid or boring topic when we chat.

Special thanks goes to Hehua, the prettiest and cutest panda ever existed. Watching her video clips has helped a lot maintaining my emotional stability, especially during the last several months of writing. It is so peaceful and satisfying watching her sitting in the sunshine eating bamboo, making me feel that pure happiness does exist.

Again, thanks to everyone who helped and supported me during this invaluable and wonderful period while I was doing PhD at UVic!

Chapter 1 – Organopalladium Catalysis as a Proving Ground for Data-Rich Approaches to Reaction Development and Quantitative Predictions

This chapter has been adapted from:

Lu J.; Leitch D. C. Organopalladium Catalysis as a Proving Ground for Data-Rich Approaches to Reaction Development and Quantitative Predictions. *ACS Catal.* **2023**, 13, 24, 15691–15707, DOI: 10.1021/acscatal.3c03864

1.1 Abstract

With the advent of high-throughput methods for both computation and experimentation, data-rich approaches to discovering and understanding chemical reactions are becoming ever more central to catalysis research. Organopalladium catalysis is at the forefront of these new approaches, providing a rich proving ground for method development and validation. This critical Perspective discusses a number of recent case studies from academic and industrial laboratories that illustrate how to generate, analyze, and correlate large data sets for quantitative predictions of reactivity and selectivity. Both the power and potential pitfalls of these approaches are discussed, as are the opportunities for both practical predictions and fundamental mechanistic insights.

1.2 Introduction

Organopalladium catalysis, and in particular palladium-catalyzed coupling, continues to be indispensable for fragment coupling in organic synthesis. Its success as a catalytic method for forming all manner of carbon-element bonds is unparalleled, especially in pursuit of complex molecule targets. Pd-catalyzed reactions are (and continue to be) central to the synthesis of myriad natural products,[1–3] active pharmaceutical ingredients,[4–6] agrochemicals,[7] and organic materials.[8] While the data was collected nearly 10 years ago, Brown and Boström's analysis of reaction classes most often used in medicinal chemistry has 3 of the top 20 as Pd-catalyzed: Suzuki-Miyaura, Sonogashira, and Buchwald-Hartwig.[6] No other examples of homogeneous organometallic catalysis are in this set. Certainly, the plethora of excellent research into non-precious metal catalysis for cross-coupling and related transformations means that these systems are emerging as viable alternatives.[9–13] However, the sheer number and variety of documented examples of Pd catalysis in complex molecule synthesis attests to its enduring importance as a reliable and effective means to access new chemical matter.

As a direct result of its fairly unique combination of wide applicability, highly variable reaction conditions, and massive amounts of published data, organopalladium catalysis is also one of the most frequently studied reaction classes in emerging data-rich methods for reaction discovery, optimization, and prediction. A 2018 survey of practitioners reveals catalytic C–C and C–N coupling as 2 of the top 3 most frequently screened reaction types by high-throughput experimentation (HTE) groups within (or affiliated with) pharmaceutical R&D.[14] And a recent comprehensive review of HTE in organometallic chemistry and catalysis from 2006-2020 reveals cross-coupling (sum of all varieties) as the most frequently reported reaction type undergoing high-throughput screening.[15] This is a stark contrast to the prior 10 years, where HTE for asymmetric hydrogenation and alkene polymerization were overwhelmingly more common.[16] Finally, due to its clear dominance as a preferred method in medicinal chemistry, the patent literature contains millions of unique examples of Pd-catalyzed transformations across a wide swath of chemical space.

Thus, organopalladium catalysis is an ideal proving ground for emerging data-rich approaches to solving chemical synthesis problems. The recent ultrafast growth of

computing and data analysis power combined with significant advancements in laboratory automation, miniaturization, and rapid chemical analysis have paved new ways to study and understand chemical systems. In organic synthesis specifically, data-driven methods have made a significant impact on accelerating the process of reaction optimization[17–19] and expanding the applicability of multivariate quantitative structure-reactivity/selectivity prediction models for synthesis planning.[20–25] A number of literature reviews published in recent years have covered a wide range of applications of the data-rich approaches in the field of organopalladium catalysis.[26–30]

Large, accurate, consistent, and comprehensive datasets from reliable sources are the foundation of any data-driven methodology, and often the determining factor for overall success. The data foundation for studies of organopalladium catalysis is built from a combination of three major sources (Figure 1.1):

1) Previously reported reaction conditions/outcomes obtained directly from the academic/patent literature, and/or from open-source/proprietary reaction databases;
2) Calculated reaction-based parameters, including structurally and/or mechanistically-relevant molecular descriptors and computed reaction energy barriers;
3) "Do-it-yourself" experimentally determined reaction data, ideally collected via automated and/or HTE approaches.

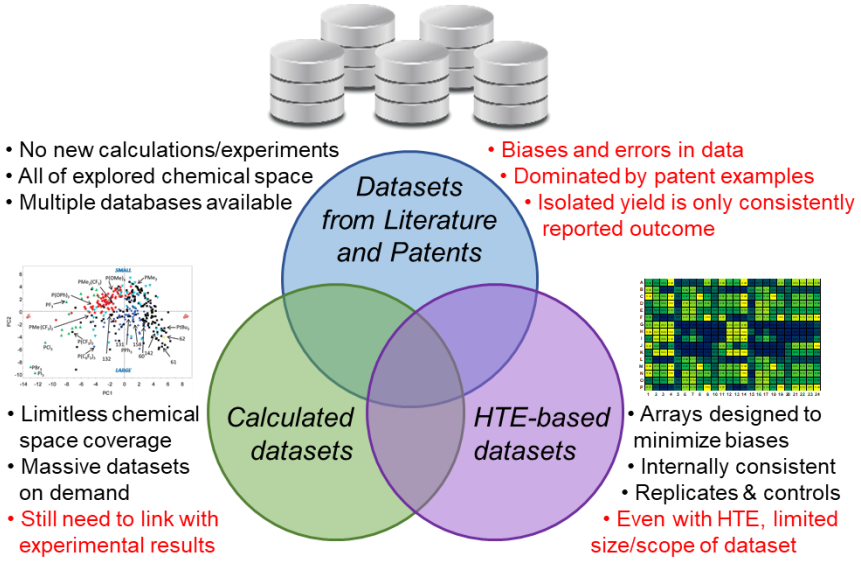


Figure 1.1. Data foundations for building quantitative models in catalyst reactivity and selectivity predictions.

This critical Perspective will cover key aspects of these three approaches to data collection and analysis, and the resulting outcomes. Illustrative examples from organopalladium catalysis show how these approaches are continuing to evolve. There are, of course, advantages and challenges specific to each approach; it is clear from the success stories thus far that a combination of approaches is the ideal strategy. In addition to accelerating catalyst and/or reaction discovery and optimization, taking a data-rich approach to studying chemical synthesis has the potential to significantly expand our understanding of the interplay between molecular structure, experimental conditions, and reaction mechanisms.

1.3 Datasets From Existing Sources

1.3.1 The Power (and Pitfalls) of Training on the Literature

Due to its major role in pharmaceutical synthesis for both discovery and manufacturing activities, there is an abundance of reported examples for Pd-catalyzed cross-coupling across an extremely broad range of chemical space. The details of these reactions are spread throughout the academic and (especially) patent literature, and are readily accessible from major chemical databases and platforms. From a data science and statistical modeling perspective, this appears to be an ideal situation: a large, diverse, and available dataset

containing “real-world” reaction systems to feed data-hungry but powerful machine learning algorithms.

Unfortunately, as noted by several experts in the field,[31–34] machine learning predictive models trained exclusively by literature/patent-derived data can suffer from several issues, including low prediction accuracy and lack of generalizability. Reaction outcome data from the literature, despite its abundance, is skewed by the objectives of each individual practitioner. A medicinal chemist making structural analogues during a lead-optimization campaign is not concerned with achieving comprehensive reaction condition coverage or collecting highly accurate yield/rate data; their job is to rapidly prepare target compounds by any means necessary. Likewise, an academic chemist developing new synthetic methods is not interested in reporting a large number of “failed” reactions; their job is to demonstrate the utility and scope of the new method. In other words, the experiments that produced the existing data were not designed or executed with an eye toward predictive statistical modeling, and thus great care must be taken before using these inputs. Glorius and co-workers have nicely addressed the common types of errors and biases in literature-based reaction datasets, which include experimental noise due to reporting errors, selection biases caused by researchers’ preference for specific reagents over others, and reporting biases toward successful results. A comparison of reported yield frequency between the *Reaxys* database and data obtained from HTE clearly reveals this latter point, with literature-derived data skewed heavily toward higher yields (illustrated in Figure 1.2 for Buchwald-Hartwig couplings).[32]

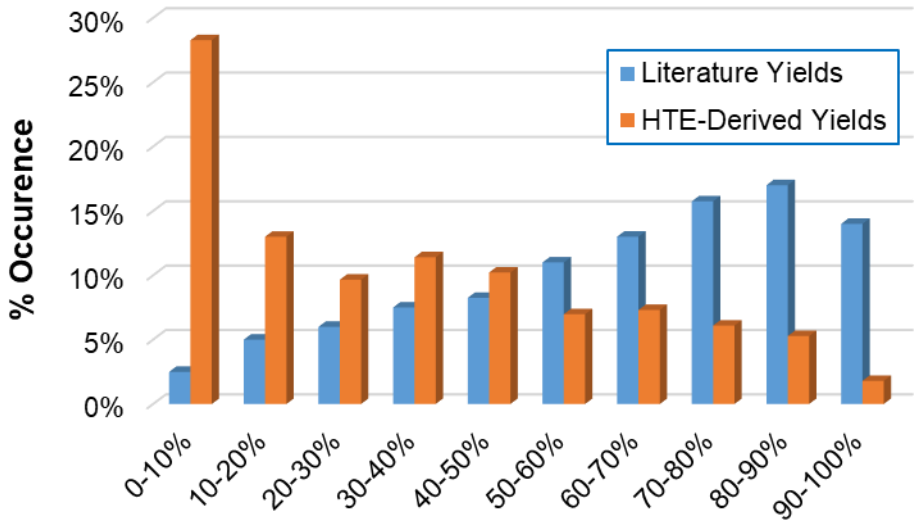


Figure 1.2. Comparison of frequency of reported yield values for Buchwald-Hartwig coupling reactions taken from the literature (Reaxys database) or from an HTE-based dataset. Values obtained from ref. 32.

Selection bias is particularly problematic in synthetic chemistry data. This is because synthetic chemists have a high tendency to select specific reagents/catalysts/conditions based on familiarity, past success on similar systems, and ease of implementation. Such practices are very successful in pursuit of new molecules, but gives rise to a highly unbalanced distribution of reaction conditions in the reported datasets. Specifically for organopalladium catalysis, the reported cross-coupling reaction conditions reveal a strong selection bias toward specific catalysts. Burke, Grzybowski and co-workers highlighted this issue when attempting to generate machine learning based models for Suzuki-Miyaura reactions.[33] They trained the models using >10,000 literature reported Suzuki couplings with an aim to predict conditions that would lead to success in a hypothetical reaction. Instead of making any chemically-meaningful predictions with respect to catalyst, solvent, or base choice, the models simply captured popularity trends[35,36] in the literature data. For example, >50% of reported Suzuki couplings use $Pd(PPh_3)_4$ as the palladium source, and therefore $Pd(PPh_3)_4$ was the top recommendation in >80% of the test cases.

Selection bias also extends to the specific target reactions being studied. Pd-catalyzed coupling reactions benefit from the large number of commercially available substrates:

organohalides, boronic acids, amines, etc. However, even with myriad readily accessible and relatively complex substrates, reported chemical space coverage of the resulting products is actually rather limited. Krska, Dreher, and co-workers at Merck revealed a large disconnect in molecular properties between the reported products of Pd-catalyzed cross-couplings and many approved small molecule drugs. The latter class of molecules generally possess more atoms, more H-bond acceptors, lower lipophilicities, and an overall higher degree of molecular complexity. Thus, prediction models trained exclusively on literature data may not be generalizable to late-stage coupling reactions on advanced pharmaceutical intermediates, where additional functional groups and molecular topologies can dramatically affect catalytic performance.

Reporting bias is another prevalent issue in the synthetic chemistry literature. Methodology studies frequently report univariate optimization within a relatively limited range of conditions, and/or a series of "deviations from standard" conditions as control studies. Furthermore, the highest achievable yields of the target products are the key outcome reported, with little/no discussion of unsuccessful reactions (though this is changing for the better more recently[37]). In addition, other information that is equally important to guide reaction prediction is infrequently disclosed. This includes aspects such as extent of reaction, rate of reaction, mass balance, and specific byproducts observed. For example, formation of homocoupling-derived byproducts is a common concern in Suzuki cross-couplings,[38,39] but the presence and/or quantity of these byproducts often goes unreported. This reporting bias results in literature-based datasets being skewed and/or incomplete, which is not ideal when applying them as training/test sets for statistical modeling. Glorius and co-workers investigated the impact of insufficient low yield data on prediction accuracy for machine learning models.[32] In a case study on Buchwald-Hartwig reaction modeling, they discovered that the literature dataset was heavily skewed toward high yielding examples (>50% of data points with >70% yield); the resulting ML model performed relatively poorly, with a mean absolute error (MAE) of ±15%. If an HTE-generated dataset was instead used, which has a more realistic distribution of yields (>25% of data points with <10% yield), the resulting model was significantly more accurate (MAE = ±10%). Similarly, the authors simulated the effect of introducing additional experimental

data to cover the low yielding reaction space, which again led to improved model performance.

Fitzner, Wuitschik, and co-workers have conducted large meta-analyses on more than 62,000 Buchwald-Hartwig coupling reactions reported in three databases (CAS, *Reaxys*, and USPTO).[34,40] Their aims were to identify and address the common issues in using literature data to guide reaction condition optimization, and to generate useful and broadly generalizable predictions. They pointed out that many reactions are reported with very limited reaction information. For example, only 75% of the C–N couplings analyzed have reaction yield reported. In addition, essential reaction conditions such as temperature, reaction time, reagent scale and catalyst loading are significantly underreported. As already discussed in the context of Glorius's study, yields that are reported are skewed toward higher values in both the patent and non-patent literature (median yield >60% in recent years).

Importantly, this meta-analysis revealed reporting biases extend beyond (high) yield and (popular) catalyst/ligand to include solvent – nearly 80% of reactions use toluene or dioxane – and base – nearly 80% of reactions use NaO*t*Bu or Cs_2CO_3. Reporting bias can also result from specific target applications becoming more prominent: the authors noted a large increase in use of P(*t*Bu)$_3$ and diarylamine nucleophiles beginning around 2014, due to an explosion of patent claims around OLED-relevant materials. Thus, while 62,000 reactions may seem like an ideal "big data" set, there are huge swaths of homogeneity in reaction conditions used, as well as overrepresented substrate/product classes based on commercial need. Thus, simply increasing the number of reactions in a dataset provides very limited effect in expanding the chemical/reaction space coverage.

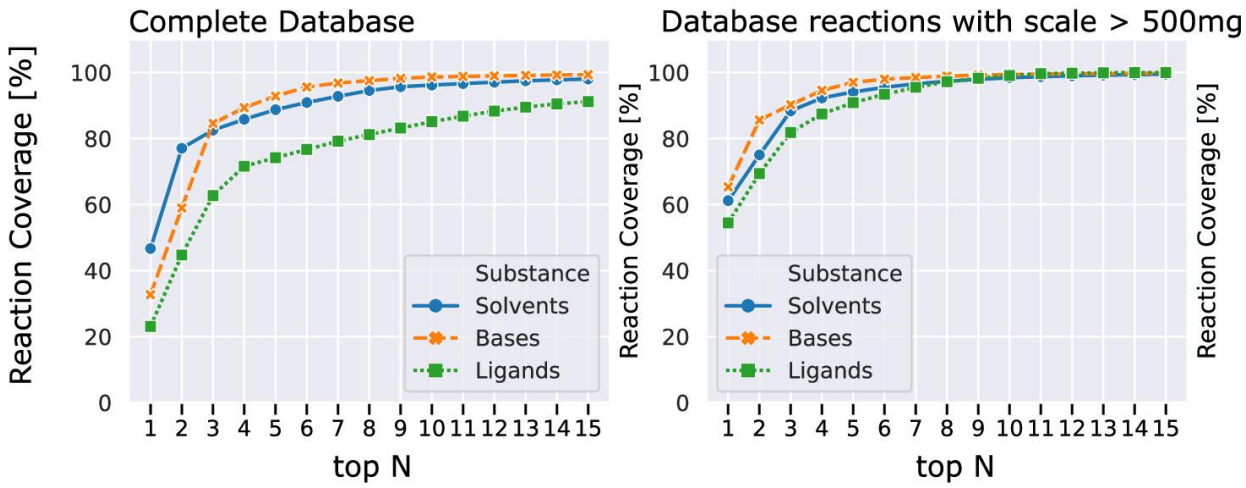


Figure 1.3. Diversity analysis of reported reaction conditions for C–N couplings from a combination of three databases (CAS, Reaxys, and USPTO). Plots represent % of reported conditions (y-axis) that contain the top N settings (x-axis). For example, ~80% of reactions in the complete database use just the top two solvents (top N = 2). Reproduced with permission from ref. 34. Copyright 2023, American Chemical Society.

These data issues have an obvious impact on model accuracy and applicability. Fitzner and Wuitschik's initial meta-analysis did result in a set of qualitative predictive tools to aid condition selection for Buchwald-Hartwig couplings based on substrate type.[40] Subsequent work from this group on quantitative machine learning models trained on this large dataset revealed how such models can be misleading when applied to new synthetic cases beyond the initial training set.[34] Although their model performs well in training/test splits within the literature data, it fails to extend its predictive power to a new reaction dataset obtained by experiment (external predictions). They attribute this lack of generality to all of the aforementioned biases in the literature data. These hidden biases retained in the literature data make it impossible to account for all the important reaction parameters that must be considered and explored when discovering or optimizing a new reaction. By presenting their failed attempt to achieve a general-purpose predictive model trained solely on literature data, they have identified a potentially general problem with such models: evaluating performance using test data taken from the overall literature data set may lead to overestimation of accuracy/generality. Thus, the authors suggest that it is important to

use lab-generated reaction data (e.g. obtained by HTE) for reaction predictive model training and assessment.

Importantly, expert curation can be a powerful approach in leveraging literature data for predictive models. A study by Schleinitz *et al.* on predicting reaction yield highlights the importance of the quality of the data, rather than the quantity, in assembling a robust training dataset. While this is not specifically organopalladium catalysis, the approach and lessons from this work are highly relevant. Among an initial dataset of >2000 Ni-catalyzed C–O couplings from the literature, the authors carefully selected a small subset of reactions (~200) that best represent the reaction information and chemical space covered by the original dataset. They then trained machine learning models using the large and curated datasets for comparison. The model trained with the small subset achieved the same predictive performance compared to the model trained on the original large dataset. Their study also reveals the importance of including failed experiments and low yield reactions extracted from optimization tables, such as those contained in Supporting Information sections. This will help to overcome reporting biases and lead to reaction datasets with a greater potential for generalization.

1.3.2 Perspectives on the Available Data.

The above examples highlight the importance of knowing the limitations and biases of literature databases as a source of training data for predictive models. Using expert chemical knowledge in dataset assembly and curation prior to model training is critical to ensure accurate and useful predictions. The success of any statistical modeling approach to reactivity predictions will rely on the quality and diversity of the training dataset. Even a relatively small dataset can be suitable if it covers a significant chemical/reaction space. The above examples also highlight how synthetic chemists can help improve the applicability of literature data toward predictive modeling. For any new method, evaluating a diverse and balanced chemical space when exploring reaction scope is enormously useful, such as with an informer library.[41] And evaluating/reporting diverse reaction conditions that result in a range of reaction yields is more useful than reporting only the “champion” conditions and yield. Recent initiatives such as the Open Reaction Database seek to standardize and streamline dataset curation to ensure each entry is complete and in

machine-readable format, which should greatly improve the quality and accuracy of resulting models.[42]

One final perspective on existing literature datasets is whether the synthetic chemists' main metric (yield) is the best response variable for predictive modeling. Whether in academic papers or patent examples, the isolated yield is the universal measure of reactivity in organic synthesis. Accordingly, all of the aforementioned models are trained to build connections between molecular structure, reaction conditions, and yield values, with the key predicted outcome being the yield.[43] However, the reliability, consistency, and reproducibility of reported yields data from different literature sources give rise to concerns for its use in predictive model training.

Due to lack of consistency in our reporting practices, a literature yield could refer to several different measurements, such as solution or assay yield by NMR, GC, or HPLC analysis (common during optimization), or isolated yield after workup/purification (for reaction scope exploration). Isolated yields also tend to suffer from reproducibility issues (as any practicing synthetic chemist is painfully aware). Importantly, isolated yields are not only a measure of reactivity, but the efficiency of the isolation procedure. Thus, there must be systematic errors in isolated yield values due to product loss during workup and purification.[44] This would be especially prevalent in cases where high purity is desired over maximum yield – such as in the pharmaceutical syntheses that populate the USPTO database! Thus, it is critical for us in the organic synthetic community to more frequently report solution yields / assay yields for reaction outcomes in addition to isolated yields, *and* for reaction databases (CAS, *Reaxys*) to include and tag these data accordingly, as opposed to only including information from reaction scope studies. Assay yields obtained using quantitative HPLC analysis are standard practice in process chemistry research. More consistent reporting of this measure of reaction outcome would definitely improve the quality of literature data for the purposes of predictive modeling.

Another concern with reaction yield as the major response variable is the small range of values it can take. Yields are bounded between 0-100%, and nearly always reported to the nearest percent. The small product masses (10-100 mg) isolated from many reactions also

mean that isolated yields have fairly significant measurement errors. The narrow reporting range and low precision of isolated yields makes it difficult to capture and model significant reactivity differences, especially in lower yielding reactions. For example, the difference between 85% and 95% yield and between 0% and 10% yield are both 10%; however, from a reactivity standpoint the former is barely significant, while the latter is very significant. Other measures of reaction outcome, such as rate or selectivity, can span multiple orders of magnitude and have distinct advantages in reactivity prediction; unfortunately, they are also much less frequently reported.

1.4 Data-Rich Computational Approaches

To go beyond the existing literature/patent database sources, alternative methods of data generation and collection are clearly required. Computational approaches are a powerful complement to experimental methods in validating mechanistic hypotheses and accelerating reaction optimization in synthetic chemistry.[45] They are also invaluable for dataset generation, especially when limited (or no) experimental data is available, and/or when collecting experimental data would be difficult (or impossible). In organopalladium chemistry and the corresponding synthetic transformations it enables, there are two types of data that are generated using computational approaches: calculated reaction coordinates and corresponding energy profiles; and calculated molecular descriptors that can be linked to reaction outcome data (whether computational or experimental) to build quantitative structure-reactivity/selectivity relationships. These are two distinct applications of computational methods, with the former having closer links to what is typically categorized as computational/theoretical chemistry, and the latter more in the realm of cheminformatics.

1.4.1 High-Throughput Computation of Reaction Energy Profiles

In the absence of large experimental data sets of catalyst activity and/or selectivity for a given transformation (or set of transformations), computational studies of catalytic reactions are invaluable as source of data. Comparing relative energies of putative

intermediates and/or transition states between different systems gives a molecular-level view of catalyst reactivity that is highly tunable. This is especially powerful where obtaining sufficient experimental data is cost and/or time prohibitive. It also provides a method to interrogate aspects of the reaction mechanism that are impossible to study using purely experimental means. While computational chemistry is often used to study specific mechanistic questions in palladium catalysis, data-rich studies require high-throughput computational methods to interrogate a large number of potential catalysts and/or substrates.

Corminboeuf and coworkers have successfully applied high-throughput computational tools to construct linear free energy scaling relationships for several catalytic reactions, and represent these using "volcano plots".[46,47] Such plots are commonly used in heterogeneous catalysis and electrochemistry, and originate from Sabatier's principle that an ideal catalyst-reactant (or catalyst-product) interaction should be neither too strong nor too weak.[48,49] By extending this concept to homogeneous catalysis, Corminboeuf's group is able to gain new insights and make predictions of catalyst performance for transition metal-catalyzed cross-coupling reactions.

They first validated the concept of molecular volcano plots using Suzuki coupling, perhaps *the* prototypical organopalladium-catalyzed reaction, as a key case study (Figure 1.4).[50] For this initial analysis, catalytic behavior was modeled using thermodynamic aspects of the catalytic cycle: namely, the free energies of calculated catalytic intermediates for a variety of transition metals, and the linear relationship between the free energies of those intermediates. This analysis effectively reproduced the experimentally validated fact that Pd-based catalysts are optimal for Suzuki cross-coupling.

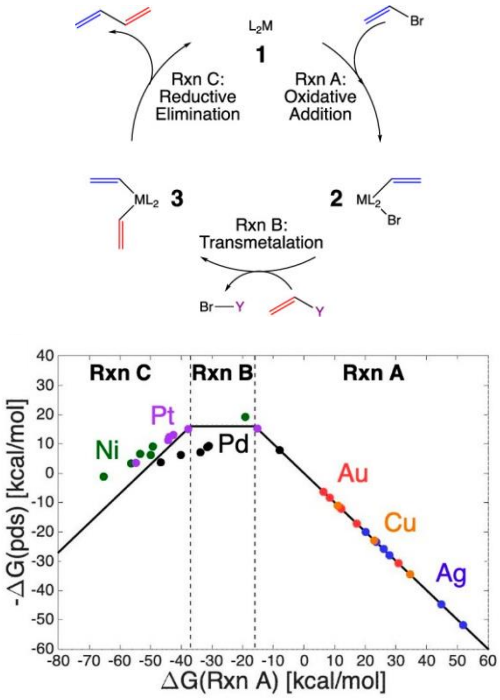


Figure 1.4. Molecular volcano plot of the calculated thermodynamics of oxidative addition (Rxn A), transmetallation (Rxn B) and reductive elimination (Rxn C) for a series of Ni, Pd, Pt, Cu, Ag, and Au catalysts Optimal performance is achieved by Pd-based catalysts, which have intermediate exergonicities of oxidative addition as compared to other transition metals Reproduced with permission from ref. 46. Copyright 2021, American Chemical Society.

In follow-up work, Corminboeuf and co-workers significantly expanded the catalyst pool under investigation to include more than 25,000 potential complexes.[51] Using DFT-obtained thermodynamic data on the free energy of oxidative addition for 7,000 of these as a training set, they were able to use machine learning analysis to predict those values for the remaining 18,000 potential catalysts. Constructing a molecular volcano plot using these thermodynamic parameters and filtering candidate catalysts by cost led to identification of 37 ideal candidates based on Pd or Cu. The authors also exploited this large dataset for further analysis through dimensionality reduction *via* clustering analysis, extracting ligand effect trends across multiple cross-coupling reaction classes.[52] Finally, in addition to studying catalyst effects on reaction performance, they have also applied the volcano plot analysis to studying how substrate structures and substituents affect the outcome.[53]

The use of calculated transition state energies is another powerful indicator of catalyst performance, though collecting sufficient data can be challenging. A recent example by Yu, Fu, and coworkers showcases this approach to study the activating effect of Brønsted acids on the Pd-catalyzed C–O bond cleavage of allyl alcohols.[54] They created a computationally derived reaction database containing 393 DFT calculated activation barriers, followed by multivariate linear regression (MLR) model training using calculated molecular descriptors. The prediction accuracy was confirmed by a small experimental test set (taken from the literature[55]), with good agreement between predicted $\Delta G^{\ddagger}$ and the experimental product yield.

Currently, accurate *and* fast mechanistic modeling of organometallic transformations remains challenging. Schoenebeck has highlighted that computational studies of large, complex systems with myriad mechanistic possibilities lead to exponentially growing search directions (and thus exponentially growing computational resources).[45] An additional challenge is that calculation outcomes often depend on the theoretical method employed,[56] leading to another layer of complexity in designing high-throughput computational studies. Therefore, amassing computational data for a large array of Pd-catalyzed reactions needs to be performed with caution, balancing both calculation speed with accuracy.

A crucial aspect of any quantitative structure-reactivity dataset is the collection of descriptors used to define molecular structures and reaction conditions. Thanks to tremendous developments in computer performance and theoretical methods, computational chemistry has made impressive progress in performing fast and reliable calculations for a wide variety of chemical systems. Especially germane to reactivity prediction, computational chemistry is extensively used to create molecular structural representations and corresponding numerical descriptors, and to compute reaction coordinates, transition states, and reaction barriers.

Whether by univariate regression or unsupervised machine learning, data-driven prediction models are developed by establishing quantitative relationships between experimentally obtained or computed molecular descriptors and measured reactivity

outcome values. The application of powerful computational tools and mechanistic insights has vastly increased the number of relevant physicochemical descriptors to quantify molecular properties. Mechanistically relevant descriptors that directly impact chemical reactivity have found great value in reactivity prediction, such as the computed notations representing the electronic,[57,58] steric,[59–61] and vibrational[62] properties of a substrate or catalyst. Activation barriers determined from calculated transition state energies have also contributed to quantitatively accurate reaction outcome predictions.[63] Finally, computational methods that combine statistical algorithms are used to develop automated workflows for tasks ranging from training data selection to error assessment of the output numbers for improved efficiency and reliability in reaction predictions.[64,65]

1.4.2 Mapping Catalyst Chemical Space

A key aspect of organopalladium catalysis where computationally-derived datasets are prevalent is in ligand parameterization. Understanding and predicting the effect that a given ligand will have on reactivity and selectivity is central to Pd-catalyzed coupling. Ligand choice is often the most consequential factor for catalyst performance (both rate and productivity), especially with less reactive but more abundant substrates (e.g. unactivated Ar–Cl).[66–69] In addition, ligand identity is known to be a crucial factor to control site selectivity for Pd-catalyzed couplings of multihalogenated substrates.[70,71] Accordingly, ligand screening from large and diverse ligand libraries is a frequent activity in industrial HTE labs. However, to adequately cover the available chemical space even within a subclass of supporting ligand is time and resource intensive. Computational tools have been invaluable in mapping this chemical space with large sets of relevant numerical descriptors.

Fey and co-workers have pioneered this approach through curation of the Ligand Knowledge Bases, a series of focused ligand descriptor datasets for specific ligand classes relevant to catalysis.[24,72] These sets include mono and bidentate ligands with P (Figure 1.5A), C (carbene), N, and O donor atoms, with more than 1,300 entries containing at least 20 descriptors each. This focused approach allows each ligand class to be described by a relevant set of descriptors; for example, descriptors for LKB-PP (bidentate P,P and P,N ligands) were adapted to account for multiple donor atoms and specific steric aspects of

chelate bite angle.[73] Likewise, LKB-C (carbene ligands) was designed to take into account differences in electronic configuration between Schrock, Fischer, and NHC type carbene ligands.[74] To highlight the LKB's relevance to organopalladium catalysis, Fey and co-workers mapped HTE-derived Pd-catalyzed C–N coupling data against the principal component analysis (PCA) map of LKB-P (monodentate P).[75] While a quantitative model was not reported, the data visualization revealed a clear structure-reactivity trend where large, electron-rich phosphines are clearly optimal. In fact, adjacent to the most reactive ligands in the PCA map are several Buchwald-type biaryl ligands (not tested in the experimental data set) that are known to be very effective in this chemistry. Thus, LKB descriptors can potentially help to guide subsequent ligand screening.

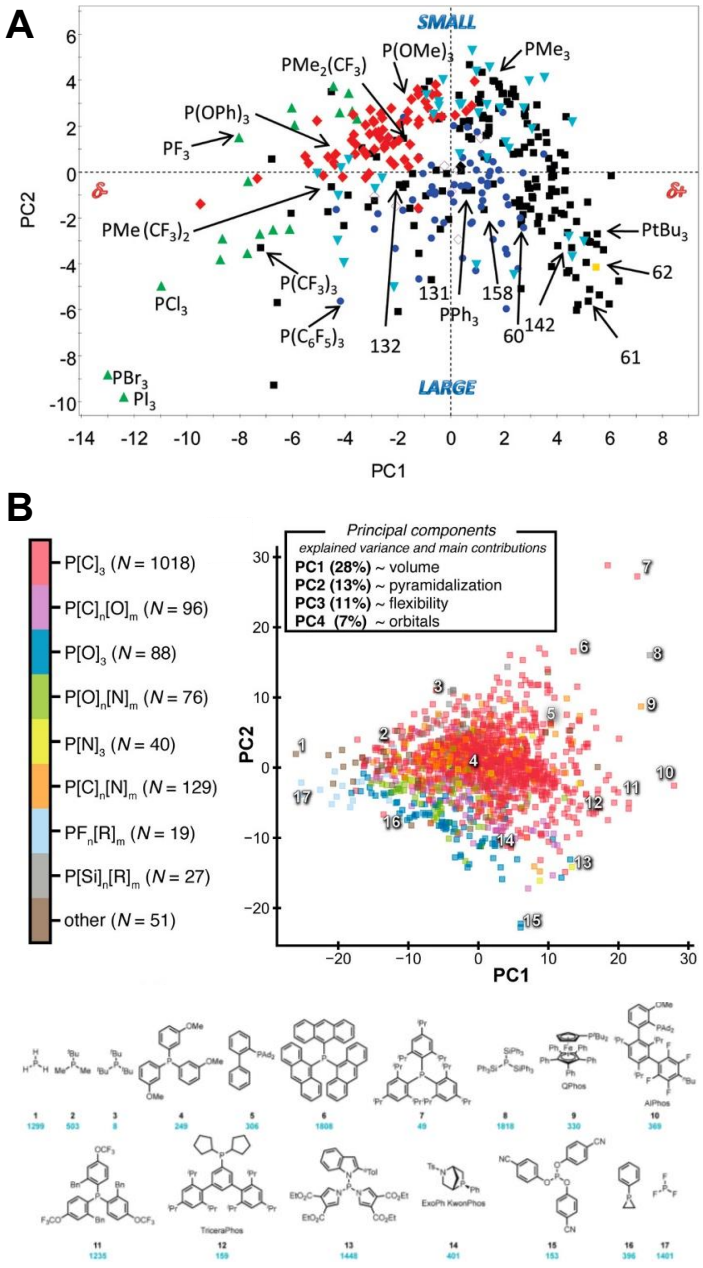


Figure 1.5. A comparison of two monodentate phosphine databases: A) LKB-P; and B) Kraken, with representative phosphines shown. Each database is represented by two-dimensional principal component analysis (PCA) plot. In the LKB-P, PC1 is predominantly electronic in nature, while PC2 is predominantly steric. In the Kraken visualization, PC1-PC4 have the indicated contributors. Figure 1.5A reproduced with permission from ref. 72, copyright 2010, American Chemical Society. Figure 1.5B reproduced with permission from ref. 77, copyright 2022, American Chemical Society.

Schoenebeck and co-workers recently reported using the LKB descriptors as a basis for ligand clustering analysis.[76] Their group's interest in dimeric Pd(I) species as precatalysts led them to investigate the ligand properties that stabilize these dimers versus those that fail to generate stable dimers. The existing knowledge of the structure features promoting Pd(I) dimer formation is extremely limited, as there are only four known dimer-inducing ligands from the LKB-P library. Furthermore, these ligands are proximal to non-dimer forming ligands (e.g. PCy_3) in the initial principal component analysis (PCA) plot of the LKB-P library. Thus, simple visual clustering of ligands with similar properties fails to reveal the key features for Pd(I) dimer formation. An alternative means of grouping ligands based on their descriptors – *k*-means clustering ($k = 8$) – led to dimer-forming and non-dimer forming ligands being grouped in different clusters. Strikingly, these clusters overlap in the PCA plot, showing that the *k*-means algorithm is capable of distinguishing apparently similar ligands; however, the sizes of the two candidate clusters were too large to be useful in predicting dimer formation for a given ligand.

To refine the ligand classification analysis, the authors introduced a group of new, "problem-specific" descriptors obtained by DFT calculations. These include some of the original LKB-P descriptors, but also many steric, electronic, and thermodynamic descriptors specific to Pd-based intermediates along the mechanism for Pd(I) dimer formation. Further clustering of the two initial candidate clusters using *k*-means ($k = 6$) and this new set of descriptors led to clear differentiation and identification of two smaller clusters that contained all of the known dimer-forming ligands. Importantly, eight predicted dimer-inducing phosphine ligands belonging to the same clusters were verified experimentally to support Pd(I) dimer formation, leading to the isolation of eight previously unknown Pd(I) dimers. This example shows how effective data analysis and descriptor generation is just as important as efficient collection and collation of computational data. While large descriptor databases like LKB provide an excellent starting point, problem-specific descriptor sets are often critical for more granular analyses and predictions.

More recently, Gensch, Sigman, Aspuru-Guzik and co-workers created an online descriptor database for monodentate phosphines (Figure 4B).[77] This database significantly

expands the chemical and descriptor space applied to this important class of ligand, and incorporates conformational effects. For an initial set of 1,558 ligands (including the 200 most commonly cited phosphines in the literature), the authors calculated 190 individual descriptors using DFT methods. From this set, they then were able to train machine learning models to predict properties for a further 300,000 potential phosphine ligands. Thus, combining DFT and ML methods is a powerful approach to rapidly generate "big data" for molecular descriptor sets.

In a follow-up study, Gensch, Sigman, and co-workers applied this phosphine descriptor database to guide the development of a diverse monodentate phosphine screening set to maximize chemical space coverage.[78] Unlike continuous reaction variables, such as temperature and reaction time, the electronic and steric properties of the PR_3 ligands are generally treated as categorical/discrete variables in experimental screening. In a typical HTE-based optimization approach, researchers will design an array using a combination of literature precedent and chemical intuition.[15] They will then determine which ligands to test next by analogy/perceived similarity to the most successful candidates in an iterative approach. Inevitably, this leads to selection biases when the search space for the discrete parameters is not well defined.

To explore the discrete ligand space systematically, and reduce these potential biases, the authors sampled all of the commercially available PR_3 ligands from the virtual library and generated four principal components space from the descriptor set. Using a computational data-driven workflow, they then winnowed down to a diverse set of 32 monophosphines intended to maximally and uniformly cover the 4D chemical space. Application of this screening set to several Pd-catalyzed C–C and C–N coupling reactions revealed the expected spread of catalytic results, with a clear positive hit (or several) observed in each case. Furthermore, given the logical design of the screening set, follow-up investigations can focus on ligands in close PC space proximity to "hit" ligands, reminiscent of the LKB-P example described previously.

The application of cheminformatics tools to organopalladium chemistry can positively impact one of the biggest challenges in using Pd-catalysis in target-based synthesis: how

does one pick the right catalyst? Practitioners, especially in pharmaceutical R&D, are moving steadily away from empirically-based screening designs into those resembling the deliberate chemical space coverage described above. Not only does this increase the likelihood of success in identifying a suitable catalyst in a given amount of time – critical in discovery chemistry – it also enables cheminformatics to combine with high-throughput experimentation to generate large, consistent, and quantitative datasets as a foundation for quantitative predictions. Finally, it also makes hypothesis-driven screening much more exact, since specific ligand features can be evaluated quantitatively rather than (at worst) superficially (e.g. investigating "similar" ligands that are not actually that similar).

1.5 Datasets from High-Throughput Experimentation

1.5.1 Big Data on Demand via Miniaturized Array-Based Experiments.

High throughput experimentation (HTE) is an increasingly common approach in both academic and industrial labs. HTE is especially well-suited to studying catalytic reactions, where the number of factors affecting reaction outcome are generally greater than in non-catalytic systems. While HTE is often used with the aim of either discovery (does this transformation work?) or optimization (what is the best set of conditions to maximize yield/selectivity?), it is also an invaluable approach to generating large internally-consistent datasets on demand. This Perspective will focus only on recent, illustrative applications of HTE in organopalladium chemistry; readers interested in HTE more generally are directed to several recent reviews on the topic.[14,15,79–85]

As HTE enables a data-rich approach to the discovery of chemical reactions – many more factors and their interactions can be studied simultaneously – it is particularly well-suited to the study of Pd-catalyzed processes. A typical cross-coupling reaction system includes numerous categorical factors (Pd source, ligand, base, solvent) and continuous factors (time, temperature, concentration, stoichiometries) that are impossible to fully assess "one factor at a time" (OFAT). Led by industrial R&D groups, particularly in pharmaceutical discovery and process chemistry, as well as academic HTE centers, data-rich HTE approaches to multifactor exploration and optimization have been demonstrated on many Pd-catalyzed reactions.[15] In addition to the direct applicability of these studies to

active pharmaceutical ingredient (API) synthesis, organopalladium catalysis is again a general proving ground for data-rich techniques.

Developing new data-rich capabilities often goes hand-in-hand with applications in organopalladium chemistry. One set of examples is the development of new methods for high-throughput reaction set-up and analysis for dense reaction arrays. Researchers at Merck used a combination of advanced liquid handling automation with rapid UPLC-MS or MALDI-TOF MS analysis to interrogate thousands of Pd-catalyzed C–N coupling reactions in plate-based formats.[86,87] Researchers at Pfizer similarly demonstrated a new flow-based HTE platform using Pd-catalyzed Suzuki-Miyaura couplings to run and analyze thousands of examples.[88] In academia, the Cernak group is continuing to push the envelope of ultra-high throughput methods in array design and execution.[89–91] In the area of laboratory automation, the Hein group, in collaboration with Merck and the Aspuru-Guzik and Sigman groups, demonstrated the power of data-dense experimentation in an autonomous optimization of a Pd-catalyzed Suzuki-Miyaura reaction.[18] The Newman group published a user-centered "how-to-HTE" guide on multiwell screening, using a Pd-catalyzed C–N coupling as the prototype reaction.[92] And finally, even undergraduate teaching lab experiments on microscale HTE employ Pd-catalyzed reactions as the exemplars.[93]

Our research group focuses not only using HTE and data-rich techniques to develop new reactions,[94–98] but also to better understand how reactivity and mechanism change as a function of the reaction system.[58,99,100] To enable these efforts in the realm of organopalladium chemistry, we identified a need for Pd precursor compounds that are specifically suited to HTE studies. In 2021 we reported an easily prepared and air-stable Pd(0) precursor, ^{DMP}DAB−Pd−MAH, which was designed specifically for HTE applications.[101]

While HTE has been widely used to conduct catalyst screening exercises as well as multifactor optimization studies, the identities of the substrates themselves are increasingly being explored *via* HTE. Sather and Martinot used a data-rich approach to find the best reaction conditions for fusing piperidine-based nucleophiles with five-membered

heteroaromatic bromides, a difficult C–N coupling type that has been rarely studied before.[102] Extensive HTE revealed Pd-PEPPSI IHept^{Cl} – an advanced precatalyst from the Organ group[103] – as the best catalyst in their model pyrazole-based system, as well as specific solvent/base combinations that prevent substrate decomposition. They subsequently tested the generality of this catalyst system in a panel of 48 unique heteroaryl bromides, with a "hit-rate" of ~50% (products observable by LCMS). This evaluation revealed several important features of the reaction, including incompatibility with certain ester functional groups, as well as protic functional groups including alcohols and amides.

In 2019 we, along with collaborators at GSK and Temple University, conducted a similar study of a challenging C–N coupling involving sulfonamide arylation.[94] Multi-substrate HTE revealed the BippyPhos class of ligands as superior, with AdBippyPhos as the optimal candidate. To evaluate the generality of this reaction, we tested an array of 288 substrate combinations: 24 heteroaryl halides and 12 primary and secondary sulfonamides. In this case, our "hit-rate" was ~25% (where coupled products were observed by LCMS in >10% area). Several substrates were ineffective when paired with any coupling partner, while others were effective with nearly all coupling partners. These kinds of substrate-focused HTE studies not only provide a map of reaction scope and highlight gaps in applicability, they are crucial as training sets for statistical analysis and reaction prediction.

1.5.2 Combining High-Throughput Experimental and Computational Approaches.

HTE screening in chemistry is often aimed at achieving the ideal solution to a synthetic task. Beyond the optimization goal, the ability to generate hundreds-to-thousands of data points in a deliberate way means HTE is ideally suited as a means to build training datasets for reaction prediction. Multivariate HTE designs create reaction data libraries that cover a broader range of chemical structure and reaction space, and have more representative outcomes than literature data (*vide supra*, Figure 1.2). Such high-density coverage of reaction space is ideal when building quantitative statistical models for reactivity and/or

selectivity, providing the experimental link to large descriptor datasets that are increasingly accessible (*vide supra*, Figure 1.5).

Catalytic reactions are known to be acutely sensitive to seemingly minor changes, where subtle alterations to catalyst structure, solvent identity, or even reaction concentration can have a dramatic impact on reaction outcomes. Understanding the underlying connections between reaction conditions, molecular structure, and chemical reactivity is both a fundamental goal of catalysis research, and is crucial to making accurate predictions. Because of the flexibility in experimental design and efficiency in data collection, HTE can be used to study how specific aspects of a reaction system impact reactivity, providing sufficient experimental data to train predictive models. Notably, the following three illustrative examples are all academic/industry collaborations!

A seminal example of combining high-throughput experimentation and computation for predictive modeling is the work of Dreher, Doyle, and co-workers on Pd-catalyzed C–N coupling.[104] Their study was focused on the effect of a specific type of inhibitory additive, isoxazole heterocycles, on the performance of various C–N couplings. Using Merck's nanoscale high-throughput platform, they collected a multidimensional experimental dataset of ~4000 reactions by combining 23 isoxazoles with different aryl halides, solvents and bases. From the computational side, a large set of molecular descriptors was generated without biasing selection with (possibly incorrect) mechanistic hypotheses. A variety of modeling approaches were evaluated, from multivariate linear regression (MLR) to machine learning (ML) algorithms, with a random forest algorithm providing the most accurate model. Even though ML-based models can be difficult to interpret mechanistically, the authors used a sensitivity analysis for key descriptors (i.e. how much is the model error increased by randomizing a given descriptor) to derive mechanistic insight. The two most sensitive descriptors are isoxazole-based, and are linked to electrophilicity (calculated ^{13}C chemical shift of the C_3 site, and LUMO energy), suggesting the most potent isoxazole inhibitors undergo competitive oxidative addition to Pd(0). This was verified experimentally by spectroscopic observation of the competitive N–O oxidative addition to $Pd(PPh_3)_4$. Notably, this work did generate commentary about the use of different feature sets,[105,106] whether chemical-based (as in Dreher and Doyle's

approach[104]) or random-valued (as discussed by Chuang and Keiser[105]). While the chemical insights and out-of-sample prediction accuracy of the initial random forest model validate the use of chemical-based descriptors in this case,[106] incorporating control procedures and best practices into ML data analysis is crucial for those looking to use this powerful technique across the chemical sciences.[107]

In another academic/industry collaboration, Sigman, Doyle, and researchers at Merck reported a quantitative workflow to identify and rationalize reactivity cliffs in Ni- and Pd-catalyzed Suzuki-Miyaura coupling (SMC) datasets using monodentate phosphine ligands.[109] Using the aforementioned Kraken organophosphorus descriptor library and SMC reaction datasets collected by high-throughput experimentation, they developed a quantitative workflow to evaluate the performance of a single ligand descriptor in identifying reactivity thresholds. Among the 3 steric ligand descriptors evaluated (cone angle, $\%V_{bur}$ (*Boltz*) and $\%V_{bur}$ (*min*)), $\%V_{bur}$ (*min*) offers the best performance in dividing the SMC datasets into active and inactive regions around a threshold value. The success of this univariate classification is based on the ligation state (formation of ML versus ML_2 species), which is intimately tied to the size of the phosphine ligand. In addition, this work highlighted an intriguing mechanistic dichotomy: in the examined Ni- and Pd-catalyzed SMC datasets, the opposite $\%V_{bur}$ (*min*) reactivity threshold directionalities indicate that the Ni-catalyzed SMCs require bisligated (NiL_2) species while Pd-catalyzed SMCs require monoligated (PdL) species. This study highlights the power of a conceptually simple univariate threshold approach to reactivity analysis and prediction.

Making predictions for multiple outcomes – such as high activity *and* selectivity – poses additional challenges for reaction modelling. This is particularly the case for regio- and/or stereoselective catalysis, where both high chemical yield and high selectivities are requirements for a successful process; thus, catalyst/ligand optimization requires a multi-objective approach. Mack, Sigman, and co-workers reported a data-driven workflow using high-throughput computation and experimentation for multi-objective ligand optimization. This approach was demonstrated for two key enantioselective steps toward an investigational API from Genentech.[108] These steps are a regio- and enantioselective Pd-

catalyzed Hayashi-Heck coupling, following by regio- and enantioselective Rh-catalyzed hydroformylation; the discussion here will focus on the first step (Figure 1.6A).

In this case, the data foundations for this study are a computational database consisting of >550 bisphosphine ligands with DFT derived descriptors, and an experimental HTE screening database using a group of selected ligands based on previous research findings. These then informed the multi-objective analysis (Figure 1.6B). The ligands screened were classified based on a reactivity threshold analysis akin to that described above, where the phosphorus lone pair occupancy was an effective single descriptor. This revealed that phosphines below a threshold value were generally active, enabling a more focused ligand set to be taken forward. Then, regioselectivity as a function of ligand structure was assessed by MLR, revealing only two descriptors – anisotropic ^{31}P NMR shielding value and P–C σ* orbital occupancy – were sufficient to build a robust correlation. Finally, a virtual screen of the entire bisphosphine database identified ligands predicted to be even more regioselective than those in the HTE dataset, leading to experimental validation of (*S*)-HexaMeO-BIPHEP as optimal from a combined activity, regioselectivity, and enantioselectivity standpoint (Figure 1.6C).

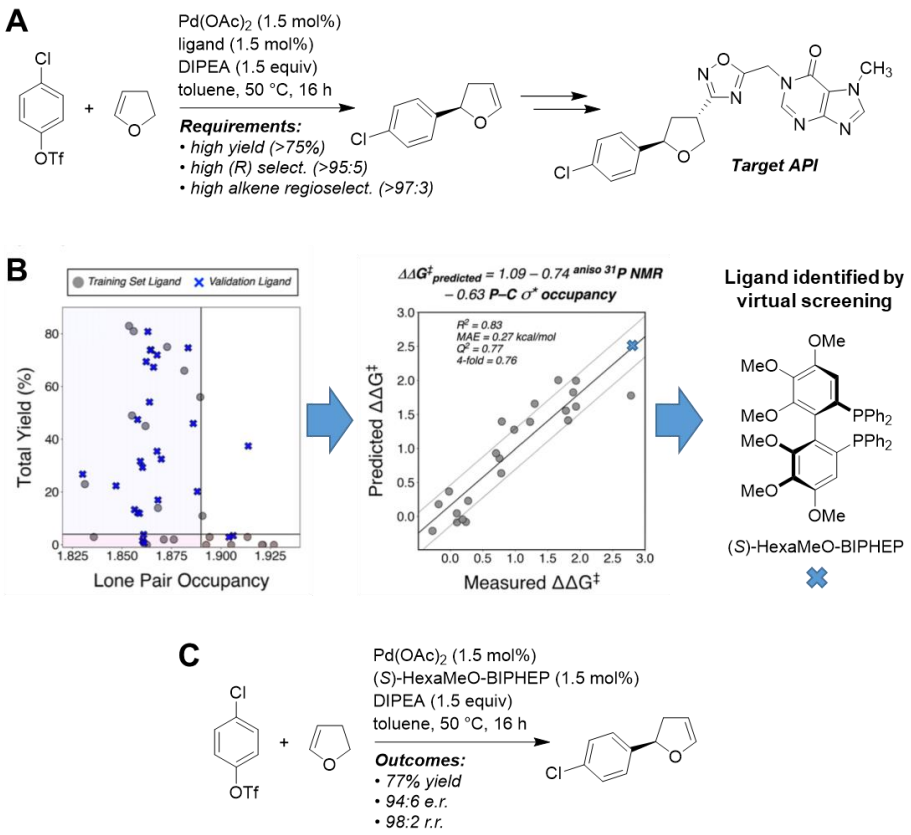


Figure 1.6. A) Hayashi-Heck coupling required in high yield, enantioselectivity, and regioselectivity. B) Step-wise multi-objective optimization via a data-rich computational / HTE approach, with univariate threshold analysis for activity, followed by MLR analysis for regioselectivity, and finally virtual screening for extrapolation to superior catalyst. C) Improved outcome with new catalyst system. Adapted with permission from ref. 108. Copyright 2023, American Chemical Society.

In another example from the Sigman/Genentech collaboration, Xu *et al.* reported the use of MLR modeling to optimize an atroposelective Pd-catalyzed Negishi reaction using chiral bisphosphine ligands. This step is key for the synthesis of a potent KRAS G12C covalent inhibitor, GDC-6036 (1).[110] Using Genentech's HTE capabilities, the authors rapidly collected a focused set of data for a variety of chiral ligands, with only 3 out of 24 giving yields >20%, and none with e.r. >70:30. One Walphos-type ligand was identified as superior in this initial set, leading to a more focused HTE campaign to generate a Walphos training set for MLR analysis. The resulting linear model facilitated virtual ligand screening and experimental verification, akin to that described above, leading to high stereoselectivity for the desired Negishi reaction using W057-2 as the optimal ligand.

Notably in this case, the excellent selectivity exhibited by W057-2 is not transferable to other coupling partners in model Negishi reactions. Even small structural changes to the substrates result in significantly diminished yield and selectivity. This lack of generality is likely due to the focused HTE-based training set, where substrate variations were not taken into account. This is obviously not an issue for the Genentech researchers, who are primarily concerned with synthesis of GDC-6036 (1) rather than simpler biaryls; however, it does reveal the broader challenges and complexities in discovering general catalytic systems, where substantially more data is required to account for interactions between substrate and catalyst.

1.5.3 Quantitative Structure-Rate Relationships (QSRRs).

Overall, combining HTE with MLR/ML approaches to link experimental outcomes to calculated descriptors is an excellent approach to building robust, accurate, and potentially generalizable predictive models. However, as discussed previously, using reaction yields as the outcome variable is not optimal for quantitative modeling. Kinetic parameters (rates, rate constants, $\Delta G^{\ddagger}$ values) are superior to yields as an outcome variable, especially for mechanism-based approaches. Many of the best-performing models – such as the two Sigman/Genentech examples above – use selectivity measurements as the outcome metric. If under kinetic control, selectivity represents a ratio of reaction rates for the formation of one product or the other, and therefore can be used to construct linear free energy relationships (LFERs). This is why selectivity values are often represented as $\Delta\Delta G^{\ddagger}$ – i.e. free energies – in the resulting models.

A classic example of LFERs in physical organic chemistry is the Hammett equation,[111] which has been widely used to study mechanistic aspects of organic and organometallic reactions.[112–114] A relevant example in organopalladium chemistry was reported by Maes, Jutand, and co-workers, who studied the oxidative addition of 2-halopyridines to $Pd(PPh_3)_4$ using Hammett analyses of experimentally measured rates.[115] The corresponding Hammett plots and DFT calculations of transition states reveals that the oxidative addition mechanism for 2-halopyridines changes depending on the identity of the halide. A polarized nucleophilic displacement mechanism was proposed for 2-chloro and 2-

bromopyridines, whereas a classic 3-centered mechanism was proposed for 2-iodopyridines.

Kinetic values can outperform yields in providing chemically-meaningful predictions because the major objective that most models are trained to predict – whether explicitly or implicitly – is kinetic in nature. Higher reaction yields are often correlated to favorable reaction kinetics. However, unlike selectivity or yield measurements – which are easily performed in HTE-type formats – determining large numbers of kinetic parameters *via* rate analysis is much more challenging. Rate data is also not as widely reported in the literature, limiting the usefulness of reaction database mining. A recent successful example of using literature rate data to train an ML QSRR model was reported by Jorner, Brinck, Norrby, and Buttar (yet another academic/industry collaboration!).[63] Using a combination of >400 experimental S_NAr rate constants from literature, the authors trained an ML model demonstrating excellent prediction accuracy across a wide range of reported S_NAr reactions. While this study focused on S_NAr chemistry and not organopalladium chemistry, their *approach* is likely generally applicable (e.g. to organopalladium reactions) provided sufficient data is available; however, as described above, rate data is much less reported in the literature, and therefore must be collected experimentally.

Traditional approaches to obtaining rate constants via reaction progress analysis, even using initial rates, is extremely time and resource intensive, and incompatible with high-throughput experimental approaches. To efficiently assemble rate-based datasets as a foundation for predictive modeling, alternative techniques are therefore required. The concept of "one-pot multi-substrate screening" was introduced by Kagan for its application in fast optimization of asymmetric catalysts for various enantioselective reactions.[116,117] In 2007-2008, Plenio and co-workers extended this concept to realize high-throughput kinetic analysis of Pd-catalyzed Sonogashira couplings.[118,119] The authors of these two studies were able to simultaneously collect initial rates for up to 25 substrates in a single reaction flask. This significantly reduces the number of measurements needed to generate large kinetic datasets. Their QSRR studies combined 29 aryl bromide substrates with 17 monophosphine ligands for a total of 410 individual reactions. This pioneering study demonstrates the power of creative experimental design in high-throughput

experimentation, where more individual experiments is not necessarily required to amass sufficient data. While the quantitative analysis provided by the authors does not attempt to create a unified multivariate model, this dataset was invaluable to our group as an external test set for catalytic predictions (*vide infra)*.

Plenio and co-workers do note some limitations of the one-pot, multi-substrate kinetic analysis approach. One challenge is measuring initial rates for very fast reactions, due to the need for offline analysis by (in their case) GC. Another challenge specific to this method is to accurately measure rates for the slowest reactions in a multi-substrate set, as the effective catalyst concentration per substrate will increase once the faster reactions finish. Thus, care must be taken to validate the results from these multi-substrate studies. Nevertheless, this approach is likely applicable to many reaction types, and should prove valuable for high-throughput kinetic analysis.

Our group has also considered the challenges of collecting large kinetic parameter datasets, and have taken a different approach to reducing the experimental burden. Rather than collect rate data from multiple measurements of concentration over time, we opted to measure *relative* rates *via* competition experiments. This effectively converts rate measurements to selectivity measurements. Using this approach, we recently reported two QSRR models for accurate prediction of S_NAr[99] and Pd-catalyzed cross-coupling reactivity.[58] By correlating calculated molecular descriptors to relative free energies of activation ($\Delta\Delta G^{\ddagger}$), we obtained MLR models that have excellent performance to predict rate and selectivity for many external datasets. A discussion of the Pd-based work is illustrative of our approach.

Rather than study a specific class of Pd-catalyzed reaction, we focused our QSRR study on a key step in the catalytic mechanism: oxidative addition. This fundamental transformation is common to myriad catalytic reactions, and is often turnover and/or selectivity determining in Pd-catalyzed transformations (Figure 1.7A). We therefore hypothesized that a QSRR model for Ar–X oxidative addition to Pd(0) would be applicable to many different Pd-catalyzed cross-coupling reactions. To collect the required rate data, we assembled a diverse set of (hetero)aryl halides with a variety of substitution patterns,

and ensuring heterocyclic substrates were well-represented. Using a model Pd(0) complex – $Pd(PCy_3)_2$ – we then performed a series of competition experiments under *pseudo* first-order conditions by having two Ar–X electrophiles in excess but equal amount compete for oxidative addition (Figure 1.7B). The product ratio, and therefore relative rate, is easily measured by quantitative ^{31}P NMR spectroscopy, and control experiments confirmed that the product ratio is kinetically controlled. This technique enables rapid assembly of a reactivity scale for substrates that spans many orders of magnitude in rate (Figure 1.7C).

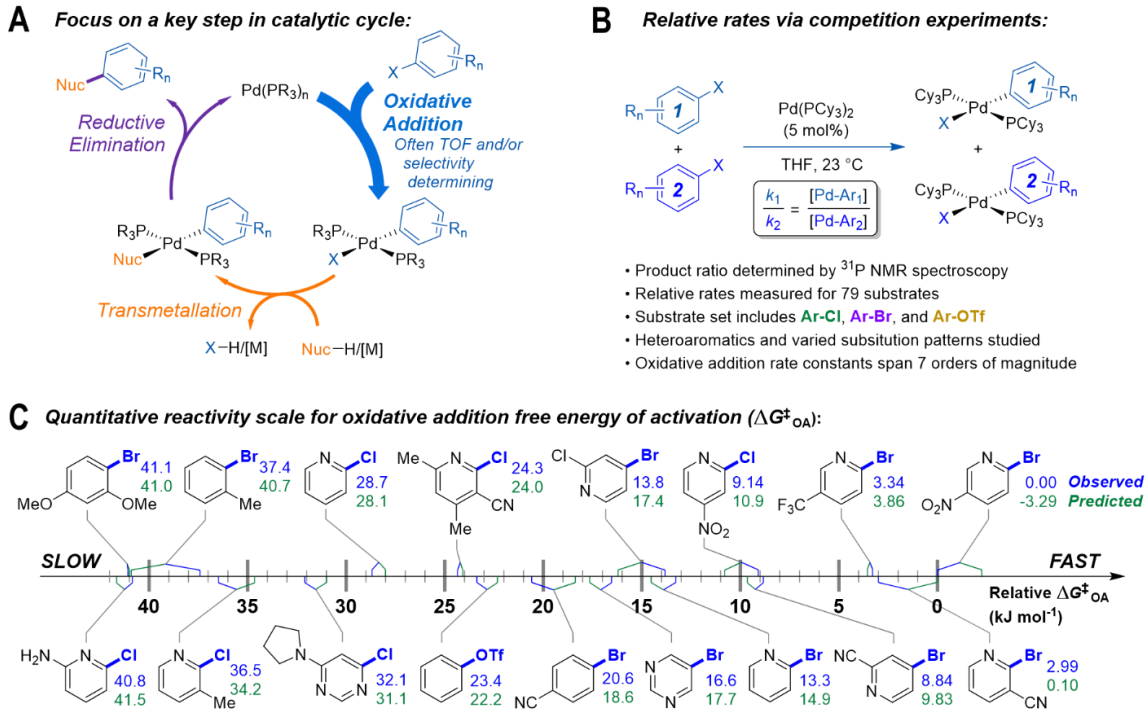


Figure 1.7. A) Generic catalytic cycle for Pd-catalyzed cross-coupling, highlighting the importance of oxidative addition. B) Experimental approach to rapid relative rate collection via competition experiments, with analysis of L2Pd(Ar)(X) complexes by 31P qNMR spectroscopy. C) Reactivity scale for representative substrates. Figure adapted with permission from ref. 58. Copyright 2022, Royal Society of Chemistry.

During feature selection for our QSRR models, we quickly determined that the average molecular electrostatic potential (*ESP*) is a particularly useful electronic descriptor for this reaction. These are the average *ESP* energy across the surface area of an atom in the molecule,[120] and are easily calculated using the wavefunction analysis application Multiwfn.[121] Molecular *ESP* has been previously used to interpret the chemical properties and reactivity of various molecular systems.[122] Germane to this Perspective, Suresh and

Koga[123] correlated minimum molecular *ESP* values at P to a variety of phosphine molecular properties, demonstrating *ESP* is an excellent descriptor for the electronic nature of phosphine ligands. Subsequent work from Anjali and Suresh correlated molecular *ESP* at Pd to computationally-determined activation barriers for oxidative addition of various Ph–X substrates to L–Pd(0) fragments.[124]

In our oxidative addition QSRR model, five descriptors are sufficient to describe the reactivity for Ar–Cl, Ar–Br, and Ar–OTf substrates (Figure 1.8). The most significant contributors are the average molecular *ESP* at the reactive carbon (ESP_1) and at the adjacent atom (C or N, ESP_2); combined, these values contribute >60% to the model output. These local descriptors not only indicate the extent of electron-deficiency, but also the polarization of the C=C or C=N bond. Steric effects were initially accounted for using simple tabulated *A* values,[125] and work is underway to incorporate calculated steric descriptors.[60] Finally, the identity of the leaving group is linked to two additional descriptors: the C–X bond strength, given by the intrinsic bond strength index (*IBSI*),[126] and the pK_a of the leaving group conjugate acid. Both of these descriptors are necessary to unify the three classes of electrophile, and subsequent work indicates this set of features enables incorporating Ar–I substrates as a fourth substrate class.[127]

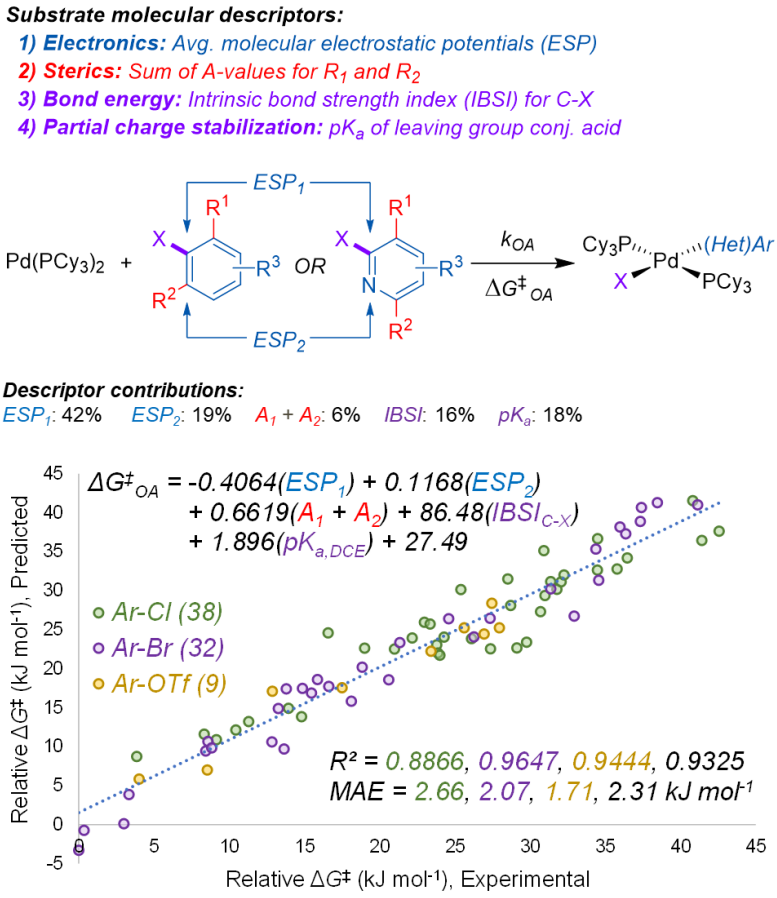


Figure 1.8. Multivariate QSRR model for Ar–X oxidative addition to Pd(PCy3)2. Figure adapted with permission from ref. 58. Copyright 2022, Royal Society of Chemistry.

To test our hypothesis that a QSRR model for oxidative addition would be applicable to multiple classes of catalytic reactions, we used predicted $\Delta G^{\ddagger}_{OA}$ values from the model to make predictions about site selectivity in multihalogenated electrophiles. There is a plethora of literature available on cross-coupling site selectivity,[128] and our analysis indicated that the MLR model in Figure 1.8 is able to predict "conventional" site selectivity for a wide range of substrates in both Suzuki and Buchwald-Hartwig reactions. This is the expected site selectivity when using "simple" catalyst systems, such as those based on PPh_3, dppf, or other common phosphines. Given we used $Pd(PCy_3)_2$ as the Pd(0) source to generate our model, this consistency with simple phosphines makes sense. Neufeldt's group have assessed factors that lead to unconventional site selectivity in Pd-catalyzed coupling, many of which are related to ligand and/or solvent identity, and their impacts on

Pd speciation.[70,71,129–133] Future work in QSRR modeling for oxidative addition must therefore endeavor to capture these ligand/solvent/speciation effects.

Finally, we tested our oxidative addition QSRR model for predicting external catalytic rate data. The aforementioned Sonogashira dataset from Plenio and co-workers was an ideal and comprehensive case study, especially since the data could be cleanly separated into two distinct sets (Figure 1.9). In one paper, the authors focused on electronic effects (*meta* and *para* substitution), with 20 different Ar–Br substrates (Substrate set #1).[119] In another report, only steric effects were studied (*ortho* substitution), with 9 additional Ar–Br substrates (Substrate set #2, which also contains Ph–Br from Substrate set #1).[118] Importantly, most of the substrates from both sets are not included in our $\Delta G^{\ddagger}_{OA}$ training set. The inclusion of 17 different monophosphine ligands also provided an opportunity to test our model's applicability beyond PCy_3-ligated catalysts.

We exclusively used Substrate set #1 as our training and test set to build an initial MLR model. For each substrate, we calculated the predicted $\Delta G^{\ddagger}_{OA}$ value from the equation in Figure 1.8; effectively, this is a pre-weighted "super-descriptor" that takes into account the steric and electronic effects of the electrophile. For the phosphine ligands, we used two descriptors – average molecular *ESP* at P (ESP_P) and % buried volume ($\%V_{bur}$)[59] – to account for electronic and steric effects, respectively. These three descriptors – $\Delta G^{\ddagger}_{OA}$, ESP_P, and $\%V_{bur}$ – lead to an accurate linear model for ln *k* of the Sonogashira reactions in Substrate set #1, with statistics given in Figure 1.9. Notably, ESP_P and the classic Tolman electronic parameter (TEP)[134] are closely correlated,[123] enabling replacement of an established and successful experimental descriptor with an easy-to-obtain computational version.

To challenge this QSRR model, we reserved data from Substrate set #2, which contains all *ortho*-substituted substrates, as an external validation set. Even though the MLR model training data had none of these substrates included, predictions for Substrate set #2 are still excellent, with a mean absolute error of 0.732 (compared to 0.529-0.542 for the training/test sets). This is possible because steric effects are already accounted for in the predicted $\Delta G^{\ddagger}_{OA}$ values, since that model was trained using a far more diverse range of

(Het)Ar–X substrates. In addition, the predicted $\Delta G^{\ddagger}_{OA}$ model appears to function equally well for all 17 phosphine ligands. Only two outlier points are observed, corresponding to 2,4,6-triisopropylphenyl bromide as a substrate with P(*t*Bu)$_3$ and P(Ad)$_2$(*t*Bu); in other words, the most sterically hindered substrate with the two largest phosphines. Individual univariate correlations of $\Delta G^{\ddagger}_{OA}$ with ln *k* for all 17 phosphines confirm this generality. Thus, even though our initial oxidative addition QSRR model was built using only Pd(PCy$_3$)$_2$, it is clearly able to make accurate predictions for many additional catalysts.

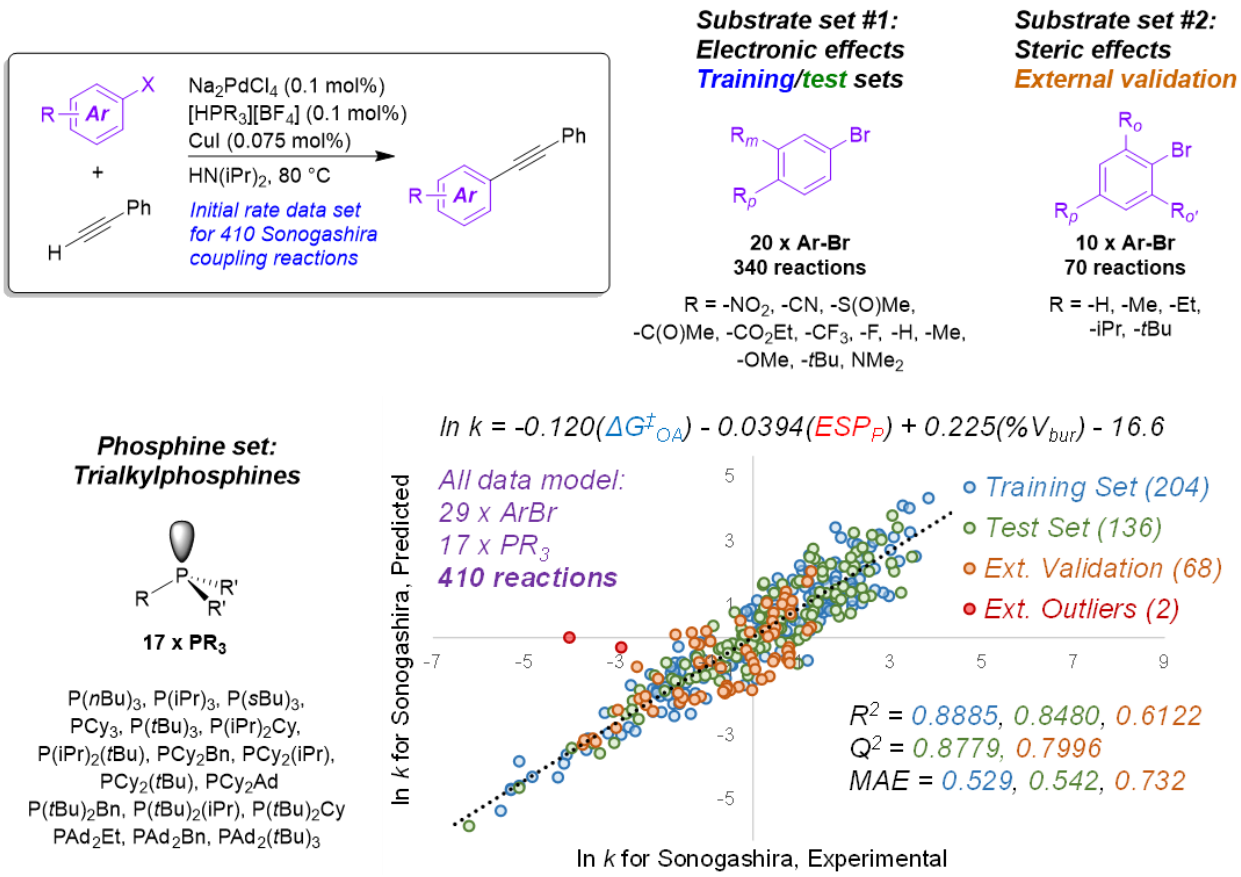


Figure 1.9. Applying predicted $\Delta G^{\ddagger}_{OA}$ values to external catalytic reaction datasets: unified QSRR of Plenio and co-workers' Sonogashira initial rates for all substrate/catalyst combinations. Figure adapted with permission from ref. 58. Copyright 2022, Royal Society of Chemistry.

1.6 Conclusions and Outlook

The art and practice of catalyst and reaction development in synthetic chemistry is in a state of change. The advent of rapid tools to collect, visualize, and analyze the plethora of existing data from the academic and patent literatures is enabling chemists to use these data like never before. Increases in computing power and speed are greatly expanding the utility of high-throughput computational chemistry approaches to mapping chemical space with molecular descriptors, as well as rapid and accurate calculation of reaction intermediates and transition states for ever more complex molecules. And advances in laboratory automation and miniaturization are changing the way chemists design and conduct experiments, moving from traditional "one-factor-at-a-time" iteration to more holistic, multifactor approaches. Organopalladium catalysis is central to each of these three aspects, providing large existing datasets, rich catalyst structure space and mechanistic diversity, and a plethora of high-throughput studies.

With these powerful new tools and approaches becoming more mainstream in catalysis research, we must be cognizant of the challenges and drawbacks of each individual approach. Literature/patent data is known to be skewed by selection and reporting biases; computational results are not acquired instantaneously, nor are they error-free; and HTE is still bottlenecked by reagent/catalyst availability, and by analysis speed/accuracy. The ideal data-rich approach to catalysis research must combine all three sources of data and the corresponding analysis methods. Importantly, massively large datasets are not required to make accurate and actionable reactivity/selectivity predictions, especially if mechanistic aspects are taken into account. As an exemplar of this fact, our work on quantitative predictions for Pd oxidative addition reactivity is broadly applicable across multiple reaction classes and accurate when presented with new data, despite being built from only ~80 relative rate constants.

While this Perspective showed how organopalladium catalysis has been the proving ground of choice for many data-rich methods, there is nothing inherent about palladium or its reactivity that makes it uniquely suited for these new approaches. Many other reaction classes are well-represented in the chemical literature, and high-throughput experimentation methods are increasingly accessible to many reaction types.[90,135] The

concepts from the case studies presented here, as well as others from the literature, are applicable to any type of synthetic chemistry. Finally, as remarked throughout, many of the seminal works on data-driven catalysis research are the result of academic/industry collaborations. These research partnerships will likely continue to be at the forefront of this emerging area.

1.7 References

(1) Trost, B. M.; Crawley, M. L. Asymmetric Transition-Metal-Catalyzed Allylic Alkylations: Applications in Total Synthesis. *Chem. Rev.* **2003**, *103*, 2921–2944. https://doi.org/10.1021/cr020027w

(2) Nicolaou, K. C.; Bulger, P. G.; Sarlah, D. Palladium-Catalyzed Cross-Coupling Reactions in Total Synthesis. *Angew. Chem. Int. Ed.* **2005**, *44*, 4442–4489. https://doi.org/10.1002/anie.200500368

(3) Bai, Y.; Davis, D. C.; Dai, M. Natural Product Synthesis via Palladium-Catalyzed Carbonylation. *J. Org. Chem.* **2017**, *82*, 2319–2328. https://doi.org/10.1021/acs.joc.7b00009

(4) Magano, J.; Dunetz, J. R. Large-Scale Applications of Transition Metal-Catalyzed Couplings for the Synthesis of Pharmaceuticals. *Chem. Rev.* **2011**, *111*, 2177–2250. https://doi.org/10.1021/cr100346g

(5) Ruiz-Castillo, P.; Buchwald, S. L. Applications of Palladium-Catalyzed C–N Cross-Coupling Reactions. *Chem. Rev.* **2016**, *116*, 12564–12649. https://doi.org/10.1021/acs.chemrev.6b00512

(6) Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **2016**, *59*, 4443–4458. https://doi.org/10.1021/acs.jmedchem.5b01409

(7) Devendar, P.; Qu, R.-Y.; Kang, W.-M.; He, B.; Yang, G.-F. Palladium-Catalyzed Cross-Coupling Reactions: A Powerful Tool for the Synthesis of Agrochemicals. *J. Agric. Food Chem.* **2018**, *66*, 8914–8934. https://doi.org/10.1021/acs.jafc.8b03792

(8) Baker, M. A.; Tsai, C.-H.; Noonan, K. J. T. Diversifying Cross-Coupling Strategies, Catalysts and Monomers for the Controlled Synthesis of Conjugated Polymers. *Chem. – Eur. J.* **2018**, *24*, 13078–13088. https://doi.org/10.1002/chem.201706102

(9) Han, F.-S. Transition-Metal-Catalyzed Suzuki–Miyaura Cross-Coupling Reactions: A Remarkable Advance from Palladium to Nickel Catalysts. *Chem. Soc. Rev.* **2013**, *42* (12), 5270–5298. https://doi.org/10.1039/C3CS35521G

(10) Tasker, S. Z.; Standley, E. A.; Jamison, T. F. Recent Advances in Homogeneous Nickel Catalysis. *Nature* **2014**, *509*, 299–309. https://doi.org/10.1038/nature13274

(11) Ananikov, V. P. Nickel: The "Spirited Horse" of Transition Metal Catalysis. *ACS Catal.* **2015**, *5*, 1964–1971. https://doi.org/10.1021/acscatal.5b00072

(12) Coombs, J. R.; Green, R. A.; Roberts, F.; Simmons, E. M.; Stevens, J. M.; Wisniewski, S. R. Advances in Base-Metal Catalysis: Development of a Screening Platform for Nickel-Catalyzed Borylations of Aryl (Pseudo)Halides with $B_2(OH)_4$. *Organometallics* **2019**, *38*, 157–166. https://doi.org/10.1021/acs.organomet.8b00307

(13) Chirik, P. J.; Engle, K. M.; Simmons, E. M.; Wisniewski, S. R. Collaboration as a Key to Advance Capabilities for Earth-Abundant Metal Catalysis. *Org. Process Res. Dev.* **2023**, *27*, 1160–1184. https://doi.org/10.1021/acs.oprd.3c00025

(14) Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; Eugenio de Diego, J.; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post, R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambeanu, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **2019**, *23*, 1213–1242. https://doi.org/10.1021/acs.oprd.9b00140

(15) Leitch, D. C.; Becica, J. 13.12 - High-Throughput Experimentation in Organometallic Chemistry and Catalysis. In *Comprehensive Organometallic Chemistry IV*; Parkin, G., Meyer, K., O'hare, D., Eds.; Elsevier: Oxford, 2022; pp 502–555. https://doi.org/10.1016/B978-0-12-820206-7.00111-6

(16) Murphy, V. High-Throughput Organometallic Chemistry: Chemical Approaches, Experimental Methods, and Screening Techniques. In *Comprehensive Organometallic Chemistry III*; Mingos, D. M. P., Crabtree, R. H., Eds.; Elsevier: Oxford, 2007; pp 341–379. https://doi.org/10.1016/B0-08-045047-4/00014-5

(17) Clayton, A. D.; Manson, J. A.; Taylor, C. J.; Chamberlain, T. W.; Taylor, B. A.; Clemens, G.; Bourne, R. A. Algorithms for the Self-Optimisation of Chemical Reactions. *React. Chem. Eng.* **2019**, *4*, 1545–1554. https://doi.org/10.1039/C9RE00209J

(18) Christensen, M.; Yunker, L. P. E.; Adedeji, F.; Häse, F.; Roch, L. M.; Gensch, T.; dos Passos Gomes, G.; Zepel, T.; Sigman, M. S.; Aspuru-Guzik, A.; Hein, J. E. Data-Science Driven Autonomous Process Optimization. *Commun. Chem.* **2021**, *4*, 112. https://doi.org/10.1038/s42004-021-00550-x

(19) Taylor, C. J.; Pomberger, A.; Felton, K. C.; Grainger, R.; Barecka, M.; Chamberlain, T. W.; Bourne, R. A.; Johnson, C. N.; Lapkin, A. A. A Brief Introduction to Chemical Reaction Optimization. *Chem. Rev.* **2023**, *123*, 3089–3126. https://doi.org/10.1021/acs.chemrev.2c00798

(20) Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A. The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res.* **2016**, *49*, 1292–1301. https://doi.org/10.1021/acs.accounts.6b00194

(21) Santiago, C. B.; Guo, J.-Y.; Sigman, M. S. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018**, *9*, 2398–2412. https://doi.org/10.1039/C7SC04679K

(22) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. QSAR without Borders. *Chem. Soc. Rev.* **2020**, *49*, 3525–3564. https://doi.org/10.1039/D0CS00098A

(23) Shen, Y.; Borowski, J. E.; Hardy, M. A.; Sarpong, R.; Doyle, A. G.; Cernak, T. Automation and Computer-Assisted Planning for Chemical Synthesis. *Nat. Rev. Methods Primer* **2021**, *1*, 23. https://doi.org/10.1038/s43586-021-00022-5

(24) Durand, D. J.; Fey, N. Building a Toolbox for the Analysis and Prediction of Ligand and Catalyst Effects in Organometallic Catalysis. *Acc. Chem. Res.* **2021**, *54*, 837–848. https://doi.org/10.1021/acs.accounts.0c00807

(25) Williams, W. L.; Zeng, L.; Gensch, T.; Sigman, M. S.; Doyle, A. G.; Anslyn, E. V. The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **2021**, *7*, 1622–1637. https://doi.org/10.1021/acscentsci.1c00535

(26) Rinehart, N. I.; Zahrt, A. F.; Henle, J. J.; Denmark, S. E. Dreams, False Starts, Dead Ends, and Redemption: A Chronicle of the Evolution of a Chemoinformatic Workflow for the Optimization of Enantioselective Catalysts. *Acc. Chem. Res.* **2021**, *54*, 2041–2054. https://doi.org/10.1021/acs.accounts.0c00826

(27) Strieth-Kalthoff, F.; Sandfort, F.; Segler, M. H. S.; Glorius, F. Machine Learning the Ropes: Principles, Applications and Directions in Synthetic Chemistry. *Chem. Soc. Rev.* **2020**, *49*, 6154–6168. https://doi.org/10.1039/C9CS00786E

(28) Lustosa, D. M.; Milo, A. Mechanistic Inference from Statistical Models at Different Data-Size Regimes. *ACS Catal.* **2022**, *12*, 7886–7906. https://doi.org/10.1021/acscatal.2c01741

(29) Jorner, K.; Tomberg, A.; Bauer, C.; Sköld, C.; Norrby, P.-O. Organic Reactivity from Mechanism to Machine Learning. *Nat. Rev. Chem.* **2021**, *5*, 240–255. https://doi.org/10.1038/s41570-021-00260-x

(30) Zahrt, A. F.; Athavale, S. V.; Denmark, S. E. Quantitative Structure–Selectivity Relationships in Enantioselective Catalysis: Past, Present, and Future. *Chem. Rev.* **2020**, *120*, 1620–1689. https://doi.org/10.1021/acs.chemrev.9b00425

(31) Artrith, N.; Butler, K. T.; Coudert, F.-X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best Practices in Machine Learning for Chemistry. *Nat. Chem.* **2021**, *13*, 505–508. https://doi.org/10.1038/s41557-021-00716-z

(32) Strieth-Kalthoff, F.; Sandfort, F.; Kühnemund, M.; Schäfer, F. R.; Kuchen, H.; Glorius, F. Machine Learning for Chemical Reactivity: The Importance of Failed Experiments. *Angew. Chem. Int. Ed.* **2022**, *61*, e202204647. https://doi.org/10.1002/anie.202204647

(33) Beker, W.; Roszak, R.; Wołos, A.; Angello, N. H.; Rathore, V.; Burke, M. D.; Grzybowski, B. A. Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki–Miyaura Coupling. *J. Am. Chem. Soc.* **2022**, *144*, 4819–4827. https://doi.org/10.1021/jacs.1c12005

(34) Fitzner, M.; Wuitschik, G.; Koller, R.; Adam, J.-M.; Schindler, T. Machine Learning C–N Couplings: Obstacles for a General-Purpose Reaction Yield Prediction. *ACS Omega* **2023**, *8*, 3017–3025. https://doi.org/10.1021/acsomega.2c05546

(35) Kowalczyk, B.; Bishop, K. J. M.; Smoukov, S. K.; Grzybowski, B. A. Synthetic Popularity Reflects Chemical Reactivity. *J. Phys. Org. Chem.* **2009**, *22*, 897–902. https://doi.org/10.1002/poc.1535

(36) Soh, S.; Wei, Y.; Kowalczyk, B.; Gothard, C. M.; Baytekin, B.; Gothard, N.; Grzybowski, B. A. Estimating Chemical Reactivity and Cross-Influence from Collective Chemical Knowledge. *Chem. Sci.* **2012**, *3*, 1497–1502. https://doi.org/10.1039/C2SC00011C

(37) Kozlowski, M. C. On the Topic of Substrate Scope. *Org. Lett.* **2022**, *24*, 7247–7249. https://doi.org/10.1021/acs.orglett.2c03246

(38) Moreno-Mañas, M.; Pérez, M.; Pleixats, R. Palladium-Catalyzed Suzuki-Type Self-Coupling of Arylboronic Acids. A Mechanistic Study. *J. Org. Chem.* **1996**, *61*, 2346–2351. https://doi.org/10.1021/jo9514329

(39) Lehmler, H.-J.; Robertson, L. W. Synthesis of Polychlorinated Biphenyls (PCBs) Using the Suzuki-Coupling. *Chemosphere* **2001**, *45*, 137–143. https://doi.org/10.1016/S0045-6535(00)00546-4

(40) Fitzner, M.; Wuitschik, G.; J. Koller, R.; Adam, J.-M.; Schindler, T.; Reymond, J.-L. What Can Reaction Databases Teach Us about Buchwald–Hartwig Cross-Couplings? *Chem. Sci.* **2020**, *11*, 13085–13093. https://doi.org/10.1039/D0SC04074F

(41) Kutchukian, P. S.; Dropinski, J. F.; Dykstra, K. D.; Li, B.; DiRocco, D. A.; Streckfuss, E. C.; Campeau, L.-C.; Cernak, T.; Vachal, P.; Davies, I. W.; Krska, S. W.; Dreher, S. D. Chemistry Informer Libraries: A Chemoinformatics Enabled Approach to Evaluate and Advance Synthetic Methods. *Chem. Sci.* **2016**, *7*, 2604–2613. https://doi.org/10.1039/C5SC04751J

(42) Kearnes, S. M.; Maser, M. R.; Wleklinski, M.; Kast, A.; Doyle, A. G.; Dreher, S. D.; Hawkins, J. M.; Jensen, K. F.; Coley, C. W. The Open Reaction Database. *J. Am. Chem. Soc.* **2021**, *143*, 18820–18826. https://doi.org/10.1021/jacs.1c09820

(43) Ertl, P.; Gerebtzoff, G.; Lewis, R.; Muenkler, H.; Schneider, N.; Sirockin, F.; Stiefl, N.; Tosco, P. Chemical Reactivity Prediction: Current Methods and Different Application Areas. *Mol. Inform.* **2022**, *41*, 2100277. https://doi.org/10.1002/minf.202100277

(44) Wernerova, M.; Hudlicky, T. On the Practical Limits of Determining Isolated Product Yields and Ratios of Stereoisomers: Reflections, Analysis, and Redemption. *Synlett* **2010**, 2701–2707. https://doi.org/10.1055/s-0030-1259018

(45) Sperger, T.; Sanhueza, I. A.; Schoenebeck, F. Computation and Experiment: A Powerful Combination to Understand and Predict Reactivities. *Acc. Chem. Res.* **2016**, *49*, 1311–1319. https://doi.org/10.1021/acs.accounts.6b00068

(46) Wodrich, M. D.; Sawatlon, B.; Busch, M.; Corminboeuf, C. The Genesis of Molecular Volcano Plots. *Acc. Chem. Res.* **2021**, *54*, 1107–1117. https://doi.org/10.1021/acs.accounts.0c00857

(47) Laplaza, R.; Das, S.; Wodrich, M. D.; Corminboeuf, C. Constructing and Interpreting Volcano Plots and Activity Maps to Navigate Homogeneous Catalyst Landscapes. *Nat. Protoc.* **2022**, *17*, 2550–2569. https://doi.org/10.1038/s41596-022-00726-2

(48) Sabatier, P. Hydrogénations et Déshydrogénations Par Catalyse. *Ber. Dtsch. Chem. Ges.* **1911**, *44*, 1984-2001. https://doi.org/10.1002/cber.19110440303

(49) Sabatier, P. *La Catalyse en Chimie Organique,* 2nd ed.; Béranger, 1913.

(50) Busch, M.; Wodrich, M. D.; Corminboeuf, C. Linear Scaling Relationships and Volcano Plots in Homogeneous Catalysis – Revisiting the Suzuki Reaction. *Chem. Sci.* **2015**, *6*, 6754–6761. https://doi.org/10.1039/C5SC02910D

(51) Meyer, B.; Sawatlon, B.; Heinen, S.; Lilienfeld, O. A. von; Corminboeuf, C. Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts. *Chem. Sci.* **2018**, *9*, 7069–7077. https://doi.org/10.1039/C8SC01949E

(52) Sawatlon, B.; Wodrich, M. D.; Meyer, B.; Fabrizio, A.; Corminboeuf, C. Data Mining the C−C Cross-Coupling Genome. *ChemCatChem* **2019**, *11*, 4096–4107. https://doi.org/10.1002/cctc.201900597

(53) Sawatlon, B.; Wodrich, M. D.; Corminboeuf, C. Probing Substrate Scope with Molecular Volcanoes. *Org. Lett.* **2020**, *22*, 7936–7941. https://doi.org/10.1021/acs.orglett.0c02862

(54) Liu, D.; Xu, Z.; Lu, X.; Yu, H.; Fu, Y. Linear Regression Model for Predicting Allyl Alcohol C–O Bond Activity under Palladium Catalysis. *ACS Catal.* **2022**, *12*, 13921–13929. https://doi.org/10.1021/acscatal.2c03847

(55) Jing, J.; Huo, X.; Shen, J.; Fu, J.; Meng, Q.; Zhang, W. Direct Use of Allylic Alcohols and Allylic Amines in Palladium-Catalyzed Allylic Amination. *Chem. Commun.* **2017**, *53*, 5151–5154. https://doi.org/10.1039/C7CC01069A

(56) Lyngvi, E.; Sanhueza, I. A.; Schoenebeck, F. Dispersion Makes the Difference: Bisligated Transition States Found for the Oxidative Addition of $Pd(PtBu_3)_2$ to Ar-OSO_2R and Dispersion-Controlled Chemoselectivity in Reactions with $Pd[P(iPr)(tBu_2)]_2$. *Organometallics* **2015**, *34* (5), 805–812. https://doi.org/10.1021/om501199t

(57) Gadre, S. R.; Suresh, C. H.; Mohan, N. Electrostatic Potential Topology for Probing Molecular Structure, Bonding and Reactivity. *Molecules* **2021**, *26*, 3289. https://doi.org/10.3390/molecules26113289

(58) Lu, J.; Donnecke, S.; Paci, I.; Leitch, D. C. A Reactivity Model for Oxidative Addition to Palladium Enables Quantitative Predictions for Catalytic Cross-Coupling Reactions. *Chem. Sci.* **2022**, *13*, 3477–3488. https://doi.org/10.1039/D2SC00174H

(59) Clavier, H.; Nolan, S. P. Percent Buried Volume for Phosphine and N-Heterocyclic Carbene Ligands: Steric Properties in Organometallic Chemistry. *Chem. Commun.* **2010**, *46*, 841–861. https://doi.org/10.1039/B922984A

(60) Brethomé, A. V.; Fletcher, S. P.; Paton, R. S. Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* **2019**, *9*, 2313–2323. https://doi.org/10.1021/acscatal.8b04043

(61) McFord, A. W.; Butts, C. P.; Fey, N.; Alder, R. W. 3× Axial vs 3× Equatorial: The ΔG_{GA} Value Is a Robust Computational Measure of Substituent Steric Effects. *J. Am. Chem. Soc.* **2021**, *143*, 13573–13578. https://doi.org/10.1021/jacs.1c04247

(62) Milo, A.; Bess, E. N.; Sigman, M. S. Interrogating Selectivity in Catalysis Using Molecular Vibrations. *Nature* **2014**, *507*, 210–214. https://doi.org/10.1038/nature13019

(63) Jorner, K.; Brinck, T.; Norrby, P.-O.; Buttar, D. Machine Learning Meets Mechanistic Modelling for Accurate Prediction of Experimental Activation Energies. *Chem. Sci.* **2021**, *12*, 1163–1175. https://doi.org/10.1039/D0SC04896H

(64) Burai Patrascu, M.; Pottel, J.; Pinus, S.; Bezanson, M.; Norrby, P.-O.; Moitessier, N. From Desktop to Benchtop with Automated Computational Workflows for Computer-Aided Design in Asymmetric Catalysis. *Nat. Catal.* **2020**, *3*, 574–584. https://doi.org/10.1038/s41929-020-0468-3

(65) Zahrt, A. F.; Rose, B. T.; Darrow, W. T.; Henle, J. J.; Denmark, S. E. Computational Methods for Training Set Selection and Error Assessment Applied to Catalyst Design: Guidelines for Deciding Which Reactions to Run First and Which to Run Next. *React. Chem. Eng.* **2021**, *6*, 694–708. https://doi.org/10.1039/D1RE00013F

(66) Martin, R.; Buchwald, S. L. Palladium-Catalyzed Suzuki−Miyaura Cross-Coupling Reactions Employing Dialkylbiaryl Phosphine Ligands. *Acc. Chem. Res.* **2008**, *41*, 1461–1473. https://doi.org/10.1021/ar800036s

(67) Fu, G. C. The Development of Versatile Methods for Palladium-Catalyzed Coupling Reactions of Aryl Electrophiles through the Use of $P(t\text{-}Bu)_3$ and PCy_3 as Ligands. *Acc. Chem. Res.* **2008**, *41*, 1555–1564. https://doi.org/10.1021/ar800148f

(68) Surry, D. S.; Buchwald, S. L. Dialkylbiaryl Phosphines in Pd-Catalyzed Amination: A User's Guide. *Chem. Sci.* **2010**, *2*, 27–50. https://doi.org/10.1039/C0SC00331J

(69) Valente, C.; Belowich, M. E.; Hadei, N.; Organ, M. G. Pd-PEPPSI Complexes and the Negishi Reaction. *Eur. J. Org. Chem.* **2010**, 4343–4354. https://doi.org/10.1002/ejoc.201000359

(70) Norman, J. P.; Larson, N. G.; Neufeldt, S. R. Different Oxidative Addition Mechanisms for 12- and 14-Electron Palladium(0) Explain Ligand-Controlled Divergent Site Selectivity. *ACS Catal.* **2022**, *12*, 8822–8828. https://doi.org/10.1021/acscatal.2c01698

(71) Norman, J. P.; Neufeldt, S. R. The Road Less Traveled: Unconventional Site Selectivity in Palladium-Catalyzed Cross-Couplings of Dihalogenated N-

Heteroarenes. *ACS Catal.* **2022**, *12*, 12014–12026. https://doi.org/10.1021/acscatal.2c03743

(72) Durand, D. J.; Fey, N. Computational Ligand Descriptors for Catalyst Design. *Chem. Rev.* **2019**, *119*, 6561–6594. https://doi.org/10.1021/acs.chemrev.8b00588

(73) Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Murray, P.; Orpen, A. G.; Osborne, R.; Purdie, M. Computational Descriptors for Chelating P,P- and P,N-Donor Ligands. *Organometallics* **2008**, *27*, 1372–1383. https://doi.org/10.1021/om700840h

(74) Fey, N.; Haddow, M. F.; Harvey, J. N.; McMullin, C. L.; Orpen, A. G. A Ligand Knowledge Base for Carbenes (LKB-C): Maps of Ligand Space. *Dalton Trans.* **2009**, 8183–8196. https://doi.org/10.1039/B909229C

(75) Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M. Expansion of the Ligand Knowledge Base for Monodentate P-Donor Ligands (LKB-P). *Organometallics* **2010**, *29*, 6245–6258. https://doi.org/10.1021/om100648v

(76) Hueffel, J. A.; Sperger, T.; Funes-Ardoiz, I.; Ward, J. S.; Rissanen, K.; Schoenebeck, F. Accelerated Dinuclear Palladium Catalyst Identification through Unsupervised Machine Learning. *Science* **2021**, *374*, 1134–1140. https://doi.org/10.1126/science.abj0999

(77) Gensch, T.; dos Passos Gomes, G.; Friederich, P.; Peters, E.; Gaudin, T.; Pollice, R.; Jorner, K.; Nigam, A.; Lindner-D'Addario, M.; Sigman, M. S.; Aspuru-Guzik, A. A Comprehensive Discovery Platform for Organophosphorus Ligands for Catalysis. *J. Am. Chem. Soc.* **2022**, *144*, 1205–1217. https://doi.org/10.1021/jacs.1c09718

(78) Gensch, T.; Smith, S. R.; Colacot, T. J.; Timsina, Y. N.; Xu, G.; Glasspoole, B. W.; Sigman, M. S. Design and Application of a Screening Set for Monophosphine Ligands in Cross-Coupling. *ACS Catal.* **2022**, *12*, 7773–7780. https://doi.org/10.1021/acscatal.2c01970

(79) Shultz, C. S.; Krska, S. W. Unlocking the Potential of Asymmetric Hydrogenation at Merck. *Acc. Chem. Res.* **2007**, *40*, 1320–1326. https://doi.org/10.1021/ar700141v

(80) Krska, S. W.; DiRocco, D. A.; Dreher, S. D.; Shevlin, M. The Evolution of Chemical High-Throughput Experimentation To Address Challenging Problems in Pharmaceutical Synthesis. *Acc. Chem. Res.* **2017**, *50*, 2976–2985. https://doi.org/10.1021/acs.accounts.7b00428

(81) J. Welch, C. High Throughput Analysis Enables High Throughput Experimentation in Pharmaceutical Process Research. *React. Chem. Eng.* **2019**, *4*, 1895–1911. https://doi.org/10.1039/C9RE00234K

(82) Allen, C. L.; Leitch, D. C.; Anson, M. S.; Zajac, M. A. The Power and Accessibility of High-Throughput Methods for Catalysis Research. *Nat. Catal.* **2019**, *2*, 2–4. https://doi.org/10.1038/s41929-018-0220-4

(83) Isbrandt, E. S.; Sullivan, R. J.; Newman, S. G. High Throughput Strategies for the Discovery and Optimization of Catalytic Reactions. *Angew. Chem. Int. Ed.* **2019**, *58*, 7180–7191. https://doi.org/10.1002/anie.201812534

(84) *The Power of High-Throughput Experimentation: General Topics and Enabling Technologies for Synthesis and Catalysis (Volume 1)*. Emmert, M. H.; Jouffroy, M.; Leitch, D. C., Eds.; ACS Symposium Series, Vol. 1419; American Chemical Society, 2022. DOI: 10.1021/bk-2022-1419. https://doi.org/10.1021/bk-2022-1419

(85) *The Power of High-Throughput Experimentation: Case Studies from Drug Discovery, Drug Development, and Catalyst Discovery (Volume 2)*; Emmert, M. H.; Jouffroy, M.; Leitch, D. C., Eds.; ACS Symposium Series, Vol. 1420; American Chemical Society, 2022. DOI: 10.1021/bk-2022-1420. https://doi.org/10.1021/bk-2022-1420

(86) Santanilla, A. B.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.-C.; Schneeweis, J.; Berritt, S.; Shi, Z.-C.; Nantermet, P.; Liu, Y.; Helmy, R.; Welch, C. J.; Vachal, P.; Davies, I. W.; Cernak, T.; Dreher, S. D. Nanomole-Scale High-Throughput Chemistry for the Synthesis of Complex Molecules. *Science* **2015**, *347*, 49–53. https://doi.org/10.1126/science.1259203

(87) Lin, S.; Dikler, S.; Blincoe, W. D.; Ferguson, R. D.; Sheridan, R. P.; Peng, Z.; Conway, D. V.; Zawatzky, K.; Wang, H.; Cernak, T.; Davies, I. W.; DiRocco, D. A.; Sheng, H.; Welch, C. J.; Dreher, S. D. Mapping the Dark Space of Chemical Reactions with Extended Nanomole Synthesis and MALDI-TOF MS. *Science* **2018**, *361*, eaar6236. https://doi.org/10.1126/science.aar6236

(88) Perera, D.; Tucker, J. W.; Brahmbhatt, S.; Helal, C. J.; Chong, A.; Farrell, W.; Richardson, P.; Sach, N. W. A Platform for Automated Nanomole-Scale Reaction Screening and Micromole-Scale Synthesis in Flow. *Science* **2018**, *359*, 429–434. https://doi.org/10.1126/science.aap9112

(89) Mahjour, B.; Shen, Y.; Cernak, T. Ultrahigh-Throughput Experimentation for Information-Rich Chemical Synthesis. *Acc. Chem. Res.* **2021**, *54*, 2337–2346. https://doi.org/10.1021/acs.accounts.1c00119

(90) Gesmundo, N.; Dykstra, K.; Douthwaite, J. L.; Kao, Y.-T.; Zhao, R.; Mahjour, B.; Ferguson, R.; Dreher, S.; Sauvagnat, B.; Saurí, J.; Cernak, T. Miniaturization of Popular Reactions from the Medicinal Chemists' Toolbox for Ultrahigh-Throughput Experimentation. *Nat. Synth.* **2023**. https://doi.org/10.1038/s44160-023-00351-1

(91) Mahjour, B.; Zhang, R.; Shen, Y.; McGrath, A.; Zhao, R.; Mohamed, O. G.; Lin, Y.; Zhang, Z.; Douthwaite, J. L.; Tripathi, A.; Cernak, T. Rapid Planning and Analysis of High-Throughput Experiment Arrays for Reaction Discovery. *Nat. Commun.* **2023**, *14*, 3924. https://doi.org/10.1038/s41467-023-39531-0

(92) Cook, A.; Clément, R.; Newman, S. G. Reaction Screening in Multiwell Plates: High-Throughput Optimization of a Buchwald–Hartwig Amination. *Nat. Protoc.* **2021**, *16*, 1152–1169. https://doi.org/10.1038/s41596-020-00452-7

(93) Lee, J.; Schmink, J. R.; Berritt, S. Introduction of Low-Barrier High-Throughput Experimentation in the Undergraduate Laboratory: Suzuki—Miyaura Reaction. *J. Chem. Educ.* **2020**, *97*, 538–542. https://doi.org/10.1021/acs.jchemed.9b00794

(94) Becica, J.; Hruszkewycz, D. P.; Steves, J. E.; Elward, J. M.; Leitch, D. C.; Dobereiner, G. E. High-Throughput Discovery and Evaluation of a General Catalytic Method for N-Arylation of Weakly Nucleophilic Sulfonamides. *Org. Lett.* **2019**, *21*, 8981–8986. https://doi.org/10.1021/acs.orglett.9b03380

(95) Becica, J.; Heath, O. R. J.; Zheng, C. H. M.; Leitch, D. C. Palladium-Catalyzed Cross-Coupling of Alkenyl Carboxylates. *Angew. Chem. Int. Ed.* **2020**, *59*, 17277–17281. https://doi.org/10.1002/anie.202006586

(96) Gaube, G.; Fernandez, N. P.; C. Leitch, D. An Evaluation of Palladium-Based Catalysts for the Base-Free Borylation of Alkenyl Carboxylates. *New J. Chem.* **2021**, *45*, 20095–20098. https://doi.org/10.1039/D1NJ04008A

(97) Pipaón Fernández, N.; Gaube, G.; Woelk, K. J.; Burns, M.; Hruszkewycz, D. P.; Leitch, D. C. Palladium-Catalyzed Direct C–H Alkenylation with Enol Pivalates Proceeds via Reversible C–O Oxidative Addition to Pd(0). *ACS Catal.* **2022**, *12*, 6997–7003. https://doi.org/10.1021/acscatal.2c01305

(98) Dhake, K.; Woelk, K. J.; Becica, J.; Un, A.; Jenny, S. E.; Leitch, D. C. Beyond Bioisosteres: Divergent Synthesis of Azabicyclohexanes and Cyclobutenyl Amines from Bicyclobutanes. *Angew. Chem. Int. Ed.* **2022**, *61*, e202204719. https://doi.org/10.1002/anie.202204719

(99) Lu, J.; Paci, I.; Leitch, D. C. A Broadly Applicable Quantitative Relative Reactivity Model for Nucleophilic Aromatic Substitution (S_NAr) Using Simple Descriptors. *Chem. Sci.* **2022**, *13*, 12681–12695. https://doi.org/10.1039/D2SC04041G

(100) Becica, J.; Glaze, O. D.; Hruszkewycz, D. P.; Dobereiner, G. E.; Leitch, D. C. The Influence of Additives on Orthogonal Reaction Pathways in the Mizoroki–Heck Arylation of Vinyl Ethers. *React. Chem. Eng.* **2021**, *6*, 1212–1219. https://doi.org/10.1039/D1RE00124H

(101) Huang, J.; Isaac, M.; Watt, R.; Becica, J.; Dennis, E.; Saidaminov, M. I.; Sabbers, W. A.; Leitch, D. C. ^{DMP}DAB–Pd–MAH: A Versatile Pd(0) Source for Precatalyst Formation, Reaction Screening, and Preparative-Scale Synthesis. *ACS Catal.* **2021**, *11*, 5636–5646. https://doi.org/10.1021/acscatal.1c00288

(102) Sather, A. C.; Martinot, T. A. Data-Rich Experimentation Enables Palladium-Catalyzed Couplings of Piperidines and Five-Membered (Hetero)Aromatic Electrophiles. *Org. Process Res. Dev.* **2019**, *23*, 1725–1739. https://doi.org/10.1021/acs.oprd.9b00233

(103) Atwater, B.; Chandrasoma, N.; Mitchell, D.; Rodriguez, M. J.; Organ, M. G. Pd-PEPPSI-IHeptCl: A General-Purpose, Highly Reactive Catalyst for the Selective Coupling of Secondary Alkyl Organozincs. *Chem. – Eur. J.* **2016**, *22*, 14531–14534. https://doi.org/10.1002/chem.201603603

(104) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186–190. https://doi.org/10.1126/science.aar5169

(105) Chuang, K. V.; Keiser, M. J. Comment on "Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning." *Science* **2018**, *362*, eaat8603. https://doi.org/10.1126/science.aat8603

(106) Estrada, J. G.; Ahneman, D. T.; Sheridan, R. P.; Dreher, S. D.; Doyle, A. G. Response to Comment on "Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning." *Science* **2018**, *362*, eaat8763. https://doi.org/10.1126/science.aat8763

(107) Lones, M. A. How to Avoid Machine Learning Pitfalls: A Guide for Academic Researchers. *arXiv* **2023**. https://doi.org/10.48550/arXiv.2108.02497

(108) Dotson, J. J.; van Dijk, L.; Timmerman, J. C.; Grosslight, S.; Walroth, R. C.; Gosselin, F.; Püntener, K.; Mack, K. A.; Sigman, M. S. Data-Driven Multi-Objective Optimization Tactics for Catalytic Asymmetric Reactions Using Bisphosphine Ligands. *J. Am. Chem. Soc.* **2023**, *145*, 110–121. https://doi.org/10.1021/jacs.2c08513

(109) Newman-Stonebraker, S. H.; Smith, S. R.; Borowski, J. E.; Peters, E.; Gensch, T.; Johnson, H. C.; Sigman, M. S.; Doyle, A. G. Univariate Classification of Phosphine Ligation State and Reactivity in Cross-Coupling Catalysis. *Science* **2021**, *374*, 301–308. https://doi.org/10.1126/science.abj4213

(110) Xu, J.; Grosslight, S.; Mack, K. A.; Nguyen, S. C.; Clagg, K.; Lim, N.-K.; Timmerman, J. C.; Shen, J.; White, N. A.; Sirois, L. E.; Han, C.; Zhang, H.; Sigman, M. S.; Gosselin, F. Atroposelective Negishi Coupling Optimization Guided by Multivariate Linear Regression Analysis: Asymmetric Synthesis of KRAS G12C Covalent Inhibitor GDC-6036. *J. Am. Chem. Soc.* **2022**, *144*, 20955–20963. https://doi.org/10.1021/jacs.2c09917

(111) Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103. https://doi.org/10.1021/ja01280a022

(112) Hansch, C.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91*, 165–195. https://doi.org/10.1021/cr00002a004

(113) Yau, H. M.; Haines, R. S.; Harper, J. B. A Robust, "One-Pot" Method for Acquiring Kinetic Data for Hammett Plots Used To Demonstrate Transmission of Substituent Effects in Reactions of Aromatic Ethyl Esters. *J. Chem. Educ.* **2015**, *92*, 538–542. https://doi.org/10.1021/ed400701m

(114) Bain, R. M.; Pulliam, C. J.; Yan, X.; Moore, K. F.; Müller, T.; Cooks, R. G. Mass Spectrometry in Organic Synthesis: Claisen–Schmidt Base-Catalyzed Condensation and Hammett Correlation of Substituent Effects. *J. Chem. Educ.* **2014**, *91*, 1985–1989. https://doi.org/10.1021/ed500288m

(115) Maes, B. U. W.; Verbeeck, S.; Verhelst, T.; Ekomié, A.; von Wolff, N.; Lefèvre, G.; Mitchell, E. A.; Jutand, A. Oxidative Addition of Haloheteroarenes to Palladium(0): Concerted versus S_NAr-Type Mechanism. *Chem. – Eur. J.* **2015**, *21*, 7858–7865. https://doi.org/10.1002/chem.201406210

(116) Gao, X.; Kagan, H. B. One-Pot Multi-Substrate Screening in Asymmetric Catalysis. *Chirality* **1998**, *10*, 120–124. https://doi.org/10.1002/chir.19

(117) Satyanarayana, T.; Kagan, H. B. The Multi-Substrate Screening of Asymmetric Catalysts. *Adv. Synth. Catal.* **2005**, *347*, 737–748. https://doi.org/10.1002/adsc.200505057

(118) Heiden, M. an der; Plenio, H. The Effect of Steric Bulk in Sonogashira Coupling Reactions. *Chem. Commun.* **2007**, 972–974. https://doi.org/10.1039/B616608C

(119) an der Heiden, M. R.; Plenio, H.; Immel, S.; Burello, E.; Rothenberg, G.; Hoefsloot, H. C. J. Insights into Sonogashira Cross-Coupling by High-Throughput Kinetics and Descriptor Modeling. *Chem. – Eur. J.* **2008**, *14*, 2857–2866. https://doi.org/10.1002/chem.200701418

(120) Lu, T.; Chen, F. Quantitative Analysis of Molecular Surface Based on Improved Marching Tetrahedra Algorithm. *J. Mol. Graph. Model.* **2012**, *38*, 314–323. https://doi.org/10.1016/j.jmgm.2012.07.004

(121) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33*, 580–592. https://doi.org/10.1002/jcc.22885

(122) Suresh, C. H.; Remya, G. S.; Anjalikrishna, P. K. Molecular electrostatic potential analysis: A powerful tool to interpret and predict chemical reactivity. *WIREs Comp. Mol. Sci.* **2022**, *12*, e1601. https://doi.org/10.1002/wcms.1601

(123) Suresh, C. H.; Koga, N. Quantifying the Electronic Effect of Substituted Phosphine Ligands via Molecular Electrostatic Potential. *Inorg. Chem.* **2002**, *41*, 1573–1578. https://doi.org/10.1021/ic0109400

(124) Anjali, B. A.; Suresh, C. H. Interpreting Oxidative Addition of Ph–X (X = CH_3, F, Cl, and Br) to Monoligated Pd(0) Catalysts Using Molecular Electrostatic Potential. *ACS Omega* **2017**, *2*, 4196–4206. https://doi.org/10.1021/acsomega.7b00745

(125) Hirsch, J. A. Table of Conformational Energies—1967. In *Topics in Stereochemistry*; John Wiley & Sons, Ltd, 1967; pp 199–222. https://doi.org/10.1002/9780470147108.ch4

(126) Klein, J.; Khartabil, H.; Boisson, J.-C.; Contreras-García, J.; Piquemal, J.-P.; Hénon, E. New Way for Probing Bond Strength. *J. Phys. Chem. A* **2020**, *124*, 1850–1860. https://doi.org/10.1021/acs.jpca.9b09845

(127) Lu, J.; Leitch, D. C. Unpublished Results.

(128) Almond-Thynne, J.; C. Blakemore, D.; C. Pryde, D.; C. Spivey, A. Site-Selective Suzuki–Miyaura Coupling of Heteroaryl Halides – Understanding the Trends for Pharmaceutically Important Classes. *Chem. Sci.* **2017**, *8*, 40–62. https://doi.org/10.1039/C6SC02118B

(129) Reeves, E. K.; Bauman, O. R.; Mitchem, G. B.; Neufeldt, S. R. Solvent Effects on the Selectivity of Palladium-Catalyzed Suzuki-Miyaura Couplings. *Isr. J. Chem.* **2020**, *60*, 406–409. https://doi.org/10.1002/ijch.201900082

(130) Reeves, E. K.; Entz, E. D.; Neufeldt, S. R. Chemodivergence between Electrophiles in Cross-Coupling Reactions. *Chem. – Eur. J.* **2021**, *27*, 6161–6177. https://doi.org/10.1002/chem.202004437

(131) Elias, E. K.; Rehbein, S. M.; Neufeldt, S. R. Solvent Coordination to Palladium Can Invert the Selectivity of Oxidative Addition. *Chem. Sci.* **2022**, *13*, 1618–1628. https://doi.org/10.1039/D1SC05862B

(132) Norman, J. P.; Larson, N. G.; Entz, E. D.; Neufeldt, S. R. Unconventional Site Selectivity in Palladium-Catalyzed Cross-Couplings of Dichloroheteroarenes under Ligand-Controlled and Ligand-Free Systems. *J. Org. Chem.* **2022**, *87*, 7414–7421. https://doi.org/10.1021/acs.joc.2c00665

(133) Ibsen, G. M.; Menezes da Silva, V. H.; Pettigrew, J. C.; Neufeldt, S. R. Triflate-Selective Suzuki Cross-Coupling of Chloro- and Bromoaryl Triflates Under Ligand-Free Conditions. *Chem. – Asian J.* **2023**, *18*, e202300036. https://doi.org/10.1002/asia.202300036

(134) Tolman, C. A. Steric effects of phosphorus ligands in organometallic chemistry and homogeneous catalysis. *Chem. Rev.* **1977**, *77*, 313–348. https://doi.org/10.1021/cr60307a002

(135) Dombrowski, A. W.; Gesmundo, N. J.; Aguirre, A. L.; Sarris, K. A.; Young, J. M.; Bogdan, A. R.; Martin, M. C.; Gedeon, S.; Wang, Y. Expanding the Medicinal Chemist Toolbox: Comparing Seven C(sp^2)–C(sp^3) Cross-Coupling Methods by Library Synthesis. *ACS Med. Chem. Lett.* **2020**, *11*, 597-604. https://doi.org/10.1021/acsmedchemlett.0c00093

Chapter 2 – A Broadly Applicable Quantitative Relative Reactivity Model for Nucleophilic Aromatic Substitution (S_NAr) Using Simple Descriptors

This chapter has been adapted from: Lu J.; Paci I.; Leitch D. C. A broadly applicable quantitative relative reactivity model for nucleophilic aromatic substitution (S_NAr) using simple descriptors. *Chem. Sci.*, **2022**, *13*, 12681-12695, DOI: 10.1039/d2sc04041g

Contributions: Experiments and computational analysis were performed by Jingru Lu.

2.1 Abstract

This chapter reports a multivariate linear regression model able to make accurate predictions for the relative rate and regioselectivity of nucleophilic aromatic substitution (S_NAr) reactions based on the electrophile structure. The model uses a diverse training/test set from experimentally determined relative S_NAr rates between benzyl alcohol and 74 unique electrophiles, including heterocycles with multiple substitution patterns. There is a robust linear relationship between the experimental S_NAr free energies of activation and three molecular descriptors that can be obtained computationally: the electron affinity (*EA*) of the electrophile; the average molecular electrostatic potential (*ESP*) at the carbon undergoing substitution; and the sum of average *ESP* values for the *ortho* and *para* atoms relative to the reactive center. Despite using only simple descriptors calculated from ground state wavefunctions, this model demonstrates excellent correlation with previously measured S_NAr reaction rates, and is able to accurately predict site selectivity for multihalogenated substrates: 91% prediction accuracy across 82 individual examples. The excellent agreement between predicted and experimental outcomes makes this easy-to-implement reactivity model a potentially powerful tool for synthetic planning.

2.2 Introduction

Making accurate quantitative predictions about reaction outcomes is highly desirable yet remains challenging in organic synthesis.[1] Qualitative reactivity predictions are relatively easy to achieve, since organic chemists are trained to predict chemical reactivity based on their knowledge of physicochemical properties of reaction partners and established reaction mechanisms, and to use these predictions in effective synthetic design.[2,3] While quantitative predictions of reaction outcomes are much more demanding tasks, which rely on sufficient high quality reaction data, extensive theoretical calculations and well-validated structure-reactivity relationships can be established using statistical methods.[4–9] Chapter 1 summarized the latest developments in quantitative reaction predictions, including applications of high-throughput experimentation, computational tools to generate molecular descriptors and machine learning to train robust QSRR and QSSR predictive models.[10–22] It also outlined the outstanding challenges, including how to collect high-quality reaction data for a large enough set of chemical space, develop new structural descriptors for complex molecular systems, and achieve general-purpose and multiple reaction objectives prediction models.[23–27]

Nucleophilic aromatic substitution (S_NAr) is one of the most important and well-studied reactions in organic synthesis.[28–30] S_NAr is a reaction type in which a nucleophile attacks an electrophilic centre, resulting in substitution of the leaving group by the nucleophile. The electrophile is usually an electron deficient (hetero)arene having one or more leaving groups. The most common leaving groups in S_NAr are halides, which follow the established reactivity trend: F > Cl ≈ Br > I.[31] Nitro,[32,33] alkoxide, and phenolate[34,35] are common non-halide leaving groups in S_NAr. Other functional groups such as cyanide[36] and five-membered nitrogen heterocycles[37,38] have also been reported as leaving groups in S_NAr reactions. S_NAr with hydride as leaving group,[39] also referred to as vicarious nucleophilic substitution in which an aromatic hydrogen is replaced by a carbon-based nucleophile, is a frequently used strategy for alkylation of (hetero)arenes.[40–42]

S_NAr is extensively used in total synthesis of natural products,[43–50] and manufacturing of active pharmaceutical and agrochemical ingredients.[51–55] For example, S_NAr is extremely powerful in synthesizing N-heterocycles, one of the most ubiquitous structural components in active pharmaceutical ingredients.[56,57] The chart in Figure 2.1 shows that S_NAr is ranked second among the most frequently occurring reactions in medicinal chemistry in 2014, only surpassed by amide bond formation.[58]

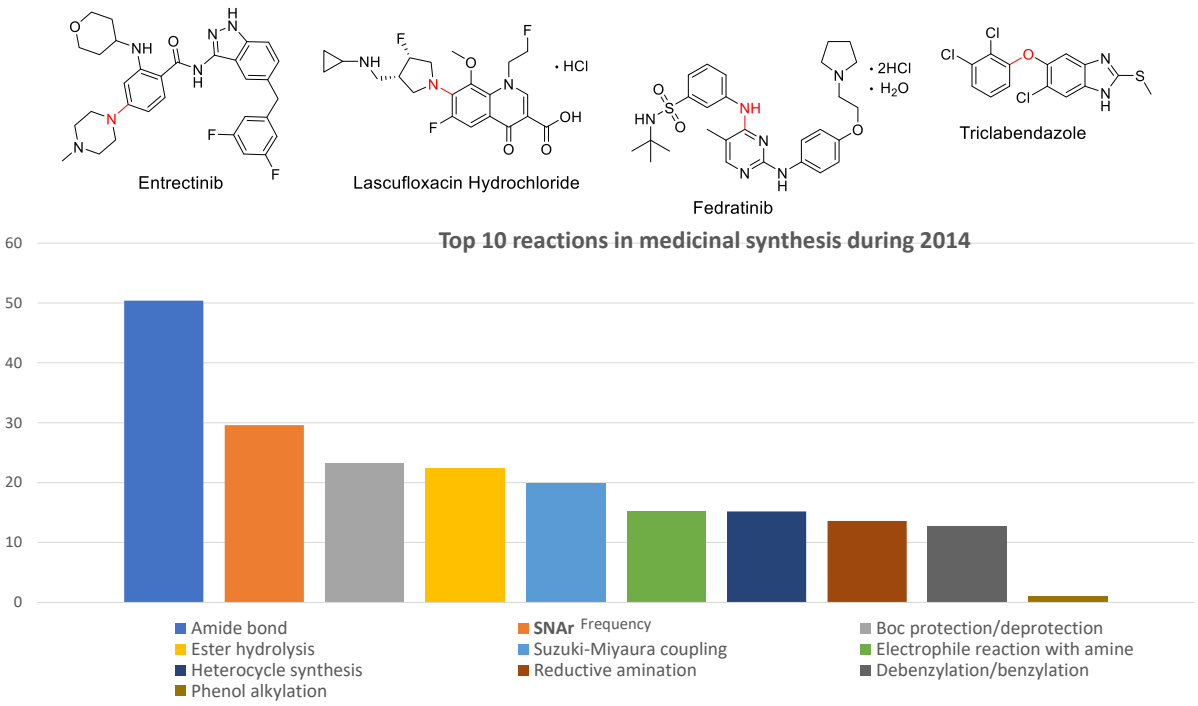


Figure 2.1. Examples of APIs prepared by multi-step synthesis that involves at least one S_NAr reaction (nucleophilic atom and bond formed by S_NAr highlighted in red). **Bottom**: Top 10 most used reactions in medicinal chemistry in 2014 ranked by frequency.[58]

Because of its importance in organic synthesis, accurate prediction of S_NAr outcomes would be invaluable in designing highly efficient and selective synthetic pathways for many important compounds. Over the past 100 years, a lot of effort has been put into developing quantitative relationships between substrate structures and S_NAr reactivity and selectivity via different approaches (Fig.2.2). For example, the linear free energy relationship described by the Hammett equation[59,60] along with the experimentally measured Hammett substituent constants are widely used in studying substituent effect in various organic reaction systems including S_NAr.[61–65]

The Mayr nucleophilicity database has been used to study S_NAr mechanisms[66–68] and develop quantitative relationships for S_NAr reactivity predictions.[22] These reactivity scales are established in relation to empirical molecular descriptors that are derived experimentally.

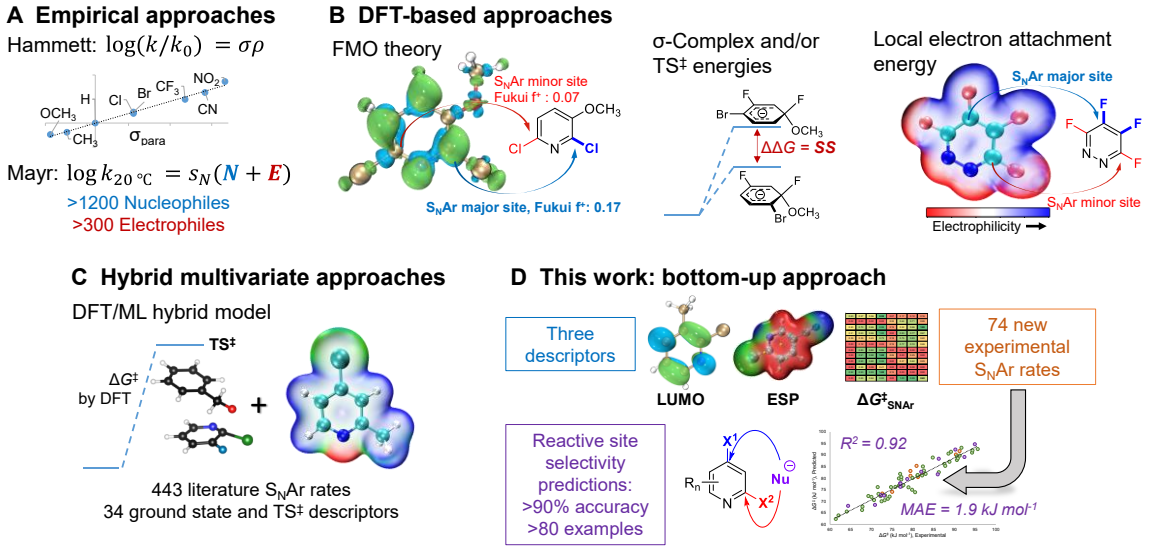


Figure 2.2. Approaches to developing quantitative structure-reactivity relationships (QSRR) for S_NAr reactions. **A)** Empirical parameters derived from experimental data. **B)** Calculated descriptors from DFT analysis (FMO = frontier molecular orbital theory; TS = transition state). **C)** Recent hybrid DFT/ML approach. **D)** Bottom-up approach combining new experimental data with simple calculated descriptors.

S_NAr predictions developed by theoretical and computational methods include calculated molecular descriptors and transition state theory (TST)-based analysis (Fig. 2.2B). For example, the Fukui indices[69] developed from Frontier Molecular Orbital (FMO) theory are widely used in predicting the nucleophilic and electrophilic sites in S_NAr reactions.[70–72] Baker and Muir[73,74] as well as Brinck, Svensson, and co-workers[75–77] have published several works on predicting regioselectivity for S_NAr reactions using DFT calculated transition state energies and/or stability of the Meisenheimer-complex intermediates. TST-based analysis by DFT is a powerful approach in exploring reaction mechanisms and providing theoretical evidence to support experimental findings. However, transition state calculations require quite high computational cost to perform the laborious potential

energy surface search in multiple directions. This is especially problematic for large molecules, where the computational cost grows exponentially with the increased molecular size. In addition, to ensure the reliability and accuracy of the calculated energy barriers, chemists need to invest additional time and effort into making correct mechanistic assumptions and evaluating different DFT methods to identify suitable approaches to account for all the essential interactions in a reaction system.

In 2016 Brinck and co-workers introduced the local electron attachment energy as a molecular descriptor for electrophilicity[78] and applied it in S_NAr reactivity and site-selectivity predictions.[79] The reaction datasets they used to evaluate the predictive ability of this descriptor are shown in Figs 2.10 to 2.12 as part of our external test sets, including reported S_NAr rates of 13 bromo-nitrobenzenes, and reported S_NAr rates and site-selectivity of 31 multi-fluorinated (hetero)arenes. While this local descriptor correlates well with experimental reactivity scales and predicts the regioselectivity in a qualitative manner, there is a need for new and more varied data and descriptors to develop broadly applicable models; for example, models that are able to predict the quantitative reactivity scale among substrates with different ring structures and halogens, and those able to predict site-selectivity for a broader scope of multihalogenated systems, including electron-rich electrophiles and substrates carrying mixed halides.

Recently, there are several publications reporting S_NAr predictive models that combine machine learning and TST-based analysis. In 2021 Buttar and coworkers reported the use of a hybrid DFT/machine learning approach to predict experimental S_NAr activation energies.[21] They collated more than 440 S_NAr rates from literature reports, and calculated 34 ground state and transition state descriptors by DFT for model training and performance evaluation. Notably, the transition state energies obtained from DFT are a crucial descriptor in the best-performing model. In 2022, McWilliams and coworkers[80] published a two-stage strategy to predict S_NAr regioselectivity of electrophiles. The first stage of their workflow is to predict the regioselectivity of an electrophile by machine learning, and to classify the results based on the predicted reactivity difference between the major and minor sites using

a confidence cutoff determined by statistical analysis. For electrophiles above the cutoff, the machine learning predictions are accepted, while for electrophiles below the cutoff with similar predicted reactivities between the major and minor sites, TST analysis by DFT is triggered to determine the site-selectivity. They trained and tested their workflow on a Pfizer internal dataset and a USPTO-derived public dataset, and achieved >90% accuracy in predicting the major sites in both datasets. The two studies demonstrate the power of combining machine learning and TST analysis in generating broadly applicable and accurate S_NAr models; however, performing the data-dependent machine learning model training and TST analysis is time and resource demanding, and requires expertise that synthetic chemists are not always trained to possess. Furthermore, both approaches rely on large, existing, high-quality datasets.

In our group's approach, the following three aspects of a predictive model have equal importance: 1) the prediction accuracy the model provides, especially for out-of-sample predictions; 2) the breadth of the substrate structures the model is applicable for; and 3) the ease and simplicity of applying the model to new systems. In previous work, the reaction datasets used to train and validate the QSSR/QSRR models are taken from literature values, which are known to have reporting and selecting biases as discussed in Chapter 1. To generate this training set, relative S_NAr rates for 74 individual electrophiles were measured. These electrophiles include many nitrogen heterocycles relevant to pharmaceutical synthesis, and electron rich heteroaryl chlorides and bromides with relatively limited examples from existing literature. Relative rates were obtained using competition experimentation, which is commonly used to generate univariate Hammett plots,[15,81–85] and is a much more efficient approach compared to traditional reaction progress kinetic analysis in measuring large set of reaction rates. Having control over the composition of the training set enables including a varied and balanced distribution of structural features, which is important in achieving a generally applicable predictive model. To make the model easy to implement, and to reduce the computational cost required, only simple and easy-to-obtain mechanistically meaningful ground state molecular descriptors were used: the electrostatic potential (*ESP*) and electron

affinity (EA). From this combination of factors, a QSRR model for S_NAr reactions was created that has excellent performance in predicting reactivity trends and site selectivity for many different electrophiles, including for multiple external test sets with significantly different molecular structures (Fig. 2.2D).

2.3 Results and Discussion

2.3.1 Creating the Training/Test Set

An efficient approach to collect a large and diverse dataset of reaction rates is critical for this experimental design. To determine a large number of reaction rates in a timely manner, high-throughput competition experimentation was conducted as shown in Fig 2.3. This experimental approach can be summarized in three steps. First, I monitored the reaction progress of three touchstone reactions under *pseudo* first order conditions to determine absolute rate constants and free energies of activation ($\Delta G^{\ddagger}_{SNAR}$) for S_NAr between benzyl alkoxide and 2-chloropyridine, 2-chloro-6-methylpyridine or 2-bromo-5-methoxypyridine in DMSO at room temperature (Fig. 2.3A). Next, I determined the relative rate constants for the electrophile library by a series of 94 individual competition experiments under analogous conditions (Fig. 2.3B, Table S4), by having two electrophiles in excess but equal amounts to compete with the nucleophile (limiting reagent). UPLC chromatograms were recorded for the reaction solutions at two time points: the start of the reaction (t_0) and completion of the reaction (t_{end}). The relative rates of two S_NAr reactions were obtained from the relative concentrations of the two remaining electrophiles at t_{end}. This method of quantification avoids the need to obtain relative response factors between the entire arrays of new S_NAr products and the internal standards. Finally, we calibrated these relative rate constants using the touchstone reactions, giving absolute rate constants and the corresponding $\Delta G^{\ddagger}_{SNAR}$ values for the entire array of S_NAr reactions (Table A5, Appendix A). We used the absolute $\Delta G^{\ddagger}_{SNAR}$ value for the 2-chloropyridine touchstone reaction (88.8 kJ mol^{-1}) as the calibration point, with the other two touchstone reactions used to confirm the validity of the competition determined $\Delta G^{\ddagger}_{SNAR}$ values. We obtain a percent

difference between the competition values and touchstone values of 1.6% for 2-chloro-6-methylpyridine and 1.3% for 2-chloro-5-methoxypyridine (Fig. A3, Appendix A).

Using this competition experimentation approach, I have assembled a library of 74 S_NAr rates for (N-hetero)aryl halides (-F, -Cl and –Br leaving groups). The substrates in the dataset contain various structural features, with different substituents at multiple positions on the heterocyclic rings. The reactivity of these substrates crosses a broad range, with reaction rates spanning 6 orders of magnitude. A quantitative reactivity scale for a group of representative electrophiles is shown in Fig. 2.3C. As an initial check on the validity of these data, I assessed the general reactivity trends against the established knowledge of S_NAr reactivity. As expected, the electron-deficient structures react much faster than the electron-rich ones; furthermore, the reactivity of the halides leaving groups follows the established trend, with rates decreasing as Ar–F >> Ar–Cl ~ Ar–Br.[31] I also constructed Hammett plots for four sets of 2-X-pyridine substrates (X=Cl, Br), giving linear correlations with ρ values of ~4-5 (Figs. A4 to A7, Appendix A). Finally, I prepared and isolated five representative S_NAr products and confirmed their structures using NMR spectroscopy and high-resolution ESI mass spectrometry (Figs. A8-A17, Appendix A).

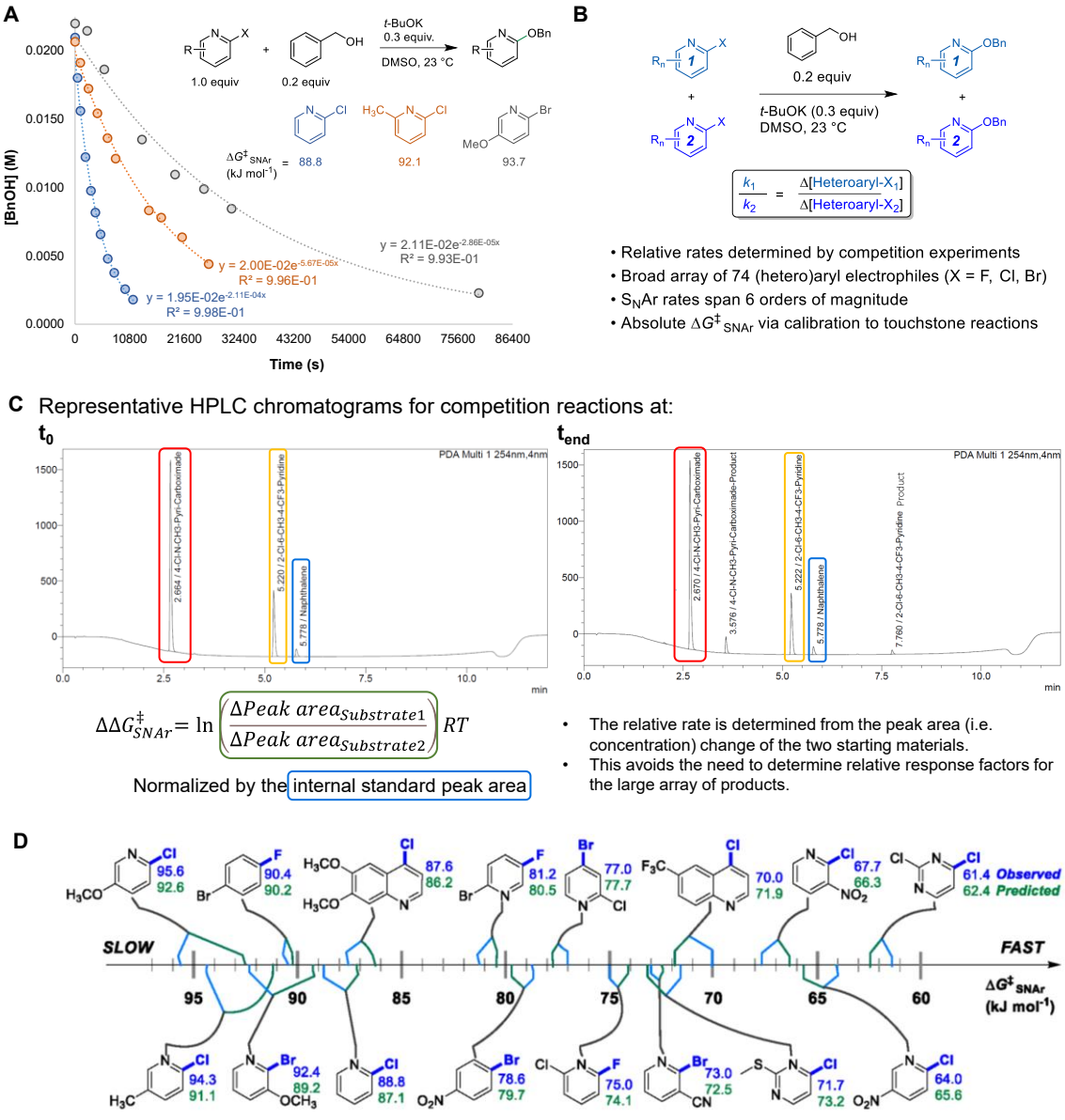


Figure 2.3. Experimental approach to collecting free energies of activation for 74 S_NAr reactions; Bn = benzyl. **A)** Touchstone reaction progress analysis under *pseudo* first order conditions. **B)** Competition experiments to establish relative rates across electrophile library. **C)** Representative primary data for determining $\Delta\Delta G^{\ddagger}_{SNAr}$ from competition experiments. **D)** Quantitative reactivity scale for representative electrophiles.

2.3.2 Model Generation and Performance.

Based on the well established S_NAr mechanism and our group's work[86] in applying ground state molecular descriptors[87] to reactivity prediction, I built a quantitative structure-reactivity model for S_NAr electrophiles using only three descriptors. These include a global descriptor in the electron affinity (*EA*) of the electrophile, and two local descriptors based on average molecular electrostatic potentials (*ESP*).[88–91] In addition to the *ESP* at the carbon undergoing substitution (ESP_1), I also discovered that the sum of *ESP* values for the *ortho* and *para* ring atoms is required for accurate predictions (ESP_2) (Fig 2.4A).

By correlating the three ground state descriptors to the experimentally obtained $\Delta G^{\ddagger}_{SNAR}$, I generated a unified structure-reactivity model able to give accurate predictions of S_NAr outcomes for substrates with various structural features and leaving groups (-Cl, -Br and -F). There is an excellent linear correlation between the predicted and actual $\Delta G^{\ddagger}_{SNAr}$ values ($R^2 = 0.92$) and a mean absolute error (MAE) of only 1.8 kJ mol^{-1} (0.43 kcal mol^{-1}) (Fig. 2.4B). Linear regression using normalized parameters (min/max normalization) gives percentage contributions for the individual descriptors, with 15.3% from *EA*, 50.2% from ESP_1 and 34.5% from ESP_2. Note that steric-based descriptors are not necessary to obtain good correlations for this data set.

The robustness of the model was assessed using cross-validation with five different random 60/40 training/test set data splits (Fig. 2.4C and Figs. A21-A24, Appendix A) and one structured split (Fig. A25, Appendix A) of 2-chloro-pyridine derivatives for training and all other substrates for testing. All of these regression analyses give essentially identical results, with excellent correlation statistics as indicated by the range of Q^2 values[92] from 0.86 to 0.93, and MAE values from 1.6 to 2.3 kJ mol^{-1} for the test sets. Finally, the model performance was assessed by analysing the distribution of residuals across the data set and identifying any possible outliers. As shown in Fig. 2.4D, the residuals are randomly distributed, almost exclusively in the range -5 to +5 kJ mol^{-1} (i.e. within an order of magnitude of the experimental rate). A box plot reveals only one significant outlier (|residual| > 5 kJ

mol^{-1}): 2-(*N*-methylcarboxamide)-4-chloropyridine. The S_NAr rate of this substrate is 6.5 kJ mol^{-1} slower than the predicted value; this may be because the acidic amide proton lowers the effective concentration of the base (potassium *tert*-butoxide) in solution by consuming it via a competitive acid-base reaction.

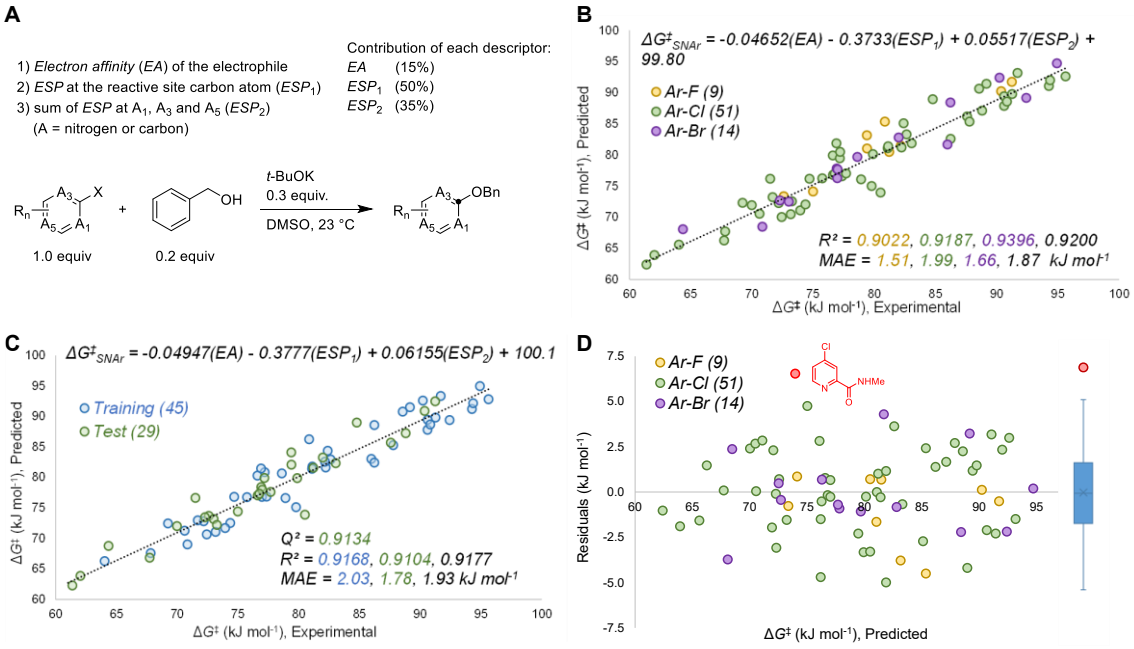


Figure 2.4. Quantitative model generation and performance. **A)** Molecular descriptors used in multivariate regression analysis, with percent contribution determined by min/max normalization. **B)** All data linear regression analysis for experimental versus predicted $\Delta G^{\ddagger}_{SNAr}$ with accompanying statistics (MAE = mean absolute error); linear correlation uses non-normalized descriptors. **C)** One of five 60/40 training/test validations, with accompanying statistics. **D)** Predicted versus residuals plot for the 74 data points, with accompanying box plot (right); one outlier is identified (|R| > 5 kJ mol^{-1} red point with accompanying structure).

The selection of the molecular descriptors was guided by the mechanistic features of nucleophilic aromatic substitution, contemporaneous work on a multivariate modeling for Pd(0) oxidative addition with (hetero)aryl halides (Chapter 3),[86] as well as iterative refinement of the included descriptors based on experimental observations and model performance. The following discussion provides more detail on creation and refinement of the model and its mechanistic basis.

FMO theory approximates nucleophile/electrophile reactivity using HOMO-LUMO interactions.[93,94] At a basic level, the HOMO of nucleophile and the LUMO of electrophile are the interacting frontier orbitals. As a result, lower LUMO energy leads to smaller HOMO-LUMO gap, resulting in improved overlap in the transition state and therefore a faster reaction. On the other hand, this simple connection between electrophilicity and LUMO energy is not necessarily valid for every system: in one recent example, Zipse, Ofial, and Mayr have demonstrated poor correlation between LUMO energy and electrophilicity for a series of Michael acceptors.[95] Nevertheless, LUMO energies were explored as a potential molecular descriptor for S_NAr reactivity. I obtained these values by calculating each substrate's electron affinity (*EA*), defined as the amount of energy released when an electron is added to the neutral system, and applying the DFT-Koopmann's theorem approximation that the LUMO energy is the negative of the *EA*.[96,97] The *EA* values were calculated using ORCA[98] (version 4.2.1) at B3LYP, def2-TVZPD level as the difference of electronic energies between the N electron and N+1 electron molecular systems. I also investigated an operationally simpler approach to calculating LUMO descriptors using Entos Envision,[99] an open online interactive platform for molecular simulation and visualization that performs rapid semi-empirical calculations using GFN1-xTB.[100] Comparing the LUMO energies computed in Envision with those from *EA* calculations at the B3LYP level revealed a strong linear correlation (R^2 = 0.88, Fig. A26, Appendix A) and a nearly identical linear regression model (Fig. A27, Appendix A) to that shown in Fig. 2.4B. While I retained the DFT-derived *EA* energies for our subsequent validation and external predictions, the values obtained from Envision could certainly be a rapid and easy to implement alternative.

Average molecular *ESP* at individual aromatic ring atoms were used as local descriptors.[88–91] The degree of electron deficiency of the aromatic ring is a key factor in determining S_NAr rates, and the *ESP* at the reactive carbon is a quantitative descriptor of this molecular feature. All electrostatic potential calculations were performed using the freely available Multiwfn program (version 3.7).[101,102]

I initially constructed a linear bivariate model using just two descriptors: *EA* and ESP_1 (at the carbon undergoing substitution) (Fig. 2.5A). The model gives good predictions for halogenated pyridines and quinolines; however, it significantly underestimates the reactivity of halogenated pyrimidines, and overestimates the reactivity of the non-heterocyclic haloarenes. The nature of these outliers led us to consider the general electronic features of the S_NAr transition state. Nucleophilic attack generates a negatively charged intermediate or $TS^{‡}$ in S_NAr, and electronegative atoms (such as nitrogen) at the *ortho* and *para* positions relative to the reactive center would stabilize the delocalized negative charge. As a result, one would expect lower activation energies and therefore faster reactions. Thus, we included the ESP_2 descriptor (sum of the ESP values at the *ortho* and *para* atoms to the reactive center) to account for these additional electronic effects, giving the superior model shown previously in Fig. 2.4B.

To highlight the importance of ESP_2 in making accurate predictions for multiple electrophile classes, I examined the two largest outliers from the bivariate (*EA* and ESP_1) model on either side of the distribution. Pairing these two outliers with halopyridines that have very similar EPS_1 values, but significantly different observed $\Delta G^{‡}_{SNAr}$ (Fig. 2.5B), reveals the importance of ESP_2. In the first case, the faster than predicted outlier 4-chloro-6-morpholinopyridine and 4-chloro-2-methylpyridine have similar *EA* values and nearly identical ESP_1 values; however, the observed S_NAr rate of the pyrimidine is ~100 times faster than the pyridine. The pyrimidine has a much more negative ESP_2 value due to the additional nitrogen at the *ortho* position and a more electron deficient ring structure as indicated by the *ESP* maps. A similar situation is present in the second case, for the slower than predicted outlier 1-bromo-3,5-bis(trifluoro-methyl)benzene and 2-chloro-5-(trifluoromethyl)pyridine: both substrates have nearly identical *EA* and ESP_1 descriptor values, but a more than 120 kJ mol^{-1} difference in ESP_2.

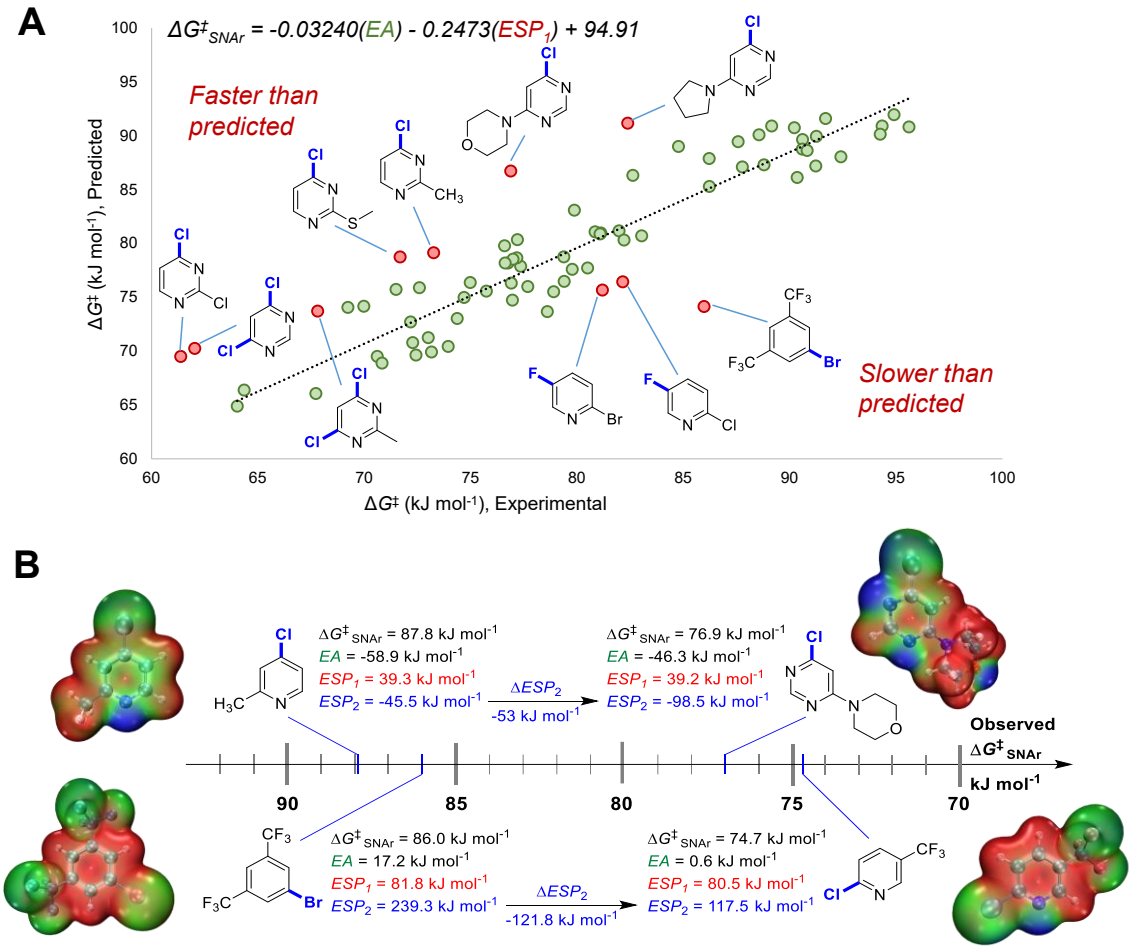


Figure 2.5. Importance of ESP_2 descriptor in predicting $\Delta G^{\ddagger}_{SNAr}$ for multiple substrate classes. **A)** Bivariate model incorporating only EA and ESP_1 descriptors, with two sets of outliers highlighted. **B)** Comparison of substrate pairs with very similar EA and ESP_1 values but significantly different $\Delta G^{\ddagger}_{SNAr}$ values, revealing the importance of ESP_2 in differentiating reactivity. *ESP* maps for each substrate structure are shown, with colour gradient indicating local *ESP* (red = maximum positive; green = 0; blue = maximum negative).

2.3.3 Predicting Site Selectivity in Multihalogenated Heterocycles

Accurate prediction of site selectivity for multifunctional substrates is of great importance in designing highly selective synthesis. Many prior efforts have been put into developing qualitative and quantitative site selectivity predictions for S_NAr.[21,42,71,73–79,103–105] I tested the model from Fig 2.4B for quantitative prediction of site selectivity with several multihalogenated heteroarenes in the training set, even though this model is not explicitly trained for this purpose. Importantly, the major

contributors to the model (*ESP*$_1$ and *ESP*$_2$) are local descriptors, which is crucial to enabling predictions at different sites within a molecule.[106]

For the 13 multihalogenated electrophiles in the training set, I compared the observed site selectivity ($\Delta\Delta G^{\ddagger}_{SNAr}$) to that predicted by the model. I also calculated the $\Delta\Delta G^{\ddagger}_{SNAr}$ for five of the substrates from transition state analysis using DFT (Fig. 2.6). In every case, the three-descriptor model from Fig. 2.4B identifies the correct major reactive site, with reasonable quantitative accuracy that is comparable to that obtained via transition state analysis; however, the predicted $\Delta\Delta G^{\ddagger}_{SNAr}$ between the major and minor sites appears to be systematically low.

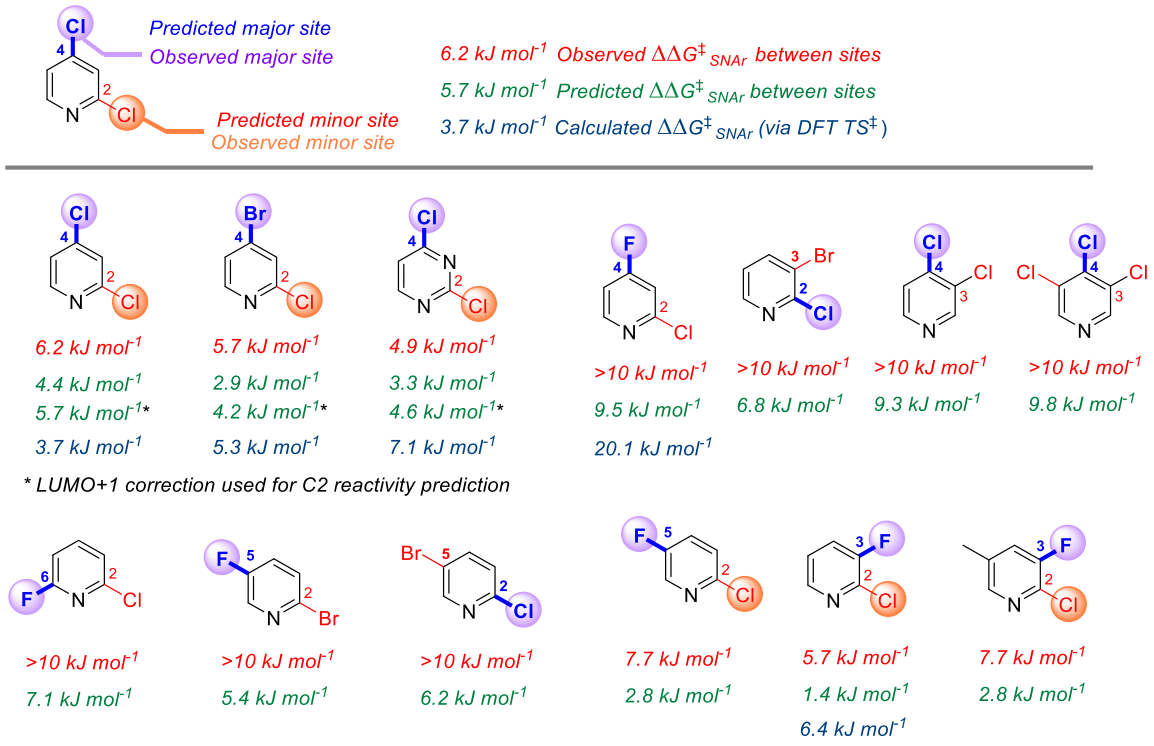


Figure 2.6. Site selectivity in multihalogenated heterocycles that are part of the training set. LUMO+1 energies are approximated by subtracting the LUMO/LUMO+1 energy gap from the *EA* value for the substrate.

To identify possible reasons for this systematic underestimation, I considered that the global *EA* descriptor may not be optimal in these cases. A transition state search on the first three substrates from Fig. 2.6 at B3LYP, def2-SVPD level in CPCM(DMSO) revealed a concerted S_NAr reaction coordinate, with transition state

structures located for each substrate. To assess the FMOs involved in these specific regioselective SNAr reactions, I examined the symmetries of the LUMO and LUMO+1 orbitals of the substrates, and visualized the HOMO orbitals of the calculated transition states. As shown for 2,4-dichloropyridines in Fig. 2.7B, the relevant antibonding orbital for nucleophilic attack at the C_4 major site is the LUMO, whereas at C_2 it is the LUMO+1; this is evident from the LUMO/LUMO+1 symmetries and the HOMO symmetries of the two transition states. By replacing the *EA* by the negative of LUMO energy for C_4 site and LUMO+1 energy for C_2 site along with the corresponding *ESP* values into the equation in Fig. 2.4B, there is improved accuracy in site selectivity prediction, with errors of 0.3-1.5 kJ mol^{-1} for $\Delta\Delta G^{\ddagger}_{SNAr}$.

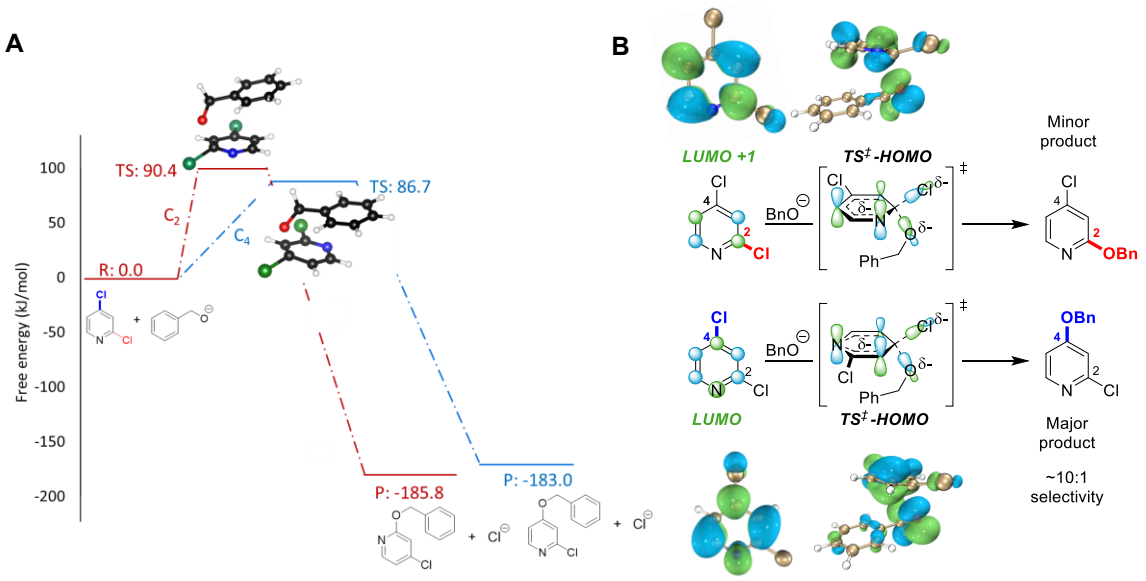


Figure 2.7. A) Calculated energy diagram for S_NAr of 2,4-dichloropyridine and the anion of benzyl alcohol in DMSO. The blue pathway is the S_NAr coordinate at C_4, the major reactive site; and the red pathway is the S_NAr coordinate at C_2, the minor reactive site. **B)** FMO analysis of S_NAr selectivity with 2,4-dichloropyridine, revealing orbital symmetry effects in the substrate (LUMO versus LUMO+1) and transition states (HOMO contributions from *ortho* and *para* sites).

For the 6 substrates with mixed halides (-F vs. -Cl/Br), the *ESP*/*EA* model correctly predicts that the Py-F is the major site in all cases; however, it consistently

underestimates the selectivity between the Py-F site and Py-Cl/Br site, especially for the three substrates where the fluorine is at an unactivated ring position (C_3 or C_5). Recent work on S_NAr mechanisms reveals that the concerted pathway is potentially more common in S_NAr reactions than the classic stepwise pathway via a Meisenheimer-type intermediate.[29,107,108] Jacobsen and coworkers[108] reported that S_NAr of (hetero)aryl-chlorides/bromides proceeds via a concerted mechanism. In contrast, S_NAr of fluorinated (hetero)arenes can be either stepwise or concerted depending on the electronic features of the electrophiles.

To examine if there is a change in reaction mechanism between sites, I performed TST analysis for two substrates (2-chloro-4-fluoropyridine and 2-chloro-3-fluoropyridine) at both the Py-F and Py-Cl site. Based on these calculations, there is not a stable Meisenheimer-type intermediate for S_NAr at the fluorine site either at the activated C_4 or the inactivated C_3 position. Instead, as the calculated energy diagrams for the two substrates show (Fig. 2.8), the Py-Cl and Py-F sites both proceed via a concerted pathway. Thus, a change in reaction mechanism is unlikely to be one of the potential reasons for the systemic underestimation of selectivity for those substrates. Instead, because the majority of the substrates in our training set are nitrogen-heterocycles with Cl or Br at the activated positions, examples of fluorinated electrophiles and substrates with leaving groups at inactivated positions are limited. This may lead to insufficient training data for this type of S_NAr electrophile, and therefore less accurate predictions.

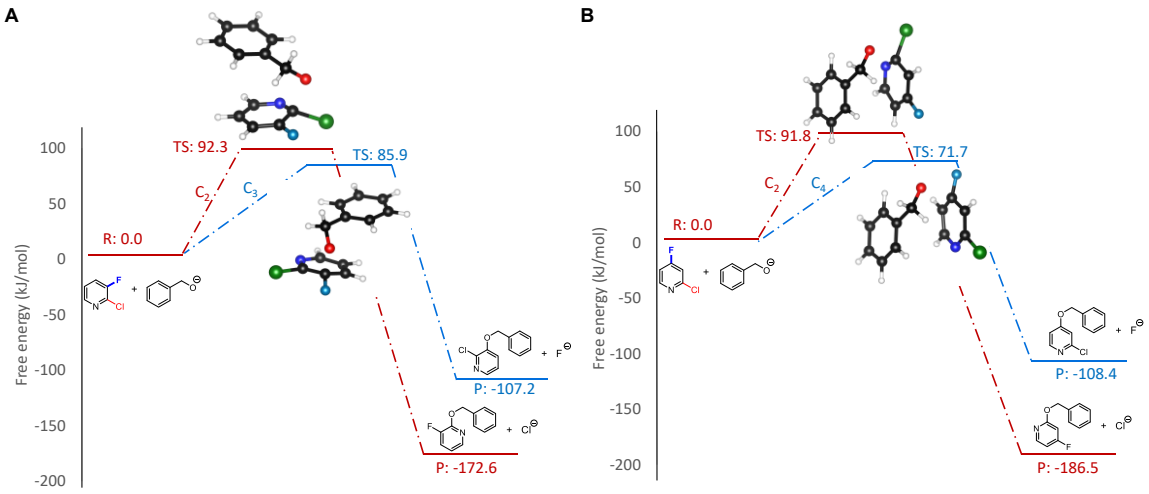


Figure 2.8. Calculated S_NAr energy diagrams of **A)** 2-chloro-3-fluoropyridine and **B)** 2-chloro-4-fluoropyridine at both reactive sites. The blue pathway is the S_NAr at the major site (Py-F); and the red pathway is the S_NAr at the minor site (Py-Cl). The concerted mechanism is found in all cases.

2.3.4 External Case Study #1: S_NAr Rate Correlations

With the three-descriptor model validated against internal data, I further challenged its performance and generality by applying it to various S_NAr reactions with different electrophiles conducted under different reaction conditions. I first examined several correlations between predicted $\Delta G^{\ddagger}_{SNAr}$ values from the model and three sets of experimental $\Delta G^{\ddagger}_{SNAr}$ values from the literature (Fig. 2.9).[64,109–111] All the electrophiles are not included in the model training/test sets. While the absolute $\Delta G^{\ddagger}_{SNAR}$ values predicted by the model are specific to the reaction conditions of the training set, there are good to excellent linear correlations between the predicted and experimental $\Delta G^{\ddagger}_{SNAr}$ values (R^2=0.72-0.99).

Group A contains 11 S_NAr rates of chloro-benzenes with an alkoxide in methanol (Fig. 2.9A). Here, I identified a potential experimental outlier the substrate **3A**, which was erroneously labelled as 4-OAc in subsequent papers rather than 4-Ac. Predicting the rate of the 4-OAc derivative led to a very large difference in predicted/experimental values, while the prediction for 4-Ac is very accurate. Overall, this substrate group exhibits excellent linear correlation ($R^2 = 0.98$) between experimental and predicted reactivity.

The 10 S_NAr rates in group B were collected from chloro(hetero)arenes and fluoro-nitrobenzenes reacting with piperidine in acetonitrile (Fig. 2.9B). The reaction rates predicted by the model give a linear correlation of $R^2 = 0.72$ to the experimental values, the weakest correlation among the 3 groups. This poorer correlation may be because the model is predominantly trained using substrates with Cl or Br leaving groups, whereas this set contains several substrates with F leaving groups. Another possible reason is a change in S_NAr mechanism. According to the aforementioned mechanistic study reported by Jacobsen and coworkers,[108] fluoro-nitrobenzenes undergo S_NAr via a stepwise mechanism, while chloro(hetero)arenes likely proceed via a concerted mechanism.

Group C contains 13 S_NAr rates for bromo-nitrobenzene substitution with piperidine acting as both nucleophile and solvent (Fig. 2.9C). Notably, the model enables accounting for solvation effects on electrophile reactivity during descriptor generation. Substrates **4C**, and **11C-13C** contains acidic or basic functional groups where the initial correlation between experimental and predicted reactivity is poor (red points). To account for potential hydrogen bonding interactions between those functional groups and piperidine, which would significantly alter the electronic features of the substrates, I introduced one explicit solvent molecule and recalculated the *ESP* values for these four electrophiles. Using these revised *ESP* values, there is excellent linear correlation across the entire substrate set. In addition to identifying the solvation effect, another experimental outlier was evident in this set (substrate **6C**). **6C** has two potentially reactive positions (Ar–Br and Ar–F). I have experimentally confirmed that reacting 6C with piperidine leads to a mixture of the two S_NAr products, in a 1.5:1 ratio, slightly favouring Ar–Br substitution (Fig. A40, Appendix A).

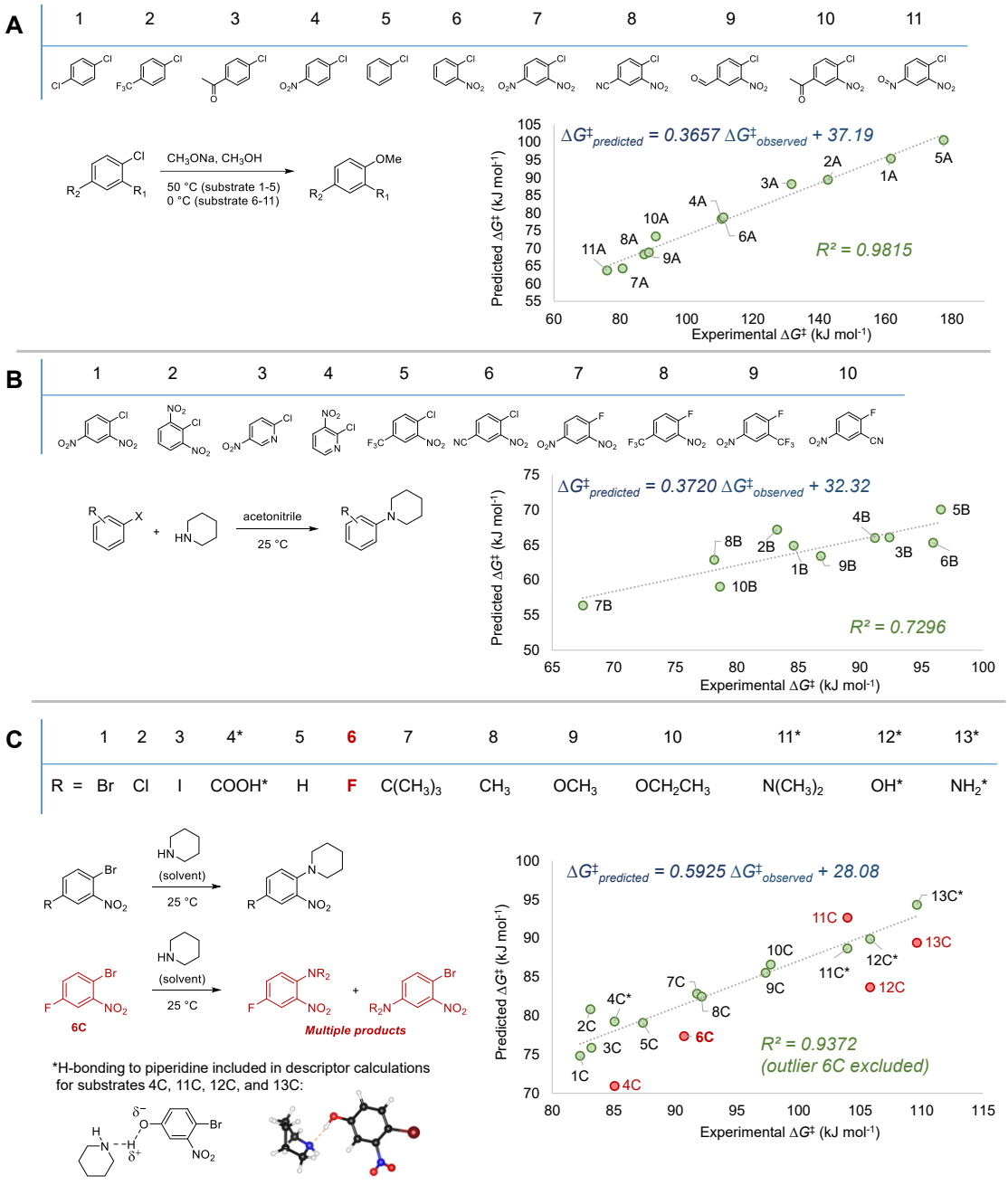


Figure 2.9. Model validation through assessing correlations between experimental $\Delta G^{\ddagger}$ values and predicted $\Delta G^{\ddagger}_{SNAr}$ for three external data sets (A, B, and C).

2.3.5 External Case Study #2: Site Selectivity Predictions

To further examine the potential applicability of the model for site selectivity predictions, I assessed 63 reported S_NAr site selectivity examples under a variety of conditions. First, I generated predictions for three data sets previously used as testing grounds for site selectivity predictions using other approaches (Figs. 2.10-2.12).[75–77,79] These data sets also contain experimentally-determined rates, providing an additional opportunity to test the model's performance.

The first set contains 7 multiply fluorinated arenes undergoing S_NAr with ammonia in dioxane/water at 80 °C, where five substrates have potential for regioisomer formation. In each case, the model correctly predicts the major reactive site, and the predicted $\Delta G^{\ddagger}_{SNAr}$ values correlate well with the experimental ln(k) values (R^2 = 0.95) for the five substrates. There are two slower than predicted outliers (**10b** and **10d**). Notably, Stenlid and Brinck also noted these examples as outliers when correlating the experimental ln(k) to their local electron attachment energy descriptor.[79] While these authors attributed this discrepancy between prediction and experiment to steric effects, there may be a different underlying reason considering the small size of both the nucleophile (ammonia) and the cyano group in **10d**.

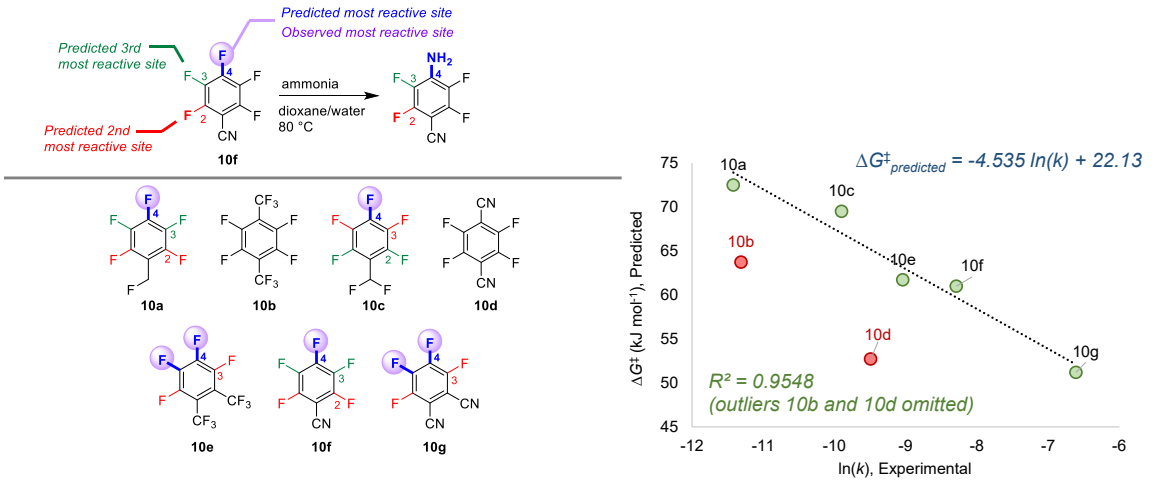


Figure 2.10. Site selectivity predictions and rate correlation for S_NAr between fluorinated arenes and ammonia. Experimental data from ref [112].

The second set involves multiple fluorinated arenes undergoing S_NAr with methoxide in methanol at 50 °C. Across the ten substrates, five have the potential to form regioisomers. In each case, the model correctly predicts the major site. For the two substrates with multiple experimental reactive sites, the predicted second most reactive site for **11d** is incorrect (C_2) based on experimental observation (C_3); however, for **11e** the predicted reactivity order from first to third is correct.

Even though there is again a quantitative underestimation of selectivity based on predicted $\Delta G^{\ddagger}_{SNAr}$ values, there is excellent linear correlation with experimental ln(k) across the entire substrate set. This is notable in the context of Stenlid and Brinck's prior work with local electron attachment energy, where the experimental ln(k) for **11g-j** does not correlate with that descriptor. Here, the *ESP*/*EA* model correctly predicts that these four substrates should have similar S_NAr rates (each within a factor of 10).

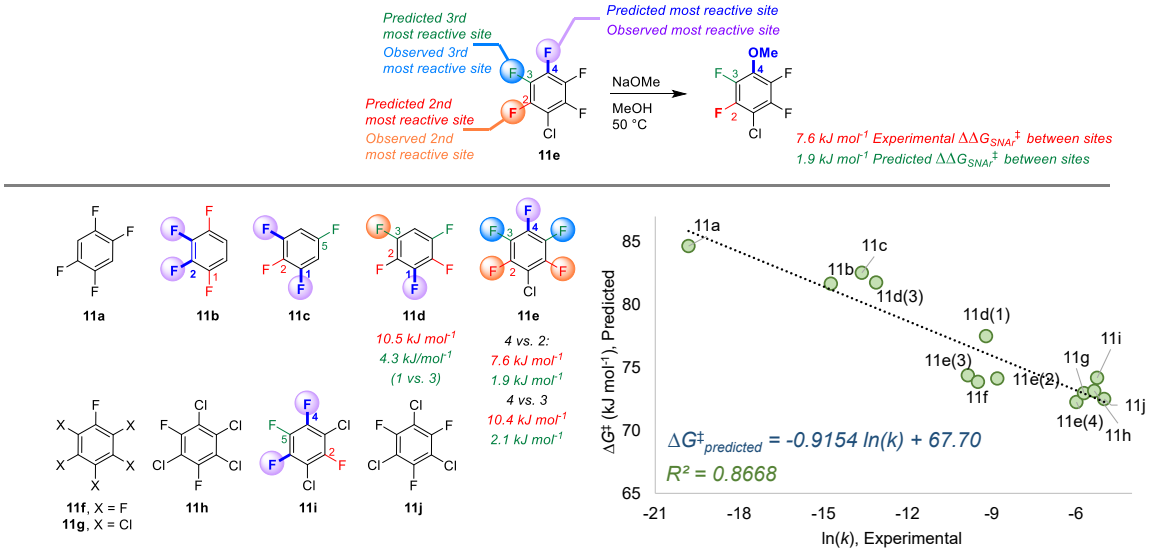


Figure 2.11. Site selectivity predictions and rate correlation for S_NAr between fluorinated arenes and methoxide. Experimental data from ref[113].

The third data set contains 18 multiply fluorinated nitrogen heterocycles undergoing S_NAr with with ammonia in dioxane/water at 80 °C, where 15 substrates have potential for regioisomer formation (Fig. 2.12).[112–114] Again, in every case the *ESP*/*EA* model correctly predicts the major reactive site, and it also correctly predicts the second most reactive site with only one exception (substrate **12l**). The quantitative selectivity predictions are also much closer to the experimental values

within this data set. There is also an excellent linear correlation between experimental ln(k) values and predicted $\Delta G^{\ddagger}_{SNAr}$ ($R^2 = 0.86$). Note that substrate **12r**, which has a rate "too fast … to measure",[114] is estimated to have a ~10^5-fold larger rate constant than **12d**; this estimated data point (blue) is not included in the linear correlation.

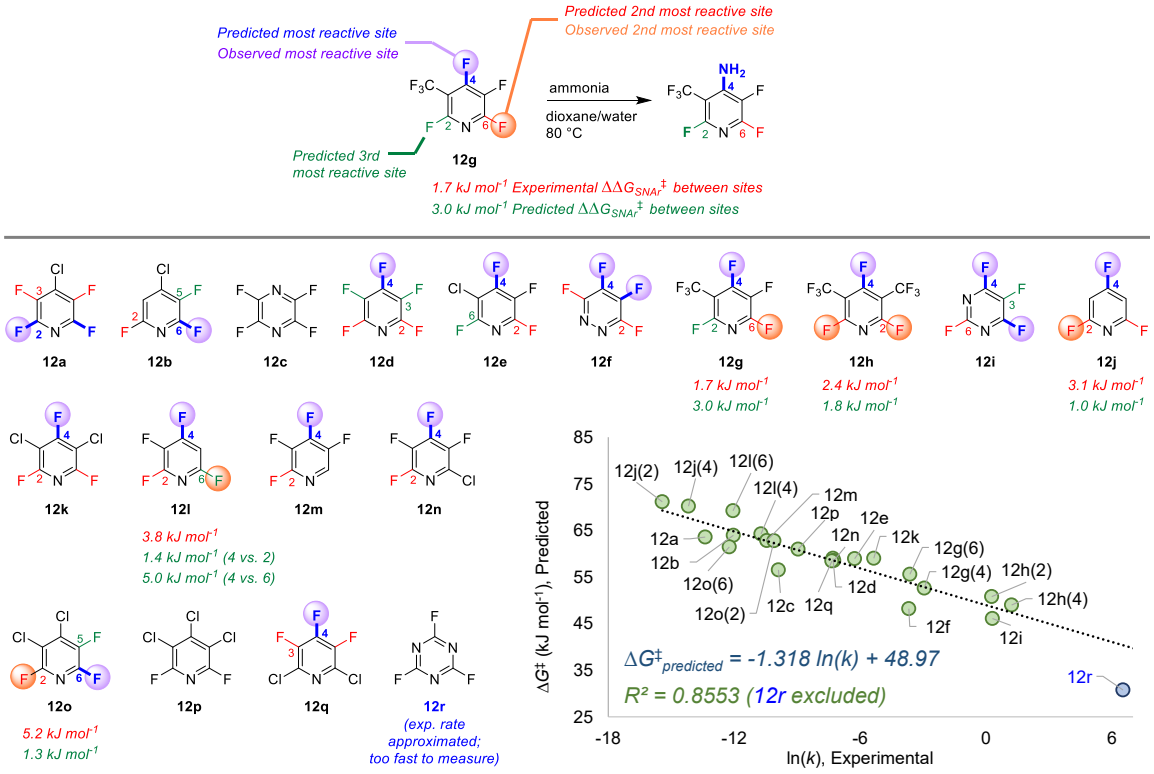


Figure 2.12. Site selectivity predictions and rate correlation for SNAr between fluorinated heterocycles and ammonia. Experimental data from refs [112–114].

Finally, to challenge the qualitative accuracy of the model, I applied it toward a series of more complex S_NAr examples with a wider variety of nucleophiles (Fig. 2.13). Sets A-D were previously collated and categorized by Brinck, Svensson, and co-workers and categorized depending on the nature of the nucleophile/electrophile pairing.[76,114–132] Using only the structure of the electrophile, the *ESP*/*EA* model is able to correctly predict the major site of reaction in 26 of the 32 cases. Within sets A and C – (hetero)aryl halides reacting with anionic nucleophiles – the two incorrect predictions are for relatively non-polar fluorinated arenes. For sets B and D, which employ neutral nucleophiles, the incorrect examples all involve secondary amine nucleophiles. In these cases, steric effects appear to play a significant role in

overriding the electronic nature of the electrophile; for example, pentachloropyridine reacts preferentially at C_4 (as predicted) with alkoxide or ammonia nucleophiles, but switches to C_2 selectivity with diethylamine. I also applied predictions to 6 mixed halide electrophiles reacting with a variety of nucleophiles in set E, drawn from examples in medicinal/agrochemical discovery.[133–138] The model is able to correctly identify the major site of reactivity for each example, except for a case where the predicted site is at an Ar–F, and the observed reactivity is at a 2-Cl-pyridine site.

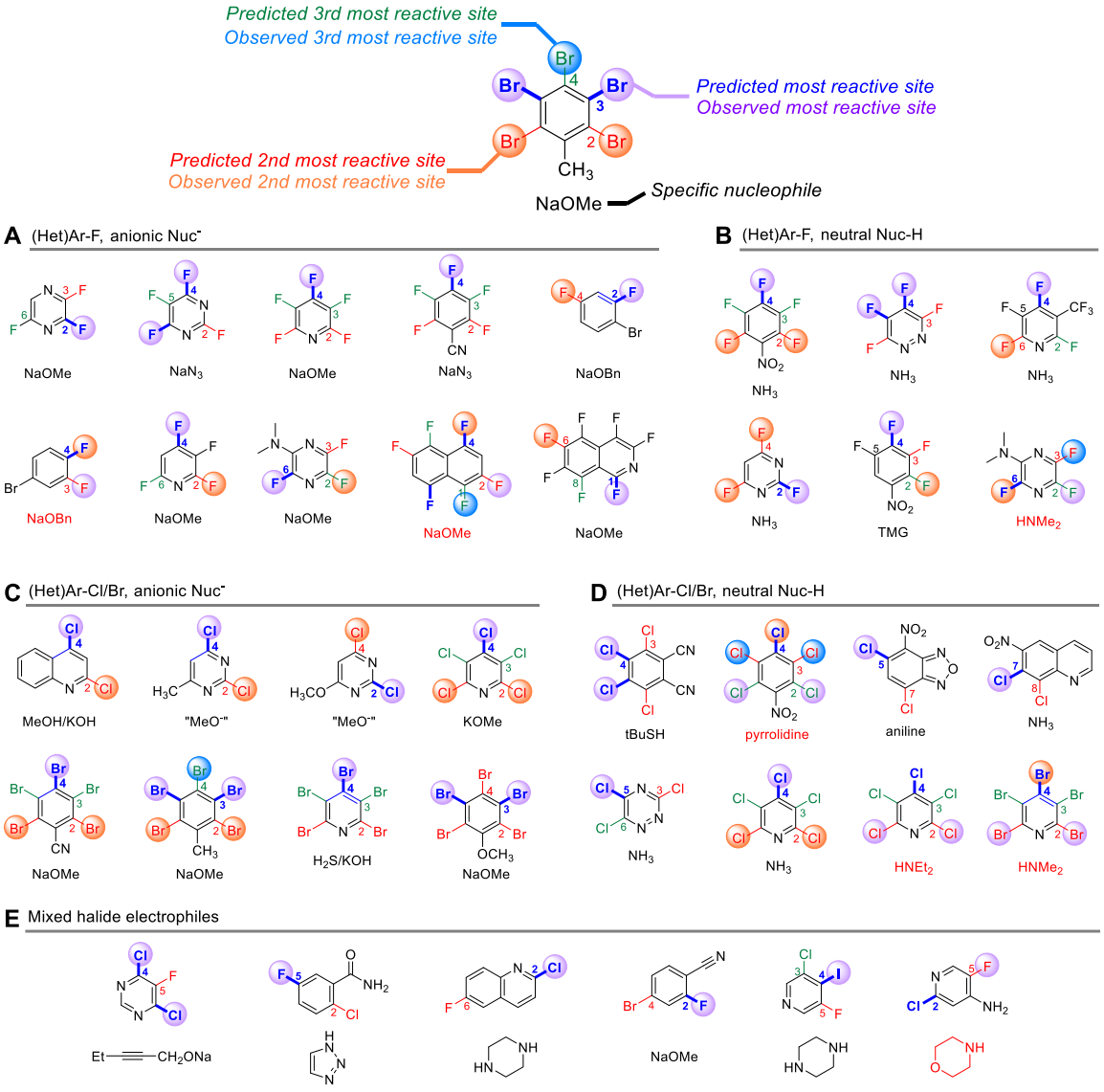


Figure 2.13. Qualitative site selectivity predictions for combinations of (hetero)aryl halides with anionic (A and C) and neutral (B and D) nucleophiles, and for mixed halide aromatics (E).

As described in Section 2.2, McWilliams and coworkers[80] recently published a two-stage workflow for S_NAr site-selectivity prediction: a machine learning (ML) model to give initial predictions followed by transition state calculations only on less confident ML-based predictions. Prediction confidence from the first step ML model is given by a Δscore value, which is the difference between the ML predicted scores of the major and minor S_NAr sites. An uncertainty threshold of 0.92 is determined by statistical analysis; when the Δscore is greater than 0.92, this outcome is considered as a high confident ML prediction without proceeding to the next step in the workflow. Otherwise, the second step DFT calculation will be triggered for ML prediction with a Δscore lower than 0.92. This workflow was tested on > 2000 regioselective S_NAr reactions from Pfizer and USPTO datasets, and excellent predictive accuracy of 96.3 and 94.7%, respectively was achieved.

Testing the *ESP*/*EA* model with the four representative examples given in the main text of the McWilliams work (**14a**–**14d**, Fig 2.14) reveals it can correctly predict the major S_NAr site in all four cases. The reported ML model gives a high confidence prediction (Δscore = 0.97) for site-selectivity of electrophile **14a**; for electrophiles **14b** (Δscore = 0.04) and **14c** (Δscore = 0.46), low confidence ML predictions triggered the second step DFT calculations, which gives high confidence predictions based on the transition state activation energies. Electrophile **14d** is a challenging case, with a low confidence ML prediction (Δscore = 0.15) followed by an incorrect predicted major site by DFT methods. Notably, the *ESP*/*EA* model correctly predicts the major site for S_NAr with **14d** (C_2). As well, the predicted selectivity between the two sites ($\Delta\Delta G^{\ddagger}_{SNAr}$ = 2.1 kJ mol^{-1}) matches quantitatively to the experimental data (C_2:C_3 = 3:1, $\Delta\Delta G^{\ddagger}_{SNAr}$ of 2.7 kJ mol^{-1} at room temp.).

Finally, I tested the *ESP*/*EA* model against ten randomly selected challenging cases from the available prediction outcomes provided in the paper. Five of these have an ML Δscore value from 0.2 to 0.5 (**14e** to **14i**) and another five have an ML Δscore lower than 0.2 (**14j** to **14n**). The *ESP*/*EA* model makes correct predictions for the S_NAr major site in all cases except for substrates **14l** to **14n**. It is worth noting that these three substrates also have high uncertainty using the reported two-stage

workflow: the ML model gives low confidence predictions (Δscore < 0.2), followed by incorrect major site predictions by DFT transition state calculations.

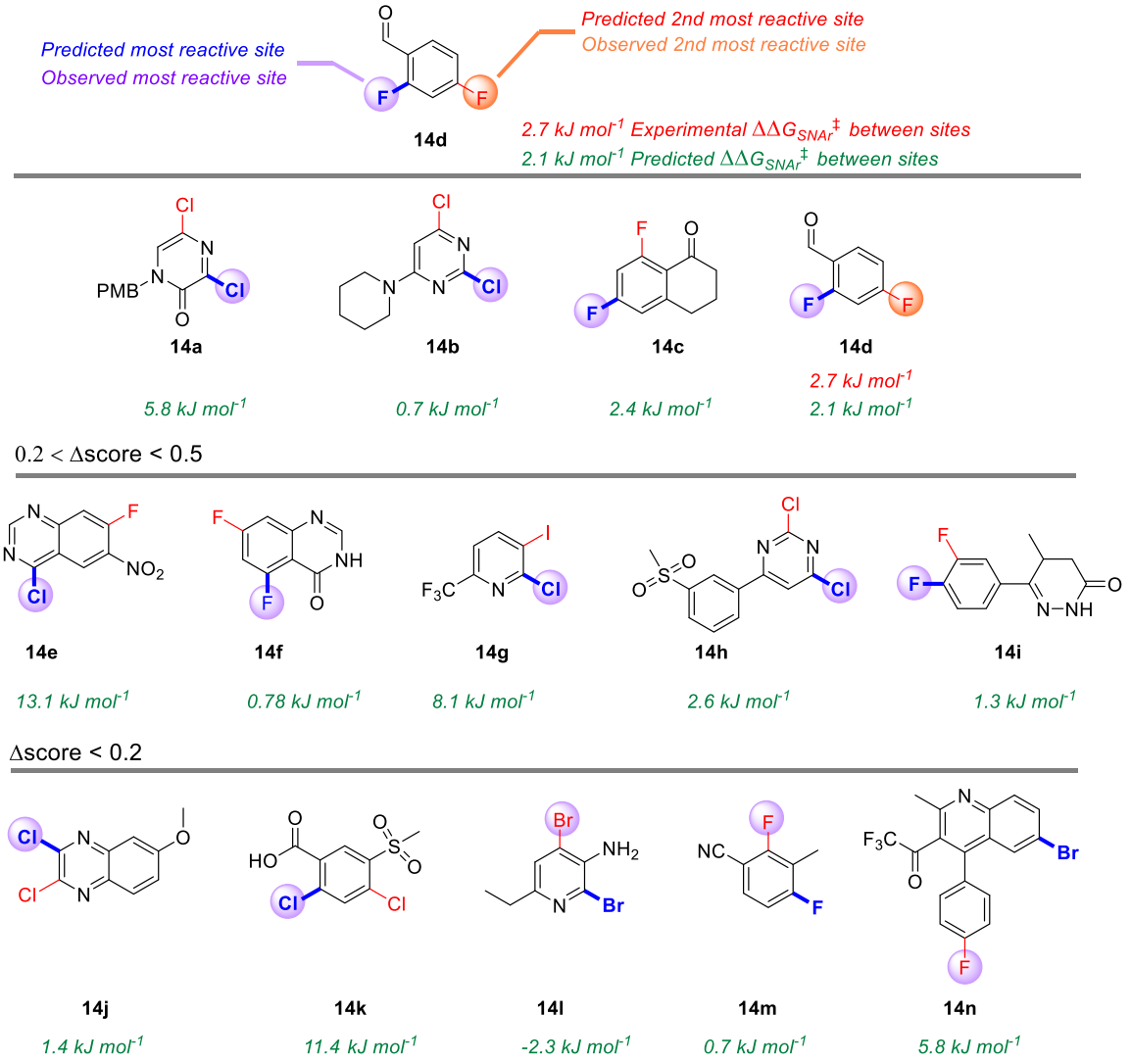


Figure 2.14. Site selectivity predictions using examples from reference [80]. The original experimental data was obtained from the public USPTO database.

2.3.6 External Case Study #3: Complex Molecule Synthetic Planning

As a final test of the potential applications of the *ESP*/*EA* model in real-world synthetic planning, I performed predictions for S_NAr reactions involved in preparing clinical candidate active pharmaceutical ingredients (APIs). These include recent reports on branebrutinib,[139] an EGFR T790 M inhibitor,[140] a Nav1.7 inhibitor,[141] a tyrosine kinase inhibitor,[142] an SRI/5-HT2A antagonist,[143] an RoRγ inverse agonist,[144] and merestinib[145] (Fig. 2.15).

The first four examples concern site selective S_NAr to generate a variety of targets from structurally complex substrates. In each case, the *ESP*/*EA* model predicts the correct reactive site in a reasonably quantitative manner based on the experimental data. For the electrophiles of the first three reactions, the small predicted $\Delta\Delta G^{\ddagger}_{SNAr}$ (2.7-4.3 kJ mol^{-1}) values are in good agreement with the reported moderate S_NAr selectivity. The predicted $\Delta\Delta G^{\ddagger}_{SNAr}$ of 3,5-dibromo-2-aminopyrazine, the electrophile of the last reaction, is 8.5 kJ mol^{-1}; this large value is consistent with the high selectivity reported for this reaction.

I also examined three examples where the substitution pattern of the S_NAr electrophile affects the site selectivity or reactivity. In the first case, synthesis of the target SRI/5-HT2A antagonist requires a site selective S_NAr to install an aryl ether *ortho* to a carbonyl functionality.[143] This was initially performed using an aldehyde moiety; however, the relatively poor site selectivity meant column chromatography was required to purify the intermediate. Further process developments identified an *N*-methylamide as a more selective alternative that retained key functionality for progressing to the target API. This improved selectivity is predicted by the *ESP*/*EA* model. A second case involves choice of either an Ar–F or Ar–Cl electrophile for S_NAr with an alkoxide nucleophile.[144] Experimental evaluation of each revealed that both substrates are viable, with the Ar–Cl version requiring slightly higher reaction temperature than the Ar–F analogue. The *ESP*/*EA* model predicts that the F for Cl switch would result in a relatively modest reactivity decrease, indicating both should be viable substrates.

The final example concerns an intramolecular S_NAr to generate an indazole *en route* to merestinib.[145] The final API contains a methoxy group para to the indazole

nitrogen; however, attempts to perform the intramolecular S_NAr with this strong electron donating group *para* to the substitution site were not successful. Instead, the researchers installed a nitro group to enable the S_NAr to proceed, but which would require multiple functional group interconversions. The predicted reactivity difference between -OMe and -NO_2 derivatives is 19.3 kJ mol^{-1} and this three-orders-of-magnitude difference in predicted rate means that the more desirable -OMe substrate could be ruled out earlier on in synthetic development. Furthermore, we examined the reactivity of a sulfonate derivative (-OSO_2Me), a potentially useful substrate that retains the required oxygen and requires fewer synthetic steps for functional group conversion. The substantial difference in predicted reactivity between -OSO_2Me and -OMe derivatives (11.8 kJ mol^{-1}) suggests that the sulfonate derivative could be a good candidate for this synthetic design.

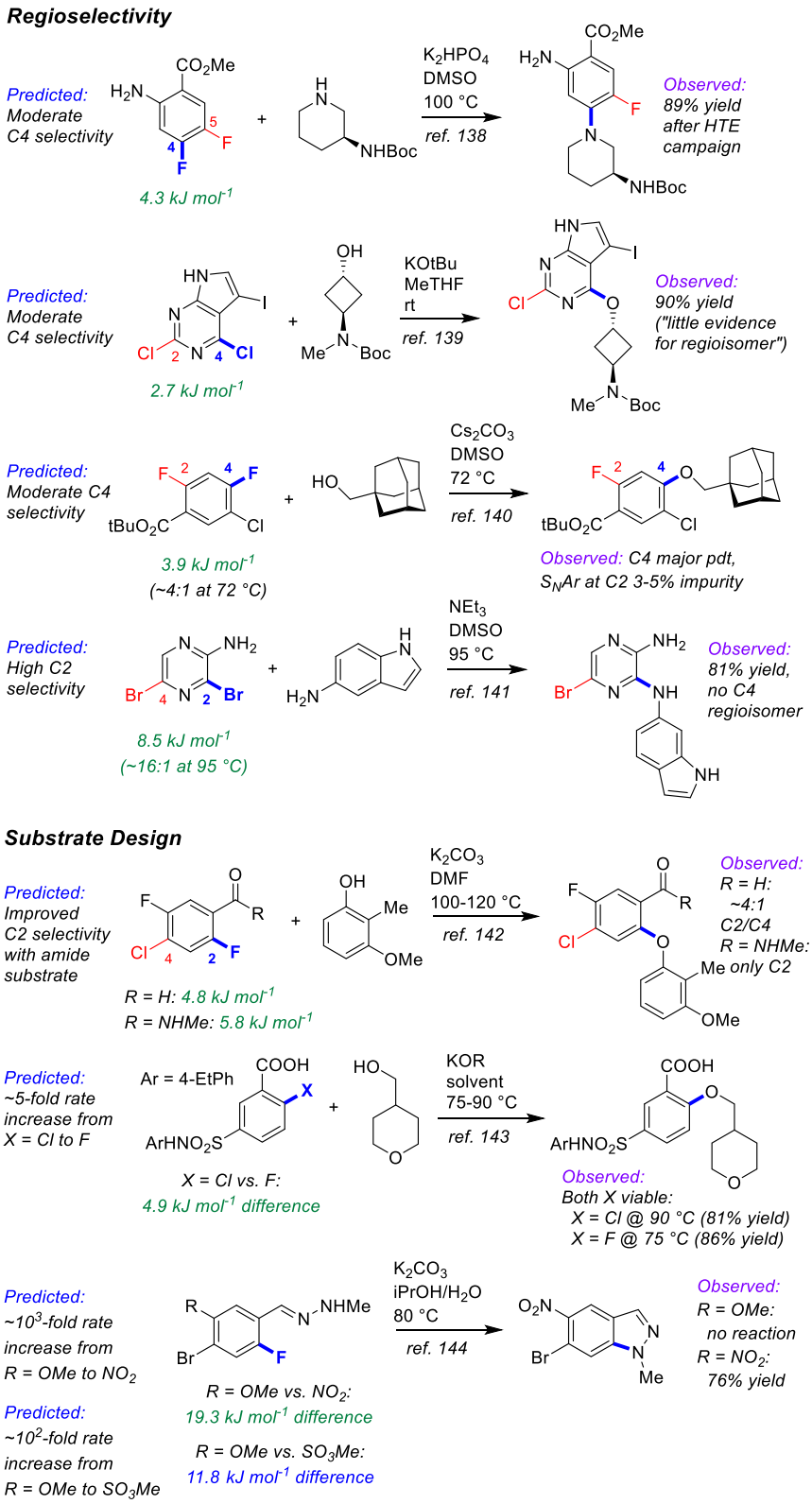


Figure 2.15. Example applications of S_NAr predictions to route development for investigational API synthesis, including regioselectivity for specific substrates, and comparison of potential substrate regioselectivity/reactivity.

2.4 Conclusions

This work in this Chapter demonstrates the power of a bottom-up approach to developing a quantitative structure-reactivity models for nucleophilic aromatic substitution reactions. Using competition experiments to rapidly collect S_NAr rates for a diverse library of electrophiles enables assembly of a reliable training set. By correlating this experimental reactivity data to simple and mechanistically meaningful ground state descriptors (ESP and EA), I developed a robust multivariate linear regression model for accurate predictions of both reactivity and site selectivity in S_NAr.

Importantly, even though the model was trained using only set of reaction conditions with a single nucleophile, it demonstrates excellent generality in predicting S_NAr reaction outcomes for a wide variety of conditions. This is evidenced by extensive external validation on a wide selection of examples. These include a >90% success rate in predicting the major site for multihalogenated (hetero)arenes (>80 cases), and examples where substrates design for API synthesis can be informed by predicted reactivity. Thus, this simple and easy-to-apply model can generate rapid and accurate predictions for complex molecule targets. There are still specific limitations to be addressed, including the inability to predict selectivity outcomes for non-halogenated leaving groups (e.g. $-NO_2$ or -OMe) and for bulky nucleophiles (with examples shown in Fig.2.13). Further work to refine the S_NAr model to account for these effects, and to build accurate predictive models for additional commonly used reaction classes, such as amide coupling, is currently underway.

2.5 References

(1) R. C. Larock Comprehensive Organic Transformations: *A Guide to Functional Group Preparations*, Wiley, 2018.

(2) Grossman, R. B. *The Art of Writing Reasonable Organic Reaction Mechanisms*, 2nd ed.; Springer: New York, 2003.

(3) E. J. Corey and X.-M. Cheng, *The Logic of Chemical Synthesis,* John Wiley & Sons, Nashville, TN, 1995.

(4) Mayr, H.; Ofial, A. R. Do General Nucleophilicity Scales Exist? *J. Phys. Org. Chem.* **2008**, *21* (7–8), 584–595. https://doi.org/10.1002/poc.1325.

(5) Mayr, H.; Ofial, A. R. Kinetics of Electrophile-Nucleophile Combinations: A General Approach to Polar Organic Reactivity. *Pure Appl. Chem.* **2005**, *77* (11), 1807–1821. https://doi.org/10.1351/pac200577111807.

(6) Mayr, H.; Ofial, A. R. Philicities, Fugalities, and Equilibrium Constants. *Acc. Chem. Res.* **2016**, *49* (5), 952–965. https://doi.org/10.1021/acs.accounts.6b00071.

(7) Mayr, H.; Ofial, A. R. Philicity, Fugality, and Equilibrium Constants: When Do Rate-Equilibrium Relationships Break Down? *Pure Appl. Chem*. **2017**, *89* (6), 729–744. https://doi.org/10.1515/pac-2017-0107.

(8) Mayr, H.; Patz, M. Scales of Nucleophilicity and Electrophilicity: A System for Ordering Polar Organic and Organometallic Reactions. *Angew. Chem. Int. Ed. Engl.* **1994**, *33* (9), 938–957. https://doi.org/10.1002/anie.199409381.

(9) Mayr, H.; Kempf, B.; Ofial, A. R. π-Nucleophilicity in Carbon−Carbon Bond-Forming Reactions. *Acc. Chem. Res.* **2003**, *36* (1), 66–77. https://doi.org/10.1021/ar020094c.

(10) Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A. The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res*. **2016**, *49* (6), 1292–1301. https://doi.org/10.1021/acs.accounts.6b00194.

(11) Niemeyer, Z. L.; Milo, A.; Hickey, D. P.; Sigman, M. S. Parameterization of Phosphine Ligands Reveals Mechanistic Pathways and Predicts Reaction Outcomes. *Nat. Chem.* **2016**, *8* (6), 610–617. https://doi.org/10.1038/nchem.2501.

(12) Wu, K.; Doyle, A. G. Parameterization of Phosphine Ligands Demonstrates Enhancement of Nickel Catalysis via Remote Steric Effects. *Nat. Chem.* **2017**, *9* (8), 779–784. https://doi.org/10.1038/nchem.2741.

(13) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360* (6385), 186–190. https://doi.org/10.1126/science.aar5169.

(14) Maryasin, B.; Marquetand, P.; Maulide, N. Machine Learning for Organic Synthesis: Are Robots Replacing Chemists? *Angew. Chem. Int. Ed*. **2018**, *57* (24), 6978–6980. https://doi.org/10.1002/anie.201803562.

(15) Engkvist, O.; Norrby, P.-O.; Selmi, N.; Lam, Y.; Peng, Z.; Sherer, E. C.; Amberg, W.; Erhard, T.; Smyth, L. A. Computational Prediction of Chemical Reactions: Current Status and Outlook. *Drug Discov. Today* **2018**, *23* (6), 1203–1218. https://doi.org/10.1016/j.drudis.2018.02.014.

(16) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363* (6424), eaau5631. https://doi.org/10.1126/science.aau5631.

(17) Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **2020**, *10* (3), 2260–2297. https://doi.org/10.1021/acscatal.9b04186.

(18) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. QSAR without Borders. *Chem. Soc. Rev.* **2020**, *49* (11), 3525–3564. https://doi.org/10.1039/D0CS00098A.

(19) Mahjour, B.; Shen, Y.; Cernak, T. Ultrahigh-Throughput Experimentation for Information-Rich Chemical Synthesis. *Acc. Chem. Res.* **2021**, *54* (10), 2337–2346. https://doi.org/10.1021/acs.accounts.1c00119.

(20) Gallegos, L. C.; Luchini, G.; St. John, P. C.; Kim, S.; Paton, R. S. Importance of Engineered and Learned Molecular Representations in Predicting Organic Reactivity, Selectivity, and Chemical Properties. *Acc. Chem. Res.* **2021**, *54* (4), 827–836. https://doi.org/10.1021/acs.accounts.0c00745.

(21) Jorner, K.; Brinck, T.; Norrby, P.-O.; Buttar, D. Machine Learning Meets Mechanistic Modelling for Accurate Prediction of Experimental Activation Energies. *Chem. Sci.* **2021**, *12* (3), 1163–1175. https://doi.org/10.1039/D0SC04896H.

(22) Orlandi, M.; Escudero-Casao, M.; Licini, G. Nucleophilicity Prediction via Multivariate Linear Regression Analysis. *J. Org. Chem.* **2021**, *86* (4), 3555–3564. https://doi.org/10.1021/acs.joc.0c02952.

(23) Shen, Y.; Borowski, J. E.; Hardy, M. A.; Sarpong, R.; Doyle, A. G.; Cernak, T. Automation and Computer-Assisted Planning for Chemical Synthesis. *Nat. Rev. Methods Primer* **2021**, *1* (1), 1–23. https://doi.org/10.1038/s43586-021-00022-5.

(24) Betinol, I. O.; Reid, J. P. A Predictive and Mechanistic Statistical Modelling Workflow for Improving Decision Making in Organic Synthesis and Catalysis. *Org. Biomol. Chem*. **2022**, *20* (30), 6012–6018. https://doi.org/10.1039/D2OB00272H.

(25) Beker, W.; Roszak, R.; Wołos, A.; Angello, N. H.; Rathore, V.; Burke, M. D.; Grzybowski, B. A. Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki–Miyaura Coupling. *J. Am. Chem. Soc.* **2022**, *144* (11), 4819–4827. https://doi.org/10.1021/jacs.1c12005.

(26) Strieth-Kalthoff, F.; Sandfort, F.; Kühnemund, M.; Schäfer, F. R.; Kuchen, H.; Glorius, F. Machine Learning for Chemical Reactivity: The Importance of Failed Experiments. *Angew. Chem. Int. Ed.* **2022**, *61* (29), e202204647. https://doi.org/10.1002/anie.202204647.

(27) Skoraczyński, G.; Dittwald, P.; Miasojedow, B.; Szymkuć, S.; Gajewska, E. P.; Grzybowski, B. A.; Gambin, A. Predicting the Outcomes of Organic Reactions via Machine Learning: Are Current Descriptors Sufficient? *Sci. Rep.* **2017**, *7* (1), 3582. https://doi.org/10.1038/s41598-017-02303-0.

(28) Bunnett, J. F.; Zahler, R. E. Aromatic Nucleophilic Substitution Reactions. Chem. Rev. **1951**, 49 (2), 273–412. https://doi.org/10.1021/cr60153a002.

(29) Rohrbach, S.; Smith, A. J.; Pang, J. H.; Poole, D. L.; Tuttle, T.; Chiba, S.; Murphy, J. A. Concerted Nucleophilic Aromatic Substitution Reactions. *Angew. Chem. Int. Ed.* **2019**, *58* (46), 16368–16388. https://doi.org/10.1002/anie.201902216.

(30) Meisenheimer, J. Ueber Reactionen Aromatischer Nitrokörper. *Justus Liebigs Ann. Chem.* **1902**, *323* (2), 205–246. https://doi.org/10.1002/jlac.19023230205.

(31) Bartoli, G.; Todesco, P. E. Nucleophilic Substitution. Linear Free Energy Relations between Reactivity and Physical Properties of Leaving Groups and Substrates. *Acc. Chem. Res.* **1977**, *10* (4), 125–132. https://doi.org/10.1021/ar50112a004.

(32) Senger, N. A.; Bo, B.; Cheng, Q.; Keeffe, J. R.; Gronert, S.; Wu, W. The Element Effect Revisited: Factors Determining Leaving Group Ability in Activated Nucleophilic Aromatic Substitution Reactions. *J. Org. Chem*. **2012**, *77* (21), 9535–9540. https://doi.org/10.1021/jo301134q.

(33) Beck, J. R. Nucleophilic Displacement of Aromatic Nitro Groups. *Tetrahedron* **1978**, *34* (14), 2057–2068. https://doi.org/10.1016/0040-4020(78)89004-8.

(34) Catalano, J. G.; Gaitonde, V.; Beesu, M.; Leivers, A. L.; Shotwell, J. B. Phenoxide Leaving Group SNAr Strategy for the Facile Preparation of 7-Amino-3-Aryl Pyrazolo[1,5-a]Pyrimidines from a 3-Bromo-7-Phenoxypyrazolo[1,5-a]Pyrimidine Intermediate. *Tetrahedron Lett.* **2015**, *56* (44), 6077–6079. https://doi.org/10.1016/j.tetlet.2015.09.068.

(35) Ibrahim, M.; Abdel-Reheem, H.; Khattab, S.; Hamed, E. Nucleophilic Substitution Reactions of 2,4-Dinitrobenzene Derivatives with Hydrazine: Leaving Group and Solvent Effects. *Int. J. Chem.* **2013**, *5* (3), p33. https://doi.org/10.5539/ijc.v5n3p33.

(36) Thompson, A. D.; Huestis, M. P. Cyanide Anion as a Leaving Group in Nucleophilic Aromatic Substitution: Synthesis of Quaternary Centers at Azine Heterocycles. *J. Org. Chem.* **2013**, *78* (2), 762–769. https://doi.org/10.1021/jo302307y.

(37) Kriķis, K.-Ē.; Novosjolova, I.; Mishnev, A.; Turks, M. 1,2,3-Triazoles as Leaving Groups in SNAr–Arbuzov Reactions: Synthesis of C6-Phosphonated Purine Derivatives. Beilstein *J. Org. Chem.* **2021**, *17* (1), 193–202. https://doi.org/10.3762/bjoc.17.19.

(38) Bakavoli, M.; Pordel, M.; Rahimizadeh, M.; Jahandari, P. A New Leaving Group in Nucleophilic Aromatic Substitution Reactions (SNAr). *J. Chem. Res.* **2008**, *2008* (8), 432–433. https://doi.org/10.3184/030823408X338693.

(39) Chupakhin, O. N.; Charushin, V. N.; Plas, H. C. van der. *Nucleophilic Aromatic Substitution of Hydrogen*; Academic Press, 2012.

(40) Owczarczyk, M. M. Z. Direct Formylation of Nitroarenes via Vicarious Nucleophilic Substitution of Hydrogen. *Tetrahedron Lett.* **1987**, *28* (26), 3021–3022. https://doi.org/10.1016/S0040-4039(00)96274-9.

(41) Antoniak, D.; Barbasiewicz, M. Alkylation of Nitropyridines via Vicarious Nucleophilic Substitution. *Org. Lett.* **2022**, *24* (2), 516–519. https://doi.org/10.1021/acs.orglett.1c03920.

(42) Ajenjo, J.; Greenhall, M.; Zarantonello, C.; Beier, P. Synthesis and Nucleophilic Aromatic Substitution of 3-Fluoro-5-Nitro-1-(Pentafluorosulfanyl)Benzene. Beilstein *J. Org. Chem.* **2016**, *12* (1), 192–197. https://doi.org/10.3762/bjoc.12.21.

(43) Evans, D. A.; Dinsmore, C. J.; Watson, P. S.; Wood, M. R.; Richardson, T. I.; Trotter, B. W.; Katz, J. L. Nonconventional Stereochemical Issues in the Design of the Synthesis of the Vancomycin Antibiotics: Challenges Imposed by Axial and Nonplanar Chiral Elements in the Heptapeptide Aglycons. *Angew. Chem. Int. Ed.* **1998**, *37* (19), 2704–2708. https://doi.org/10.1002/(SICI)1521-3773(19981016)37:19<2704::AID-ANIE2704>3.0.CO;2-1.

(44) Evans, D. A.; Wood, M. R.; Trotter, B. W.; Richardson, T. I.; Barrow, J. C.; Katz, J. L. Total Syntheses of Vancomycin and Eremomycin Aglycons. *Angew. Chem. Int. Ed.* **1998**, *37* (19), 2700–2704. https://doi.org/10.1002/(SICI)1521-3773(19981016)37:19<2700::AID-ANIE2700>3.0.CO;2-P.

(45) Nicolaou, K. C.; Natarajan, S.; Li, H.; Jain, N. F.; Hughes, R.; Solomon, M. E.; Ramanjulu, J. M.; Boddy, C. N. C.; Takayanagi, M. Total Synthesis of Vancomycin Aglycon—Part 1: Synthesis of Amino Acids 4–7 and Construction of the AB-COD Ring Skeleton. *Angew. Chem. Int. Ed.* **1998**, *37* (19), 2708–2714. https://doi.org/10.1002/(SICI)1521-3773(19981016)37:19<2708::AID-ANIE2708>3.0.CO;2-E.

(46) Nicolaou, K. C.; Jain, N. F.; Natarajan, S.; Hughes, R.; Solomon, M. E.; Li, H.; Ramanjulu, J. M.; Takayanagi, M.; Koumbis, A. E.; Bando, T. Total Synthesis of Vancomycin Aglycon—Part 2: Synthesis of Amino Acids 1–3 and Construction of the AB-COD-DOE Ring Skeleton. Angew. *Chem. Int. Ed*. **1998**, *37* (19), 2714–2716. https://doi.org/10.1002/(SICI)1521-3773(19981016)37:19<2714::AID-ANIE2714>3.0.CO;2-#.

(47) Nicolaou, K. C.; Takayanagi, M.; Jain, N. F.; Natarajan, S.; Koumbis, A. E.; Bando, T.; Ramanjulu, J. M. Total Synthesis of Vancomycin Aglycon—Part 3: Final Stages. *Angew. Chem. Int. Ed*. **1998**, *37* (19), 2717–2719. https://doi.org/10.1002/(SICI)1521-3773(19981016)37:19<2717::AID-ANIE2717>3.0.CO;2-I.

(48) Zhang, A. J.; Burgess, K. Total Syntheses of Vancomycin. *Angew. Chem. Int. Ed.* **1999**, *38* (5), 634–636. https://doi.org/10.1002/(SICI)1521-3773(19990301)38:5<634::AID-ANIE634>3.0.CO;2-G.

(49) Cheng, L.-J.; Xie, J.-H.; Chen, Y.; Wang, L.-X.; Zhou, Q.-L. Enantioselective Total Synthesis of (−)-Δ8-THC and (−)-Δ9-THC via Catalytic Asymmetric Hydrogenation and SNAr Cyclization. *Org. Lett.* **2013**, *15* (4), 764–767. https://doi.org/10.1021/ol303351y.

(50) Yamashita, K.; Kume, Y.; Ashibe, S.; Puspita, C. A. D.; Tanigawa, K.; Michihata, N.; Wakamori, S.; Ikeuchi, K.; Yamada, H. Total Synthesis of Mallotusinin. *Chem. – Eur. J.* **2020**, *26* (69), 16408–16421. https://doi.org/10.1002/chem.202002753.

(51) Flick, A. C.; Leverett, C. A.; Ding, H. X.; McInturff, E.; Fink, S. J.; Helal, C. J.; DeForest, J. C.; Morse, P. D.; Mahapatra, S.; O'Donnell, C. J. Synthetic Approaches to New Drugs Approved during 2018. *J. Med. Chem.* **2020**, *63* (19), 10652–10704. https://doi.org/10.1021/acs.jmedchem.0c00345.

(52) Jeanmart, S.; Edmunds, A. J. F.; Lamberth, C.; Pouliot, M. Synthetic Approaches to the 2010–2014 New Agrochemicals. Bioorg. *Med. Chem.* **2016**, *24* (3), 317–341. https://doi.org/10.1016/j.bmc.2015.12.014.

(53) Jeanmart, S.; Edmunds, A. J. F.; Lamberth, C.; Pouliot, M.; Morris, J. A. Synthetic Approaches to the 2015–2018 New Agrochemicals. *Bioorg. Med. Chem.* **2021**, *39*, 116162. https://doi.org/10.1016/j.bmc.2021.116162.

(54) Baumann, M.; Baxendale, I. R. An Overview of the Synthetic Routes to the Best Selling Drugs Containing 6-Membered Heterocycles. Beilstein *J. Org. Chem.* **2013**, *9* (1), 2265–2319. https://doi.org/10.3762/bjoc.9.265.

(55) Flick, A. C.; Leverett, C. A.; Ding, H. X.; McInturff, E.; Fink, S. J.; Mahapatra, S.; Carney, D. W.; Lindsey, E. A.; DeForest, J. C.; France, S. P.; Berritt, S.; Bigi-Botterill, S. V.; Gibson, T. S.; Liu, Y.; O'Donnell, C. J. Synthetic Approaches to the New Drugs Approved during 2019. *J. Med. Chem.* **2021**, *64* (7), 3604–3657. https://doi.org/10.1021/acs.jmedchem.1c00208.

(56) Vitaku, E.; Smith, D. T.; Njardarson, J. T. Analysis of the Structural Diversity, Substitution Patterns, and Frequency of Nitrogen Heterocycles among U.S. FDA Approved Pharmaceuticals. *J. Med. Chem.* **2014**, *57* (24), 10257–10274. https://doi.org/10.1021/jm501100b.

(57) Delost, M. D.; Smith, D. T.; Anderson, B. J.; Njardarson, J. T. From Oxiranes to Oligomers: Architectures of U.S. FDA Approved Pharmaceuticals Containing Oxygen Heterocycles. *J. Med. Chem.* **2018**, *61* (24), 10996–11020. https://doi.org/10.1021/acs.jmedchem.8b00876.

(58) Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **2016**, *59* (10), 4443–4458. https://doi.org/10.1021/acs.jmedchem.5b01409.

(59) Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59* (1), 96–103. https://doi.org/10.1021/ja01280a022.

(60) Hansch, Corwin.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91* (2), 165–195. https://doi.org/10.1021/cr00002a004.

(61) Kang, T.-A.; Cho, H.-J.; Um, I.-H. Kinetic Study on SNAr Reaction of 1-Y-Substituted-Phenoxy-2,4-Dinitrobenzenes with Hydroxide Ion: Effect of Substituent Y on Reactivity and Reaction Mechanism. Bull. *Korean Chem. Soc.* **2014**, *35* (7), 2135–2138. https://doi.org/10.5012/bkcs.2014.35.7.2135.

(62) Yau, H. M.; Haines, R. S.; Harper, J. B. A Robust, "One-Pot" Method for Acquiring Kinetic Data for Hammett Plots Used To Demonstrate Transmission of Substituent Effects in Reactions of Aromatic Ethyl Esters. *J. Chem. Educ.* **2015**, *92* (3), 538–542. https://doi.org/10.1021/ed400701m.

(63) Bain, R. M.; Pulliam, C. J.; Yan, X.; Moore, K. F.; Müller, T.; Cooks, R. G. Mass Spectrometry in Organic Synthesis: Claisen–Schmidt Base-Catalyzed Condensation and Hammett Correlation of Substituent Effects. *J. Chem. Educ.* **2014**, *91* (11), 1985–1989. https://doi.org/10.1021/ed500288m.

(64) Miller, J.; Kai-Yan, W. 655. The SN Mechanism in Aromatic Compounds. Part XXIX. Some Para-Substituted Chlorobenzenes. *J. Chem. Soc.*, 1963, 3492-3495. https://doi.org/10.1039/JR9630003492.

(65) Fry, S. E.; Pienta, N. J. Effects of Molten Salts on Reactions. Nucleophilic Aromatic Substitution by Halide Ions in Molten Dodecyltributylphosphonium Salts. *J. Am. Chem. Soc*. **1985**, *107* (22), 6399–6400. https://doi.org/10.1021/ja00308a045.

(66) Um, I.-H.; Im, L.-R.; Kang, J.-S.; Bursey, S. S.; Dust, J. M. Mechanistic Assessment of SNAr Displacement of Halides from 1-Halo-2,4-Dinitrobenzenes by Selected Primary and Secondary Amines: Brønsted and Mayr Analyses. *J. Org. Chem.* **2012**, *77* (21), 9738–9746. https://doi.org/10.1021/jo301862b.

(67) ElGuesmi, N.; Berionni, G.; Asghar, B. H. Mechanistic Studies and Quantification of the Electrophilicity of Aromatic Triflones in σ-Complexation and S_NAr Reactions. *J. Fluor. Chem*. **2014**, *160*, 41–47. https://doi.org/10.1016/j.jfluchem.2014.01.003.

(68) Mahdhaoui, F.; Zaier, R.; Dhahri, N.; Ayachi, S.; Boubaker, T. SNAr Reactions of Substituted Pyridines with Secondary Amines in Aqueous Solution: Kinetic and Reactivity Indices in Density Functional Theory. *Int. J. Chem.* Kinet. **2019**, *51* (4), 249–257. https://doi.org/10.1002/kin.21247.

(69) Yang, W.; Parr, R. G. Hardness, Softness, and the Fukui Function in the Electronic Theory of Metals and Catalysis. Proc. *Natl. Acad. Sci.* **1985**, *82* (20), 6723–6726. https://doi.org/10.1073/pnas.82.20.6723.

(70) Scales, S.; Johnson, S.; Hu, Q.; Do, Q.-Q.; Richardson, P.; Wang, F.; Braganza, J.; Ren, S.; Wan, Y.; Zheng, B.; Faizi, D.; McAlpine, I. Studies on the Regioselective Nucleophilic Aromatic Substitution (SNAr) Reaction of 2-Substituted 3,5-Dichloropyrazines. *Org. Lett.* **2013**, *15* (9), 2156–2159. https://doi.org/10.1021/ol4006695.

(71) Epiotis, N. D.; Cherry, W. An MO Study of the Orientational Selectivity in Nucleophilic Substitution Reactions of Polyhalobenzenes. *J. Am. Chem. Soc.* **1976**, *98* (18), 5432–5435. https://doi.org/10.1021/ja00434a003.

(72) O'Donnell, M.; Charrier, J.-D. A Complementary Route to Diaminopyrimidines through Regioselective SNAr Amination Reactions. *Tetrahedron* **2015**, *71* (10), 1515–1522. https://doi.org/10.1016/j.tet.2015.01.043.

(73) Muir, M.; Baker, J. A Simple Calculational Model for Predicting the Site for Nucleophilic Substitution in Aromatic Perfluorocarbons. *J. Fluor. Chem.* **2005**, *126* (5), 727–738. https://doi.org/10.1016/j.jfluchem.2005.02.018.

(74) Baker, J.; Muir, M. The Meisenheimer Model for Predicting the Principal Site for Nucleophilic Substitution in Aromatic Perfluorocarbons — Generalization to Include Ring-Nitrogen Atoms and Non-Fluorine Ring Substituents. *Can. J. Chem.* **2010**, *88* (7), 588–597. https://doi.org/10.1139/V10-047.

(75) Liljenberg, M.; Brinck, T.; Rein, T.; Svensson, M. Utilizing the σ-Complex Stability for Quantifying Reactivity in Nucleophilic Substitution of Aromatic Fluorides. Beilstein *J. Org. Chem.* **2013**, *9* (1), 791–799. https://doi.org/10.3762/bjoc.9.90.

(76) Liljenberg, M.; Brinck, T.; Herschend, B.; Rein, T.; Tomasi, S.; Svensson, M. Predicting Regioselectivity in Nucleophilic Aromatic Substitution. *J. Org. Chem.* **2012**, *77* (7), 3262–3269. https://doi.org/10.1021/jo202569n.

(77) Liljenberg, M.; Brinck, T.; Herschend, B.; Rein, T.; Rockwell, G.; Svensson, M. A Pragmatic Procedure for Predicting Regioselectivity in Nucleophilic Substitution of Aromatic Fluorides. *Tetrahedron Lett.* **2011**, *52* (24), 3150–3153. https://doi.org/10.1016/j.tetlet.2011.04.032.

(78) Brinck, T.; Carlqvist, P.; Stenlid, J. H. Local Electron Attachment Energy and Its Use for Predicting Nucleophilic Reactions and Halogen Bonding. *J. Phys. Chem. A* **2016**, *120* (50), 10023–10032. https://doi.org/10.1021/acs.jpca.6b10142.

(79) Stenlid, J. H.; Brinck, T. Nucleophilic Aromatic Substitution Reactions Described by the Local Electron Attachment Energy. *J. Org. Chem.* **2017**, *82* (6), 3072–3083. https://doi.org/10.1021/acs.joc.7b00059.

(80) Guan, Y.; Lee, T.; Wang, K.; Yu, S.; McWilliams, J. C. SNAr Regioselectivity Predictions: Machine Learning Triggering DFT Reaction Modeling through Statistical Threshold. *J. Chem. Inf. Model*. **2023**, *63* (12), 3751–3760. https://doi.org/10.1021/acs.jcim.3c00580.

(81) Mack, J. B. C.; Bedell, T. A.; DeLuca, R. J.; Hone, G. A. B.; Roizen, J. L.; Cox, C. T.; Sorensen, E. J.; Du Bois, J. Rhodium-Catalyzed C–H Amination: A Case Study of Selectivity in C–H Functionalization Reactions. *J. Chem. Educ.* **2018**, *95* (12), 2243–2248. https://doi.org/10.1021/acs.jchemed.7b00697.

(82) Man Yau, H.; K. Croft, A.; B. Harper, J. 'One-Pot' Hammett Plots: A General Method for the Rapid Acquisition of Relative Rate Data. *Chem. Commun.* **2012**, *48* (71), 8937–8939. https://doi.org/10.1039/C2CC34074G.

(83) Yau, H. M.; Haines, R. S.; Harper, J. B. A Robust, "One-Pot" Method for Acquiring Kinetic Data for Hammett Plots Used To Demonstrate Transmission of Substituent Effects in Reactions of Aromatic Ethyl Esters. *J. Chem. Educ.* **2015**, *92* (3), 538–542. https://doi.org/10.1021/ed400701m.

(84) Fenwick, N. W.; Telford, R.; Saidykhan, A.; Martin, W. H. C.; Bowen, R. D. Hammett Correlation in the Accelerated Formation of 2,3-Diphenylquinoxalines in Nebulizer Microdroplets. *Molecules* **2021**, *26* (16), 5077. https://doi.org/10.3390/molecules26165077.

(85) Fiori, K. W.; Du Bois, J. Catalytic Intermolecular Amination of C−H Bonds: Method Development and Mechanistic Insights. *J. Am. Chem. Soc*. **2007**, *129* (3), 562–568. https://doi.org/10.1021/ja0650450.

(86) Lu, J.; Donnecke, S.; Paci, I.; Leitch, D. C. A Reactivity Model for Oxidative Addition to Palladium Enables Quantitative Predictions for Catalytic Cross-Coupling Reactions. *Chem. Sci.* **2022**, *13* (12), 3477–3488. https://doi.org/10.1039/D2SC00174H.

(87) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003**, *103* (5), 1793–1874. https://doi.org/10.1021/cr990029p.

(88) Suresh, C. H.; Alexander, P.; Vijayalakshmi, K. P.; Sajith, P. K.; Gadre, S. R. Use of Molecular Electrostatic Potential for Quantitative Assessment of Inductive Effect. *Phys. Chem. Chem. Phys.* **2008**, *10* (43), 6492–6499. https://doi.org/10.1039/B809561B.

(89) Sayyed, F. B.; Suresh, C. H. Quantification of Substituent Effects Using Molecular Electrostatic Potentials: Additive Nature and Proximity Effects. *New J. Chem.* **2009**, *33* (12), 2465–2471. https://doi.org/10.1039/B9NJ00333A.

(90) Remya, G. S.; Suresh, C. H. Quantification and Classification of Substituent Effects in Organic Chemistry: A Theoretical Molecular Electrostatic Potential Study. *Phys. Chem. Chem. Phys.* **2016**, *18* (30), 20615–20626. https://doi.org/10.1039/C6CP02936A.

(91) Gadre, S. R.; Suresh, C. H.; Mohan, N. Electrostatic Potential Topology for Probing Molecular Structure, Bonding and Reactivity. *Molecules* **2021**, *26* (11), 3289. https://doi.org/10.3390/molecules26113289.

(92) Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the Definition of the Q2 Parameter for QSAR Validation. *J. Chem. Inf. Model.* **2009**, *49* (7), 1669–1678. https://doi.org/10.1021/ci900115y.

(93) Houk, K. N. Frontier Molecular Orbital Theory of Cycloaddition Reactions. *Acc. Chem. Res.* **1975**, *8* (11), 361–369. https://doi.org/10.1021/ar50095a001.

(94) Fukui, K.; Yonezawa, T.; Nagata, C.; Shingu, H. Molecular Orbital Theory of Orientation in Aromatic, Heteroaromatic, and Other Conjugated Molecules. *J. Chem. Phys.* **2004**, *22* (8), 1433–1442. https://doi.org/10.1063/1.1740412.

(95) Allgäuer, D. S.; Jangra, H.; Asahara, H.; Li, Z.; Chen, Q.; Zipse, H.; Ofial, A. R.; Mayr, H. Quantification and Theoretical Analysis of the Electrophilicities of Michael Acceptors. *J. Am. Chem. Soc.* **2017**, *139* (38), 13318–13329. https://doi.org/10.1021/jacs.7b05106.

(96) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L. Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy. *Phys. Rev. Lett.* **1982**, *49* (23), 1691–1694. https://doi.org/10.1103/PhysRevLett.49.1691.

(97) Koopmans, T. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. *Physica* **1934**, *1* (1), 104–113. https://doi.org/10.1016/S0031-8914(34)90011-2.

(98) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA Quantum Chemistry Program Package. *J. Chem. Phys*. **2020**, *152* (22), 224108. https://doi.org/10.1063/5.0004608.

(99) Interactive Chemistry | Envision. https://www.entos.ai/envision (accessed 2023-09-17).

(100) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All Spd-Block Elements (Z = 1–86). *J. Chem. Theory Comput.* **2017**, *13* (5), 1989–2009. https://doi.org/10.1021/acs.jctc.7b00118.

(101) Lu, T.; Chen, F. Quantitative Analysis of Molecular Surface Based on Improved Marching Tetrahedra Algorithm. *J. Mol. Graph. Model.* **2012**, *38*, 314–323. https://doi.org/10.1016/j.jmgm.2012.07.004.

(102) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33* (5), 580–592. https://doi.org/10.1002/jcc.22885.

(103) Wang, B.; Rong, C.; Chattaraj, P. K.; Liu, S. A Comparative Study to Predict Regioselectivity, Electrophilicity and Nucleophilicity with Fukui Function and Hirshfeld Charge. *Theor. Chem. Acc.* **2019**, *138* (12), 124. https://doi.org/10.1007/s00214-019-2515-1.

(104) Burdon, J. A Rationalization of Orientation and Reactivity in the Nucleophilic Replacement Reactions of Aromatic Polyhalo-Compounds. *Tetrahedron* **1965**, *21* (12), 3373–3380. https://doi.org/10.1016/S0040-4020(01)96958-3.

(105) Burdon, J.; Parsons, I. W. A Recent Attempt to Apply Frontier Orbital Theory to Nucleophilic Aromatic Substitution. *J. Am. Chem. Soc.* **1977**, *99* (23), 7445–7447. https://doi.org/10.1021/ja00465a007.

(106) Roy, R. K.; Saha, S. Studies of Regioselectivity of Large Molecular Systems Using DFT Based Reactivity Descriptors. *Annu. Rep. Sect. C Phys. Chem.* **2010**, *106* (0), 118–162. https://doi.org/10.1039/B811052M.

(107) Neumann, C. N.; Hooker, J. M.; Ritter, T. Concerted Nucleophilic Aromatic Substitution with 19F− and 18F−. *Nature* **2016**, *534* (7607), 369–373. https://doi.org/10.1038/nature17667.

(108) Kwan, E. E.; Zeng, Y.; Besser, H. A.; Jacobsen, E. N. Concerted Nucleophilic Aromatic Substitutions. *Nat. Chem*. **2018**, *10* (9), 917–923. https://doi.org/10.1038/s41557-018-0079-7.

(109) Miller, J.; Parker, A. J. The SN Mechanism in Aromatic Compounds. 25. Substituent Effects of Multiple-Bond Nitrogen. *Aust. J. Chem*. **1958**, *11* (3), 302–308. https://doi.org/10.1071/ch9580302.

(110) Berliner, E.; Monack, L. C. Nucleophilic Displacement in the Benzene Series [1]. *J. Am. Chem. Soc.* **1952**, *74* (6), 1574–1579. https://doi.org/10.1021/ja01126a069.

(111) Crampton, M. R.; Emokpae, T. A.; Isanbor, C. The Effects of Ring Substituents and Leaving Groups on the Kinetics of SNAr Reactions of 1-Halogeno- and 1-Phenoxy-Nitrobenzenes with Aliphatic Amines in Acetonitrile. Eur. *J. Org. Chem.* **2007**, *2007* (8), 1378–1383. https://doi.org/10.1002/ejoc.200600968.

(112) Bolton, R.; Sandall, J. P. B. Nucleophilic Displacement in Polyhalogenoaromatic Compounds. Part 1. Kinetics of Reaction of Polychlorofluorobenzene Derivatives. *J. Chem. Soc.* Perkin Trans. 2 **1976**, No. 13, 1541–1545. https://doi.org/10.1039/P29760001541.

(113) Chambers, R. D.; Martin, P. A.; Sandford, G.; Williams, D. L. H. Mechanisms of Reactions of Halogenated Compounds: Part 7. Effects of Fluorine and Other Groups as Substituents on Nucleophilic Aromatic Substitution. *J. Fluor. Chem.* **2008**, *129* (10), 998–1002. https://doi.org/10.1016/j.jfluchem.2008.04.009.

(114) Chambers, R. D.; Martin, P. A.; Waterhouse, J. S.; Williams, D. L. H.; Anderson, B. Mechanisms for Reactions of Halogenated Compounds. Part 4.[1] Activating Influences of Ring-Nitrogen and Trifluoromethyl in Nucleophilic Aromatic Substitution. *J. Fluor. Chem.* **1982**, *20* (4), 507–514. https://doi.org/10.1016/S0022-1139(00)82276-9.

(115) Chambers, R. D.; Musgrave, W. K. R.; Urben, P. G. Polyhalogenoheterocyclic Compounds. Part XXVI. Nucleophilic Substitution in Trifluoropyrazines. *J. Chem. Soc., Perkin Trans*. *1*, 1974, 2580-2584. https://doi.org/10.1039/P19740002580.

(116) Banks, R. E.; Prakash, A.; Venayak, N. D. Studies in Azide Chemistry. Part IX [1]. Investigations Involving Fluorinated Azidopyrimidines and 4-Azido-3-Chloro-2,5,6-Trifluoropyridine. *J. Fluor. Chem.* **1980**, *16* (4), 325–338. https://doi.org/10.1016/S0022-1139(00)82318-0.

(117) Chambers, R. D.; Seabury, M. J.; Williams, D. L. H.; Hughes, N. Mechanisms for Reactions of Halogenated Compounds. Part 5. Orientating Effects of Fluorine Substituents on Nucleophilic Substitution in Naphthalene and Other Polycyclic Systems. *J. Chem. Soc. Perkin 1* **1988**, No. 2, 251–254. https://doi.org/10.1039/P19880000251.

(118) Chambers, R. D.; Seabury, M. J.; Williams, D. L. H.; Hughes, N. Mechanisms for Reactions of Halogenated Compounds. Part 6. Investigations into the Activating Effect of Ortho-Fluorine in Nucleophilic Aromatic Substitution. *J. Chem. Soc. Perkin 1* **1988**, No. 2, 255–257. https://doi.org/10.1039/P19880000255.

(119) Keana, J. F. W.; Cai, S. X. New Reagents for Photoaffinity Labeling: Synthesis and Photolysis of Functionalized Perfluorophenyl Azides. *J. Org. Chem.* **1990**, 55 (11), 3640–3647. https://doi.org/10.1021/jo00298a048.

(120) Dirr, R.; Anthaume, C.; Désaubry, L. Regioselectivity of Fluorine Substitution by Alkoxides on Unsymmetrical Difluoroarenes. *Tetrahedron Lett.* **2008**, *49* (31), 4588–4590. https://doi.org/10.1016/j.tetlet.2008.05.091.

(121) Brooke, G. M.; Chambers, R. D.; Drury, C. J.; Bower, M. J. Remarkable Orientational Effects in the Displacement of the Fluorine from Heptafluoro-Isoquinoline and -Quinoline towards Sulfur Nucleophiles. Further Reactions with Oxygen Nucleophiles. *J. Chem. Soc. Perkin 1* **1993**, No. 18, 2201–2209. https://doi.org/10.1039/P19930002201.

(122) Bunnett, J. F.; Morath, R. J. The Ortho: Para Ratio in Activation of Aromatic Nucleophilic Substitution by the Nitro Group1. *J. Am. Chem. Soc.* **1955**, *77* (19), 5051–5055. https://doi.org/10.1021/ja01624a033.

(123) Belli, M. L.; Illuminati, G.; Marino, G. Meta-Substituent Effects on the Kinetics of Methoxydechlorination of Some 2- and 4-Chloroquinolines. *Tetrahedron* **1963**, 19 (2), 345–355. https://doi.org/10.1016/S0040-4020(01)98537-0.

(124) 美穂湯川; 時寛新矢; 良宣後藤; 尚夫坂本; 博吉沢; 敦子渡辺; 宏山中. MNDO (Modified Neglect of Diatomic Overlap) Study of the Nucleophilic Substitution Reactions of Chloropyrimidines. *Chem. Pharm. Bull.* (Tokyo) **1989**, *37* (11), 2892–2896. https://doi.org/10.1248/cpb.37.2892.

(125) Flowers, W. T.; Haszeldine, R. N.; Majid, S. A. Synthesis and Reactions of Pentachloropyridine. *Tetrahedron Lett.* **1967**, *8* (26), 2503–2505. https://doi.org/10.1016/S0040-4039(00)90842-6.

(126) Collins, I.; Suschitzky, H. Polyhalogeno-Aromatic Compounds. Part VIII. Reactions of Hexabromobenzene with Nucleophiles. *J. Chem. Soc. C Org.* **1969**, No. *18*, 2337–2341. https://doi.org/10.1039/J39690002337.

(127) Volkov, K. A.; Avramenko, G. V.; Negrimovskii, V. M.; Luk'yanets, E. A. Phthalocyanines and Related Compounds: XLIII. Synthesis of Poly[Phenyl(Alkyl)Sulfanyl]-Substituted Phthalonitriles and Some Phthalocyanines Based Thereon. *Russ. J. Gen. Chem.* **2007**, *77* (6), 1108–1116. https://doi.org/10.1134/S107036320706028X.

(128) Zou, X.-Z.; Toshikaz, I. High Pressure Synthesis of Nitrophenylamines Type Nonlinear Optical Compounds. *Chin. J. Org. Chem.* **2002**, *22*, 778–781.

(129) Bijev, A. New Heterocyclic Hydrazones in the Search for Antitubercular Agents: Synthesis and In Vitro Evaluations. *Lett. Drug Des. Discov.* **2006**, *3* (7), 506–512. https://doi.org/10.2174/157018006778194790.

(130) Collins, I.; Suschitzky, H. Polyhalogeno-Aromatic Compounds. Part XIV. Nucleophilic Substitution and Peroxy-Acid Oxidation of Pentabromopyridine and Some of Its NN-Dialkylamino- and Bis-(NN-Dialkylamino)-Derivatives. *J. Chem. Soc. C Org.* **1970**, No. *11*, 1523–1530. https://doi.org/10.1039/J39700001523.

(131) Tanaka, K.; Deguchi, M.; Iwata, S. Ab Initio Study of Nucleophilic Aromatic Substitution of Polyfluorobenzene. *J. Chem. Res.* **1999**, *23* (9), 528–529. https://doi.org/10.1177/174751989902300905.

(132) Delia, T. J.; Anderson, D. P.; Schomaker, J. M. 2,4,6-Trifluoropyrimidine. Reactions with Nitrogen Nucleophiles. *J. Heterocycl. Chem.* **2004**, *41* (6), 991–993. https://doi.org/10.1002/jhet.5570410623.

(133) J. M. Allen , R. J. Butlin , C. Green , W. Mccoull , G. R. Robb and J. M. Wood , *Benzothiazoles as Ghrelin Receptor Modulators*, WO2009047558A1, 2009.

(134) O. Weber , V. Voehringer , H.-G. Lerchen , F.-T. Hafner , J. Keldenich , K.-H. Schlemmer , U. Krenz and B. Riedl *, Benzofuran and Benzothiophene Derivatives Useful in the Treatment of Cancers of the Central Nervous System* , . WO2008025509A1, 2008.

(135) N. Ahmad , D. Boyall , J.-D. Charrier , C. Davis , R. Davis , S. Durrant , G. E. I. Jardi , D. Fraysse , J.-M. Jimenez , D. Kay , R. Knegtel , D. Middleton , M. O'Donnell , M. Panesar , F. Pierard , J. Pinder , D. Shaw , P.-H. Storck , J. Studley and H. Twin , *Compounds Useful as Inhibitors of Atr Kinase* , . WO2014089379A1, 2014.

(136) Zhou, D.; Stack, G. P.; Lo, J.; Failli, A. A.; Evrard, D. A.; Harrison, B. L.; Hatzenbuhler, N. T.; Tran, M.; Croce, S.; Yi, S.; Golembieski, J.; Hornby, G. A.; Lai, M.; Lin, Q.; Schechter, L. E.; Smith, D. L.; Shilling, A. D.; Huselton, C.; Mitchell, P.; Beyer, C. E.; Andree, T. H. Synthesis, Potency, and in Vivo Evaluation of 2-Piperazin-1-Ylquinoline Analogues as Dual Serotonin Reuptake Inhibitors and Serotonin 5-HT1A Receptor Antagonists. *J. Med. Chem.* **2009**, *52* (15), 4955–4959. https://doi.org/10.1021/jm900374r.

(137) H. Mizuno and A. Manabe , *Pyrimidine Compounds and Pests Controlling Composition Containing the Same* , . WO2004099160A1, 2004.

(138) R. Al-Awar , M. Isaac , A. M. Chau , A. Mamai , I. Watson , G. Poda , P. Subramanian , B. Wilson , D. Uehling , M. Prakesch , B. Joseph and J.-A. Morin , *Inhibitors of the Bcl6 Btb Domain Protein-Protein Interaction and Uses Thereof*. WO2019153080A1, 2019.

(139) Stevens, J. M.; Simmons, E. M.; Tan, Y.; Borovika, A.; Fan, J.; Forest, R. V.; Geng, P.; Guerrero, C. A.; Lou, S.; Skliar, D.; Steinhardt, S. E.; Strotman, N. A. Leveraging High-Throughput Experimentation to Drive Pharmaceutical Route Invention: A Four-Step Commercial Synthesis of Branebrutinib (BMS-986195). *Org. Process Res. Dev*. **2022**, *26* (4), 1174–1183. https://doi.org/10.1021/acs.oprd.1c00443.

(140) Tao, Y.; Keene, N. F.; Wiglesworth, K. E.; Sitter, B.; McWilliams, J. C. Early Process Development of an Irreversible Epidermal Growth Factor Receptor (EGFR) T790 M Inhibitor. *Org. Process Res. Dev.* **2019**, *23* (3), 382–388. https://doi.org/10.1021/acs.oprd.8b00437.

(141) Stumpf, A.; Cheng, Z. K.; Beaudry, D.; Angelaud, R.; Gosselin, F. Improved Synthesis of the Nav1.7 Inhibitor GDC-0276 via a Highly Regioselective S_NAr Reaction. *Org. Process Res. Dev.* **2019**, *23* (9), 1829–1840. https://doi.org/10.1021/acs.oprd.9b00082.

(142) Bremberg, U.; Eriksson-Bajtner, J.; Lehmann, F.; Oltner, V.; Sölver, E.; Wennerberg, J. Development of a Synthesis of Kinase Inhibitor AKN028. *Org. Process Res. Dev.* **2018**, *22* (10), 1360–1364. https://doi.org/10.1021/acs.oprd.8b00092.

(143) Tao, Y.; Widlicka, D. W.; Hill, P. D.; Couturier, M.; Young, G. R. A Scalable Synthesis of CE-157119 HCl Salt, an SRI/5-HT $_{2A}$ Antagonist. *Org. Process Res. Dev.* **2012**, *16* (11), 1805–1810. https://doi.org/10.1021/op3002273.

(144) Barcan, G. A.; Conde, J. J.; Mokhallalati, M. K.; Nilson, M. G.; Xie, S.; Allen, C. L.; Andemichael, Y. W.; Calandra, N. A.; Leitch, D. C.; Li, L.; Morris, M. J. Nucleophilic Aromatic Substitutions of 2-Halo-5-(Sulfamoyl)Benzoic Acids and N,O-Bis-Alkylation via Phase Transfer Catalysis: Synthesis of RoRγ Inverse Agonist GSK2981278A. *Org. Process Res. Dev.* **2019**, *23* (7), 1396–1406. https://doi.org/10.1021/acs.oprd.9b00147.

(145) Lu, Y.; Cole, K. P.; Fennell, J. W.; Maloney, T. D.; Mitchell, D.; Subbiah, R.; Ramadas, B. An Alternative Indazole Synthesis for Merestinib. *Org. Process Res. Dev.* **2018**, *22* (3), 409–419. https://doi.org/10.1021/acs.oprd.8b00016.

Chapter 3 – A Reactivity Model for Oxidative Addition to Palladium Enables Quantitative Predictions for Catalytic Cross-Coupling Reactions

This chapter has been adapted from: Lu J.; Donnecke S.; Paci I.; Leitch D. C. A reactivity model for oxidative addition to palladium enables quantitative predictions for catalytic cross-coupling reactions. *Chem*. *Sci*., **2022**, *13*, 3477—3488, DOI: 10.1039/D2SC00174H

Contributions:

Experiments and computational analysis were performed by Jingru Lu.

Computational analysis was performed by Sofia Donnecke.

3.1 Abstract

This chapter reports a quantitative structure-reactivity model for oxidative addition, a key step in palladium-catalyzed cross couplings. This multivariate linear model links simple molecular descriptors to experimentally determined relative rates of oxidative addition for 79 electrophiles, including chloride, bromide and triflate leaving groups. Because oxidative addition is often the rate and/or selectivity determining step in palladium-catalyzed reactions, this model can be used to make quantitative predictions about catalytic reaction outcomes. Demonstrated applications include a multivariate linear model for the initial rate of Sonogashira coupling reactions, and successful site-selectivity predictions for Suzuki, Buchwald-Hartwig, and Stille reactions of multihalogenated substrates relevant to the synthesis of pharmaceuticals and natural products. In addition to its predictive power, this model is also useful for gaining mechanistic insights into how structurally different substrates undergo oxidative addition.

3.2 Introduction

Palladium-catalyzed cross coupling is among the most important reaction classes in modern organic synthesis. It provides versatile and powerful synthetic strategies to link molecular fragments,[1] with applications ranging from the manufacture of active pharmaceutical ingredients,[2] to the selective modification of biomolecules,[3] to the creation of new functional materials.[4] The rapid advancement in computing power and high-throughput experimentation (HTE) techniques have enabled data-rich approaches in Pd-catalyzed cross-coupling reaction discovery and quantitative predictions. These were discussed in Chapter 1 with a focus on recent developments in this field. Using quantitative structure-selectivity relationships (QSSR) and quantitative structure-reactivity relationships (QSRR) for reaction prediction in Pd-catalyzed processes has led to impressive progress, especially with the aid of advanced statistical analysis tools including machine learning[5–14] and various approaches for large-sized reaction data collection;[15–23] however, current predictive models face limitations such as interpretability, insufficient coverage of chemical space, and inability to make out-of-sample predictions.[24–27]

Here, I describe an approach to predict reaction outcomes of multiple Pd-catalyzed cross-couplings by focusing on the structure-reactivity relationships for one key mechanistic step: oxidative addition of the electrophile to Pd(0) (Fig. 3.1). Oxidative addition is a fundamental organometallic transformation common to many catalytic reactions;[28] it is particularly important in Pd-catalyzed cross-couplings because it is often the rate and/or selectivity determining step[29] (Fig. 3.1A). Thus, a hypothetical quantitative structure-reactivity model for oxidative addition[30–32] of aryl electrophiles to a Pd(0) complex should enable predictions for Pd-catalyzed cross-couplings under various conditions. I developed just such a predictive model by correlating experimental relative rate data to mechanistically meaningful ground state descriptors using multivariate linear regression (MLR) analysis, the same approach as for S_NAr prediction (Chapter 2). To achieve this, I measured oxidative addition rates for a substrate library containing 79 diverse aryl electrophiles (Fig. 3.2B), with an emphasis on incorporating pharmaceutically relevant heterocycles (Fig. 3.1C).

As hypothesized, this MLR model can be applied to many Pd-catalyzed cross-couplings using different substrates and under various reaction conditions. It can predict the catalytic reaction rate based on the substrate structure, correctly predict the most reactive site when multiple sites are available, and can also help to guide synthetic route planning. Importantly, the predictive ability of this model extends well beyond the molecular structures and reaction conditions included in the initial training set. In addition to making predictions to individual substrates, I trained a structure-reactivity model to predict the substrate and ligand effect for Sonogashira coupling using >400 external reaction data from literature. The success of this model training shows how combining MLR and ground state descriptors can lead to powerful prediction tools applicable to multiple reaction variables for diverse reaction systems. Finally, this work demonstrates the potential of focusing reaction prediction efforts on fundamental mechanistic steps in catalytic mechanisms.[33]

While Section 3.3 is part of my published work referred to at the start of this Chapter, Section 3.4 covers additional results in Pd-oxidative addition prediction obtained after publication. This includes: absolute rate measurements by reaction progress analysis; structure-reactivity studies on new substrates including aryl chlorides, (*N*-hetero)aryl iodides and tosylates; investigations of the oxidative addition mechanism of aryl triflates; and development of a simple four-descriptor (ESP_1, A_1, *IBSI*, pK_a) predictive model for aryl halides and triflates.

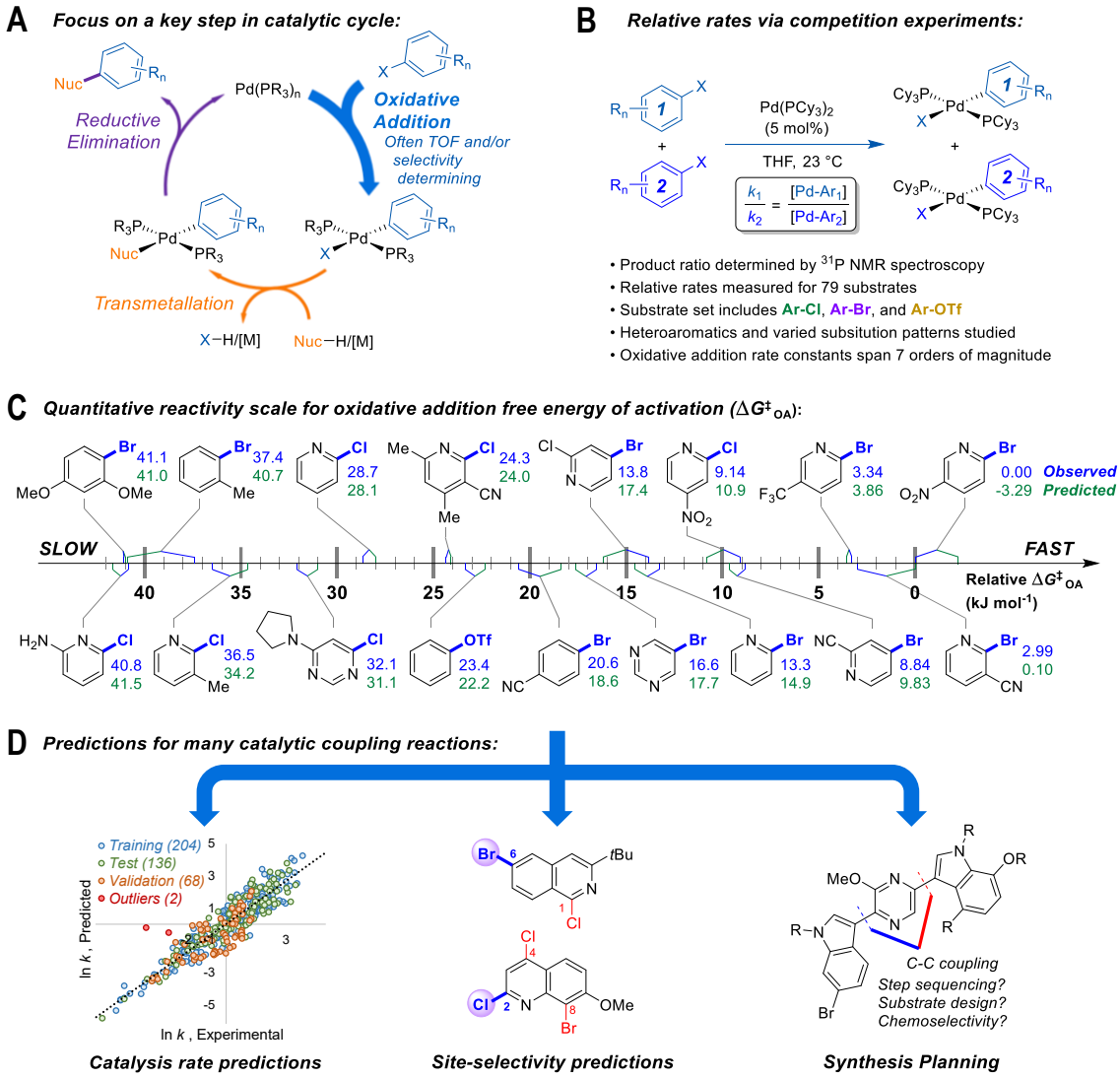


Figure 3.1. (A) Simplified cross-coupling mechanism, highlighting oxidative addition as the rate and/or selectivity determining step. **(B)** Competition experiment approach to determining relative rates of oxidative addition by quantifying ratios of Pd(II) products via ^{31}P NMR spectroscopy. **(C)** Relative reactivity scale for oxidative addition to $Pd(PCy_3)_2$ for selected substrates; observed $\Delta G^{\ddagger}_{OA}$ for 2-bromo-5-nitropyridine set to 0 kJ mol^{-1}. **(D)** Utility of reactivity model in predictions for cross-coupling in synthesis.

3.3 Results and Discussion

3.3.1 Development of the Oxidative Addition Reactivity Model

A reliable and diverse training set of reaction data is fundamental to the success of any structure-reactivity predictive model. Using the competition experimentation approach introduced in Chapter 2, I determined relative rates for 79 (hetero)aryl chlorides, bromides, and triflates undergoing oxidative addition with $Pd(PCy_3)_2$ in THF by measuring the Pd(II) product ratio by ^{31}P NMR spectroscopy (Fig. 3.1B). These 79 experimental $\Delta G^{\ddagger}_{OA}$ values represent a unified reactivity scale that spans more than 7 orders of magnitude in rate, containing substrates with a wide variety of steric and electronic parameters, as well as the three leaving groups (-Cl, -Br and triflate). The most reactive substrate, 2-chloro-5-nitropyridine, is the zero point for this scale ($\Delta G^{\ddagger}_{OA}$ set to 0 kJ mol^{-1}), and all other $\Delta G^{\ddagger}_{OA}$ values are given relative to this substrate (Fig 3.1C).

Before using these relative rates to build predictive models, I carried out several checks on these data. First, I constructed Hammett plots for five sets of substrates, obtaining ρ values consistent with previous reports[34] (Figs. B31 to B37, Appendix B). Next, I ruled out the possibility of reversible oxidative addition leading to thermodynamic product ratios. I selected 6 pairs of substrates leading to Pd(II) product ratios across a range from 3:1 to 30:1 in competition experiments. For each pair of substrates, I mixed equimolar amounts of the two Pd(II) products from individual oxidative addition reactions and recorded the product ratio after 24 hrs; in every case, the initial 1:1 mole ratio was retained, indicating no reversibility under these conditions. Finally, I isolated and characterized six new representative Pd(II) oxidative addition complexes to confirm their structures as *trans*-$Pd(PCy_3)_2(Ar)(X)$ by NMR and HRMS analysis (compounds P1-P6, Figs. B7-B30, Appendix B).

The mechanistic features of oxidative addition to Pd(0) informed the selection of mechanistically meaningful descriptors (Fig. 3.2A).[30] Mechanisms for aryl halide oxidative addition to Pd(0) have been extensively studied both computationally[35–38] and experimentally,[34,39–42] and are generally proposed to proceed via initial coordination of the aromatic π-system to Pd. The aromatic ring can coordinate to

Pd(0) via either a C=C bond or C=N bond (for heterocycles) to form a π-complex intermediate, in which the charge distribution is influenced by the polarization of the coordinated bond. From the intermediate structure, two types of oxidative addition transition state have been proposed: a 3-centered, relatively non-polar TS involving simultaneous Pd–C and Pd–X bond formation, and a polarized TS with C–X bond cleavage occurring alongside Pd–C bond formation; this latter pathway shares some features with the S_NAr mechanism.[34,40] While other coordination modes between the (hetero)aryl substrate and Pd are possible (more discussion in Chapter 4), here the analysis is restricted to the two types of π-complex intermediate that would immediately precede oxidative addition.

Based on the structural and electronic features of the transition states, I built the initial reactivity model for oxidative addition of (hetero)aryl–X (X = Cl, Br) using the following simple molecular descriptors: the average electrostatic potentials (*ESP*) as electronic descriptors for specific atoms;[31,43–47] A-values as steric descriptors;[48] and the intrinsic bond strength index (*IBSI*) as a bond energy descriptor (Fig. 3.2B).[49] The MLR model shown in Fig. 3.2C reveals a strong correlation between the measured and predicted $\Delta G^{\ddagger}_{OA}$ across this substrate library.

The model incorporates two *ESP* values: one for the carbon undergoing substitution (ESP_1), the inclusion of which is guided by chemical intuition; and a second for an atom adjacent to the reactive site (ESP_2), which is crucial to making predictions for substrates with different ring structures. For 2-halopyridines, ESP_2 corresponds to the nitrogen atom. For aryl bromides with inequivalent adjacent carbons, ESP_2 is the value of the more electronegative carbon. This is the site with the most negative *ESP* value, leading to a smaller $\Delta G^{\ddagger}_{OA}$ from the MLR equation and therefore faster oxidative addition (due to the positive sign of the ESP_2 coefficient in that equation).

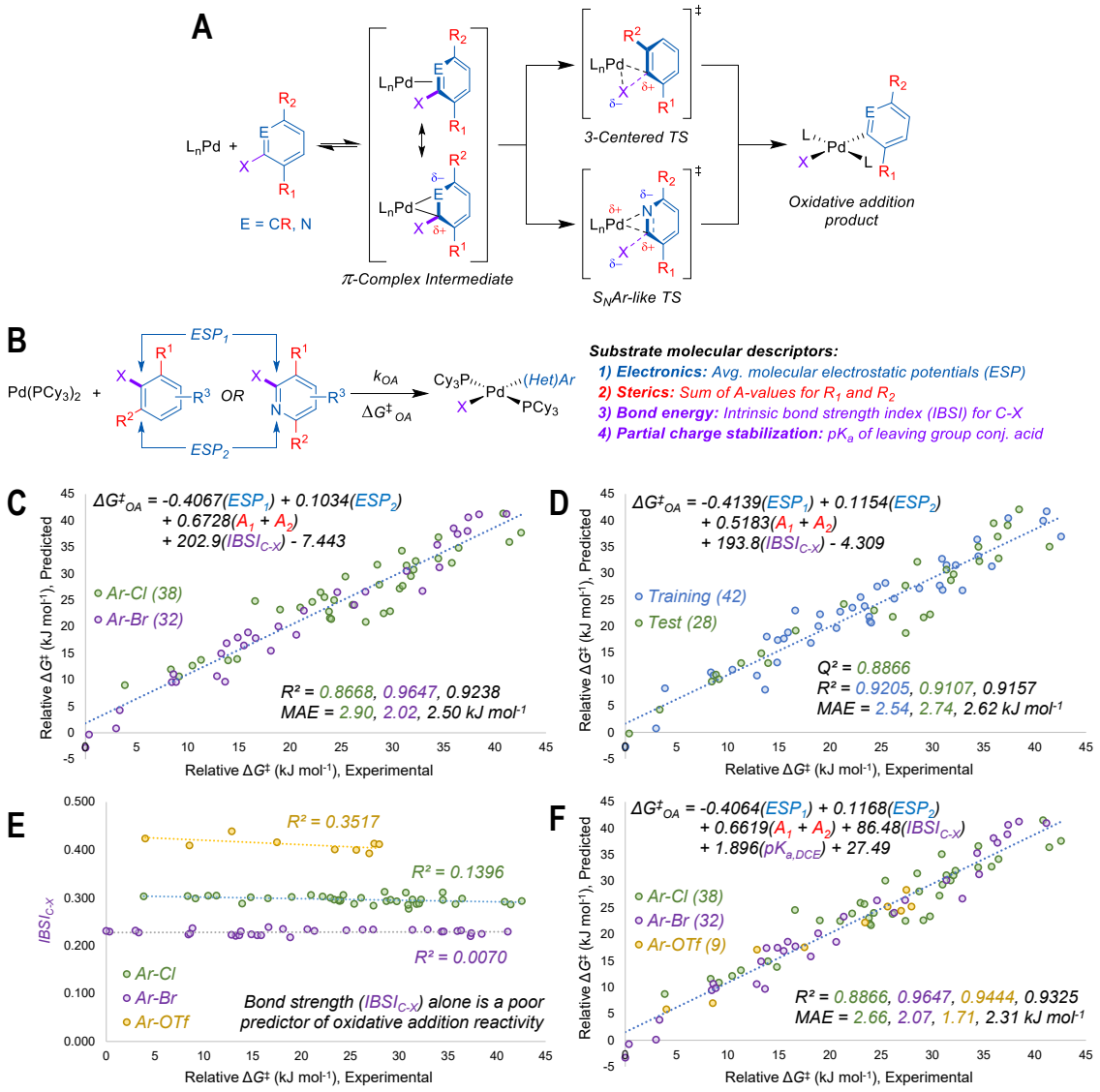


Figure 3.2. Design and performance of a quantitative reactivity model for oxidative addition to Pd(0). **(A)** General mechanism for oxidative addition to $L_nPd(0)$, with π-complex intermediate preceding either Pd insertion into C–X bond, or an S_NAr-like displacement of X. **(B)** Molecular descriptors used to model oxidative addition reactivity as a function of substrate structure. **(C)** Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 70 Ar–Cl and Ar–Br substrates in THF, including all data points in regression analysis. **(D)** Representative multivariate linear regression model generated using a 60/40 training/test split. **(E)** Univariate plot of $IBSI_{C-X}$ versus $\Delta G^{\ddagger}_{OA}$ for Ar–Cl, Ar–Br, and Ar–OTf, revealing that bond strength is poorly correlated to $\Delta G^{\ddagger}_{OA}$ within each leaving group set. **(F)** Unified linear regression model of $\Delta G^{\ddagger}_{OA}$ for Ar–Cl, Ar–Br, and Ar–OTf substrates in THF, including all data points in regression analysis. MAE = mean absolute error. Colour-coding on R^2, Q^2, and MAE values corresponds to the matching data subset, values in black are for all data.

The model also incorporates two substituent A-values for groups R_1 and R_2 to account for the steric effects on the oxidative addition rate. While the steric effect of groups *ortho* to the reactive site is intuitive, the effect of R_2 for 2-halopyridines is not initially obvious; however, from experimental observations, the steric effect of R_2 is approximately equal to that of R_1 throughout the 2-halopyridine library. As these two terms have very similar coefficients when treated separately during multivariate linear regression analysis, I opted to use the summed value (A_1+A_2) as a single descriptor. Finally, I incorporated the *IBSI* to create a unified model applicable to both (hetero)aryl chlorides and bromides.

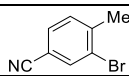
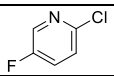
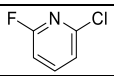
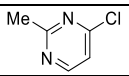
I evaluated the robustness of this linear model by regression analysis of five random 60/40 training/test splits, two structural splits and two rate splits, and comparing the mean absolute errors (MAE) and predictive squared correlation coefficients (Q^2)[50]; one example of a random 60/40 split is shown in Fig. 3.2D and the remainder in Figs. B60 to B63 (Appendix B). All of these analyses give similar linear equations and excellent predictive accuracy on the test set, with MAEs from 2.2-3.0 kJ mol^{-1} and Q^2 from 0.86-0.93. To validate that all the four molecular descriptors (*ESP_1, ESP_2, IBSI, A_1+A_2*) are contributory and necessary to achieve an excellent-performed predictive model, I evaluated alternative models from univariate regression and multivariate regression using combinations of fewer descriptors (Table B8, Appendix B); all of these simpler models lead to diminished predictive ability, giving significantly worse performance statistics.

Initial attempts to incorporate aryl triflate electrophiles into this initial four-descriptor model (Fig. 3.2C) were unsuccessful, leading to poor correlations and inaccurate predictions. I attribute this to C–X bond strength being an insufficient descriptor to differentiate between the reactivity of various leaving groups. Bond strength arguments, for example using calculated BDEs, are often used to rationalize relative oxidative addition reactivity for different substrates, such as for site selectivity in multihalogenated systems.[51–53] In the initial model (Fig. 3.2C), the *IBSI* term is essentially a step function: there are relatively narrow value distributions within Ar–Br (0.23 ± 0.01) and Ar–Cl (0.29 ± 0.02) data sets, but a large gap

between those data sets that reflects the weaker C–Br bond. Strikingly, there is zero correlation between *IBSI* and $\Delta G^{\ddagger}_{OA}$, either within each electrophile class or across the entire data set (Fig. 3.2E). Furthermore, the bond strength order (C–O > C–Cl > C–Br) is inconsistent with the observed reactivity order (Ar–OTf > Ar–Br > Ar–Cl, if Ar is identical) in oxidative addition. Clearly, bond strength on its own is a poor predictor of oxidative addition reactivity when comparing two substrates with either the same[54] or different leaving groups. Note that the ESP_1 values of aryl triflates are much more positive than the identically substituted aryl bromides, which would lead to faster reaction rates; however, this *ESP* argument alone is insufficient to overcome the two-orders-of-magnitude difference in oxidative addition rate between Ar–OTf and Ar–Br (where Ar is identical).

To account for additional transition state stabilization by the leaving group itself, which builds up negative charge as the C–C bond is broken (Fig. 3.2A), I used a simple descriptor of anion stability: the pK_a of the leaving group's conjugate acid. Adding the pK_a values for HOTf, HBr and HCl (-11.3, -4.4 and 0.2 respectively, previously measured in DCE as a non-polar solvent[55]) results in a unified predictive model (Fig. 3.2F). This model provides, for the first time, a reliable method to quantitatively evaluate the relative reactivity of a hypothetical multisubstituted (hetero)aryl triflate, bromide or chloride toward oxidative addition with a Pd(0) complex without the need of transition state analysis. Cross-validation with five random 60/40 training/test splits gave excellent agreement between experimental and predicted values (Figs. B69 to B73, Appendix B). Of the 79 substrates, there are only four examples identified as predicted outliers (standardized residual ≥ 2, Table 3.1), and there are no obvious features of these outliers that would point to systematic prediction errors for a particular substructure.

Table 3.1. Identified outliers from the regression model in Fig. 3.2F.

$\Delta G^{\ddagger}_{OA}$ (kJ mol^{-1})				
Observed	32.92	29.76	29.12	16.57
Predicted	26.74	23.29	22.67	24.59
Residual	6.18	6.37	6.45	-8.02

Finally, I examined the relative contribution of each descriptor to the prediction accuracy of the two models (Figs. 3.2C and 3.2F) by performing MLR using normalized descriptors by min-max rescaling. The coefficients of the normalized models and their relative contributions are shown in Table 3.2. In both models, ESP_1 is the top-ranked contributor, followed by ESP_2 at roughly half the weighing. In contrast, steric effects account for <10% of the predicted $\Delta G^{\ddagger}_{OA}$. The C–X bond strength by *IBSI* is also a relatively small contributor (16-20%), which is consistent with the observed poor correlation between *IBSI* itself and $\Delta G^{\ddagger}_{OA}$ (Fig. 3.2E). In fact, the p*Ka* of HX is just as consequential for predicting $\Delta G^{\ddagger}_{OA}$ as the *IBSI* in the unified model from Fig. 3.2F (X = Cl, Br, OTf).

Table 3.2. Linear regression coefficients and % contribution to predicted $\Delta G^{\ddagger OA}$ for min/max scaled descriptors.

Model	ESP_1	ESP_2	(A_1+A_2)	*IBSI*	pK_a	Int.
Fig.	-48.36	20.97	7.037	19.25	-	30.28
3.2C	51%	22%	7%	20%	-	-
Fig.	-51.65	23.68	6.924	19.14	21.81	16.94
3.2F	42%	19%	6%	16%	18%	-

3.3.2 Mechanistic Aspects of Oxidative Addition Linked to Molecular Descriptors

The predictive power of this reactivity model is a direct result of its mechanistic foundation. ESP_1 is related to the electrophilicity of the reactive carbon, reflecting the degree of partial positive charge in the π-complex intermediate and transition state: larger positive ESP_1 values lead to lower activation energy ($\Delta G^{\ddagger}_{OA}$) and thus faster oxidative addition. ESP_1 and ESP_2 together reflect the polarization degree of the π bond (C=C or C=N), where a more electronegative atom (larger negative ESP_2) leads to a more polarized π bond and faster oxidative addition. DFT calculations reveal that the *ESP* at Pd in the intermediate/transition state also correlates with oxidative addition rate. As shown in Fig. 3.3A, there is a linear correlation between ESP_{Pd} for the 11 calculated π-complexes and $\Delta G^{\ddagger}_{OA}$, revealing the significant effect that substrate-catalyst bonding has on the electronic structure, and therefore the reactivity of Pd. Importantly, this linear relationship between ESP_{Pd} and $\Delta G^{\ddagger}_{OA}$ is

evidently captured by the ESP_1 and ESP_2 values, which combined provide a reliable and easy-to-calculate alternative to using ESP_{Pd} in predicting oxidative addition rates.

Analyzing how the substrate binds (which will be discussed in greater detail in Chapter 4) to the Pd(0) complex also helps to rationalize the observed equal steric effect from R_1 and R_2 to the $\Delta G^{\ddagger}_{OA}$ for the 2-halopyridine substrate series. Larger substituents in these positions destabilize the π-complex intermediate and oxidative addition TS through steric repulsion between the bound substrate and the ancillary phosphine ligands. Because the Pd center coordinates to the C=N bond to form the π-complex intermediate, R_1 and R_2 will occupy roughly equivalent positions with respect to the phosphine (Fig. 3.3B). Importantly, this mechanistic feature of the equal steric effects from R_1 and R_2 flows directly from the quantitative reactivity model.

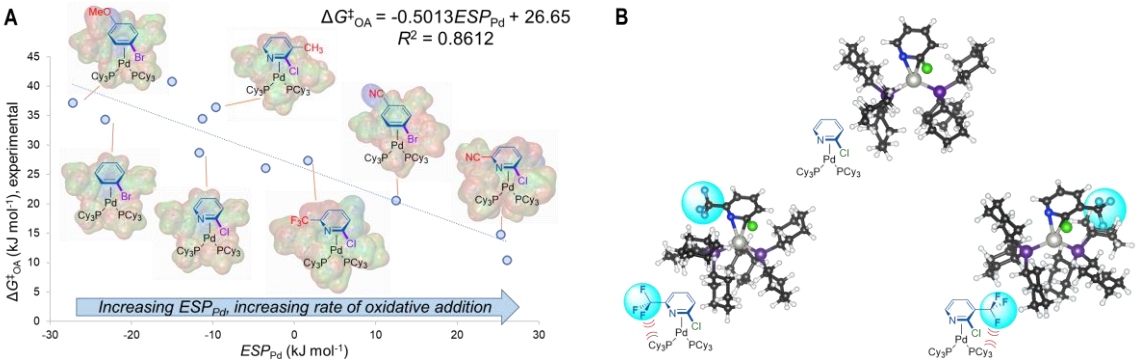


Figure 3.3. Electronic and steric features of oxidative addition. **(A)** ESP_{Pd} for calculated π-complex intermediate structures correlates with oxidative addition rates; structures for 7 of 11 examples shown; electrostatic potential maps for each intermediate are overlaid onto the line structures. **(B)** Calculated structures of π-complex intermediates reveal how steric strain induced by R_1 and R_2 (here, $-CF_3$ groups) in 2-halopyridines affect oxidative addition reactivity in equal proportions.

3.3.3 Case Study #1: Predicting Reaction Rates

To test the hypothesis that an oxidative addition reactivity model is able to predict reaction outcomes of various Pd-catalyzed cross-couplings, I carried out three case studies for $\Delta G^{\ddagger}_{OA}$ prediction on external reaction data. The first case involves structure-reactivity modeling for Sonogashira couplings using 410 literature reported initial rates (29 aryl bromides and 17 catalyst, Fig. 3.4A).[56,57] I first predicted $\Delta G^{\ddagger}_{OA}$ for each of the 29 aryl bromides in the reaction sets, using the equation from Fig. 3.2C (the simpler model when considering only halide-based electrophiles); among the 29 aryl bromides, only 9 of those are included in the training set. Remarkably, the predicted $\Delta G^{\ddagger}_{OA}$ values are linearly correlated with the experimental ln(k) values for all 17 investigated phosphine ligands (3 correlations are shown in Fig. 3.4B and the rest in Figs. B74 and B75, Appendix B). These are good-to-excellent correlations (R^2 from 0.79 to 0.93) despite the fact that these Sonogashira reactions are conducted under different conditions (higher temperature and different solvent) than the oxidative addition experiments, and the fact that the predicted $\Delta G^{\ddagger}_{OA}$ is derived from a reactivity model trained only using PCy_3.

Following the excellent performance of the model in predicting relative reaction rates for substrates, I then combined the predicted $\Delta G^{\ddagger}_{OA}$ with descriptors for the phosphine ligands to assemble a unified MLR model to accurately predict ln(k) for the entire array of 410 Sonogashira reactions. Using two simple descriptors for the free phosphines – the average *ESP* at phosphorus and the percent buried volume ($\%V_{bur}$) at phosphorus[58] – in combination with the predicted $\Delta G^{\ddagger}_{OA}$ for aryl bromides from the equation in Fig. 3.4C, I obtained a three-descriptor model able to predict initial rates spanning 10 orders of magnitude. These phosphine descriptors outperform the analogous descriptors calculated for the corresponding mono or bis(phosphine) Pd(0) complexes. The training and test sets used to build this model are from a random 60/40 split of the substrate set #1, which contains 20 3-substituted and 4-substituted bromobenzenes (Fig. 3.4A). This substrate set was assembled by Plenio and co-workers to assess electronic effects on Sonogashira rates.

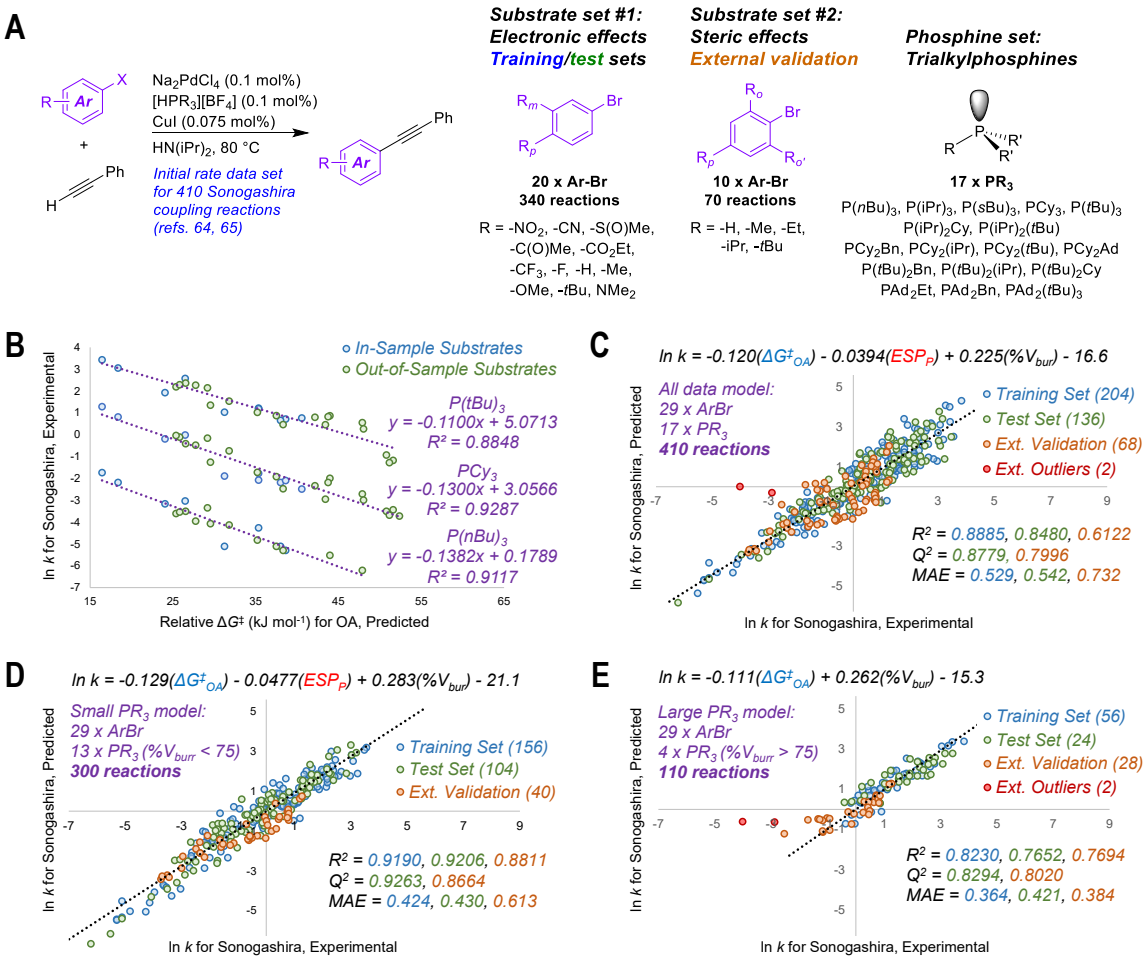


Figure 3.4. Translating oxidative addition predictions to quantitative models of catalytic reactivity. **(A)** General reaction scheme and chemical space explored for 410 Sonogashira reactions, with two distinct substrate sets; initial rates determined previously.[56,57] **(B)** Univariate linear correlations between predicted $\Delta G^{\ddagger}_{OA}$ for oxidative addition to $Pd(PCy_3)_2$ and ln *k* for Sonogashira coupling with three phosphines; out-of-model substrates are Ar–Br molecules not included in $\Delta G^{\ddagger}_{OA}$ training set. **(C)** Unified three-descriptor model for predicting ln *k* for the entire set of 410 reactions (29 substrates, 17 ligands), with data partitioned into training (60% of set #1), test (40% of set #1), and external validation (set #2); two external outlier points (red) are not included in the external validation statistics. **(D)** Subset of the model with 13 "small" phosphines ($\%V_{bur} < 75$). (E) Subset of the model with 4 "large" phosphines ($\%V_{bur} > 75$); two external outlier points (red) are not included in the external validation statistics. MAE = mean absolute error. Colour-coding on R^2, Q^2, and MAE values corresponds to the matching data subset, values in black are for all data.

To challenge the model, I reserved substrate set #2 as an external validation set. This set contains 9 *ortho*-substituted bromobenzenes, which focuses on steric effects. Despite the fact that the Sonogashira rate prediction model is trained only with substrate set #1, which contains no *ortho*-substituted substrates and therefore no information about steric effects on reaction rate, the model is still able to predict ln(*k*) for substrate set #2 with an MAE of 0.732 (Q^2 = 0.7996). Only two reactions (highlighted as red points in Figs. 3.4C and 3.4E) are identified as significant outliers: 2,4,6-triisopropylphenylbromide with P(*t*Bu)$_3$ and PAd$_2$(*t*Bu) (the most sterically hindered substrate with the two largest ligands), which are much slower than the model predicts.

The ability of the Sonogashira model to predict ln(*k*) of this external substrate class is due to the fact that steric effects are already accounted for in the predicted $\Delta G^{\ddagger}_{OA}$ from the oxidative addition reactivity model. In other words, the mechanistically relevant molecular properties of substrates – electronics, sterics, and C–X bond strength – are all encoded within the predicted $\Delta G^{\ddagger}_{OA}$ values. As a result, applying it as a single, pre-weighted "super" descriptor for the substrate enables accurate prediction on catalytic reactivity even when the available training data is not comprehensive.

Improved prediction accuracy can be achieved by separating the 410-member dataset into two subsets based on phosphine ligand size. One set includes reactions using 13 phosphines with a $\%V_{bur} < 75$, and the other includes the 4 largest phosphines with a $\%V_{bur} > 75$. For the small phosphine set, a slight reweighting of the three descriptors leads to smaller MAEs and higher R^2 and Q^2 values for training, test, and external sets (Fig. 3.4D). For the large phosphine set, I used 70/30 random training/test split due to the smaller number of data points. The three-descriptor reactivity model gives smaller MAEs for the training and test sets while the contribution from the ESP_P descriptor is almost zero. Thus, a simpler, two-descriptor ($\Delta G^{\ddagger}_{OA}$ and $\%V_{bur}$) model is superior in making accurate predictions for the large phosphine set. This partition into small and large phosphines is consistent with prior studies that describe a change in mechanism, where oxidative addition occurs from

either bis(phosphine) or mono(phosphine) Pd(0) intermediates.[59–61] Notably, the predicted $\Delta G^{\ddagger}_{OA}$ is effective in both cases. This case study demonstrates that $\Delta G^{\ddagger}_{OA}$ can not only be used to predict reactivity of substrates with different substitution patterns, but also can be used together with catalyst-based descriptors to account for both substrate and catalyst effects on rates.

3.3.4 Case Study #2: Predicting Site Selectivity

The second case study tests the use of predicted $\Delta G^{\ddagger}_{OA}$ for site-selectivity predictions of multi-halogenated heteroarenes in Pd-catalyzed cross-couplings. All of the descriptors in the $\Delta G^{\ddagger}_{OA}$ prediction model are local rather than global, enabling it to predict distinct $\Delta G^{\ddagger}_{OA}$ values for each reactive site in a molecule. I assessed a series of substrates (Fig. 3.5) with reported experimental selectivities for Suzuki-Miyaura[51] and Buchwald-Hartwig[59–63] coupling. The model correctly identifies the major reactive site for 22 out of 24 cases, including many heterocycle classes (isoquinolines, diazines, and several 5-membered heterocycles) not included in the initial training set, as well as substrates with non-identical halides (-Cl vs. -Br).

The two exceptions are isothiazoles, for which the experimental major site is the C_5 (adjacent to sulfur) while the predicted major site is C_3 (adjacent to nitrogen). To further probe this, I evaluated the model's predicted $\Delta G^{\ddagger}_{OA}$ for sulfur containing electrophiles against experimental values using 2,5-dibromothiophene and 2-iodothiophene (the reactivity of (hetero)aryl iodides will be discussed in Section 3.4). For both substrates, the predicted rates are much slower than the observed rates from competition experiments, with prediction errors greater than 11 kJ mol^{-1}. This is much larger than the error of the most significant outlier in the original dataset (Table 3.1). These unsuccessful predictions for halo-thiophenes (and by extension, the isothiazoles) are likely due to a lack of sulfur-containing substrates in the training set, and a possible change in oxidative addition mechanism caused by sulfur coordination to palladium.[62,63] Further work is needed to assess this more completely.

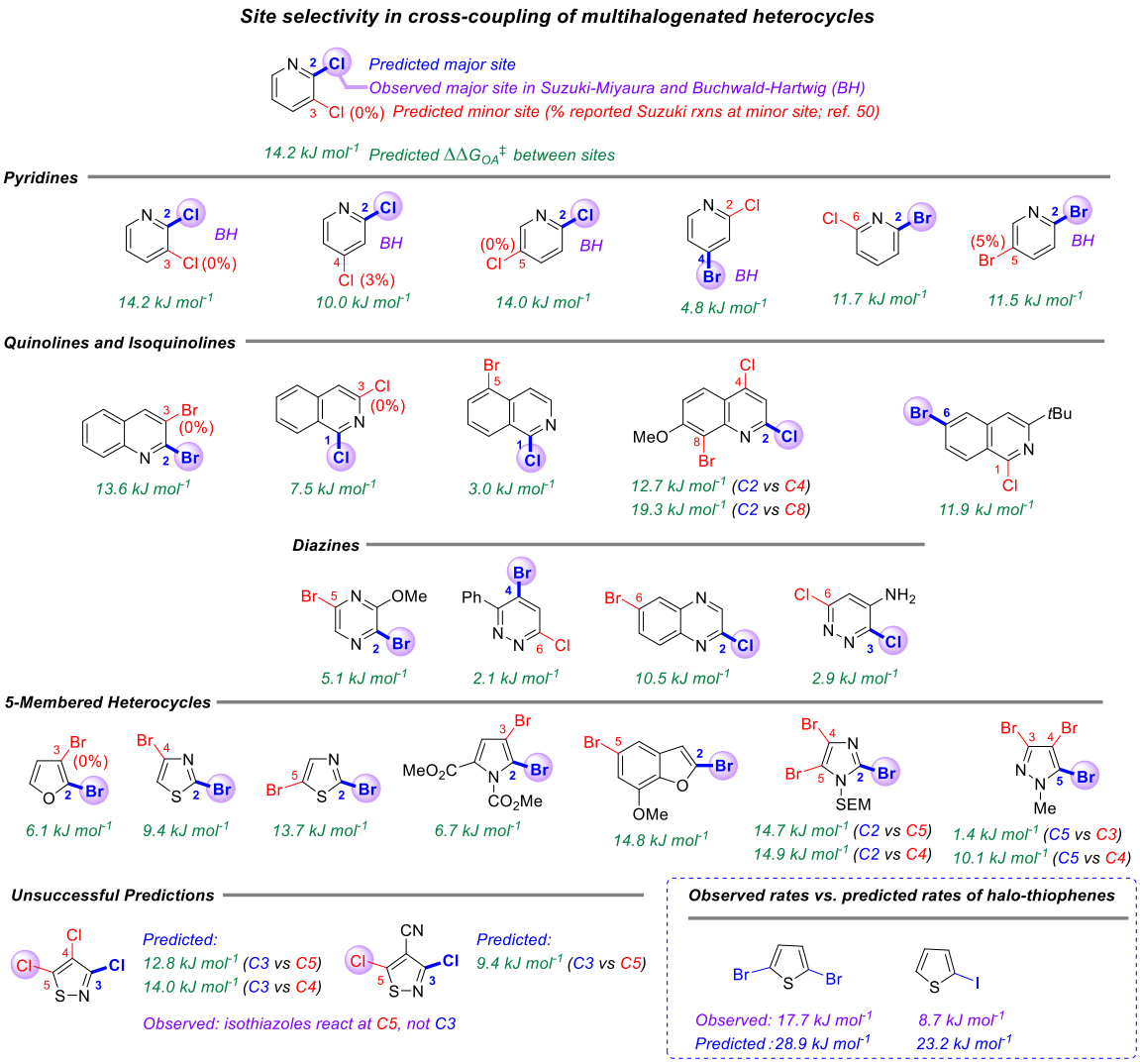


Figure 3.5. Predicted and reported selectivities for multihalogenated heterocycles in Suzuki-Miyaura and Buchwald-Hartwig cross-coupling reactions (examples of Buchwald-Hartwig substrates denoted with "BH"). Coloured labels on the heterocycles correspond to predicted major site (blue), predicted minor site (red, along with percentage of exceptions as reported in ref. [51]), and observed site (purple sphere). The magnitude of $\Delta\Delta G^{\ddagger}_{OA}$ between the two sites is given in green. Contents circled in dashed blue line are the predicted and observed oxidative addition rates from my experimental data.

In addition to identifying the major site of reactivity, these predictions are also quantitative (Fig. 3.6). $\Delta\Delta G^{\ddagger}_{OA}$ values here are the difference of $\Delta G^{\ddagger}_{OA}$ between the two C–X sites. For the Suzuki-Miyaura coupling of methyl 2,6-dichloronicotinate (Fig. 3.6A), the predicted $\Delta\Delta G^{\ddagger}_{OA}$ is 5.4 kJ mol^{-1}, favoring C_6 by ~7:1 at 65 °C. The

reported selectivity using $Pd(PPh_3)_4$ at 65 °C is 5:1 C_6 to C_3 ($\Delta\Delta G^{\ddagger}_{OA}$ = 4.5 kJ mol^{-1}).[64] The model predicts poor selectivity for 3,6-dichloro-4-methoxypyridazine, with a $\Delta\Delta G^{\ddagger}_{OA}$ of only 0.7 kJ mol^{-1}, favoring C_6 by ~1.3:1 at 100 °C. The observed selectivity for Suzuki-Miyaura coupling using $Pd(PPh_3)_4$ at 100 °C is 3:1 C_6 to C_3, ($\Delta\Delta G^{\ddagger}_{OA}$ = 3.4 kJ mol^{-1}).[65] For the Buchwald-Hartwig coupling of 2,4-dichloropyridine, the model predicts $\Delta\Delta G^{\ddagger}_{OA}$ = 10.0 kJ mol^{-1}, favoring C_2 by 25:1 at 100 °C; the reported selectivity of this substrate with a set of amine coupling partners using a Xantphos-based catalyst ranges from 20:1 to 50:1 C_2 to C_4 at 100 °C ($\Delta\Delta G^{\ddagger}_{OA}$ = 9.3-12.1 kJ mol^{-1}).[66]

Site selectivity is known to be influenced by reaction conditions such as catalyst and solvent.[51,67,68] The $\Delta G^{\ddagger}_{OA}$ prediction model is based on a simple monodentate phosphine (PCy_3) using non-polar solvents, which represents a standard combination of reaction conditions used in synthetic applications. To illustrate how the model could guide the development of chemo/regioselective coupling reactions, I have applied predictions to substrates known to have tunable selectivity (Fig. 3.6B). First, 3,5-dichloropyridazine is predicted to have good selectivity favoring C_3 with $\Delta\Delta G^{\ddagger}_{OA}$ = 10.6 kJ mol^{-1}, consistent with the observed excellent selectivity (C_3:C_5 > 99:1) using simple ligands (PPh_3, dppf); therefore, inverting this selectivity should require extensive catalyst/solvent screening. Researchers at Merck conducted high-throughput screening using a set of 20 phosphine ligands, and discovered that the very large Qphos ligand gives high C_5 selectivity.[69]

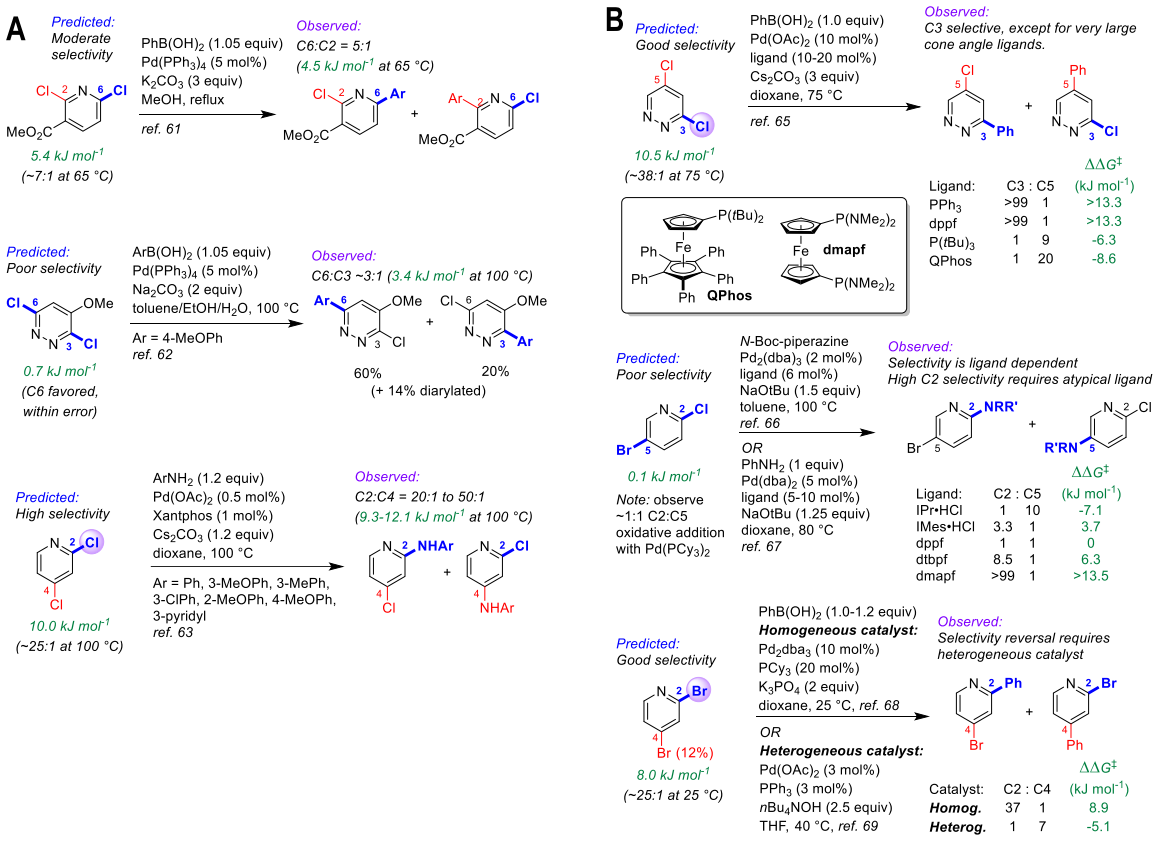


Figure 3.6. (A) Quantitative selectivity predictions for dihalogenated heterocycles with small-to-medium $\Delta\Delta G^{\ddagger}_{OA}$ between two sites, and observed product ratios. **(B)** Predictions for substrates with observed tunable selectivity, demonstrating that "simple" catalysts are quantitatively consistent with predicted selectivities; overriding predicted reactivity requires targeted screening and/or catalyst design. Coloured labels on the heterocycles correspond to predicted major site (blue), predicted minor site (red, along with percentage of exceptions as reported in ref. [51]), and observed site (purple sphere). The magnitude of $\Delta\Delta G^{\ddagger}_{OA}$ between the two sites is given in green.

Second, the predicted $\Delta\Delta G^{\ddagger}_{OA}$ for 2-chloro-5-bromopyridine is 0.2 kJ mol^{-1}, indicating effectively no selectivity; I confirmed this by oxidative addition of this substrate with $Pd(PCy_3)_2$ which gives 1:1 mixture of C_2 and C_5 products. Thus, the predicted $\Delta\Delta G^{\ddagger}_{OA}$ indicates that achieving good selectivity at one site over the other will require more complex systems. Consistent with this, two previous studies reported ligand-controlled reactivity: high C_5 selectivity can be achieved using simple ligands (Xantphos, dppf, P(*t*Bu)$_3$, Sphos, etc.); while high C_2 selectivity requires extensive screening effort and use of an unconventional diaminophosphine (dmapf).[70,71]

Finally, 2,4-dibromopyridine is predicted to have high selectivity at C_2 ($\Delta\Delta G^{\ddagger}_{OA}$ = 8.0 kJ mol^{-1}, C_2:C_4 ~25:1 at 25 °C), which is consistent with the previously reported selectivity from experiments using homogenous Pd catalysis.[72] A recent report revealed that a heterogenous, nanoparticle-based Pd catalyst is capable of inverting the selectivity to ~1:7 C_2:C_4.[73] As for the examples above, overriding the predicted selectivity requires a dramatically different catalyst system.

3.3.5 Case Study #3: Retrospective Synthesis Planning

As a final case study, I applied $\Delta G^{\ddagger}_{OA}$ predictions to two reported synthesis toward dragmacidin D (Fig. 3.7).[74–76] The general sequence relies on two regioselective cross-coupling reactions to a dihalogenated pyrazine core (Fig. 3.7A). Among the key design questions are how to achieve selective, sequential couplings, and how to ensure compatibility with the existing Ar–Br in indole 1 (which is present in the natural product).[76]

Two approaches to this problem have been reported. Garg, Sarpong, and Stoltz (Fig. 3.7B)[74] reported an approach using 2-iodo-5-chloro-3-methoxypyrazine to maximize site-selectivity, though they observed that the nature of the protecting group on the indole also has a significant effect on reactivity. With an *N*–Ts protecting group, bis(arylation) is readily achieved at elevated temperature, whereas with *N*–TIPS, the intermediate pyrazine is deactivated, preventing a second coupling. Applying $\Delta G^{\ddagger}_{OA}$ predictions to the two intermediates clearly reveals not only the existence of this remote electronic effect from the *N*–Ts protecting group, but the magnitude of this effect, leading to ~10-fold higher predicted reactivity between *N*–Ts and *N*–TIPS substrates. Thus, a subtle but important reactivity difference could be anticipated based on predicted $\Delta G^{\ddagger}_{OA}$ prior to experimental work, and help guide protecting group selection.

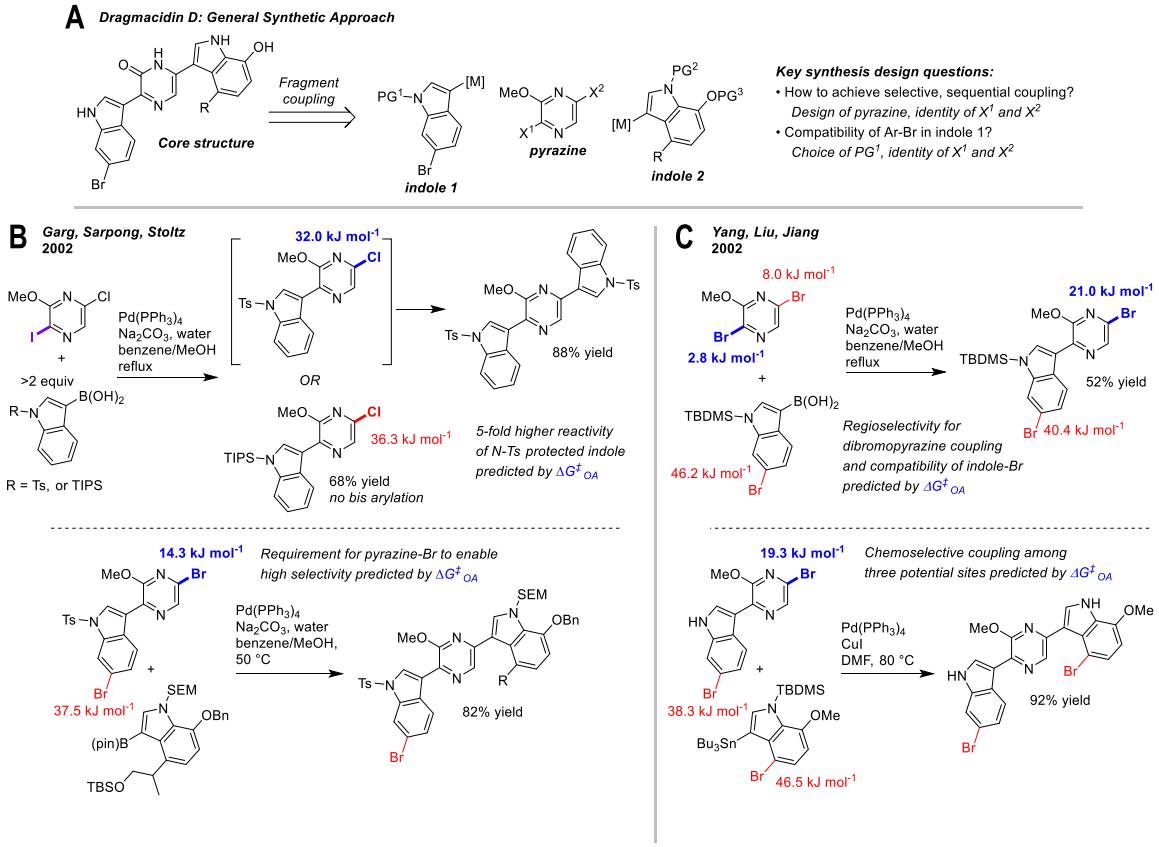


Figure 3.7. Retrospective analysis of applying $\Delta G^{\ddagger}_{OA}$ predictions to synthesis design: dragmacidin D. **(A)** Retrosynthesis of core structure, involving selective fragment coupling to a dihalogenated pyrazine. **(B)** Approach involving differential halogenation, tosylate protecting group on indole 1, and switch from 5-chloro to 5-bromopyrazine intermediates for selective coupling.[74] **(C)** Approach involving regioselective coupling to dibromopyrazine, TBS protecting group on indole 1, and regioselective Stille coupling.[75]

The second coupling required a switch to the 2-iodo-5-bromo-3-methoxypyrazine to maximize site-selectivity, which again can be predicted by the model. The $\Delta\Delta G^{\ddagger}_{OA}$ for the two Ar–Br positions in the pyrazine-indole intermediate used in the synthesis is 23.2 kJ mol^{-1}, consistent with the observed high selectivity. Prediction for the corresponding 5-chloropyrazine intermediate gives a $\Delta\Delta G^{\ddagger}_{OA}$ of <5 kJ mol^{-1} (~6:1 at 50 °C), which while still selective for the desired position, would likely lead to overreaction at the indole Ar–Br position and yield loss. Again, this type of prediction could help to guide synthetic design and subsequent experimental investigations.

Yang, Liu, and Jiang reported a similar approach (Fig. 3.7C)[75] that employed 2,5-dibromo-3-methoxypyrazine as the pyrazine building block; a series of model studies established that regioselective C_2-coupling is possible. Despite the fact that both sites appear very similar, the model predicts that C_2 is the major site, though with moderate selectivity ($\Delta\Delta G^{\ddagger}_{OA}$ = 5.1 kJ mol^{-1}, 6:1 at 80 °C); this lower selectivity may be reflected in the lower isolated yield relative to the differential halogenation strategy of Garg *et al*. In the subsequent Stille coupling, Yang *et al*. use the unprotected indole derivative, which the model predicts is ~10-fold less reactive than the corresponding *N*–Ts substrate; nevertheless, the desired reaction site is heavily favored relative to the other two Ar–Br positions, consistent with the observed regioselectivity.

3.4 Recent Progress on Pd Oxidative Addition Prediction

3.4.1 Reactivity Prediction of Aryl Chlorides in Pd Oxidative Addition

This section describes further exploration of the structure-reactivity relationship in Pd oxidative addition for new substrate classes beyond those initially published in 2022. First, I extended the applicability of the reactivity model from Fig. 3.2F to aryl chlorides. The aryl chloride substrate set contains 10 4-subsituted-chlorobenzenes, with the observed relative rates from competition experiments spanning 3 orders of magnitude (Fig. 3.8A). Overall, aryl chlorides are the least reactive species among the substrates in the training data set. The Hammett plot for these 10 substituted aryl chlorides gives an excellent linear fit to the σ_{para}^{-} constants, with ρ = +1.9. This ρ value is much smaller than that obtained from prior work by Portnoy and Milstein,[40] who studied the oxidative addition mechanism of aryl chlorides to Pd(dippp)$_2$. They reported a reaction pathway where the bisligated Pd undergoes reversible dissociation of one dippp ligand, followed by oxidative addition of aryl chlorides to the monoligated (dippp)Pd(0) complex via a polarized nucleophilic displacement (S_NAr-like) mechanism. Key evidence for this mechanism is the large positive Hammett ρ value (ρ = +5.2) they obtained from a similar set of 4-substituted chlorobenzenes. With Pd(PCy$_3$)$_2$, the smaller ρ values I obtained for aryl chlorides (ρ = +2.0) and aryl bromides (ρ = +2.3) suggest that the

oxidative addition of these substrates may occur via a non-polar concerted 3-centered mechanism.

These 10 aryl chlorides comprise an external test set, and I examined the ability of the model from Fig. 3.2F to make predictions for this new substrate class. As shown in Fig. 3.8B, the model predicts the relative rates of aryl chlorides with high accuracy ($Q^2 = 0.94$ and MAE = 2.27 kJ mol^{-1} for the test set), despite the fact that the $\Delta G^{\ddagger}_{OA}$ values fall far beyond the range of those covered by the training set. Including the relative rates of the 10 aryl chlorides as part of the training set results in the model shown in Fig. 3.8C. The new model is nearly identical to the one from Fig. 3.2F, with a slightly lower MAE (2.54 kJ mol^{-1} vs. 2.66 kJ mol^{-1}) for the (hetero)aryl chlorides. The successful inclusion of these aryl chlorides extends the reactivity scale of the oxidative addition dataset, which now spans 9 orders of magnitude in reaction rate. In addition, it shows that the model is compatible with unactivated (i.e. slow to react) substrates in oxidative addition.

Hartwig and coworkers[42] reported that oxidative addition mechanism of aryl halides to $L_2Pd(0)$ complexes is more dependent on the halide type other than the ligand type, and their kinetic study suggests that aryl bromides and iodides oxidative addition to $Pd(PCy_3)_2$ is via a 14-electron (PdL_2) pathway while aryl chlorides is via a 12-electron (PdL) pathway. The fact that aryl chlorides along with substrates that presumably react via 14-electron pathway all fit well to the same model brings new mechanistic insights. Hartwig kineitic analysis was conducted at 70 °C while all my reaction rates were collected at room temperature; one possibility would be that the oxidative addition mechanism is also dependent on the reaction condition: high temperature promotes dissociation of one PCy_3 ligand which enables the 12-electron pathway, while at low temperature aryl chlorides react directly with the 14-electron $Pd(PCy_3)_2$. My study shows that same as other substrate types, oxidative addition rates of aryl chlorides are linearly related to the substrates' electronic feature. Further investigation is needed for study of the electronic effect of the ligands and the steric effect of both the ligands and substrates to better understand the oxidative addition mechanism of aryl chlorides.

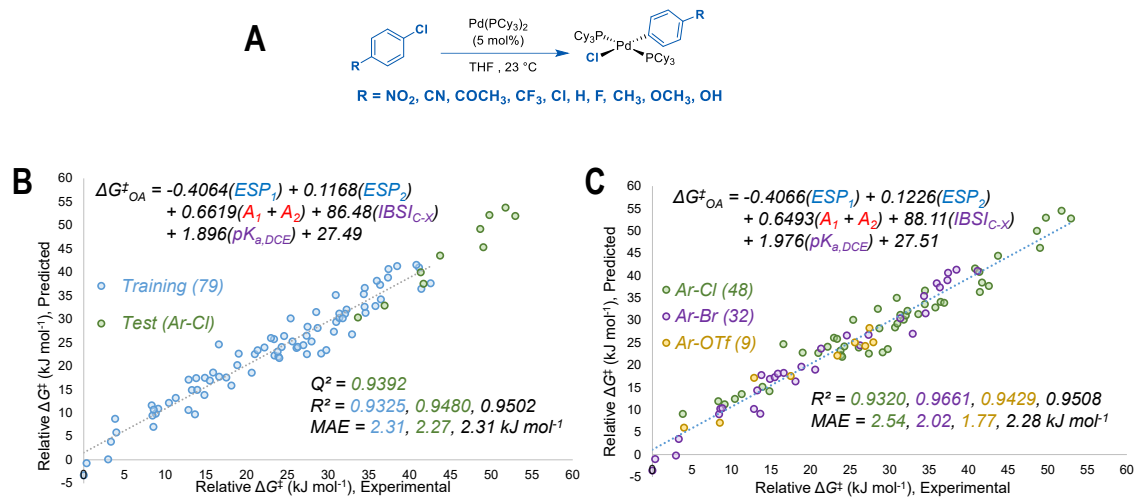


Figure 3.8. Performance of the quantitative model for oxidative addition of aryl chlorides to Pd(0). **(A)** Structures of the 10 4-substituted aryl chlorides in my reaction dataset. **(B)** Multivariate linear regression model using all data points except for the 10 aryl chlorides for training set, and the 10 aryl chlorides for external test set. **(C)** Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 89 Ar–Cl, Ar–Br and Ar–OTf substrates in THF, including all data points in regression analysis. MAE = mean absolute error. Colour-coding on R^2, Q^2, and MAE values corresponds to the matching data subset, values in black are for all data.

3.4.2 A Four-descriptor Model for Halobenzene Oxidative Addition

Depending on the nature of the aromatic ring, the electrophiles in the dataset can be classified into two subsets: 2-halopyridines, which have a nitrogen next to the reactive site; and halobenzenes, which include non-heteroaryl (*pseudo*)halides and 3-, 4-, or 5-halopyridines where the adjacent atom to the reactive site is a carbon. The difference in the putative transition state structures (Fig. 3.2A) between these two types of substrates leads us to think about the importance of the adjacent atom to the reaction rates in each case. In a nucleophilic displacement (S_N-type) mechanism, palladium-nitrogen coordination occurs at the formation of both the π-complex intermediate and transition state. The negative charge in the pyridyl ring is stabilized by the electronegative nitrogen, which helps to lower the energy of this polarized transition state. In contrast, in a 3-centered mechanism, the adjacent carbon is only involved in forming the π-complex, where the palladium center coordinates to the C=C bond.

While the model from Fig. 3.8C is able to make accurate prediction for both types of substrate, with MAEs of 2.5 kJ mol^{-1} for the 2-halopyridines set and 2.0 kJ mol^{-1} for the halobenzenes set, I considered the possibility that a simpler model could be applicable to the halobenzene subset. This substrate set contains 43 electrophiles, from one of the most reactive (electron-deficient aryl triflates) to the least reactive (electron-rich aryl chlorides), with reaction rates spanning 9 orders of magnitude (Fig. 3.9A).

A four-descriptor (*ESP*$_1$, *A*$_1$, *IBSI* and p*K*$_a$) model seemed feasible, based on the fact that the adjacent atom is of less relevance in the 3-centered mechanism. This MLR model (Fig. 3.9B) exhibits a strong linear relationship between the experimental and predicted $\Delta G^{\ddagger}_{OA}$ ($R^2 = 0.96$, MAE = 2.0 kJ mol^{-1}); the residual plot (Fig. 3.9C) shows that none of the substrates is classified as predicted outlier (standardized residual < 2 in all cases). One example of the five 70/30 random splits shown in Fig. 3.9D (the remainder in Figs. B81A – B81D, Appendix B) gives a similar linear equation and excellent predictive power on the test set, with an MAE of 2.7 kJ mol^{-1} and a Q^2 of 0.91.

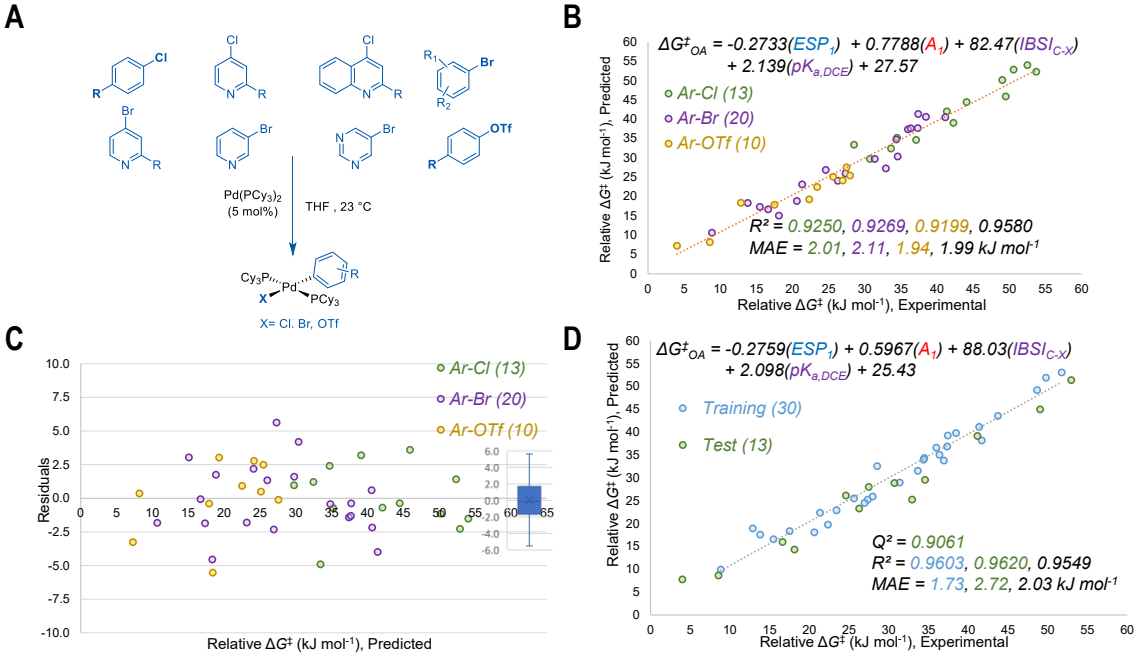


Figure 3.9. Performance of the four-descriptor (*ESP*$_1$, *A*$_1$, *IBSI*, p*K*$_a$) model for oxidative addition of the halobenzene subset. **(A)** Structural features included in the 43-member halo-benzene subset.

(B) Multivariate linear regression model using all the 43 data points training set. **(C)** Residual plot for the predicted $\Delta G^{\ddagger}_{OA}$ using the equation from Fig. 3.9B. **(D)** Multivariate linear regression model generated using a 70/30 training/test split. MAE = mean absolute error. Colour-coding on R^2, Q^2, and MAE values corresponds to the matching data subset, values in black are for all data.

3.4.3 Reactivity of 2-Pyridyl Triflate in Oxidative Addition to Pd(0)

2-Pyridyl triflate, with a nitrogen next to the reactive site, is predicted to be an extremely reactive species using the equation from Fig. 3.8C. Surprisingly, its experimental rate, measured and confirmed by three competition experiments with different competing partners, is much slower than what the general model predicts ($\Delta G^{\ddagger}_{OA\_predicted}$ = 2.2 kJ mol^{-1} versus $\Delta G^{\ddagger}_{OA\_observed}$ = 12.9 kJ mol^{-1}, Fig. 3.10A). However, this substrate is no longer an outlier when using the equation from the four-descriptor model in Fig. 3.9B ($\Delta G^{\ddagger}_{OA_predicted}$ = 15.0 kJ mol^{-1} versus $\Delta G^{\ddagger}_{OA\_observed}$ = 12.9 kJ mol^{-1}, Fig. 3.10B). The fact that the four-descriptor model accurately predicts the oxidative addition rate for this 2-halopyridine-type substrate prompted an investigation into possible mechanistic causes for this counterintuitive result.

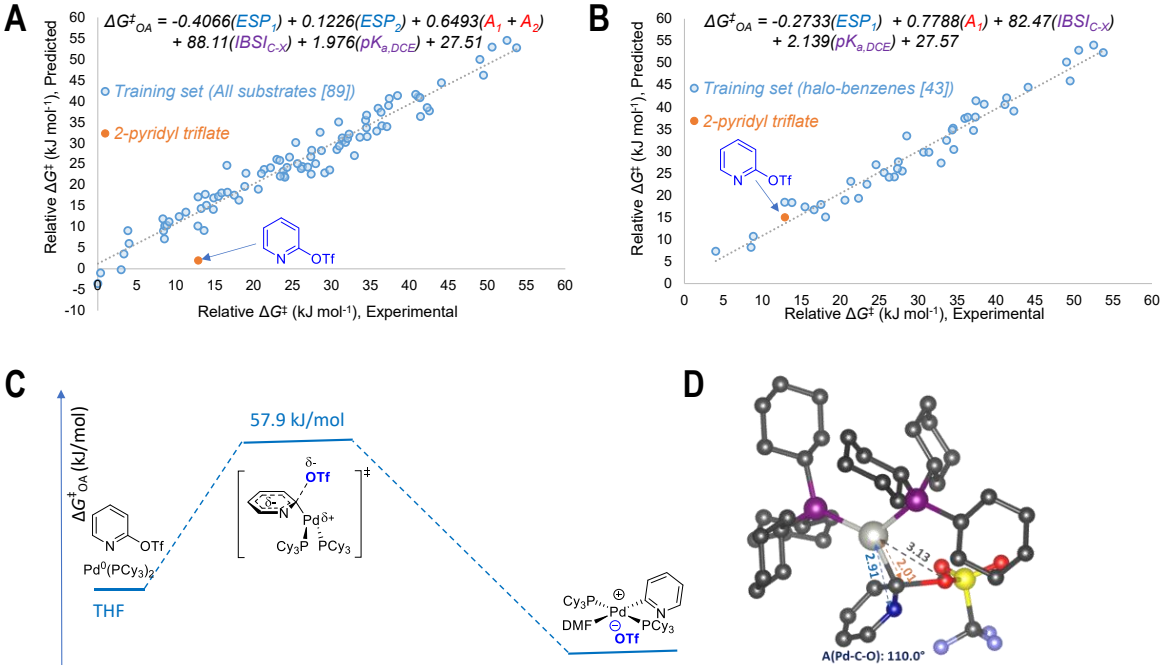


Figure 3.10. Prediction of the relative oxidative addition rate of 2-pyridyl-triflate to Pd(0) and the DFT calculated oxidative addition mechanism. **(A)** Relative rate prediction using the five-descriptor model from Fig. 3.8C. **(B)** Relative rate prediction using the four-descriptor model from Fig. 3.9B. **(C)** DFT calculated reaction mechanism of 2-pyridyl-triflate oxidative addition to Pd(0)$(PCy_3)_2$ in THF. TS and INT search, Gibbs energy correction at RI BP86/def2-SVP, def2/J, D3BJ/def2-TZVP(Pd)/CPCM; Single point energy at RI-B2PLYP/def2-TZVP, def2-TZVP/C, D3/CPCM. **(D)** Calculated transition state with bond length (Pd-N, Pd-C and Pd-O) and bond angle (Pd-C-O) measurements.

TST analysis by DFT reveals that the oxidative addition of 2-pyridyl triflate potentially follows a reaction mechanism resembling the concerted pathway in a genuine S_NAr reaction, which is different from the nucleophilic displacement or the 3-centered mechanisms discussed in Section 3.3.1. The calculated TS complex (Fig. 3.10C and D) is highly polarized, with large partial negative charge distributed in the pyridyl ring and the leaving group triflate. Analysis of the bond lengths indicates that the palladium center coordinates to the reactive carbon without interacting with the adjacent nitrogen, and the triflate is far away from the palladium center as it dissociates from carbon. In addition, no stable π-complex intermediate could be located along the reaction coordinate from the starting reagents to the TS. This computed mechanism is consistent with the outcome from the model prediction, both of which indicate that the pyridyl nitrogen is not crucial in determining the oxidative addition rate of this substrate (more discussion in Chapter 4). This example demonstrates how investigating outliers to the reactivity models can yield additional mechanistic insights, with more examples to be discussed in Chapters 4 and 5.

3.4.4 Reactivity of 2-Pyridyl Tosylate in Oxidative Addition to Pd(0)

Aryl tosylates are another commonly used pseudohalide electrophile class in Pd catalyzed cross-coupling. In general, aryl tosylates are much less reactive than their triflate and bromide analogues,[77] and Pd catalyzed reactions with tosylates as coupling partners are usually conducted at high temperature.[78–80] A prior synthetic study by Hartwig[81] reported the oxidative addition of aryl tosylate to Pd(0) at room temperature enabled by a Pd(0) complex with sterically hindered phosphine ligands:

(Pd(CyPF-t-Bu)[P(o-tolyl)$_3$]). This is proposed to proceed via a pathway initiated by dissociation of P(o-tolyl)$_3$, followed by oxidative addition of aryl tosylate to the 14-electron Pd(CyPF-t-Bu) complex.

As a preliminary test of the reactivity of aryl tosylates toward Pd(PCy$_3$)$_2$, I conducted individual oxidative addition experiments in THF for three electron deficient (hetero)aryl tosylates (2-cyano-phenyl tosylate, 4-cyano-phenyl tosylate and 2-pyridyl tosylate), which are predicted to have similar relative rates using the equation from Fig. 3.8C (the five-descriptor model). Reactions of the two cyano-phenyl tosylates resulted in low conversion of Pd(0) to Pd(II) complex at room temperature and at 40 °C overnight. However, oxidative addition of 2-pyridyl tosylate to Pd(0) reached completion at room temperature overnight; moreover, the observed relative oxidative addition rate of 2-pyridyl tosylate from competition experiment is in good agreement with its predicted rate ($\Delta G^{\ddagger}_{OA\_predicted}$ = 37.9 kJ mol^{-1} versus $\Delta G^{\ddagger}_{OA\_observed}$ = 35.5 kJ mol^{-1}, Fig. 3.11). This observation provides mechanistic insight indicating that the pyridyl nitrogen plays an important role in enabling oxidative addition of this substrate to Pd(0) at mild reaction conditions, and also indicates another potential mechanistic change that is preventing the substituted phenyl tosylate derivatives from reacting.

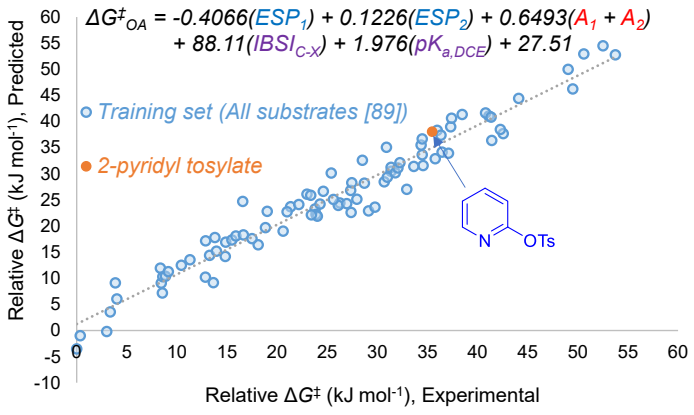


Figure 3.11. Relative rate prediction for oxidative addition of 2-pyridyl tosylate to Pd(PCy$_3$)$_2$ using the five-descriptor model from Fig. 3.8C.

3.4.5 Reactivity of (Hetero)aryl Iodide in Oxidative Addition to Pd(0)

Aryl iodides are generally considered as the most reactive electrophile class based on the established reactivity order for oxidative addition to Pd(0): $ArI > ArBr \approx ArOTf \gg ArOTs > ArCl$.[77] A mechanistic study by Pflüger and coworkers[82,83] suggested a nonpolar 3-centered transition state for oxidative addition of substituted iodobenzenes to $Pd(PPh_3)_4$ based on their kinetic analysis. They obtained similar Hammett ρ values (ρ = +2 in THF, ρ = +2.3 ± 0.2 in toluene) indicating relative insensitivity to substituents (electron donating or withdrawing). They also obtained similar activation energies for reactions in THF and toluene, two solvents which are significantly different in polarity. Later, Jutand and coworkers[34] investigated the oxidative addition mechanism of 2-iodopyridine derivatives to $Pd(PPh_3)_4$ using the same experimental approach of constructing Hammett plots and performing DFT calculations of transition states, and their study also revealed that 2-iodopyridines proceed via a nonpolar 3-centered transition state.

I measured the relative oxidative addition rates of seven (hetero)aryl iodides (Ar–I) to $Pd(PCy_3)_2$ in THF for a preliminary study of the structure-reactivity relationship of this electrophile class. Using the 7 iodide-based electrophiles as an external test to assess the model performance revealed the predicted rates from both models (Fig. 3.8C and Fig. 3.9B) are much slower than the observed rates (Tables 3.3, 3.4). Most of the (hetero)aryl iodides are predicted outliers. Interestingly, the predicted $\Delta G^{\ddagger}_{OA}$ of 2-iodopyridine by the 5-descriptor model (Fig. 3.8C) is within the acceptable error (standardized residual < 2), while this substrate is the largest outlier when using the four-descriptor model (Fig. 3.9B, error = -15.5 kJ mol^{-1}).

Table 3.3. $\Delta G^{\ddagger}_{OA}$ prediction for the seven iodide electrophiles using the equation from Fig. 3.8C.

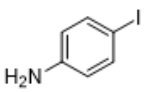
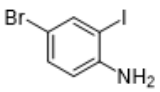
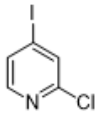
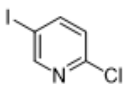
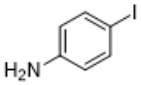
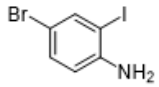
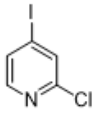
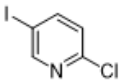
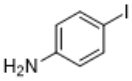
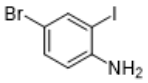
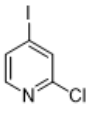
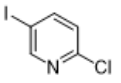
$\Delta G^{\ddagger}_{OA}$ (kJ mol^{-1})							
Observed	16.5	21.2	19.3	9.7	-1.5	1.4	1.3
Predicted	24.7	34.9	26.2	14.5	7.9	11.2	5.4
Residual	-8.2	-13.7	-6.9	-4.7	-9.4	-9.8	-4.2

Table 3.4. $\Delta G^{\ddagger}_{OA}$ prediction for the seven iodide electrophiles using the equation from Fig. 3.9B.

$\Delta G^{\ddagger}_{OA}$ (kJ mol^{-1})							
Observed	16.5	21.2	19.3	9.7	-1.5	1.4	1.3
Predicted	24.3	33.3	27.1	13.7	8.8	10.4	16.8
Residual	-7.8	-12.2	-7.8	-4.0	-10.3	-9.0	-15.5

Improved prediction accuracy from both models is possible when including the reaction data of the (hetero)aryl iodides as part of the training set. The four-descriptor model makes better predictions (Table 3.5) for the six iodobenzene substrates, with an MAE of 3.2 kJ mol^{-1} (compare to 5.3 kJ mol^{-1} from the 5-descriptor model). This is consistent with the reported 3-centered mechanism for the Ar–I substrates in oxidative addition to Pd(0). For the 2-iodopyridine, the 5-descriptor model gives an accurate prediction (Table 3.5) with a residual of -0.6 kJ mol^{-1}, while it is still a predicted outlier (residual = -9.2 kJ mol^{-1}) in the four-descriptor model. This comparison in prediction accuracy suggests that with the more electron rich PCy_3 ligand compared to the PPh_3 ligand used in Jutand's work[34], the oxidative addition of 2-iodopyridine to $Pd(0)(PCy_3)_2$ may proceed via a more polarized transition state. More investigation is necessary to assess this possibility.

Table 3.5. $\Delta G^{\ddagger}_{OA}$ prediction after including iodide electrophiles in the training set.

	4-descriptor model						5-descriptor model
$\Delta G^{\ddagger}_{OA}$ (kJ mol^{-1})							
Observed	16.5	21.2	19.3	9.7	-1.5	1.4	1.3
Predicted	18.1	26.9	22.5	8.8	3.1	4.7	1.9
Residual	-1.7	-5.7	-3.2	0.9	-4.6	-3.3	-0.6

Linear regression analysis using the min/max rescaled descriptors enables comparison of the contribution of each descriptor to predicted $\Delta G^{\ddagger}_{OA}$ before and after including the Ar–I reaction set. Table 3.6 shows the results from the four-

descriptor model (2-iodopyridine is not included in the training set): by including Ar–I set, the contribution from the C–X bond strength (*IBSI*) increases by 6%, which agrees with the mechanistic argument that the weaker C–I bond leads to faster oxidative addition. In the meantime, there is a 4% decrease in ESP_1 contribution and a slight 2% decrease in pK_a contribution.

The four-descriptor model shows good compatibility for the Ar–I substrates, as evidenced by the MAEs for the Ar–I prediction before (7.7 kJ mol^{-1}) and after (3.5 kJ mol^{-1}) a small number (6) of Ar-I datapoints were added to the training set. Better accuracy and broader scope for predicting Ar–I reactivity can likely be achieved by incorporating more Ar–I substrates in the training set.

Table 3.6. Linear regression coefficients %contribution (min/max normalization).

4-descriptor model	ESP_1	A_1	IBSI	pK_a
Before including Ar-I	41%	10%	21%	29%
After including Ar-I	34%	10%	29%	27%

3.5 Conclusions

A quantitative structure-reactivity model for oxidative addition, a key step in many catalytic mechanisms, enables accurate predictions for the outcome of various cross-coupling reactions. By correlating relative reaction rates with simple ground state molecular descriptors, an excellent linear relationship between the experimental and predicted rates can be realized. Importantly, this model links various substrate classes into one unified reactivity scale. Predictions of $\Delta G^{\ddagger}_{OA}$ are based on the molecular structure of the substrates, with applicability well beyond the oxidative addition training set. These $\Delta G^{\ddagger}_{OA}$ values can be used to predict rates and selectivities for many different catalytic reactions under various conditions, including Sonogashira, Suzuki, Buchwald-Hartwig, and Stille couplings.

Given the importance of palladium-catalyzed reactions in the synthesis of complex organic molecules, quantitative reactivity predictions could be used during synthetic planning to design substrates with high intrinsic selectivity, and/or to identify where achieving the desired selectivity is likely to be more challenging/resource intensive. Applying $\Delta G^{\ddagger}_{OA}$ predictions to hypothetical synthetic sequences and potential intermediates could therefore be used to design more selective routes and/or prioritize different potential routes prior to commencing experimental investigations. It could also be used to identify where extensive reaction development is necessary (e.g. overriding predicted selectivity), or where "simple" systems are likely to be successful.

Finally, as I expanded the model to a broader range of substrate classes, including new leaving groups, it became evident that this model also provides a useful tool for mechanistic elucidation. In return, the mechanistic insights revealed by this quantitative model will help to refine its predictive ability, by recognizing additional mechanistically relevant molecular descriptors and incorporating those into the model. This aspect will be further discussed in Chapter 5.

3.6 References

(1) Campeau, L.-C.; Hazari, N. Cross-Coupling and Related Reactions: Connecting Past Success to the Development of New Reactions for the Future. *Organometallics* **2019**, *38* (1), 3–35. https://doi.org/10.1021/acs.organomet.8b00720.

(2) Magano, J.; Dunetz, J. R. Large-Scale Applications of Transition Metal-Catalyzed Couplings for the Synthesis of Pharmaceuticals. *Chem. Rev.* **2011**, *111* (3), 2177–2250. https://doi.org/10.1021/cr100346g.

(3) Vinogradova, E. V.; Zhang, C.; Spokoyny, A. M.; Pentelute, B. L.; Buchwald, S. L. Organometallic Palladium Reagents for Cysteine Bioconjugation. *Nature* **2015**, *526* (7575), 687–691. https://doi.org/10.1038/nature15739.

(4) Wang, K.; Osuka, A.; Song, J. Pd-Catalyzed Cross Coupling Strategy for Functional Porphyrin Arrays. *ACS Cent. Sci.* **2020**, *6* (12), 2159–2178. https://doi.org/10.1021/acscentsci.0c01300.

(5) Meyer, B.; Sawatlon, B.; Heinen, S.; Lilienfeld, O. A. von; Corminboeuf, C. Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts. *Chem. Sci.* **2018**, *9* (35), 7069–7077. https://doi.org/10.1039/C8SC01949E.

(6) Santiago, C. B.; Guo, J.-Y.; Sigman, M. S. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018**, *9* (9), 2398–2412. https://doi.org/10.1039/C7SC04679K.

(7) Maryasin, B.; Marquetand, P.; Maulide, N. Machine Learning for Organic Synthesis: Are Robots Replacing Chemists? *Angew. Chem. Int. Ed.* **2018**, *57* (24), 6978–6980. https://doi.org/10.1002/anie.201803562.

(8) Sandfort, F.; Strieth-Kalthoff, F.; Kühnemund, M.; Beecks, C.; Glorius, F. A Structure-Based Platform for Predicting Chemical Reactivity. *Chem* **2020**, *6* (6), 1379–1390. https://doi.org/10.1016/j.chempr.2020.02.017.

(9) Wodrich, M. D.; Fabrizio, A.; Meyer, B.; Corminboeuf, C. Data-Powered Augmented Volcano Plots for Homogeneous Catalysis. *Chem. Sci.* **2020**, *11* (44), 12070–12080. https://doi.org/10.1039/D0SC04289G.

(10) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360* (6385), 186–190. https://doi.org/10.1126/science.aar5169.

(11) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363* (6424), eaau5631. https://doi.org/10.1126/science.aau5631.

(12) Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **2020**, *10* (3), 2260–2297. https://doi.org/10.1021/acscatal.9b04186.

(13) Newman-Stonebraker, S. H.; Smith, S. R.; Borowski, J. E.; Peters, E.; Gensch, T.; Johnson, H. C.; Sigman, M. S.; Doyle, A. G. Univariate Classification of Phosphine Ligation State and Reactivity in Cross-Coupling Catalysis. *Science* **2021**, *374* (6565), 301–308. https://doi.org/10.1126/science.abj4213.

(14) Hueffel, J. A.; Sperger, T.; Funes-Ardoiz, I.; Ward, J. S.; Rissanen, K.; Schoenebeck, F. Accelerated Dinuclear Palladium Catalyst Identification through Unsupervised Machine Learning. *Science* **2021**, *374* (6571), 1134–1140. https://doi.org/10.1126/science.abj0999.

(15) Schleinitz, J.; Langevin, M.; Smail, Y.; Wehnert, B.; Grimaud, L.; Vuilleumier, R. Machine Learning Yield Prediction from NiCOlit, a Small-Size Literature Data Set of Nickel Catalyzed C–O Couplings. *J. Am. Chem. Soc.* **2022**, *144* (32), 14722–14730. https://doi.org/10.1021/jacs.2c05302.

(16) Fitzner, M.; Wuitschik, G.; J. Koller, R.; Adam, J.-M.; Schindler, T.; Reymond, J.-L. What Can Reaction Databases Teach Us about Buchwald–Hartwig Cross-Couplings? *Chem. Sci.* **2020**, *11* (48), 13085–13093. https://doi.org/10.1039/D0SC04074F.

(17) Krska, S. W.; DiRocco, D. A.; Dreher, S. D.; Shevlin, M. The Evolution of Chemical High-Throughput Experimentation To Address Challenging Problems in Pharmaceutical Synthesis. *Acc. Chem. Res.* **2017**, *50* (12), 2976–2985. https://doi.org/10.1021/acs.accounts.7b00428.

(18) Allen, C. L.; Leitch, D. C.; Anson, M. S.; Zajac, M. A. The Power and Accessibility of High-Throughput Methods for Catalysis Research. *Nat. Catal.* **2019**, *2* (1), 2–4. https://doi.org/10.1038/s41929-018-0220-4.

(19) Isbrandt, E. S.; Sullivan, R. J.; Newman, S. G. High Throughput Strategies for the Discovery and Optimization of Catalytic Reactions. *Angew. Chem. Int. Ed.* **2019**, *58* (22), 7180–7191. https://doi.org/10.1002/anie.201812534.

(20) Buitrago Santanilla, A.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.-C.; Schneeweis, J.; Berritt, S.; Shi, Z.-C.; Nantermet, P.; Liu, Y.; Helmy, R.; Welch, C. J.; Vachal, P.; Davies, I. W.; Cernak, T.; Dreher, S. D. Nanomole-Scale High-Throughput Chemistry for the Synthesis of Complex Molecules. *Science* **2015**, *347* (6217), 49–53. https://doi.org/10.1126/science.1259203.

(21) Shevlin, M. Practical High-Throughput Experimentation for Chemists. *ACS Med. Chem. Lett.* **2017**, *8* (6), 601–607. https://doi.org/10.1021/acsmedchemlett.7b00165.

(22) Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; Eugenio de Diego, J.; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post, R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambeanu, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **2019**, *23* (6), 1213–1242. https://doi.org/10.1021/acs.oprd.9b00140.

(23) Mahjour, B.; Shen, Y.; Cernak, T. Ultrahigh-Throughput Experimentation for Information-Rich Chemical Synthesis. *Acc. Chem. Res.* **2021**, *54* (10), 2337–2346. https://doi.org/10.1021/acs.accounts.1c00119.

(24) Fitzner, M.; Wuitschik, G.; Koller, R.; Adam, J.-M.; Schindler, T. Machine Learning C–N Couplings: Obstacles for a General-Purpose Reaction Yield Prediction. *ACS Omega* **2023**, *8* (3), 3017–3025. https://doi.org/10.1021/acsomega.2c05546.

(25) Beker, W.; Roszak, R.; Wołos, A.; Angello, N. H.; Rathore, V.; Burke, M. D.; Grzybowski, B. A. Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki–Miyaura Coupling. *J. Am. Chem. Soc.* **2022**, *144* (11), 4819–4827. https://doi.org/10.1021/jacs.1c12005.

(26) Singh, S.; Sunoj, R. B. Molecular Machine Learning for Chemical Catalysis: Prospects and Challenges. *Acc. Chem. Res.* **2023**, *56* (3), 402–412. https://doi.org/10.1021/acs.accounts.2c00801.

(27) Strieth-Kalthoff, F.; Sandfort, F.; Kühnemund, M.; Schäfer, F. R.; Kuchen, H.; Glorius, F. Machine Learning for Chemical Reactivity: The Importance of Failed Experiments. *Angew. Chem. Int. Ed.* **2022**, *61* (29), e202204647. https://doi.org/10.1002/anie.202204647.

(28) Labinger, J. A. Tutorial on Oxidative Addition. *Organometallics* **2015**, *34* (20), 4784–4795. https://doi.org/10.1021/acs.organomet.5b00565.

(29) Reeves, E. K.; Entz, E. D.; Neufeldt, S. R. Chemodivergence between Electrophiles in Cross-Coupling Reactions. *Chem. – Eur. J.* **2021**, *27* (20), 6161–6177. https://doi.org/10.1002/chem.202004437.

(30) Sandford, C.; Fries, L. R.; Ball, T. E.; Minteer, S. D.; Sigman, M. S. Mechanistic Studies into the Oxidative Addition of Co(I) Complexes: Combining Electroanalytical Techniques with Parameterization. *J. Am. Chem. Soc.* **2019**, *141* (47), 18877–18889. https://doi.org/10.1021/jacs.9b10771.

(31) Anjali, B. A.; Suresh, C. H. Interpreting Oxidative Addition of Ph–X (X = CH3, F, Cl, and Br) to Monoligated Pd(0) Catalysts Using Molecular Electrostatic Potential. *ACS Omega* **2017**, *2* (8), 4196–4206. https://doi.org/10.1021/acsomega.7b00745.

(32) Kashihara, M.; Gordon, C. P.; Copéret, C. Reactivity of Substituted Benzenes toward Oxidative Addition Relates to NMR Chemical Shift of the Ipso-Carbon. *Org. Lett.* **2020**, *22* (22), 8910–8915. https://doi.org/10.1021/acs.orglett.0c03300.

(33) Busch, M.; Wodrich, M. D.; Corminboeuf, C. A Generalized Picture of C–C Cross-Coupling. *ACS Catal.* **2017**, *7* (9), 5643–5653. https://doi.org/10.1021/acscatal.7b01415.

(34) Maes, B. U. W.; Verbeeck, S.; Verhelst, T.; Ekomié, A.; von Wolff, N.; Lefèvre, G.; Mitchell, E. A.; Jutand, A. Oxidative Addition of Haloheteroarenes to Palladium(0): Concerted versus SNAr-Type Mechanism. *Chem. – Eur. J.* **2015**, *21* (21), 7858–7865. https://doi.org/10.1002/chem.201406210.

(35) Senn, H. M.; Ziegler, T. Oxidative Addition of Aryl Halides to Palladium(0) Complexes: A Density-Functional Study Including Solvation. *Organometallics* **2004**, *23* (12), 2980–2988. https://doi.org/10.1021/om049963n.

(36) Ahlquist, M.; Norrby, P.-O. Oxidative Addition of Aryl Chlorides to Monoligated Palladium(0): A DFT-SCRF Study. *Organometallics* **2007**, *26* (3), 550–553. https://doi.org/10.1021/om0604932.

(37) L. McMullin, C.; Jover, J.; N. Harvey, J.; Fey, N. Accurate Modelling of Pd(0) + PhX Oxidative Addition Kinetics. *Dalton Trans.* **2010**, *39* (45), 10833–10836. https://doi.org/10.1039/C0DT00778A.

(38) Kashihara, M.; Gordon, C. P.; Copéret, C. Reactivity of Substituted Benzenes toward Oxidative Addition Relates to NMR Chemical Shift of the Ipso-Carbon. *Org. Lett.* **2020**, *22* (22), 8910–8915. https://doi.org/10.1021/acs.orglett.0c03300.

(39) Amatore, C.; Jutand, A.; Khalil, F.; M'Barki, M. A.; Mottier, L. Rates and Mechanisms of Oxidative Addition to Zerovalent Palladium Complexes Generated in Situ from Mixtures of Pd0(Dba)2 and Triphenylphosphine. *Organometallics* **1993**, *12* (8), 3168–3178. https://doi.org/10.1021/om00032a045.

(40) Portnoy, M.; Milstein, D. Mechanism of Aryl Chloride Oxidative Addition to Chelated Palladium(0) Complexes. *Organometallics* **1993**, *12* (5), 1665–1673. https://doi.org/10.1021/om00029a026.

(41) Mitchell, E. A.; Jessop, P. G.; Baird, M. C. A Kinetics Study of the Oxidative Addition of Bromobenzene to Pd(PCy3)2 (Cy = Cyclohexyl) in a Nonpolar Medium: The Influence on Rates of Added PCy3 and Bromide Ion. *Organometallics* **2009**, *28* (23), 6732–6738. https://doi.org/10.1021/om900679w.

(42) Barrios-Landeros, F.; Carrow, B. P.; Hartwig, J. F. Effect of Ligand Steric Properties and Halide Identity on the Mechanism for Oxidative Addition of Haloarenes to Trialkylphosphine Pd(0) Complexes. *J. Am. Chem. Soc.* **2009**, *131* (23), 8141–8154. https://doi.org/10.1021/ja900798s.

(43) Gadre, S. R.; Suresh, C. H.; Mohan, N. Electrostatic Potential Topology for Probing Molecular Structure, Bonding and Reactivity. *Molecules* **2021**, *26* (11), 3289. https://doi.org/10.3390/molecules26113289.

(44) Grimmel, S. A.; Reiher, M. The Electrostatic Potential as a Descriptor for the Protonation Propensity in Automated Exploration of Reaction Mechanisms. *Faraday Discuss.* **2019**, *220* (0), 443–463. https://doi.org/10.1039/C9FD00061E.

(45) Remya, G. S.; Suresh, C. H. Quantification and Classification of Substituent Effects in Organic Chemistry: A Theoretical Molecular Electrostatic Potential Study. *Phys. Chem. Chem. Phys.* **2016**, *18* (30), 20615–20626. https://doi.org/10.1039/C6CP02936A.

(46) Sayyed, F. B.; Suresh, C. H. Quantification of Substituent Effects Using Molecular Electrostatic Potentials: Additive Nature and Proximity Effects. *New J. Chem.* **2009**, *33* (12), 2465–2471. https://doi.org/10.1039/B9NJ00333A.

(47) Suresh, C. H.; Alexander, P.; Vijayalakshmi, K. P.; Sajith, P. K.; Gadre, S. R. Use of Molecular Electrostatic Potential for Quantitative Assessment of Inductive Effect. *Phys. Chem. Chem. Phys.* **2008**, *10* (43), 6492–6499. https://doi.org/10.1039/B809561B.

(48) *J. A. Hirsch Topics in Stereochemistry , John Wiley & Sons, Ltd, 1967, Pp. 199–222*.

(49) Klein, J.; Khartabil, H.; Boisson, J.-C.; Contreras-García, J.; Piquemal, J.-P.; Hénon, E. New Way for Probing Bond Strength. *J. Phys. Chem. A* **2020**, *124* (9), 1850–1860. https://doi.org/10.1021/acs.jpca.9b09845.

(50) Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the Definition of the Q2 Parameter for QSAR Validation. *J. Chem. Inf. Model.* **2009**, *49* (7), 1669–1678. https://doi.org/10.1021/ci900115y.

(51) Almond-Thynne, J.; Blakemore, D. C.; Pryde, D. C.; Spivey, A. C. Site-Selective Suzuki–Miyaura Coupling of Heteroaryl Halides – Understanding the Trends for Pharmaceutically Important Classes. *Chem. Sci.* **2016**, *8* (1), 40–62. https://doi.org/10.1039/C6SC02118B.

(52) Palani, V.; Perea, M. A.; Sarpong, R. Site-Selective Cross-Coupling of Polyhalogenated Arenes and Heteroarenes with Identical Halogen Groups. *Chem. Rev.* **2022**, *122* (11), 10126–10169. https://doi.org/10.1021/acs.chemrev.1c00513.

(53) Garcia, Y.; Schoenebeck, F.; Legault, C. Y.; Merlic, C. A.; Houk, K. N. Theoretical Bond Dissociation Energies of Halo-Heterocycles: Trends and Relationships to Regioselectivity in Palladium-Catalyzed Cross-Coupling Reactions. *J. Am. Chem. Soc.* **2009**, *131* (18), 6632–6639. https://doi.org/10.1021/ja9004927.

(54) Legault, C. Y.; Garcia, Y.; Merlic, C. A.; Houk, K. N. Origin of Regioselectivity in Palladium-Catalyzed Cross-Coupling Reactions of Polyhalogenated Heterocycles. *J. Am. Chem. Soc.* **2007**, *129* (42), 12664–12665. https://doi.org/10.1021/ja075785o.

(55) Paenurk, E.; Kaupmees, K.; Himmel, D.; Kütt, A.; Kaljurand, I.; Koppel, I. A.; Krossing, I.; Leito, I. A Unified View to Brønsted Acidity Scales: Do We Need Solvated Protons? *Chem. Sci.* **2017**, *8* (10), 6964–6973. https://doi.org/10.1039/C7SC01424D.

(56) Heiden, M. an der; Plenio, H. The Effect of Steric Bulk in Sonogashira Coupling Reactions. *Chem. Commun.* **2007**, No. 9, 972–974. https://doi.org/10.1039/B616608C.

(57) an der Heiden, M. R.; Plenio, H.; Immel, S.; Burello, E.; Rothenberg, G.; Hoefsloot, H. C. J. Insights into Sonogashira Cross-Coupling by High-Throughput Kinetics and Descriptor Modeling. *Chem. – Eur. J.* **2008**, *14* (9), 2857–2866. https://doi.org/10.1002/chem.200701418.

(58) Clavier, H.; Nolan, S. P. Percent Buried Volume for Phosphine and N-Heterocyclic Carbene Ligands: Steric Properties in Organometallic Chemistry. *Chem. Commun.* **2010**, *46* (6), 841–861. https://doi.org/10.1039/B922984A.

(59) Niemeyer, Z. L.; Milo, A.; Hickey, D. P.; Sigman, M. S. Parameterization of Phosphine Ligands Reveals Mechanistic Pathways and Predicts Reaction Outcomes. *Nat. Chem.* **2016**, *8* (6), 610–617. https://doi.org/10.1038/nchem.2501.

(60) Schoenebeck, F.; Houk, K. N. Ligand-Controlled Regioselectivity in Palladium-Catalyzed Cross Coupling Reactions. *J. Am. Chem. Soc.* **2010**, *132* (8), 2496–2497. https://doi.org/10.1021/ja9077528.

(61) Newman-Stonebraker, S. H.; Smith, S. R.; Borowski, J. E.; Peters, E.; Gensch, T.; Johnson, H. C.; Sigman, M. S.; Doyle, A. G. Univariate Classification of Phosphine Ligation State and Reactivity in Cross-Coupling Catalysis. *Science* **2021**, *374* (6565), 301–308. https://doi.org/10.1126/science.abj4213.

(62) Ananikov, V. P.; Piroyan, A. O.; Gaiduk, K. A.; Beletskaya, I. P.; Khrustalev, V. N.; Antipin, M. Yu. Sulfur-Containing Alkenes—A New Class of Chelating Ligands: Synthesis, Coordination to Palladium, and Structure of the Resulting Complexes. *Russ. J. Org. Chem.* **2009**, *45* (12), 1743–1754. https://doi.org/10.1134/S107042800912001X.

(63) Sharma, K.; Swami, M.; Singh, R.; Fahmi, N.; Singh, R. V. Sulfur-Bonded Coordination Compounds of Palladium(II) and Platinum(II) and Their Antimicrobial Activity. *Phosphorus Sulfur Silicon Relat. Elem.* **2009**, *184* (8), 1964–1974. https://doi.org/10.1080/10426500802417133.

(64) Yang, W.; Wang, Y.; Corte, J. R. Efficient Synthesis of 2-Aryl-6-Chloronicotinamides via PXPd2-Catalyzed Regioselective Suzuki Coupling. *Org. Lett.* **2003**, *5* (17), 3131–3134. https://doi.org/10.1021/ol035188g.

(65) Blaise, E.; Kümmerle, A. E.; Hammoud, H.; de Araújo-Júnior, J. X.; Bihel, F.; Bourguignon, J.-J.; Schmitt, M. Access to 4-Alkylaminopyridazine Derivatives via Nitrogen-Assisted Regioselective Pd-Catalyzed Reactions. *J. Org. Chem.* **2014**, *79* (21), 10311–10322. https://doi.org/10.1021/jo501930s.

(66) Burton, R. J.; Crowther, M. L.; Fazakerley, N. J.; Fillery, S. M.; Hayter, B. M.; Kettle, J. G.; McMillan, C. A.; Perkins, P.; Robins, P.; Smith, P. M.; Williams, E. J.; Wrigley, G. L. Highly Regioselective Buchwald–Hartwig Amination at C-2 of 2,4-Dichloropyridine Enabling a Novel Approach to 2,4-Bisanilinopyridine (BAPyd) Libraries. *Tetrahedron Lett.* **2013**, *54* (50), 6900–6904. https://doi.org/10.1016/j.tetlet.2013.10.035.

(67) Reeves, E. K.; Bauman, O. R.; Mitchem, G. B.; Neufeldt, S. R. Solvent Effects on the Selectivity of Palladium-Catalyzed Suzuki-Miyaura Couplings. *Isr. J. Chem.* **2020**, *60* (3–4), 406–409. https://doi.org/10.1002/ijch.201900082.

(68) Reeves, E. K.; Entz, E. D.; Neufeldt, S. R. Chemodivergence between Electrophiles in Cross-Coupling Reactions. *Chem. – Eur. J.* **2021**, *27* (20), 6161–6177. https://doi.org/10.1002/chem.202004437.

(69) Dai, X.; Chen, Y.; Garrell, S.; Liu, H.; Zhang, L.-K.; Palani, A.; Hughes, G.; Nargund, R. Ligand-Dependent Site-Selective Suzuki Cross-Coupling of 3,5-Dichloropyridazines. *J. Org. Chem.* **2013**, *78* (15), 7758–7763. https://doi.org/10.1021/jo401096u.

(70) Ji, J.; Li, T.; Bunnelle, W. H. Selective Amination of Polyhalopyridines Catalyzed by a Palladium−Xantphos Complex. *Org. Lett.* **2003**, *5* (24), 4611–4614. https://doi.org/10.1021/ol0357696.

(71) Keylor, M. H.; Niemeyer, Z. L.; Sigman, M. S.; Tan, K. L. Inverting Conventional Chemoselectivity in Pd-Catalyzed Amine Arylations with Multiply Halogenated Pyridines. *J. Am. Chem. Soc.* **2017**, *139* (31), 10613–10616. https://doi.org/10.1021/jacs.7b05409.

(72) Sicre, C.; Alonso-Gómez, J.-L.; Cid, M. M. Regioselectivity in Alkenyl(Aryl)-Heteroaryl Suzuki Cross-Coupling Reactions of 2,4-Dibromopyridine. A Synthetic and Mechanistic Study. *Tetrahedron* **2006**, *62* (48), 11063–11072. https://doi.org/10.1016/j.tet.2006.09.040.

(73) Scott, N. W. J.; Ford, M. J.; Jeddi, N.; Eyles, A.; Simon, L.; Whitwood, A. C.; Tanner, T.; Willans, C. E.; Fairlamb, I. J. S. A Dichotomy in Cross-Coupling Site Selectivity in a Dihalogenated Heteroarene: Influence of Mononuclear Pd, Pd Clusters, and Pd Nanoparticles—the Case for Exploiting Pd Catalyst Speciation. *J. Am. Chem. Soc.* **2021**, *143* (25), 9682–9693. https://doi.org/10.1021/jacs.1c05294.

(74) Garg, N. K.; Sarpong, R.; Stoltz, B. M. The First Total Synthesis of Dragmacidin D. *J. Am. Chem. Soc.* **2002**, *124* (44), 13179–13184. https://doi.org/10.1021/ja027822b.

(75) Yang, C.-G.; Liu, G.; Jiang, B. Preparing Functional Bis(Indole) Pyrazine by Stepwise Cross-Coupling Reactions: An Efficient Method to Construct the Skeleton of Dragmacidin D. *J. Org. Chem.* **2002**, *67* (26), 9392–9396. https://doi.org/10.1021/jo026450m.

(76) Garg, N. K.; Stoltz, B. M. A Unified Synthetic Approach to the Pyrazinone Dragmacidins. *Chem. Commun.* **2006**, No. 36, 3769–3779. https://doi.org/10.1039/B605929E.

(77) Roy, A. H.; Hartwig, J. F. Oxidative Addition of Aryl Sulfonates to Palladium(0) Complexes of Mono- and Bidentate Phosphines. Mild Addition of Aryl Tosylates and the Effects of Anions on Rate and Mechanism. *Organometallics* **2004**, *23* (2), 194–202. https://doi.org/10.1021/om034187p.

(78) Bhayana, B.; Fors, B. P.; Buchwald, S. L. A Versatile Catalyst System for Suzuki−Miyaura Cross-Coupling Reactions of C(Sp2)-Tosylates and Mesylates. *Org. Lett.* **2009**, *11* (17), 3954–3957. https://doi.org/10.1021/ol9015892.

(79) Zhou, J.; Fu, G. C. Palladium-Catalyzed Negishi Cross-Coupling Reactions of Unactivated Alkyl Iodides, Bromides, Chlorides, and Tosylates. *J. Am. Chem. Soc.* **2003**, *125* (41), 12527–12530. https://doi.org/10.1021/ja0363258.

(80) Fu, X.; Zhang, S.; Yin, J.; McAllister, T. L.; Jiang, S. A.; Tann, C.-H.; Thiruvengadam, T. K.; Zhang, F. First Examples of a Tosylate in the Palladium-Catalyzed Heck Cross Coupling Reaction. *Tetrahedron Lett.* **2002**, *43* (4), 573–576. https://doi.org/10.1016/S0040-4039(01)02240-7.

(81) Roy, A. H.; Hartwig, J. F. Oxidative Addition of Aryl Tosylates to Palladium(0) and Coupling of Unactivated Aryl Tosylates at Room Temperature. *J. Am. Chem. Soc.* **2003**, *125* (29), 8704–8705. https://doi.org/10.1021/ja035835z.

(82) Fauvarque, J.-F.; Pflüger, F.; Troupel, M. Kinetics of Oxidative Addition of Zerovalent Palladium to Aromatic Iodides. *J. Organomet. Chem.* **1981**, *208* (3), 419–427. https://doi.org/10.1016/S0022-328X(00)86726-1.

(83) Amatore, C.; Pfluger, F. Mechanism of Oxidative Addition of Palladium(0) with Aromatic Iodides in Toluene, Monitored at Ultramicroelectrodes. *Organometallics* **1990**, *9* (8), 2276–2282. https://doi.org/10.1021/om00158a026.

Chapter 4 – Solvent Effects on Oxidative Addition to $Pd(PCy_3)_2$

Contributions:

Experiments and computational analysis were performed by Jingru Lu.

Additional supporting experiments were performed by Holly Celuszak.

4.1 Abstract

This Chapter describes solvent effects on the rate of oxidative addition of (hetero)aryl halides and triflates to the bisligated $Pd(PCy_3)_2$. First, quantitative reactivity models for oxidative addition in different solvents revealed very little solvent effect to the overall model features. Second, specific substrates carrying an H-bond donor do have different rates of oxidative addition due to solvent basicity hydrogen bonding (pK_{HB}) between solvent and substrate. Next, specific substrates exhibit improved site-selectivity in polar solvents, which is likely due to two different oxidative addition mechanisms operating (nucleophilic displacement versus 3-centered). Finally, as described in Chapter 3, transition state analysis of the oxidative addition of 2-pyridyltriflate identified a concerted, S_NAr-like mechanism exhibiting greater charge polarization compared to the nucleophilic displacement mechanism for 2-halopyridines. In this Chapter, that mechanistic distinction led to realizing better triflate/halide (Cl or Br) selectivity in polar solvents for a group of 2-halopyridyl triflates. This is consistent with the very polarized transition state proposed for aryltriflate oxidative addition based on DFT calculations.

4.2 Introduction

Selection of an appropriate solvent is one of the most important factors in reaction optimization in homogenous catalysis. Solvent can significantly affect the rate and selectivity of a reaction,[1–4] as well as the stability and performance of a catalyst.[5–7] In palladium-catalyzed cross-couplings, solvent plays many roles in reaction outcomes.[8,9] For example, as a medium it can stabilize polar transition states and the resulting intermediates from an elementary step.[10] It can also directly participate in the reaction mechanism, for example by coordinating to a palladium center to influence the catalyst selectivity.[11]

In this Chapter, I report how solvent effects influence the rate and selectivity of oxidative addition to palladium(0), with the results and discussion organized into four sub-sections (Fig. 4.1). The first section focuses on how solvent affects the quantitative prediction models, in terms of both the predictive accuracy and the relevant weighing of each molecular descriptor. In each of 3 solvent systems – toluene, THF and 1:1 THF/DMF – I determined oxidative addition rates for a diverse library of substrates under otherwise identical reaction conditions. These led to individual reactivity predictive models for each solvent, enabling comparison of model performance as well as solvent-dependent reactivity trends (Fig. 4.1A).

The second section describes the role of solvent as related to the solvent hydrogen bonding basicity (pK_{HB}) and its impact on the electronic properties of the electrophiles. An intermolecular H-bond between a solvent with a large pK_{HB} (e.g. DMF) and a substrate carrying a good H-bond donor (e.g. –NH_2) will increase the electron density around the reactive centre, resulting in a significant effect on both the reaction rate and site-selectivity of these substrates in oxidative addition (Fig. 4.1B).

The third section concerns the impact of solvent polarity on site-selectivity of specific 3-substituted-2,6-dichloropyridine derivatives. This study revealed that the high C_6/C_2 selectivity between two electronically similar sites is due to different reaction mechanisms controlled by frontier molecular orbital symmetries. The C_6 site reacts via a polarized nucleophilic displacement $TS^{‡}$ and the C_2 site via a

relatively nonpolar 3-centered TS‡. C_6/C_2 selectivity improves as the solvent polarity increases, which is consistent with the distinct mechanisms proposal (Fig. 4.1C). Note that there is more discussion of orbital symmetry effects in Chapter 5.

The final section expands the discussion of the oxidative addition mechanism for 2-pyridyl triflate, highlighting key differences to the mechanisms of 2-halopyridine oxidative addition. DFT calculations reveal a concerted S_NAr-like pathway is viable, leading to a very polarized TS‡. This is consistent with experimental observations that aryl triflates react much faster in polar solvents. This extends to solvent effects on the site-selectivity of 2-halopyridyl triflate derivatives, where using polar solvents improves selectivity at the triflate site (Fig. 4.1D).

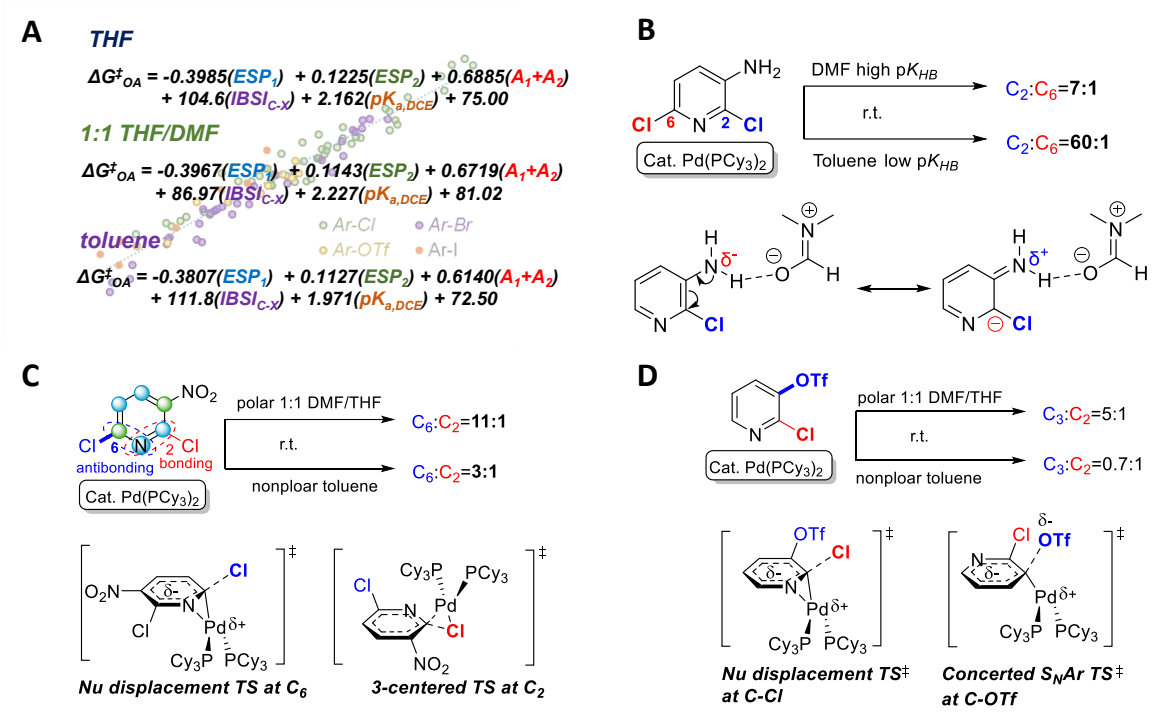


Figure 4.1. Summary of the solvent effect cases discussed in this Chapter. **A)** Multivariate linear regression models of $\Delta G^{\ddagger}_{OA}$ in THF, 1:1 THF/DMF and toluene. **B)** Influence of solvent hydrogen bonding on oxidative addition outcomes, here shows the site-selectivity of 2,6-dichloro-3-aminopyridine as a representative example. **C)** Solvent effects on site-selectivity via one type of mechanism being favoured over another, here shows 2,6-dichloro-3-nitropyridine as an example. **D)** Mechanistic study on oxidative addition of 2-pyridyl triflate and solvent effect on site-selectivity between a triflate and a chloride/bromide site, here shows 2-chloropyridyl-3-triflate as an example.

4.3 Results and Discussion

4.3.1 Oxidative Addition Reactivity Models in Different Solvents

As a starting point for this solvent effect study, I collected relative rates of oxidative addition for various electrophiles to $Pd(PCy_3)_2$ in three different solvent systems: toluene, THF and a mixture of 1:1 THF/DMF. In the latter case, adding THF was necessary to ensure complete dissolution of $Pd(PCy_3)_2$ and the oxidative addition products. These three solvents are among the frequently used in palladium-catalyzed reactions, and represent a range of solvent polarity. Using the same competition experimentation approach described in Chapter 3, I collected relative rate data for (hetero)aryl chlorides, bromides, iodides and triflates in each solvent. Finally, I measured the absolute rates for touchstone reactions in each solvent, and used those values to convert the relative rates into a scale of absolute rates for the entire array of reactions.

To calibrate the relative rate data against absolute rates, I used the same approach as described in Chapter 2 for S_NAr modelling. In each solvent system, I first measured the absolute rates of two touchstone reactions by monitoring the consumption of $Pd(PCy_3)_2$ using ^{31}P NMR spectroscopy (the rate constant for each touchstone reaction was measured in duplicate, and the $\Delta G^{\ddagger}_{OA}$ is given as the average value). The absolute $\Delta G^{\ddagger}_{OA}$ value of one touchstone reaction is the calibration point to obtain the absolute rates of the entire array of substrates, and the other touchstone reaction is a check on the validity of the competition determined $\Delta G^{\ddagger}_{OA}$ values. The percent difference between the $\Delta G^{\ddagger}_{OA}$ values by competition experiments and kinetic analysis is 0.5% in toluene, 2.3% in THF, and 3.0% in THF/DMF. The results of the kinetic analysis for reactions in toluene are shown in Fig. 4.2 and the rest are in Figs.C1 and C2, Appendix C. This approach produced absolute rate scales containing 99 substrates in THF, 49 substrates in 1:1 THF/DMF and 50 in toluene.

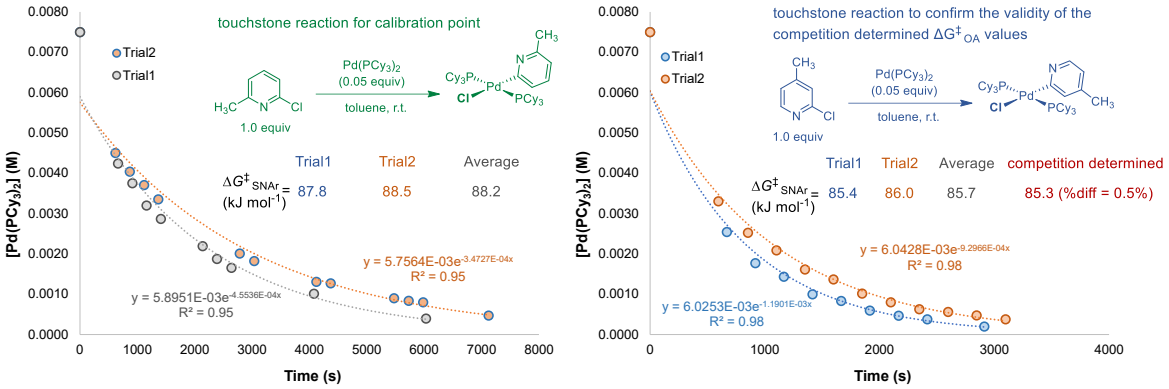


Figure 4.2. Touchstone reaction progress analysis under pseudo first order conditions in toluene. The $\Delta G^{\ddagger}_{OA}$ of 2-chloro-6-methylpyridine **(left)** was used as the calibration point to obtain the absolute rates of the entire array of substrates, and the $\Delta G^{\ddagger}_{OA}$ of 2-chloro-4-methylpyridine **(right)** was used to confirm the validity of the competition determined $\Delta G^{\ddagger}_{OA}$ for this substrate.

Using these rate data in the 3 solvents, I built individual MLR models by correlating the measured rates in each solvent to the same molecular descriptors introduced in Chapter 3 (*ESP*$_1$, *ESP*$_2$, A_1+A_2, *IBSI* and pK_a). As the results in Fig. 4.3 show, there is excellent correlation for each substrate class (Ar–Br, Ar–Cl, Ar–OTf and Ar–I) from all the 3 five-descriptor predictive models, as indicated by the R^2 (0.92-0.94) and MAE (2.1 – 2.5 kJ mol^{-1}) values. The robustness of the models was validated using train-test splits and the results are shown in Figs. C7 – C9, Appendix C. Relative contributions from each descriptor in different solvents (obtained using min-max normalization) are quite similar, with *ESP*$_1$ being the top-ranked variable (36 – 39%), followed by *ESP*$_2$, *IBSI* and pK_a (~20% for each), and steric A-values (5–6%) as least contributing. Analysis of the relative contribution from each descriptor indicates no evidence of solvent being a major impacting factor on the prediction outcomes in a general sense. The similar linear equations and correlation statistics for the 3 models show, surprisingly, that these three solvents have little effect on the absolute rates of oxidative addition using $Pd(PCy_3)_2$. As a result, I made no attempt to incorporate solvent-based descriptors into the model.

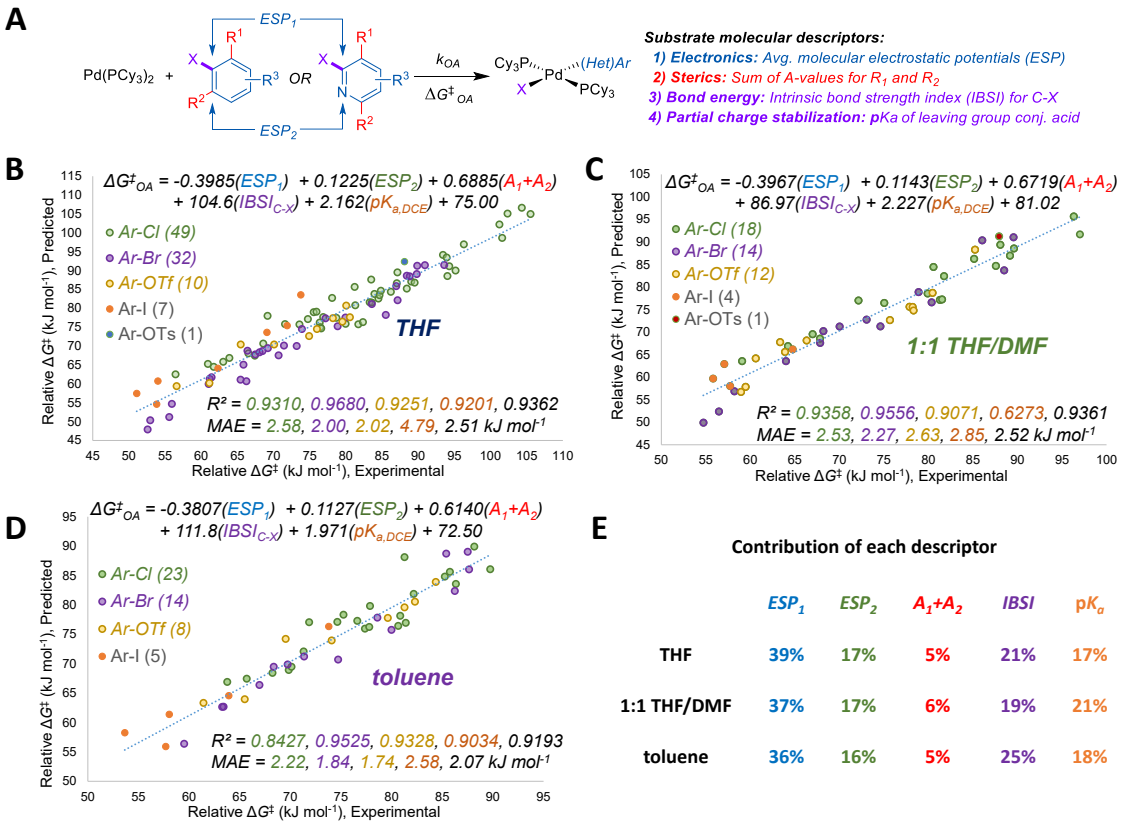


Figure 4.3. A) Molecular descriptors used to model oxidative addition reactivity as a function of substrate structure. **B)** Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 97 substrates in THF. **C)** Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 49 substrates in 1:1 THF/DMF. **D)** Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 50 substrates in toluene. **E)** relative contribution of each descriptor to the predictive accuracy of the 3 multivariate linear regression models.

While the *overall* rate correlations are very similar across the three solvents, I examined the data more closely to identify more subtle solvent effects on specific classes of substrate or on individual substrates. Comparing the results for overlapping substrates (rates measured in all 3 solvents) by stacking the rate scales in different solvents in one plot visualizes solvent-based differences for specific Ar–Br, Ar–Cl and Ar–OTf examples. Assessing solvent effects on Ar–I reactivity remains for future work due to the small number of datapoints currently available. A comparison of the 11 overlapping Ar–Br rates is shown in Fig. 4.4. Overall, Ar–Br substrates react slightly faster in THF and 1:1 THF/DMF than in toluene; however, there is no significant solvent effect within this set.

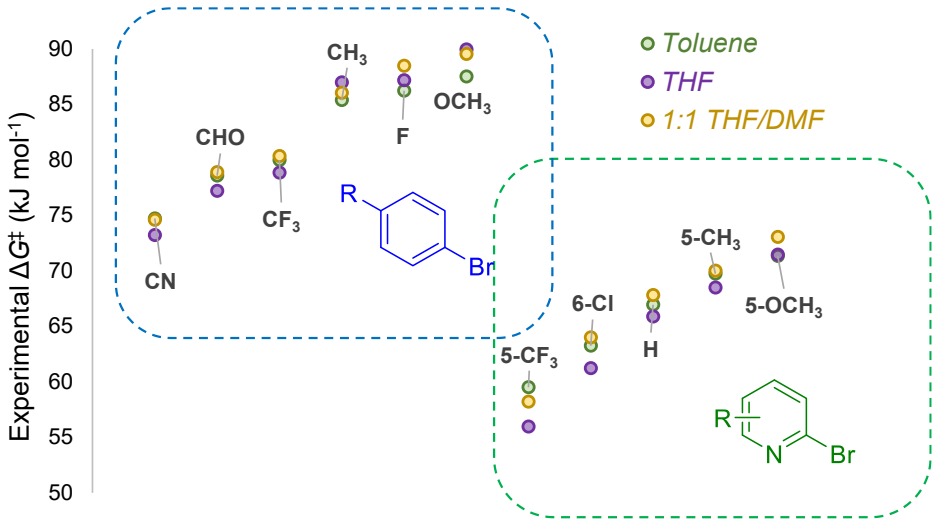


Figure 4.4. Experimental rates in three solvents for oxidative addition of the 11 overlapping Ar-Br substrates, including five 4-substituted bromobenzenes and five substituted 2-bromopyridines.

4.3.2 Solvent Effect via Solvent Hydrogen Bond Basicity (pK_{HB})

The rate scales of the 10 overlapping Ar–Cl are shown in Fig. 4.5. Overall, the $\Delta G^{\ddagger}_{OA}$ values are slightly lower in THF and 1:1 THF/DMF compared to those in toluene as observed for Ar–Br substrates, with one exception: 2-chloro-3-aminopyridine, which reacts much faster in toluene ($\Delta G^{\ddagger}_{OA}$ = 81.3 kJ mol^{-1}) and in THF (83.5 kJ mol^{-1}) than in 1:1 THF/DMF (87.6 kJ mol^{-1}). This reversed reactivity trend and the significant differences among the measured $\Delta G^{\ddagger}_{OA}$ values in different solvents make this substrate a clear outlier from the Ar–Cl subset.

As discussed in Chapter 3, there are two general mechanisms for oxidative addition of Ar–X substrates to Pd(0): a relatively non-polar 3-centered insertion mechanism, and a more polar nucleophilic displacement mechanism. The nucleophilic displacement mechanism is generally favoured for 2-halopyridines, including for 2-chloro-3-aminopyridine (more evidence for this mechanism will be presented in Chapter 5). As a result, polar solvents should *accelerate* oxidative addition for this substrate by stabilizing the polar transition state; however, 2-chloro-3-aminopyridine is more reactive in *less* polar solvents (toluene>THF>THF/DMF), indicating solvent properties other than the dielectric constant may be responsible.

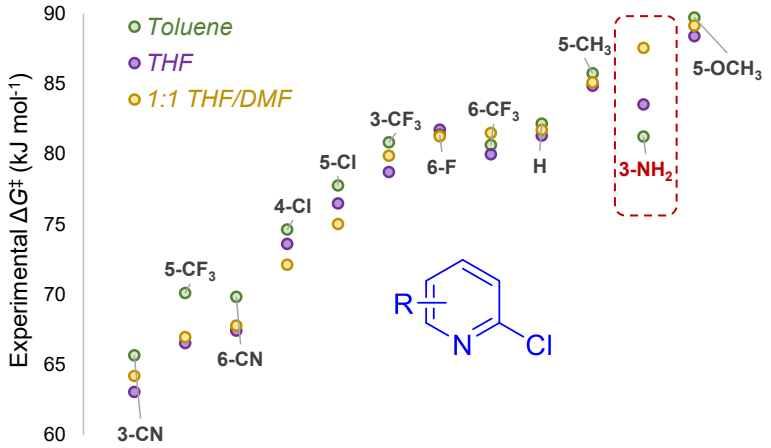


Figure 4.5. Experimental rates of the 12 overlapping substituted 2-chloropyridine derivatives oxidative addition to $Pd(PCy_3)_2$ in toluene, THF and 1:1 THF/DMF.

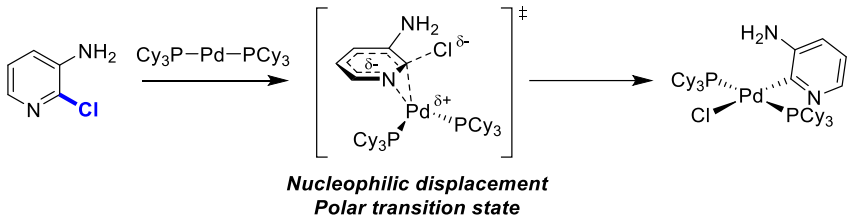


Figure 4.6. Oxidative addition of 2-chloro-3-aminopyridine to $Pd(PCy_3)_2$ via the nucleophilic displacement mechanism.

Computational and experimental studies show that the solvent hydrogen bond basicity (SHBB, quantified by pK_{HB})[12], instead of the solvent polarity, is the primary factor for the observed reactivity differences for 2-chloro-3-aminopyridine in these three solvents. The solvent-solute hydrogen bond (H-bond) interaction plays an important role in determining a wide range of chemical properties, and a lot of studies have been taken to understand its impact in various molecular systems.[13–15] Relevant to this work, Wang and coworkers reported a dramatic effect of solvent H-bonding on the regioselectivity in an S_NAr reaction;[16] however, to the best of my knowledge solvent-substrate H-bonding has not been used to rationalize rate and/or selectivity outcomes in palladium-catalyzed cross couplings.

To further probe if SHBB is a factor affecting other substrates, oxidative addition rates for 2-chloro-3-hydroxypyridine – which contains a strong H-bond donor – were measured in different solvents (Fig. 4.7A). The measured $\Delta G^{\ddagger}_{OA}$ of this 3-OH derivative is 79.9 kJ mol^{-1} in toluene, 82.1 kJ mol^{-1} in THF, and 91.9 kJ mol^{-1} in DMF/THF, which follows the same trend as for the 3-NH_2 derivative. In contrast, unsubstituted 2-chloropyridine has nearly identical $\Delta G^{\ddagger}_{OA}$ values in all three solvents.

Solvents with high pK_{HB} are good H-bond acceptors, and the intermolecular interaction between the solvent molecule and a substrate containing hydrogen bond donor will distort the electron density of the substrate, affecting its reactivity. The calculated structure of 2-chloro-3-aminopyridine with one explicit DMF molecule has the carbonyl oxygen acting as an H-bond acceptor from the –NH_2 group (Fig. 4.7B). This will generate partial negative charge at the amino nitrogen, turning it into a stronger EDG and thus deactivating the substrate toward oxidative addition.

To quantitatively account for the SHBB on the electronic features of these substrates, I performed DFT calculations on each chloropyridine in the presence of explicit solvent molecule (THF or DMF) with the CPCM solvation model to further account for dielectric effects. The starting structure contains the substrate and one solvent molecule, with the atoms acting as H-bond donor and acceptor placed close to each other. The optimized structure at the minimum of the potential energy surface was located via unconstrained geometry optimization at B3LYP def2-SVP level. Then a single point energy calculation was performed for this optimized structure at B3LYP def2-TZVPD level, and the resulting wavefunction used for partial charge and ESP calculations.

From these H-bonded structures, calculated electrostatic potentials (*ESP*) at the C_2 sites of the 3-amino and 3-hydroxy derivatives, along with the unsubstituted 2-chloropyridine, reveal the SHBB effects (Fig. 4.7C). The ΔESP_{C2} values shown are calculated as the difference between the ESP_{C2} of the substrate-solvent H-bonded complex in the CPCM solvation model and the ESP_{C2} of the substrate in the CPCM solvation model. This ensures ΔESP_{C2} only accounts for the influence from the

substrate-solvent H-bond. In all cases, the presence of the explicit solvent molecule leads to larger negative *ESP* at the C_2 site, corresponding to increased electron density. Comparing the ΔESP_{C2} of the 3-NH_2 and 3-OH derivatives to 2-chloropyridine, the magnitude of the ΔESP_{C2} values indicates a much greater electronic change at the C_2 site when the solvent-substrate H-bond is present. A plot of the *ESP* isosurfaces for 2-chloro-3-aminopyridine and its DMF H-bonded complex provides a direct visualization of the solvent H-bond effect on the electron density around the pyridyl ring (Fig. 4.7D).

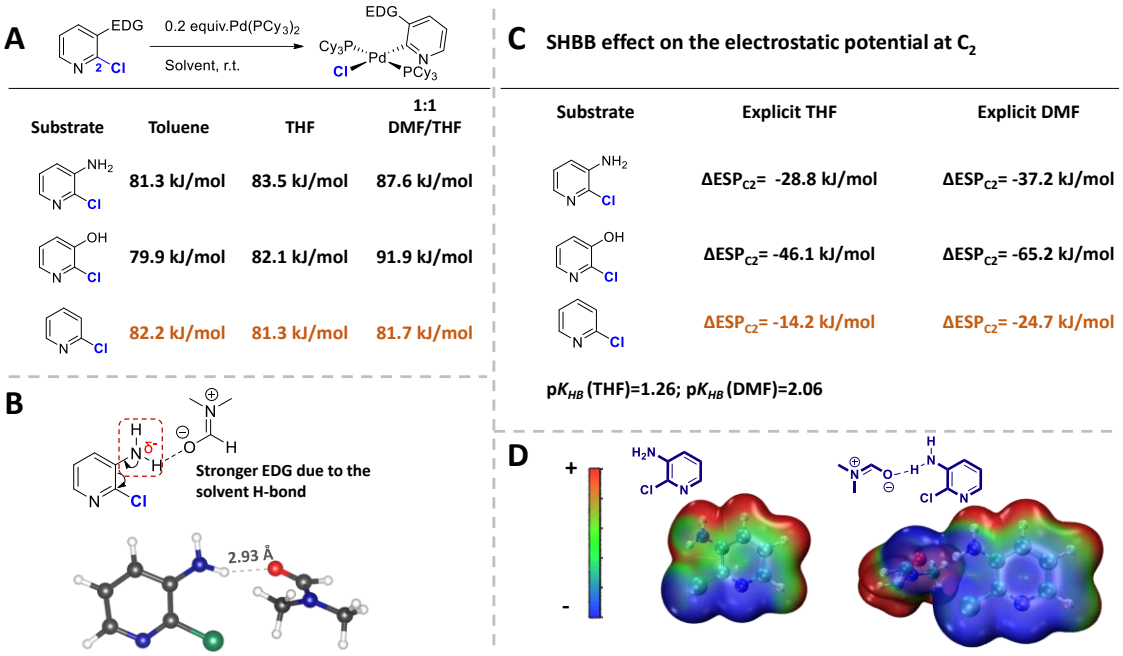


Figure 4.7. A) Experimental $\Delta G^{\ddagger}_{OA}$ for oxidative addition of 2-chloro-3-aminopyridine, 2-chloro-3-hydroxypyridine and 2-chloropyridine to $Pd(PCy_3)_2$ in toluene, THF and 1:1 THF/DMF. **B)** Intermolecular H-bond between DMF and the –NH_2 group, and its deactivating effect on the C_2 site. **C)** Computational analysis of the solvent hydrogen bond basicity (SHBB) effect on the *ESP* at the C_2 site; ΔESP_{C2} values are the difference between ESP_{C2} of the substrate-solvent H-bond complex in the CPCM solvation model and ESP_{C2} of the substrate itself in the CPCM solvation model. **D)** *ESP* isosurfaces at 0.002 au of 2-chloro-3-aminopyridine and the 2-chloro-3-aminopyridine–DMF H-bond complex; blue represents negative *ESP*, and red represents positive *ESP*.

The structure of the substrate also impacts the ΔESP_{C2} values. Taking the THF series as an example: for 2-chloropyridine, the distance between the oxygen of the THF and the C_3–H is 2.3 Å. This weak interaction, which is also indicated by the

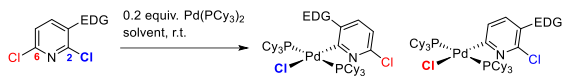
low ΔESP_{C2} value (-14.2 kJ mol^{-1}), likely arises from intermolecular dipole-induced dipole attraction. A distance of 1.9 Å between the oxygen of THF and the N–H of the 3-NH_2 derivative, and a larger ΔESP_{C2} value, indicate a stronger interaction consistent with a hydrogen bond. For the 3-OH derivative, the O_{THF} to H–OPy distance is only 1.6 Å, and the ΔESP_{C2} (-46.1 kJ mol^{-1}) is the largest among the three substrates. Here, the considerably more acidic hydroxyl proton results in a stronger solvent-substrate H-bond interaction.

Importantly, the proposed solvent H-bond effect is also consistent with observed oxidative addition site-selectivity with the three 2,6-dichloro-3-EDG-pyridine derivatives in Fig. 4.8 (EDG = NH_2, NHCOMe and OH). Comparing the measured site-selectivity in different solvent systems, there is a clear trend that the best C_2/C_6 selectivity is achieved in toluene, a low pK_{HB} solvent; as the (co-)solvent pK_{HB} increases, there is a consistent decrease to the C_2/C_6 selectivity with all three substrates. Additional experimental evidence of the solvent H-bond effect is given by the site-selectivity data of the 3-OMe derivative in different solvents: for this substrate, which has no H-bond donor, C_2 is always the only observed reactive site (C_2/C_6 > 99:1) regardless of the solvent type; this observation is in clear contrast to the 3-OH derivative case, where there is a decrease in C_2/C_6 selectivity from >99:1 to 6:1 as the pK_{HB} of the (co-)solvent increases.

A computational analysis on the electronic features of the 2-chloro-3-EDG-pyridines (EDG = NH_2, OH) in Fig. 4.7 demonstrates that the solvent H-bond deactivates the reactive centre by turning the -NH_2 and -OH groups into stronger EDGs. Intuitively, this should deactivate both C_2 (*ortho*) and C_6 (*para*) sites. Quantitative assessment of the increased electron donation via DFT calculations (conducted in the same way as for the 2-chloro-3-EDG-pyridine derivatives discussed above) is summarized in Fig. 4.8B. This analysis shows that both sites are in fact strongly deactivated by the solvent H-bond, as indicated by the large negative ΔESP values; however, the C_2 sites are deactivated to a greater degree than the C_6 sites in all cases: from the substrate itself to it with an explicit H-bond acceptor solvent molecule (THF or DMF), the difference between the ESP values from C_2 to

C_6 site either gets significantly smaller, or turned into a negative number as shown in Figure 5B. This is consistent with the observed decrease in C_2 selectivity in solvents with high pK_{HB}.

A



Entry	Solvent	EDG = NH_2 C_2:C_6	EDG = NHCOMe C_2:C_6	EDG = OH C_2:C_6	EDG = OMe C_2:C_6	EDG = OAc C_2:C_6
1	Toluene	60:1	80:1	>>	>>	1:1
2	THF	9:1	8:1	6:1	>>	1:1
3	1:1 THF/DMF	7:1	2:1	6:1	>>	1:1
4	1:1THF/DMA	6:1	--	--	--	--
5	1:1 THF/Et_3N	4:1	--	--	--	--
6	1:1 THF/Propylene carbonate	--	3:1	--	--	--

B SHBB effect on the electrostatic potentials at C_2 and C_6

Substrate EDG =	No solvent H-bond	Explicit THF	Explicit DMF
	ΔESP (C_2-C_6) (kJ/mol)		
NH_2	24.6	19.5	10.7
NHCOMe	18.0	7.5	-9.0
OH	0.6	-7.3	-12.9

C

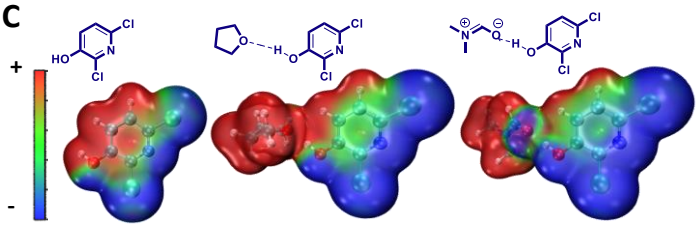


Figure 4.8. A) Experimental C_2/C_6 selectivity for oxidative addition of 2,6-dichloro-3-EDG-pyridines (EDG = NH_2, NHCOMe, OH) to $Pd(PCy_3)_2$ in different solvents (>> indicates only C_2 is observed). **B)** Computational analysis of the solvent hydrogen bond basicity (SHBB) effect on the electrostatic potential (*ESP*) at the C_2 and C_6 sites; the Δ*ESP* values at C_2 and C_6 are the difference between the *ESP* values of the substrate- solvent H-bonding complex in the CPCM solvation model and the *ESP* values of the substrate in the CPCM solvation model. **C)** *ESP* isosurfaces at 0.002 au of 2,6-dichloro-3-hydroxypyridine, the substrate–THF complex, and the substrate–DMF complex; blue represents negative *ESP*, and red represents positive *ESP*.

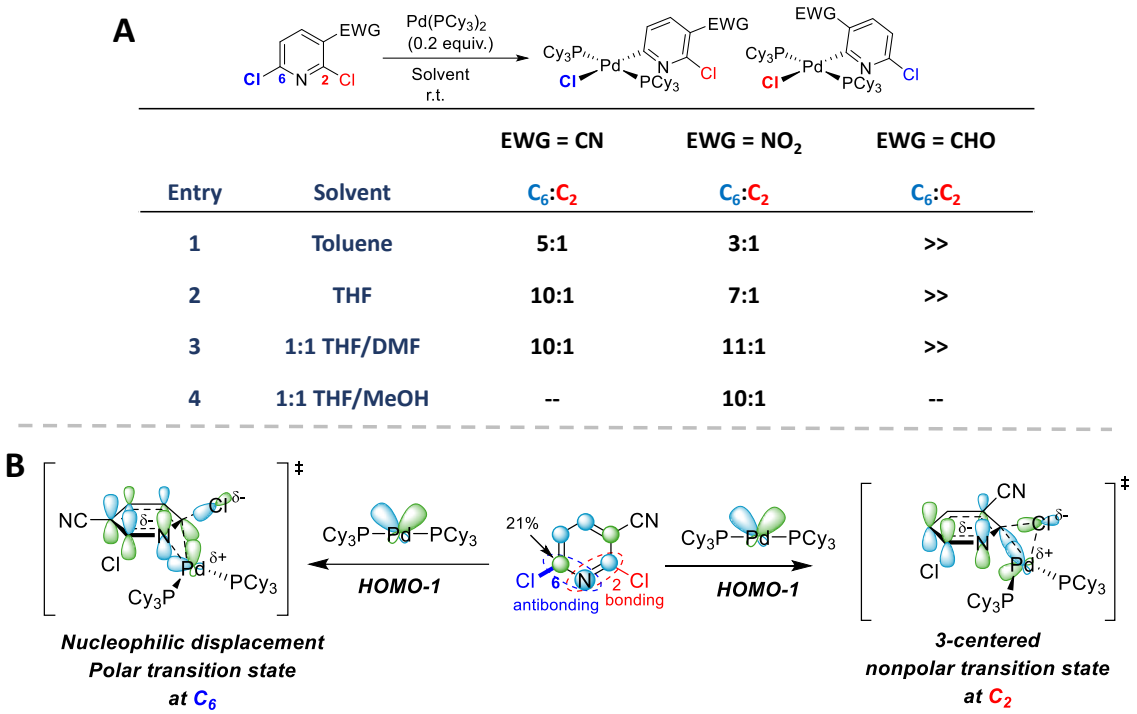
In summary, solvent hydrogen bond basicity (pK_{HB}) influences the rate and selectivity of oxidative addition for substrates containing a hydrogen bond donor EDG. Intermolecular H-bonding between solvents with high pK_{HB} and substituents like -NH_2, -NHCOMe and -OH increases the electron density of the pyridine rings, and results in a strong deactivating effect toward oxidative addition reactivity. This effect also alters the site-selectivity of 2,6-dichloro-3-EDG-pyridines, where the best C_2/C_6 selectivity occurs in toluene, following a trend that solvents with high pK_{HB}

give lower C_2/C_6 selectivity. Computational analysis supports this experimental observation by showing how the solvent H-bond with the EDGs at the C_3 position deactivates the C_2 site more strongly than the C_6 site.

4.3.3 Influence of Solvent Polarity on Site-Selectivity

In addition to solvent effects on the oxidative addition of EDG-substituted 2-chloropyridines, I also studied site selectivity as a function of solvent for several EWG-substituted derivatives; the results shown in Fig. 4.9A. For the 3-CN and 3-NO_2 2,6-dichloropyridine derivatives, there is good C_6/C_2 selectivity (~10:1) in solvents with high dielectric constants (THF, 1:1 mixtures of THF/DMF and THF/MeOH), while the ratios between the C_6/C_2-isomers drop to 5:1 (3-CN) and 3:1 (3-NO_2) in the nonpolar toluene. These observations are consistent with a change in reaction mechanism between the two sites, where oxidative addition at C_6–Cl proceeds through a polarized nucleophilic displacement pathway, whereas reaction at C_2 proceeds through a non-polar 3-centered mechanism (Fig 4.9B). Higher polarity solvents will stabilize the nucleophilic displacement $TS^{‡}$, resulting in faster oxidative addition at C_6. Further evidence in support of this mechanism switch is presented in Chapter 5.

Notably, 3-CHO-2,6-dichloropyridine exhibits excellent C_6/C_2 selectivity (>99:1) in all solvent systems, regardless of the solvent dielectric constant. A possible reason for solvent polarity not having a significant influence on site selectivity is that this substrate is intrinsically heavily biased toward oxidative addition at C_6. The *ESP*-based prediction model from Chapter 3 predicts high C_6/C_2 selectivity ($\Delta\Delta G^{‡}_{OA}$ = 6.8 kJ mol^{-1}) for the 3-CHO derivative, but only moderate to poor selectivity for the 3-CN and 3-NO_2 derivatives ($\Delta\Delta G^{‡}_{OA}$ = 0.0 kJ mol^{-1} and 3.4 kJ mol^{-1}, respectively). The smaller $\Delta\Delta G^{‡}_{OA}$ values for 3-CN and 3-NO_2 indicate that the electronic differentiation of these two sites is minimal; as a result, other reaction parameters, such as solvent polarity, are likely to exert an observable effect on site-selectivity.



Entry	Solvent	EWG = CN C_6:C_2	EWG = NO_2 C_6:C_2	EWG = CHO C_6:C_2
1	Toluene	5:1	3:1	>>
2	THF	10:1	7:1	>>
3	1:1 THF/DMF	10:1	11:1	>>
4	1:1 THF/MeOH	--	10:1	--

Figure 4.9. A) Observed C_6/C_2 selectivity of 2,6-dichloro-3-EWG-pyridines (EWG = CN, NO_2, CHO) in different solvents. **B)** A proposed hypothesis that the increased C_6/C_2 selectivity in polar solvents is caused by change in mechanism at the two sites; this will be further discussed and proved in Chapter 5.

4.3.4 Mechanistic Study of Oxidative Addition of 2-Pyridyl Triflate and Solvent Effects on Triflate versus Chloride/Bromide Selectivity

One substrate class where there is a clear trend between solvent identity and oxidative addition rate is the aryl triflates. Comparing $\Delta G^{\ddagger}_{OA}$ values of the 8 overlapping aryl triflates in different solvents reveals that oxidative addition is faster as the solvent polarity increases (Fig. 4.10). In addition, there are much larger differences between the $\Delta G^{\ddagger}_{OA}$ values of the same substrate measured in different solvents relative to other substrate types. For example, for the 10 (hetero)aryl bromides in Fig. 4.4, the average difference between the $\Delta G^{\ddagger}_{OA}$ values in toluene and 1:1 THF/DMF is only 0.7 kJ mol^{-1}, while for the 8 aryl triflates, the average difference is 5.0 kJ mol^{-1} (approximately a factor of 10 rate increase at room temperature). I hypothesize the solvent effect in this case is again due to solvent

polarity stabilizing highly polar transition states, where oxidative addition of aryl triflates proceeds by a nucleophilic attack-type mechanism.

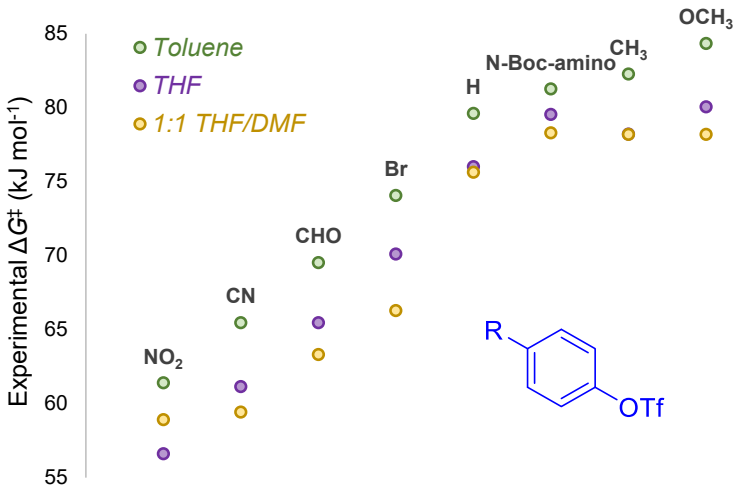


Figure 4.10. Experimental rates of the 8 overlapping para-substituted aryl triflates oxidative addition to $Pd(PCy_3)_2$ in toluene, THF and 1:1 THF/DMF.

In Chapter 3, I discussed an interesting observation of the prediction accuracy of different models for oxidative addition of 2-pyridyl triflate. Studies on the structurally similar 2-chloro/bromopyridine derivatives revealed that the electronic influence of the neighbouring atom plays an important role in determining the oxidative addition rate: electronegative nitrogen leads to faster reaction rates for pyridines compared to the less electronegative carbon in analogous benzene derivatives. However, 2-pyridyl triflate is a significant outlier using the five-descriptor model (ESP_1, ESP_2, A_1+A_2, *IBSI*, pK_a), with the predicted $\Delta G^{\ddagger}_{OA}$ value 10 kJ mol^{-1} lower than the experimental value. In contrast, the predicted value from the four-descriptor (ESP_1, A_1+A_2, *IBSI*, pK_a) model is in excellent agreement with experiment. These observations – of solvent effects and model predictions – are consistent with oxidative addition of Ar–Cl and Ar–OTf proceed via distinct mechanistic paths.

Differences between oxidative addition of aryl triflates and other substrate types have been studied by several research groups. Fu and coworkers[17] reported ligand-

controlled site-selectivity for the Suzuki coupling of chlorophenol triflate, observing that the bisligated $Pd(PCy_3)_2$ favours the C–OTf site while the monoligated $Pd[P(tBu)_3]$ favours the C–Cl site. Later, solvent effects on the site-selectivity of this Suzuki coupling catalyzed by $Pd/P(tBu)_3$ was investigated by Proutiere and Schoenebeck (Fig. 4.11A).[1] It was found that C–Cl is the exclusive reacting site in less polar solvents (toluene and THF). In contrast, in polar solvents the Cl/OTf selectivity is reversed, and high selectivity at the C–OTf site is observed in DMF and MeCN. Their computational study reveals that the selectivity is controlled by the active catalytic species present in different solvents. In polar solvent, $Pd[P(tBu)_3]$ coordinates to anions in the reaction system to form the palladate species $[Pd(PtBu_3)X]^-$ (X=F or $ArBO_2H$), which prefers C–OTf insertion via a lower energy transition state.

In 2022, Neufeldt and coworkers[11] investigated the same Suzuki coupling reaction and proposed that solvent effect to the C–Cl versus C–OTf selectivity arises from solvent coordination to palladium. They observed that the Suzuki coupling of the 4-chlorophenol triflate (Fig. 4.11B) in polar solvents like MeCN, DMF and DMSO favour reaction at the C–OTf bond; however, in other polar solvents like MeOH, acetone and H_2O, the major product is from C–Cl substitution. They re-evaluated the transition state energies using methods including dispersion correction, which was not taken into account by Proutiere and Schoenebeck, and the results suggest that instead of solvent polarity, the dramatic difference in site-selectivity is caused by the solvent coordinating ability. In coordinating solvent like DMF, the reaction is catalyzed by $Pd[P(tBu)_3](DMF)$, and the solvent-coordinated transition state for the rate-limiting oxidative addition is lower in energy for C–OTf insertion.

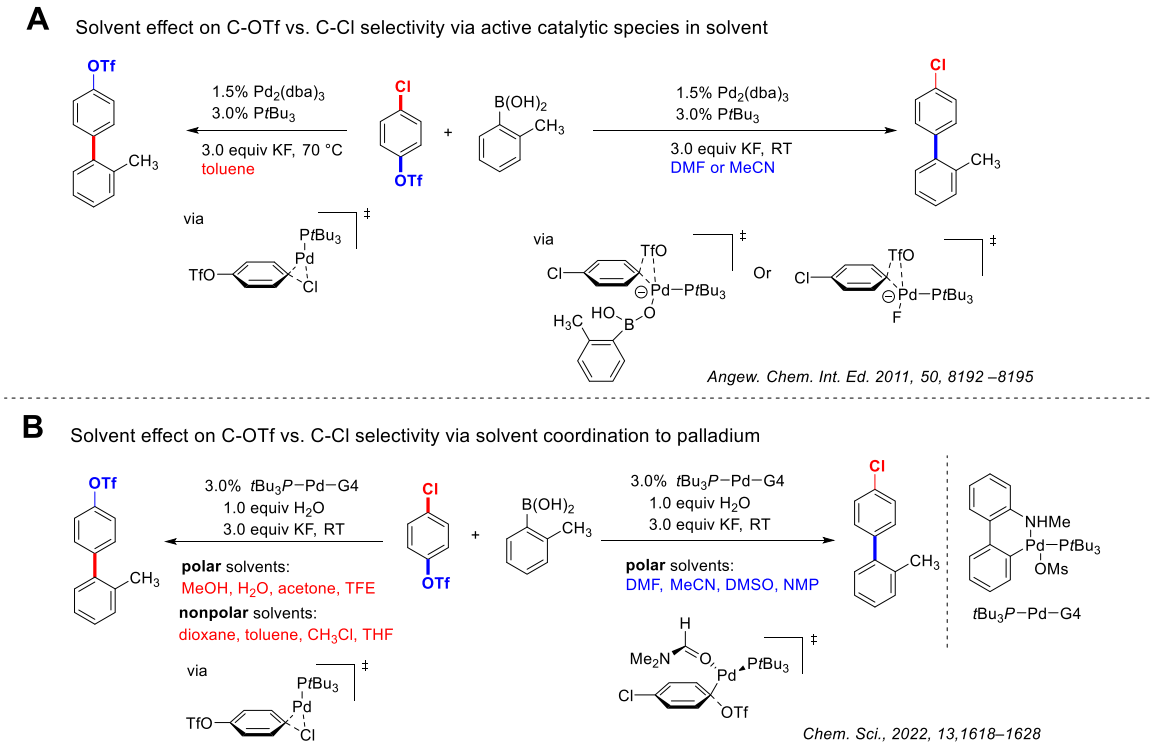
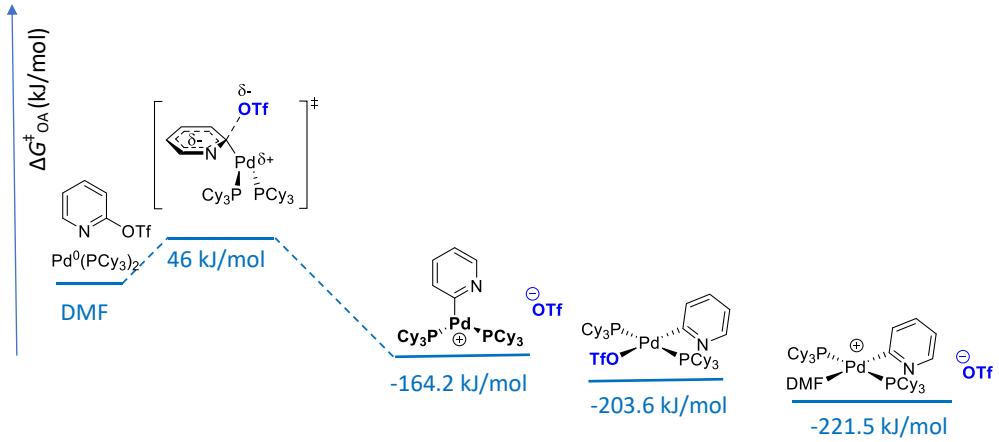


Figure 4.11. Solvent effect on C-OTf vs. C-Cl selectivity reported in literature.

For my studies, the oxidative addition of (hetero)aryl (pseudo)halides to $Pd(PCy_3)_2$ is done with no added halide anions or bases. Furthermore, solvent coordination to a bisligated Pd(0) complex should be disfavoured due to the steric hindrance, and would also generate a less reactive $L_3Pd(0)$ species. Nevertheless, I investigated the possibility of solvent-coordinated or charged species being involved in oxidative addition of 2-pyridyl triflate to $Pd(PCy_3)_2$ in DMF. DFT calculations failed to locate any stable DMF-coordinated $Pd(PCy_3)_2$ complex or DMF-coordinated π-complex intermediate prior to the oxidative addition transition state.

I did locate cationic Pd(II) species with triflate anions post-oxidative addition: Py-$Pd(PCy_3)_2^+$ and the solvent-coordinated trans-Py-Pd(DMF)$(PCy_3)_2^+$ (Fig. 4.12). The summed single point energies of the cationic Py-$Pd(PCy_3)_2^+$ and the triflate anion is 39.4 kJ mol^{-1} higher than the uncharged *trans*-Py-Pd(OTf)$(PCy_3)_2$ complex, while the summed single point energies of the *trans*-Py-Pd(DMF)$(PCy_3)_2^+$ and triflate anion is 17.9 kJ mol^{-1} lower than the summed energies of the uncharged *trans*-Py-Pd(OTf)$(PCy_3)_2$ complex and a DMF. In a mechanistic study of aryl triflate

oxidative addition to $Pd(PPh_3)_4$ in DMF by Jutand and Mosleh,[18] formation of the cationic Ar-$Pd(PPh_3)_2^+$ complex was confirmed experimentally. I located structurally similar cationic complexes by DFT calculations, and the electronic energy profile suggests that in DMF, the Pd(II) product is likely the solvent-coordinated cationic species, with the triflate anion stabilized by solvation with DMF.



TS search and geometry optimization of the possible product species at RI BP86/def2-SVP, def2/J, D3BJ/def2-TZVP(Pd)/CPCM; Single point energy at RI-B2PLYP/def2-TZVP, def2-TZVP/C, D3/CPCM.

Figure 4.12. Free energies of the oxidative addition product of 2-pyridyl triflate in DMF (CPCM) in the form of cationic Py-$Pd(PCy_3)_2^+$, uncharged trans-Py-Pd(OTf)$(PCy_3)_2$ and solvent-coordinated trans-Py-Pd(DMF)$(PCy_3)_2^+$.

Transition state analysis for oxidative addition of 2-chloropyridine and 2-pyridyl-triflate in toluene, THF and DMF (CPCM) enables further contrast between the mechanisms and resulting solvent effects. As the energy diagrams in Fig. 4.13 show, changing solvent model does not dramatically affect the transition state characteristics for each substrate. For 2-chloropyridine, there is a stable π-complex intermediate on the reaction coordinate, where the Pd(0) center coordinates to the C=N bond of the pyridine. The oxidative addition $TS^‡$ is characteristic of a nucleophilic displacement mechanism (metrical parameters for $TS^‡$ calculated in DMF CPCM): first, the short Pd–N (2.36 Å) length, the long Pd–Cl (3.01 Å) length, and the relatively obtuse C–Pd–Cl angle (92.4°) indicate an early transition state, with minimal Pd–Cl bonding; second, the $TS^‡$ has considerable partial negative charge delocalized on the pyridine (-0.41e) as well as partial positive charge at Pd

(+0.29e); third, there is a large increase in *ESP* at the Pd center from $Pd(PCy_3)_2$ to the transition state (ΔESP = +110.8 kJ mol^{-1}). These features of the nucleophilic displacement mechanism were previously outlined in Chapter 3.

For the 2-pyridyl triflate, I could not locate any stable pre-coordination intermediates, either to the C=N/C=C π-bond or elsewhere on the pyridine ring. In addition, calculations revealed a transition state structure different from either the nucleophilic displacement or the 3-centered transition states (Fig. 4.13B). Metrical parameters for the transition state in DMF CPCM indicate almost no coordination between Pd and N (Pd–N length = 2.91 Å). Second, the long Pd–OTf length (3.13 Å) and the obtuse Pd–C–OTf angle (110.0°) indicate minimal Pd–OTf bonding, inconsistent with a 3-centered TS‡. Third, this TS‡ exhibits an even greater degree of polarization than the nucleophilic displacement TS‡ structures, with a large partial negative charge on the pyridine (-0.50e), considerable partial positive charge at Pd (+0.32e), as well as a substantial increase in *ESP* at Pd center from $Pd(PCy_3)_2$ to the TS‡ (ΔESP = +140.4 kJ mol^{-1}). The calculated TS‡ energies follow the order of DMF < THF < toluene, which is consistent with the experimental trend.

This concerted oxidative addition pathway for 2-pyridyl triflate via a heavily charged transition state shares similar mechanistic features to the concerted S_NAr mechanism discussed in the introduction of Chapter 2. Here, the nucleophilic $Pd(PCy_3)_2$ attacks the electrophilic C–OTf site with simultaneous departure of the triflate anion. Notably, the calculated structure indicates no direct involvement of the adjacent C=N bond. This highly polarized S_NAr-like TS‡ for oxidative addition of 2-pyridyl triflate to $Pd(PCy_3)_2$ provides computational support to the hypothesis that polar solvents increase the rate of Ar–OTf oxidative addition to $L_2Pd(0)$ by stabilizing the polar TS‡.

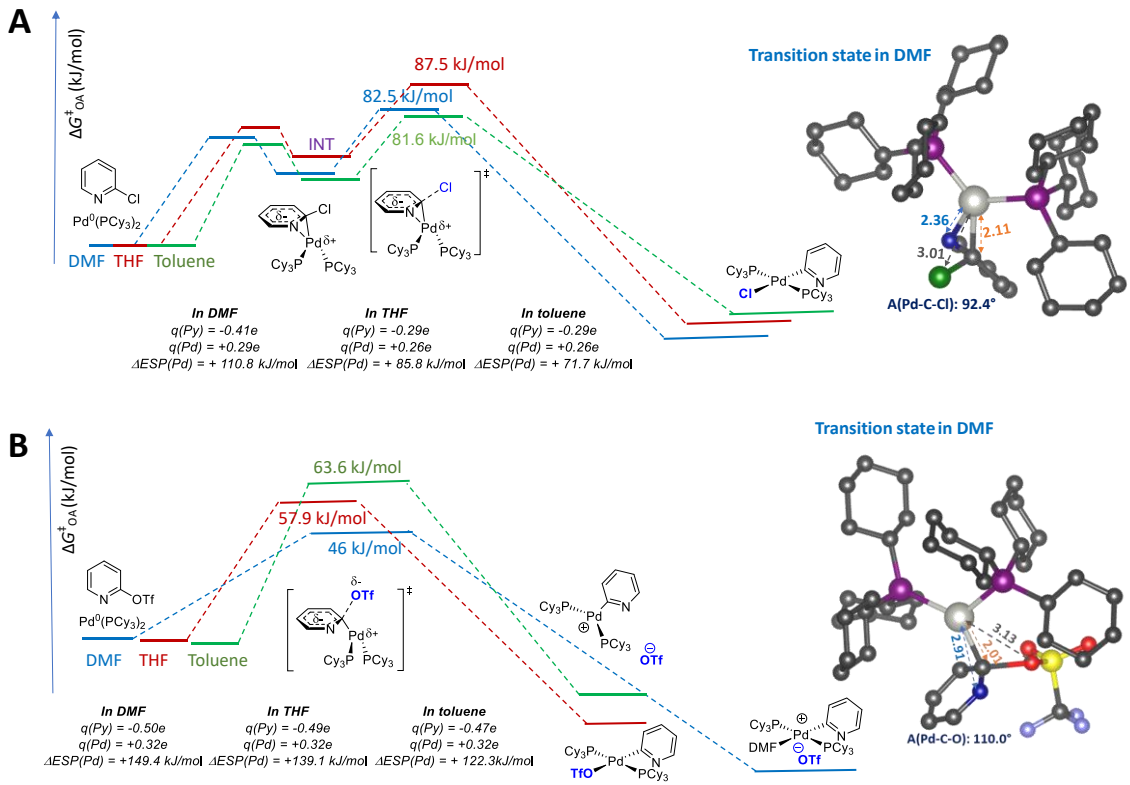


Figure 4.13. A) Calculated energy diagram for oxidative addition of 2-chloropyridine to $Pd(PCy_3)_2$ in DMF, THF and toluene. **B)** Calculated energy diagram for oxidative addition of 2-pyridyl triflate to $Pd(PCy_3)_2$ in DMF, THF and toluene. The blue/red/green pathways are the calculated reaction coordinate in DMF, THF and toluene, respectively. Transition state geometries were optimized in implicit solvation model CPCM at RI BP86 def2-SVP def2/J level with D3BJ dispersion for all atoms except for Pd, for which a def2-TZVP basis set was used; and the single point energy in implicit solvation model CPCM at RI-B2PLYP D3 def2-TZVP def2-TZVP/C level.

Finally, I studied solvent effects on C–OTf versus C-Br/Cl oxidative addition site-selectivity with the five substrates in Table 4.1. As described above, ligand controlled Pd-catalyzed Suzuki coupling of chloroaryl triflate was reported by Fu and coworkers,[17] and later studied computationally by Schoenebeck and Houk,[19] with PCy_3 favouring coupling at triflate and the bulky P*t*Bu$_3$ favouring coupling at chloride. For 2-chloro-5-triflatopyridine and 2-chloro-6-triflatopyridine, oxidative addition to $Pd(PCy_3)_2$ takes place exclusively at triflate regardless of the solvent

(entries 1-2). In contrast, selectivity is influenced by solvent for 2-chloro-3-triflatopyridine: oxidative addition is favoured at triflate in 1:1 THF/DMF, but there is reversed in nonpolar toluene, with oxidative addition at C–Cl slightly favoured (entry 3).

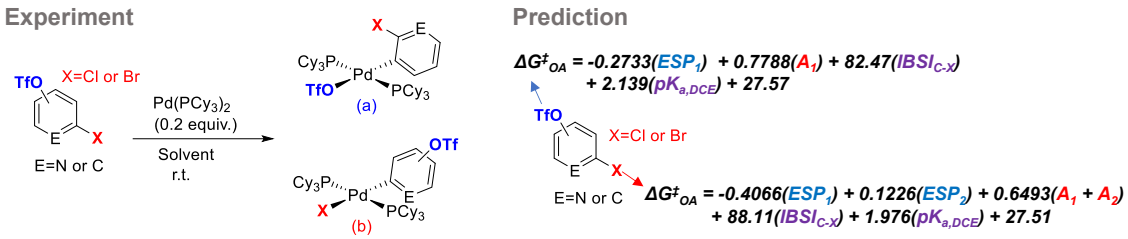
In 2011, Wang and coworkers[20] demonstrated ligand controlled selectivity between triflate and bromide sites in Pd-catalyzed Suzuki coupling. Analogous to Fu's work, bulky P*t*Bu$_3$ leads to reaction at bromide while the smaller PPh$_3$ leads to reaction at triflate. In 2017, Wu and coworkers[10] reported a combined ligand and solvent effect on a site-selective palladium-catalyzed carbonylation of bromoaryl triflates. Their study reveals that monodentate ligands P(*n*Bu)Ad$_2$ and P*t*Bu$_3$ prefer reaction at C–Br regardless of the solvent, while bidentate ligands require specific solvents to be selective: Xantphos and toluene leads to highly selective carbonylation at C–Br, while dppf and polar solvents such as DMF, DMSO and NMP achieve high selectivity at triflate.

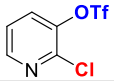
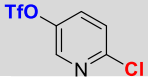
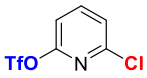
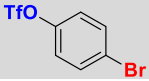
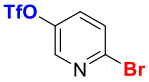
For the oxidative addition of 4-bromoaryl triflate to Pd(PCy$_3$)$_2$, triflate is the exclusive site for oxidative addition in all 3 solvents (Table 4.1, entry 4). In contrast, oxidative addition of 2-bromopyridyl-5-triflate exhibits a significant solvent effect. Oxidative addition at the bromide site is preferred in toluene (Br/OTf = ~7:1) while reaction at triflate is favoured in polar 1:1 THF/DMF (OTf/Br = 2:1, entry 5).

The site-selectivity predictions for the five substrates are also given in Table 4.1. The $\Delta G^{\ddagger}_{OA}$ values at the Cl/Br sites are from the five-descriptor model (Fig. 3.9C) and the $\Delta G^{\ddagger}_{OA}$ values at the triflate sites are from the four-descriptor model (Fig. 3.10B); both models are trained using reaction rate data measured in THF. For all five substrates, the predicted site-selectivity qualitatively matches the observed selectivity in THF by correctly identifying the major sites, though quantitative accuracy is lacking. Notably, when triflate is the predicted major site (entries 2 – 4), selectivity at Cl/Br is unlikely to be realized by just switching the solvent; however, for the substrates with poor predicted selectivity (entry 1) or Br/Cl as the predicted major site (entry 5), solvent-influenced selectivity is observed where increased Br/Cl selectivity is realized in nonpolar solvents. While more work is needed to expand

the number of investigated substrates and to refine quantitative accuracy, these predictions provide potentially useful information in guiding reaction condition selection and synthesis design for maximum selectivity.

Table 4.1. Experimental and predicted site-selectivity for oxidative addition of halopyridyl triflates to $Pd(PCy_3)_2$ in three solvent systems.



	Experimental a (C-OTf) : b(C-X)			Substrate	Predicted a (C-OTf) : b(C-X)
Entry	1:1 DMF/THF	THF	Toluene		
1	5:1	2:1	1:1.3		2:1
2	> 99:1	> 99:1	> 99:1		4:1
3	> 99:1	> 99:1	> 99:1		12:1
4	> 99:1	> 99:1	> 99:1		10:1
5	2:1	1:2	1:7		1:24

4.4 Conclusions

In summary, I expanded the initial quantitative structure-reactivity relationship for oxidative addition into the area of solvent effects. First, oxidative addition rates collected in three solvent systems enabled construction of individual reactivity models to evaluate potential solvent effects. Similar multivariate linear equations and general reactivity trends regardless of solvent identity indicate minimal changes to overall oxidative addition reactivity. Closer analysis of the quantitative reactivity scales in different solvents revealed several outliers, either as individual substrates or a substrate class with significant variation in reaction rates in different solvents. The outlier 2-chloro-3-aminopyridine revealed the importance of solvent H-bond basicity (pK_{HB}) when H-bond donating EDGs are present on the substrate. Computational analysis revealed that intermolecular H-bonding between solvent and substrate will deactivate the adjacent C–X sites toward oxidative addition. Thus, faster oxidative addition rates are observed in solvents with small pK_{HB}.

In addition, solvent polarity affects oxidative addition site-selectivity for pyridine derivatives with EWG substituents. Improved C_6/C_2 selectivity for 2,6-dichloro-3-EWG-pyridines in polar solvents is likely due to different mechanisms operating for reactions at the two sites. Higher C_6 selectivity in polar solvents is due to solvent stabilization of the polar nucleophilic displacement TS‡, while oxidative addition at C_2 likely proceeds via a 3-centered mechanism. This mechanistic dichotomy will be further explored in Chapter 5.

Finally, transition state analysis of the oxidative addition of 2-pyridyl triflate to $Pd(PCy_3)_2$ by DFT revealed a concerted S_NAr-like TS‡ that is even more polarized than the nucleophilic displacement TS‡ for 2-halpopyridines. Solvent effects on triflate/halide site selectivity for a group of 2-halopyridyl triflates are also consistent with highly polar transition states for C–OTf oxidative addition. Improved selectivity for oxidative addition at triflate occurs in polar solvents, though additional computational evidence is needed to confirm the viability of the S_NAr-like TS‡ for other Ar–OTf substrates. All of these insights should aid synthetic chemists in better predicting reaction conditions and designing selective syntheses.

4.5 References

(1) Proutiere, F.; Schoenebeck, F. Solvent Effect on Palladium-Catalyzed Cross-Coupling Reactions and Implications on the Active Catalytic Species. *Angew. Chem. Int. Ed.* **2011**, *50* (35), 8192–8195. https://doi.org/10.1002/anie.201101746.

(2) Hatanaka, Y.; Hiyama, T. Stereochemistry of the Cross-Coupling Reaction of Chiral Alkylsilanes with Aryl Triflates: A Novel Approach to Optically Active Compounds. *J. Am. Chem. Soc.* **1990**, *112* (21), 7793–7794. https://doi.org/10.1021/ja00177a048.

(3) Grimster, N. P.; Gauntlett, C.; Godfrey, C. R. A.; Gaunt, M. J. Palladium-Catalyzed Intermolecular Alkenylation of Indoles by Solvent-Controlled Regioselective C□H Functionalization. *Angew. Chem. Int. Ed.* **2005**, *44* (20), 3125–3129. https://doi.org/10.1002/anie.200500468.

(4) Yang, C.; Lee, H. M.; Nolan, S. P. Highly Efficient Heck Reactions of Aryl Bromides with N-Butyl Acrylate Mediated by a Palladium/Phosphine−Imidazolium Salt System. *Org. Lett.* **2001**, *3* (10), 1511–1514. https://doi.org/10.1021/ol015827s.

(5) Ohrai, K.; Kondo, K.; Sodeoka, M.; Shibasaki, M. Effects of Solvents and Additives in the Asymmetric Heck Reaction of Alkenyl Triflates: Catalytic Asymmetric Synthesis of Decalin Derivatives and Determination of the Absolute Stereochemistry of (+)-Vernolepin. *J. Am. Chem. Soc.* **1994**, *116* (26), 11737–11748. https://doi.org/10.1021/ja00105a014.

(6) Marset, X.; Khoshnood, A.; Sotorríos, L.; Gómez-Bengoa, E.; Alonso, D. A.; Ramón, D. J. Deep Eutectic Solvent Compatible Metallic Catalysts: Cationic Pyridiniophosphine Ligands in Palladium Catalyzed Cross-Coupling Reactions. *ChemCatChem* **2017**, *9* (7), 1269–1275. https://doi.org/10.1002/cctc.201601544.

(7) Cárdenas, J. C.; Fadini, L.; Sierra, C. A. Triphenylphosphite and Ionic Liquids: Positive Effects in the Heck Cross-Coupling Reaction. *Tetrahedron Lett.* **2010**, *51* (52), 6867–6870. https://doi.org/10.1016/j.tetlet.2010.10.104.

(8) Sherwood, J.; H. Clark, J.; S. Fairlamb, I. J.; M. Slattery, J. Solvent Effects in Palladium Catalysed Cross-Coupling Reactions. *Green Chem.* **2019**, *21* (9), 2164–2213. https://doi.org/10.1039/C9GC00617F.

(9) Huxoll, F.; Jameel, F.; Bianga, J.; Seidensticker, T.; Stein, M.; Sadowski, G.; Vogt, D. Solvent Selection in Homogeneous Catalysis—Optimization of Kinetics and Reaction Performance. *ACS Catal.* **2021**, *11* (2), 590–594. https://doi.org/10.1021/acscatal.0c04431.

(10) Shen, C.; Wei, Z.; Jiao, H.; Wu, X.-F. Ligand- and Solvent-Tuned Chemoselective Carbonylation of Bromoaryl Triflates. *Chem. – Eur. J.* **2017**, *23* (54), 13369–13378. https://doi.org/10.1002/chem.201702015.

(11) Elias, E. K.; Rehbein, S. M.; Neufeldt, S. R. Solvent Coordination to Palladium Can Invert the Selectivity of Oxidative Addition. *Chem. Sci.* **2022**, *13* (6), 1618–1628. https://doi.org/10.1039/D1SC05862B.

(12) Kamlet, M. J.; Taft, R. W. The Solvatochromic Comparison Method. I. The .Beta.-Scale of Solvent Hydrogen-Bond Acceptor (HBA) Basicities. *J. Am. Chem. Soc.* **1976**, *98* (2), 377–383. https://doi.org/10.1021/ja00418a009.

(13) Laurence, C.; Brameld, K. A.; Graton, J.; Le Questel, J.-Y.; Renault, E. The PKBHX Database: Toward a Better Understanding of Hydrogen-Bond Basicity for Medicinal Chemists. *J. Med. Chem.* **2009**, *52* (14), 4073–4086. https://doi.org/10.1021/jm801331y.

(14) Du, M.-X.; Yuan, Y.-F.; Zhang, J.-M.; Liu, C.-Y. Hydrogen-Bonding Interactions in Polymer–Organic Solvent Mixtures. *Macromolecules* **2022**, *55* (11), 4578–4588. https://doi.org/10.1021/acs.macromol.2c00799.

(15) Abraham, M. H.; Lieb, W. R.; Franks, N. P. Role of Hydrogen Bonding in General Anesthesia. *J. Pharm. Sci.* **1991**, *80* (8), 719–724. https://doi.org/10.1002/jps.2600800802.

(16) Wang, X.; Salaski, E. J.; Berger, D. M.; Powell, D. Dramatic Effect of Solvent Hydrogen Bond Basicity on the Regiochemistry of SNAr Reactions of Electron-Deficient Polyfluoroarenes. *Org. Lett.* **2009**, *11* (24), 5662–5664. https://doi.org/10.1021/ol902353t.

(17) Littke, A. F.; Dai, C.; Fu, G. C. Versatile Catalysts for the Suzuki Cross-Coupling of Arylboronic Acids with Aryl and Vinyl Halides and Triflates under Mild Conditions. *J. Am. Chem. Soc.* **2000**, *122* (17), 4020–4028. https://doi.org/10.1021/ja0002058.

(18) Jutand, A.; Mosleh, A. Rate and Mechanism of Oxidative Addition of Aryl Triflates to Zerovalent Palladium Complexes. Evidence for the Formation of Cationic (.Sigma.-Aryl)Palladium Complexes. *Organometallics* **1995**, *14* (4), 1810–1817. https://doi.org/10.1021/om00004a038.

(19) Schoenebeck, F.; Houk, K. N. Ligand-Controlled Regioselectivity in Palladium-Catalyzed Cross Coupling Reactions. *J. Am. Chem. Soc.* **2010**, *132* (8), 2496–2497. https://doi.org/10.1021/ja9077528.

(20) Wang, J.; Seefeld, M. A.; Luengo, J. Unusual Ligand-Dependent Chemoselective Suzuki–Miyaura Cross-Coupling Reactions of 3-Bromo-4-Trifloyl-Thiophenes. *Tetrahedron Lett.* **2011**, *52* (48), 6346–6348. https://doi.org/10.1016/j.tetlet.2011.09.026.

Chapter 5 – Mechanism and Site-Selectivity of Halopyridine Oxidative Addition to $L_2Pd(0)$ Are Controlled by Frontier Molecular Orbital Symmetry

Contributions: Experiments and computational analysis were performed by Jingru Lu. Single-crystal X-ray diffraction experiments and analysis were conducted by Dr. Nathan D. Schley of Vanderbilt University.

5.1 Abstract

This Chapter reports how the reaction mechanism and site-selectivity of 2-halopyridine oxidative addition to $L_2Pd(0)$ are both controlled by frontier molecular orbital symmetry. There are substantial oxidative addition rate differences between pairs of 2-chloro-3-EDG-pyridine / 2-chloro-5-EDG-pyridine ($\Delta\Delta G^{\ddagger}_{OA}$ = 10.4-11.6 kJ mol^{-1} favouring the 3 isomer), where EDG = NH_2, OMe and F. A mechanistic study reveals that the LUMO symmetries of those substrates control the oxidative addition mechanism. For the 3-EDG derivatives, high LUMO orbital coefficients at the reactive carbon and antibonding LUMO symmetries through the C_2=N bond enable oxidative addition to $L_2Pd(0)$ via the nucleophilic displacement mechanism, which is calculated to have a lower activation energy than the classic 3-centered mechanism; conversely, the LUMO of the 5-EDG derivatives has a node through the C_5–C_2 plane. Similarly, site-selectivity for a series of 2,6-dichloro-3-EDG-pyridines and 2-chloro-5-bromo-3-EDG-pyridines is controlled by the LUMO features. Incorporating simple and mechanistically meaningful frontier orbital based molecular descriptors to the oxidative addition model leads to improved prediction accuracy for both relative rates and site-selectivity of substituted 2-halopyridine oxidative addition to $L_2Pd(0)$.

5.2 Introduction

Palladium-catalyzed cross-coupling is a powerful synthetic method to form new bonds between two molecular fragments. Its wide application in pharmaceutical and chemical industries makes it among the most important reactions in modern organic synthesis.[1,2] Oxidative addition is often the rate and selectivity determining step of the catalytic cycle,[3] and understanding its mechanism is crucial for optimizing reaction conditions and improving selectivity in synthesis cases. The mechanism of palladium oxidative addition has been extensively studied among different types of organic (*pseudo*)halides and palladium complexes.[4–7] The oxidative addition mechanism is determined by a number of factors, including the specific structure of the electrophile, the identity of the (*pseudo*)halide, the coordination number of the Pd(0) species, and the structure of the ancillary ligand(s).

One common strategy to control site-selectivity in metal-catalyzed reactions is the incorporation of a coordinating directing group. Pre-coordination of the substrate to these groups places the catalytic center proximal to a specific reactive site, leading to selective functionalization. While many definitive characterized examples of directing group coordination are reported, with a few shown in Fig. 5.1,[8–13] there are also many cases where pre-coordination is invoked to explain site-selectivity with no additional evidence. This includes several cases of site-selective cross-coupling reactions. Herein, I offer an alternative explanation for specific substituent directing effects in Pd-catalyzed cross-coupling, based on changes to the oxidative addition mechanism controlled by frontier molecular orbital symmetry.

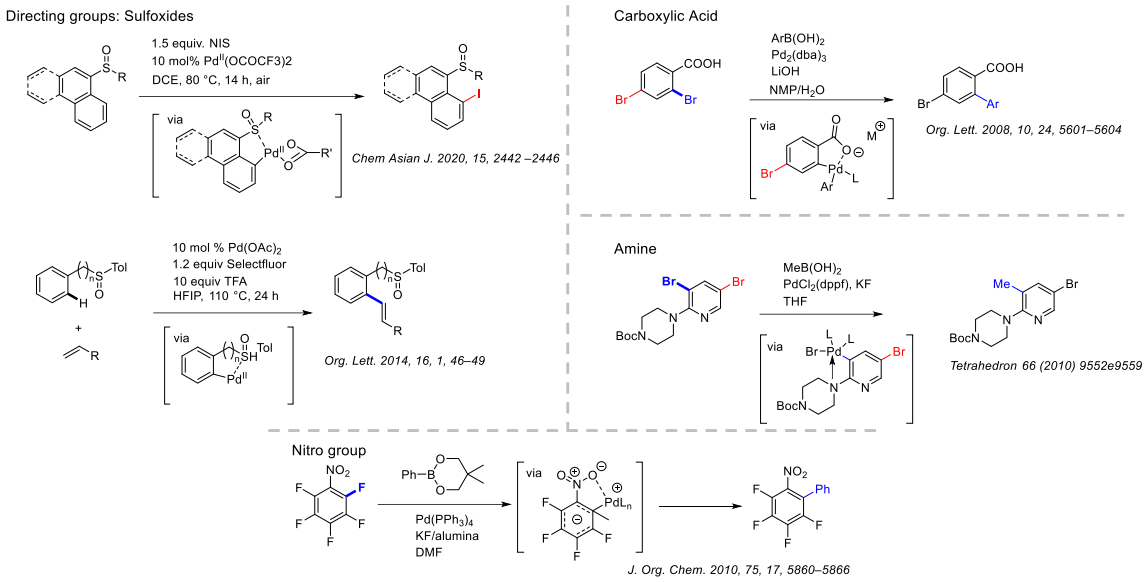


Figure 5.1. Representative literature examples of directing group pre-coordination in palladium-catalyzed cross-coupling reactions.

Frontier molecular orbital (FMO) theory was introduced by Fukui in the 1950s.[14] It is a good approximation for describing chemical bonding and reactions in terms of the interaction between the frontier orbitals of the reactants (e.g. HOMO and LUMO). FMO theory is now widely accepted by the organic chemistry community, finding many successful applications in mechanistic elucidation.[15–17] FMO considerations are often an essential impacting factor to the stereo- or regioselectivity in organic reactions. The frontier orbitals of a molecule determine a series of molecular properties, including the nucleophilicity and electrophilicity, which are among the fundamental concepts in organic chemistry reactivity. Two major attributes of the FMOs impact reaction outcomes: the energy level and the orbital localization. The energy level of FMOs indicate the accessibility for the corresponding electrons to interact in a reaction, and the orbital localization estimates which site is most suitable to undergo a specific type of transformation. For example, an electrophile with high LUMO density at one reactive site over another will undergo site-selective nucleophilic aromatic substitution (S_NAr). Computational chemists have developed a series of FMO-based molecular descriptors[18–21] that are often used to describe organic reactions involving a

nucleophile and an electrophile.[22–24] For example, the orbital localization-based descriptors, Fukui functions[25,26] and orbital coefficients, have been widely used in site-selectivity prediction in S_NAr.[27,28]

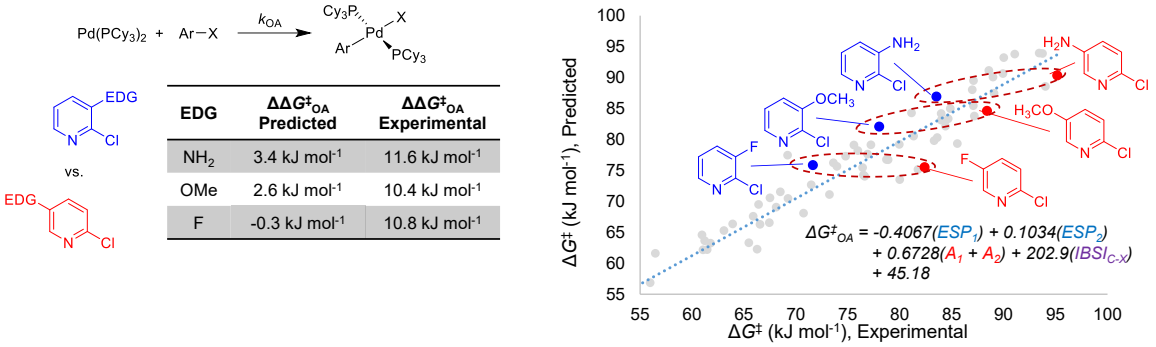
Reaction rate and/or regioselectivity controlled by frontier molecular orbital symmetry has been reported for many organic reaction systems. In a classic example, orbital overlap between diene and dienophile in Diels-Alder reactions leads to regio and stereoselective products.[29] In Chapter 2, I addressed several site-selectivity cases for S_NAr reactions where the two sites on a dihalopyridine undergo S_NAr using different antibonding orbitals (LUMO or LUMO+1) according to orbital symmetry.[30] Recently, Neufeldt's group reported site-selectivity in palladium-catalyzed cross-couplings controlled by the symmetry of the Pd(0) HOMO,[31] which will be further discussed below. In this Chapter, I consider the frontier molecular orbital properties of the substrates undergoing oxidative addition. Through experimental evidence and transition state analysis by DFT calculations, I demonstrate the importance of these frontier orbitals in determining the favoured reaction mechanism, and accordingly affecting the rate/site-selectivity in oxidative addition to Pd(0).

5.3 Results and Discussion

5.3.1 Mechanism of 2-Halopyridine Oxidative Addition Is Controlled by Frontier Orbital Symmetry

In Chapter 3, I reported a structure-reactivity relationship for oxidative addition to $L_2Pd(0)$ that is able to quantitatively predict relative reactivity and conventional site selectivity for a variety of substrates. This model uses simple, mechanistically-relevant molecular descriptors to account for electronic and steric effects, as well as (*pseudo*)halide identity. These include average molecular electrostatic potential (*ESP*) at specific atoms in the substrate, steric A-values[32], the intrinsic bond strength index (*IBSI*)[33] for the C–X bond strength, and pK_a values for the leaving group conjugate acid. The resulting quantitative reactivity and selectivity predictions are accurate not only within the training/test dataset, but also in external case studies with complex molecules. Importantly for the present work, the model parameters do not directly include any molecular orbital features.

While interrogating the few clear outlier substrates for possible evidence of mechanistic changes, I noted an apparent systematic discrepancy between predicted and experimental oxidative addition relative rates (given as $\Delta\Delta G^{\ddagger}_{OA}$) between three related substrate pairs: 2-chloro-3/5-aminopyridine, 2-chloro-3/5-methoxypyridine and 2-chloro-3/5-fluoropyridine (Fig. 5.2). These substrate pairings have very similar predicted oxidative addition rates (based on very similar ESP values at the key atoms), regardless of whether the electron-donating group (EDG) is in the 3- or 5-position ($\Delta\Delta G^{\ddagger}_{OA}$ = 0.3-3.4 kJ mol^{-1}); however, the difference in experimental oxidative addition rates is substantial ($\Delta\Delta G^{\ddagger}_{OA}$ = 10.4-11.6 kJ mol^{-1}). In each case, the 3-EDG substrates are faster than predicted, while the 5-EDG substrates are slower than predicted. One possibility is that the 3-amino and 3-methoxy groups could be coordinating to Pd(0) to accelerate the oxidative addition rate; however, this is unlikely for the 3-fluoro case. I also used DFT calculations in an attempt to locate putative intermediate Pd(0) structures and Pd(II) oxidative addition products that exhibit coordination to the 3-amino or 3-methoxy groups; however, no stable structures could be located.



EDG	$\Delta\Delta G^{\ddagger}_{OA}$ Predicted	$\Delta\Delta G^{\ddagger}_{OA}$ Experimental
NH_2	3.4 kJ mol^{-1}	11.6 kJ mol^{-1}
OMe	2.6 kJ mol^{-1}	10.4 kJ mol^{-1}
F	-0.3 kJ mol^{-1}	10.8 kJ mol^{-1}

Figure 5.2. Systematic discrepancy between predicted and experimental oxidative addition relative rates (given as $\Delta\Delta G^{\ddagger}_{OA}$, where $\Delta\Delta G^{\ddagger}_{OA}$ = $\Delta G^{\ddagger}_{OA}$ (2-Cl-5-EDG-py) – $\Delta G^{\ddagger}_{OA}$ (2-Cl-3-EDG-py)) between three related substrate pairs: 2-chloro-3/5-EDG-pyridine (EDG = NH_2, OMe, F).

Recently, Neufeldt and coworkers reported that site-selectivity in Pd-catalyzed cross-coupling of 2,4-dihalogenated pyridines is controlled by the HOMO symmetry of the Pd(0) species undergoing oxidative addition.[31] Specifically, a change in Pd-based HOMO symmetry from π in 14e$^-$ $L_2Pd(0)$ species to σ in 12e$^-$ LPd(0) species can override conventional site-selectivity, favouring oxidative addition at the 4-position of the pyridine rather than the 2-position. Inspired by this work, I hypothesized that orbital symmetry – which is not accounted for in the predictive models from Chapters 3 or 4 – could be playing a key role in the rate divergence between substrate pairings in Fig. 5.2. Specifically, changes to the LUMO symmetry of the substrates could result in different oxidative addition mechanisms.

To assess this hypothesis, I conducted a combined computational and experimental study of oxidative addition rate and site-selectivity for substituted halopyridines. I first analyzed the oxidative addition mechanisms/transition states for the series 2-chloro-3/4/5-EDG-pyridine (EDG = NH_2, OMe, F); the case of the –NH_2 derivatives is shown in Fig. 5.3 (–OMe and –F derivatives are in Table D6, Appendix D). These results are consistent with experimental observations, with $TS^{\ddagger}$ energies in the order 3-NH_2 < 4-NH_2 < 5-NH_2. Furthermore, the calculated $\Delta\Delta E^{\ddagger}_{OA}$ of 13.3 kJ mol^{-1} between the 3-NH_2 and 5-NH_2 transition states is quantitatively consistent with experiment ($\Delta\Delta G^{\ddagger}_{OA}$ = 11.6 kJ mol^{-1}). Importantly, these calculations reveal that two distinct mechanisms operate depending on the substituent location.

For the 3-NH_2 derivative, a discrete intermediate is observed along the reaction coordinate ($INT_{3\text{-}NH2}$), where the Pd(0) center coordinates to the C=N bond of the pyridine. The subsequent oxidative addition $TS^{‡}$ is characteristic of a nucleophilic displacement mechanism:[6] first, the short Pd–N (2.29 Å) length, the long Pd–Cl (3.10 Å) length, and the relatively obtuse C–Pd–Cl angle (95.4°) indicate an early transition state, with minimal Pd–Cl bonding; second, the $TS^{‡}$ has considerable partial negative charge delocalized on the pyridine (-0.34e) as well as partial positive charge at Pd (+0.29e); third, there is a large increase in *ESP* at the Pd center from $Pd(PCy_3)_2$ to the transition state (ΔESP = +89 kJ mol^{-1}). These characteristics are shared by the oxidative addition $TS^{‡}$ structures for the 3-OMe and 3-F derivatives (Figs D65-D70, Appendix D).

In contrast, the 4-NH_2 and 5-NH_2 pyridine derivatives undergo a more traditional 3-centered oxidative addition mechanism, with higher $\Delta E^{‡}_{OA}$ values than the 3-NH_2 oxidative addition $TS^{‡}$. For 4-NH_2, I located a pre-coordination intermediate with Pd(0) coordinated to a C=C π-bond rather than the C=N bond; this is also the case for the 4-OMe and 4-F derivatives. For the 5-NH_2 and 5-OMe derivatives, no stable pre-coordination intermediates appear along the reaction coordinate, either to a C=N or C=C π-bond or elsewhere on the pyridine ring; however, I did locate an intermediate for the 5-F derivative, with Pd(0) coordinated to the C=N π-bond.

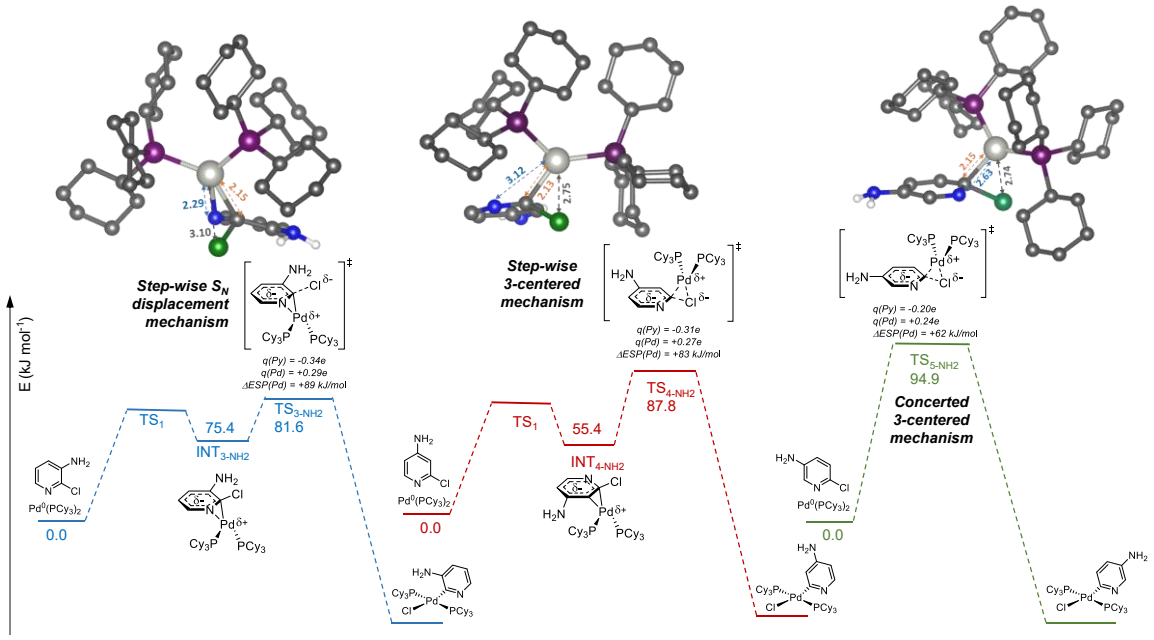


Figure 5.3. Calculated energy diagram for oxidative addition of 2-chloro-3/4/5-NH_2-pyridine to $Pd(PCy_3)_2$ in THF. The blue/red/green pathways are the oxidative addition coordinates of 3-NH_2, 4-NH_2 and 5-NH_2, respectively. Transition state geometries were optimized with CPCM(THF) at RI BP86 def2-SVP def2/J level with D3BJ dispersion for all atoms except for Pd, for which a def2-TZVP basis set was used; and the single point energy at RI-B2PLYP D3 def2-TZVP def2-TZVP/C level with CPCM(THF).

The structures of the 3-centered transition states have distinct characteristics, including long Pd–N (2.58-3.12 Å) and shorter Pd–Cl (2.74-2.81 Å) distances, more acute C–Pd–Cl angles (81.8-84.6°), and lesser partial charge magnitude on both the pyridine ring and the Pd center. In general, there is a smaller increase in *ESP* at Pd between $Pd(PCy_3)_2$ and these TS‡ structures relative to the corresponding nucleophilic displacement TS‡ (e.g. ΔESP = +83 and +62 kJ mol^{-1} for the 4-NH_2 and 5-NH_2 TS‡ structures respectively). All of these features point to a more symmetric, less polarized TS‡ than for the nucleophilic displacement TS‡.[6] In all cases, attempts to locate the alternative oxidative addition TS‡ type (3-centered for 3-EDG substrates, nucleophilic displacement for 4/5-EDG substrates) were not successful.

This change in oxidative addition mechanism for the 2-chloro-3/4/5-EDG-pyridine systems flows directly from frontier molecular orbital considerations. The 14e$^-$ $Pd(PCy_3)_2$ has a π-symmetric HOMO-1 orbital, which becomes the HOMO as the P–Pd–P bond is bent away from a linear geometry.[32] This enables Pd to

synergistically bond to the π systems of aromatic substrates, activating them toward oxidative addition. For the substrates, the -NH_2 substituted pyridines are illustrated as a representative case (Fig. 5.4). The LUMO of the 3-NH_2 derivative has a nodal plane cutting through the C_2=N bond, leading to antibonding symmetry between these two atoms, and the LUMO coefficient is large at both C_2 (15%) and N (22%). These LUMO properties are well-matched with the HOMO-1 of $L_2Pd(0)$, leading to strong bonding between Pd and the C_2=N bond in the π-complex intermediate, and further stabilization of the nucleophilic displacement transition state.

For the 4-NH_2 derivative, the LUMO coefficient is 13% at C_2; however, C_2 and N have π-bonding symmetry between the two atoms in the LUMO, which prevents $Pd(PCy_3)_2$ pre-coordination to the C_2=N bond. Instead, C_2 and C_3 are π-antibonding in the LUMO, making the C_2=C_3 bond the preferred site for $Pd(PCy_3)_2$ coordination. This is consistent with DFT calculations that reveal just such a C_2=C_3 pre-coordination intermediate for the 4-NH_2 derivative, followed by a 3-centered transition state with the $Pd(PCy_3)_2$ unit bonded to only the C_2 atom.

For the 5-NH_2 derivative, the LUMO coefficient at C_2 is very small (5%) due to a nodal plane cutting through C_2 and C_5. Such low orbital density at the reactive center indicates that this LUMO may not be the effective frontier orbital involved in the transformation at C_2. In contrast, the LUMO+1 coefficient at C_2 is much higher, which accounts for 19% of the orbital density. Visualizing the HOMO of the 3-centered transition state, which is formed from interaction between the bonding orbital of $Pd(PCy_3)_2$ and antibonding orbital of the substrate, reveals the LUMO+1 is the participating orbital for the 5-NH_2 derivative. In addition, the LUMO and LUMO+1 orbitals of pyridines are not degenerate due to the lower symmetry of these molecules. As LUMO+1 is higher in energy, it is expected that the $TS^{‡}$ for the oxidative addition of the 5-NH_2 derivative will also be higher in energy, leading to a slower rate. These characteristics of the frontier orbitals and their connections to the reaction mechanism and reactivity trend (3-EDG > 4-EDG > 5-EDG) are shared by the -OMe and -F groups (see Appendix D for specific details).

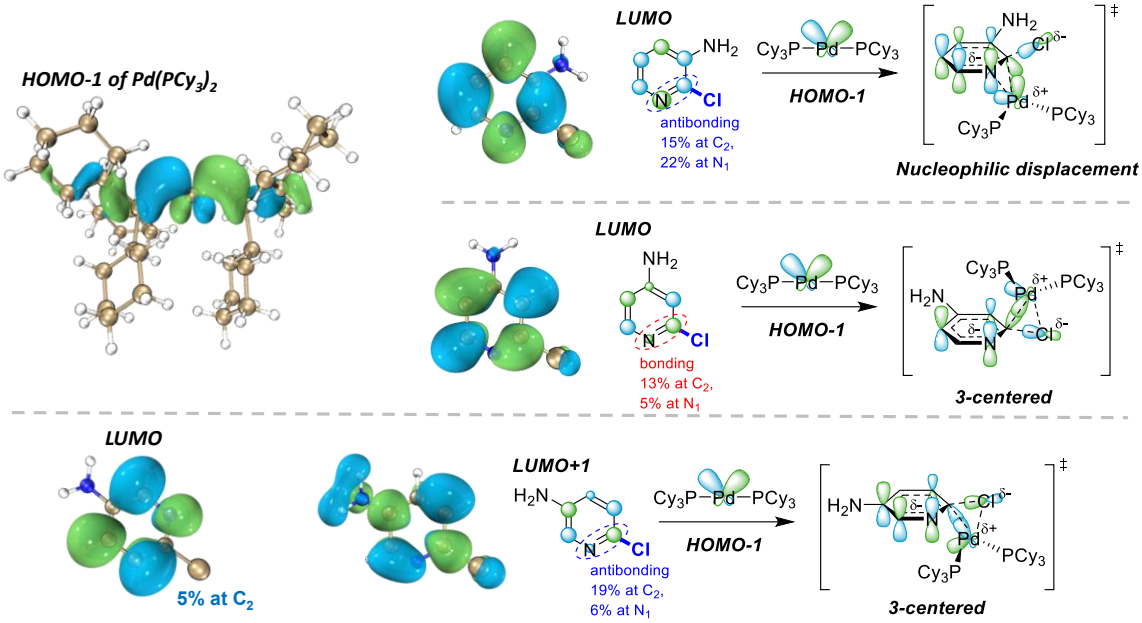


Figure 5.4. Key frontier orbitals of the $Pd(PCy_3)_2$ and the -NH_2 group that are involved in oxidative addition.

To further test this hypothesis, I studied the impact of frontier orbital properties to the oxidative addition rates with three more pairs of 2-chloro-3/5-EDG-pyridines (Fig. 5.5). The 3-EDG substrates, with high LUMO coefficients at C_2 and antibonding orbital symmetry at the C_2=N bond, are expected to proceed via the nucleophilic displacement TS‡. The 5-EDG substrates, with low LUMO coefficients at C_2, are expected to undergo a higher energy 3-centered pathway. The observed relative rates agree well with this mechanistic hypothesis, where $\Delta\Delta G^{\ddagger}_{OA}$ is 8.9 kJ mol^{-1} for the -O*i*Pr pair, and 7.6 kJ mol^{-1} for the -NHCOMe pair. The -O*t*Bu pair is notable: even with a sterically large group at the C_3 site, the 3-O*t*Bu is still 1.9 kJ mol^{-1} faster than the 5-O*t*Bu. This is an example where the frontier orbital impact is significant enough to override the steric effect, though sterics do cause the 3-O*t*Bu derivative to be slower than the 3-OiPr substrate.

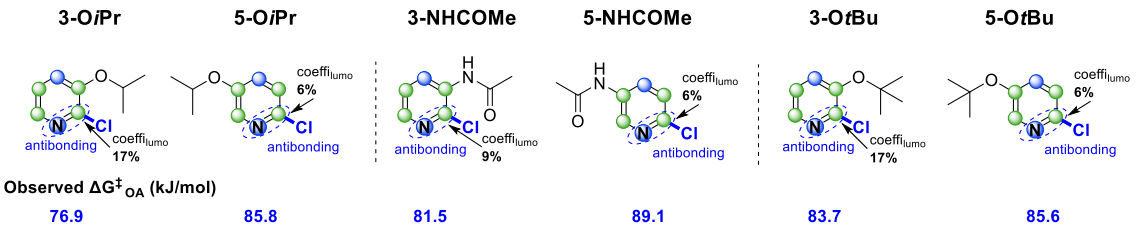


Figure 5.5. Experimental $\Delta G^{\ddagger}_{OA}$ values for 3-EDG / 5-EDG pairs in oxidative addition to $Pd(PCy_3)_2$.

Next, I investigated the influence of frontier orbital symmetry to site-selectivity with two groups of dihalogenated heterocycles. Group A contains eight 2,6-dichloro-3-R-pyridine/pyrazine substrates (R = EDG or EWG). The C_2 and C_6 sites have similar *ESP* values and therefore similar electronic characteristics; however, the different frontier orbital features at the two sites make them distinguishable from one another. For the five substrates in Fig. 5.6A with an EDG at the C_3 position, the C_2 site has the right LUMO properties for a lower energy oxidative addition pathway, via the nucleophilic displacement $TS^{‡}$, while the C_6 site would undergo oxidative addition via a 3-centered $TS^{‡}$ that is higher in energy. Once again, the observed site-selectivities match these qualitative predictions: C_2 is the major reactive site in all cases. For 4 out of 5 of the substrates, the structures of the major oxidative addition products were confirmed by single crystal X-ray diffraction experiments; the major product from 2,6-dichloro-3-OMe-pyridine was confirmed by NMR characterization data given in the supporting information of this Chapter (Appendix D).

In contrast, the three substrates in Fig. 5.6B have an EWG group at the C_3 position. The electronic effect from the EWG rearranges the frontier orbital density surface, and in this case makes C_6 the suitable site for the lower-in-energy nucleophilic displacement pathway. The observed site-selectivities agree with this frontier molecular orbital argument: C_6 is the major site in all three cases. The oxidative addition products were characterized by NMR, which is available in the supporting information of this Chapter (Appendix D).

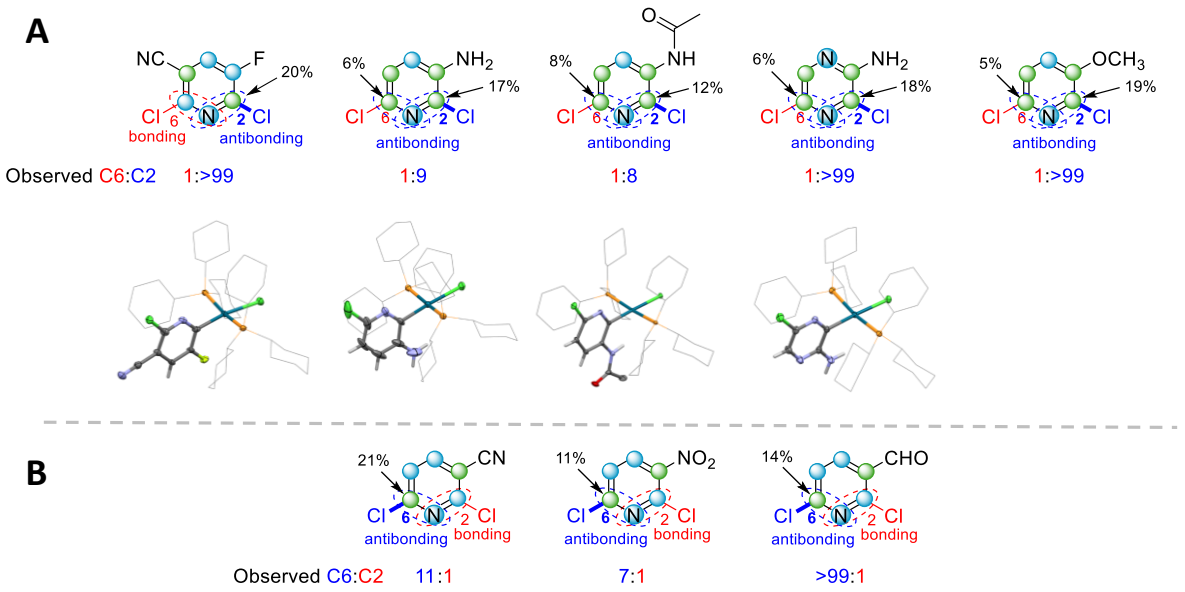


Figure 5.6. A) *Top:* LUMO features and observed site-selectivity for 2,6-dichloro-3-EDG-pyridines/pyrazine. *Bottom:* Solid-state molecular structures (via X-ray crystallography) of the major oxidative addition products. **B)** LUMO features and observed site-selectivity for 2,6-dichloro-3-EWG-pyridines/pyrazine

The four 2-chloro-5-bromo-3-EDG-pyridines in Group B (Fig. 5.7) reveal how intramolecular selectivity for mixed halide substrates can be affected by frontier orbital symmetry. Generally speaking, substrates with both -Br and -Cl leaving groups tend to react at the Br site, as documented extensively in the literature.[34–36] Select cases where C–Cl is the major reactive site are also reported; however, inverting the selectivity to favour Cl often requires extensive reaction development, such as by ligand screening.[37] For these substrates, I sought to determine whether FMO considerations alone could make the unconventional C-Cl the favoured site. For 2-chloro-5-bromopyridine, which has no additional substituents, the LUMO coefficient at C_2 is 7%. As discussed in Chapter 3, I observe no selectivity between the two sites in oxidative addition to $Pd(PCy_3)_2$ (C_2/C_5=1:1). For the latter three substrates with an EDG at the C_3 position, the LUMO coefficient at C_2 is increased to 14% ~ 17%. With the C_2 site accounting for more LUMO density, I do observe that C_2 becomes the dominant site in oxidative addition to $Pd(PCy_3)_2$, from a moderate C_2/C_5 selectivity of 2:1 for the 3-NH_2 derivative, to good C_2/C_5 selectivities of 6:1 for the 3-F derivative and 8:1 for the 3-OMe derivative.

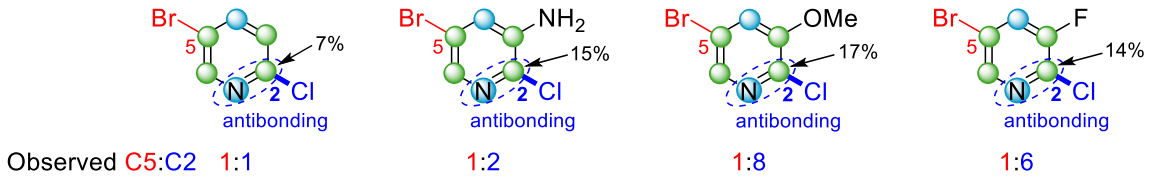


Figure 5.7. LUMO features and observed site-selectivity of 2-chloro-5-bromo-3-EDG-pyridines.

5.3.2 Investigating the Oxidative Addition Mechanism of 2-Chloro-6-EDG-Pyridines

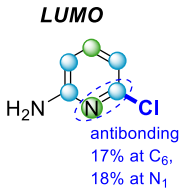
My investigation into the influence of frontier orbitals on the oxidative addition rates and reaction mechanisms for 2-chloro-6-EDG-pyridines raises questions. I compared the relative rates between the three 3-EDG and 6-EDG pairs (EDG = NH_2, OMe and F) and the results are shown in Table 5.1. The 3-EDG is much faster than its 6-EDG analogue experimentally ($\Delta\Delta G^{\ddagger}_{OA}$ = 9.9-11.3 kJ mol^{-1}); however, their oxidative addition rates are much closer by prediction, that the differences ($\Delta\Delta G^{\ddagger}_{OA}$) between the 3/6-$NH_2$ pair is 7.0 kJ mol^{-1}, the 3/6-OMe is 4.6 kJ mol^{-1}, and the predicted rates between the 3/6-F pair is almost identical ($\Delta\Delta G^{\ddagger}_{OA}$ = 0.7 kJ mol^{-1}). In each case, the 3-EDG substrates are faster than predicted, while the 6-EDG substrates are slower than predicted.

The 2-chloro-6-EDG pyridines share similar LUMO characteristics to their 3-EDG analogues: the C_6 site with high LUMO density and antibonding orbital symmetry cutting through the C_6=N bond; however, I did not observe the same rate accelerating effect in the 6-EDG derivatives as what I saw in the 3-EDG derivatives.

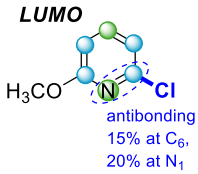
I conducted transition state analysis by DFT to study the reaction mechanism of the 2-chloro-6-EDG-pyridines. The nucleophilic displacement transition states were found in all three cases (6-NH_2, 6-OMe and 6-F). However, when I compare the calculated $\Delta E^{\ddagger}_{OA}$ of the 6-EDG to the 3/4/5-EDG in the same group, this number falls off the trend of the observed rates. For the 6-F derivative, although the $\Delta G^{\ddagger}_{OA}$ does not match the order of the observed rates, the value 82.4 kJ mol^{-1} is still in an acceptable error range. While the 6-NH_2 and 6-OMe have extremely low computed $\Delta G^{\ddagger}_{OA}$ values (66.1 and 66.7 kJ mol^{-1}), which make the two substrates the

outstanding outliers in each of the group. Investigation on the 6-EDG will be carried on for additional molecular features that may impact the oxidative addition rates.

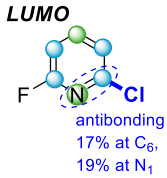
Table 5.1. LUMO features of the 2-chloro-6-EDG-pyridines; experimental $\Delta G^{\ddagger}_{OA}$ and computed $\Delta G^{\ddagger}_{OA}$ of the 3,4,5,6-EDG substrates in each group (EDG = NH_2, OMe and F).



	3-NH_2	4-NH_2	5-NH_2	6-NH_2
$\Delta G^{\ddagger}_{OA}$ Experimental (kJ/mol)	83.5	94.0	95.2	93.4
$\Delta G^{\ddagger}_{OA}$ Computed (kJ/mol)	81.6	87.8	94.9	66.1



	3-OMe	4-OMe	5-OMe	6-OMe
$\Delta G^{\ddagger}_{OA}$ Experimental (kJ/mol)	78.0	84.4	88.4	89.3
$\Delta G^{\ddagger}_{OA}$ Computed (kJ/mol)	80.4	88.8	87.2	66.7



	3-F	4-F	5-F	6-F
$\Delta G^{\ddagger}_{OA}$ Experimental (kJ/mol)	71.6	76.6	82.4	81.7
$\Delta G^{\ddagger}_{OA}$ Computed (kJ/mol)	75.8	84.9	85.1	82.4

5.3.3 Incorporating Frontier Orbital Based Descriptors into a Predictive Model

I have shown that the frontier orbital properties of the substrates are important in determining the reaction rate and site-selectivity of halo-pyridines in oxidative addition. The examples I discussed above demonstrate the predictive power of the two LUMO features: LUMO density at the reactive carbon, and the orbital phase between that carbon and its neighbouring nitrogen. I sought to find molecular orbital-based descriptors which I can incorporate into my predictive model, and I have found two of the LUMO descriptors showing great performance in quantitatively accounting for the frontier orbital influence and improving the predictive accuracy of my model.

One descriptor is the LUMO coefficient at the reactive carbon (*%lumo(C)*), a positive number for the LUMO density at that carbon; another one is to represent the phase – whether the orbital interaction between that carbon and its adjacent nitrogen is bonding or antibonding. Here I use the LUMO coefficient of the nitrogen (*%lumo(N)*) with signs (+/-) to represent the orbital phase and the LUMO density at the nitrogen: a positive number indicates that there is a bonding interaction between the C and N, and a negative number for antibonding interaction.

I put together a subset containing 48 2-chloro/bromo-pyridine derivatives from my reaction dataset for a multivariate linear model fit with the additional two LUMO descriptors. The model (Fig. 5.8B) constructed by six descriptors in total including the electronic and steric effect, bond strength and LUMO features, demonstrates excellent predictive ability based on the R^2 (0.93) and MAE (2.3 kJ mol^{-1}) values. I evaluated the robustness of this linear model by regression analysis of five random 70/30 training/test splits; one example is shown in Fig. 5.8C and the remainder in Fig. D71, Appendix D. The signs of the coefficients of the two LUMO descriptors are in good agreement with the mechanistic features: the negative coefficient in front of *%lumo(C)* indicate that higher LUMO density at the reactive carbon leading to faster reaction (lowering the reaction $\Delta G^{\ddagger}_{OA}$), while the positive coefficient in front of *%lumo(N)* indicates that antibonding symmetry through the C=N bond accelerates the reaction (lowers the reaction $\Delta G^{\ddagger}_{OA}$). I also examined the relative contribution of each descriptor to the prediction accuracy of this six-descriptor model by performing MLR using normalized descriptors by min-max rescaling. The relative contribution from each descriptor is shown in Table 5.2. Same as the predictive models I demonstrated in Chapter 3, ESP_1 here is still the top-ranked contributor weighing 52%. Contributions from the LUMO descriptors are small, with 2.8% from *%lumo(C)*, and 4.1% from *%lumo(N)*.

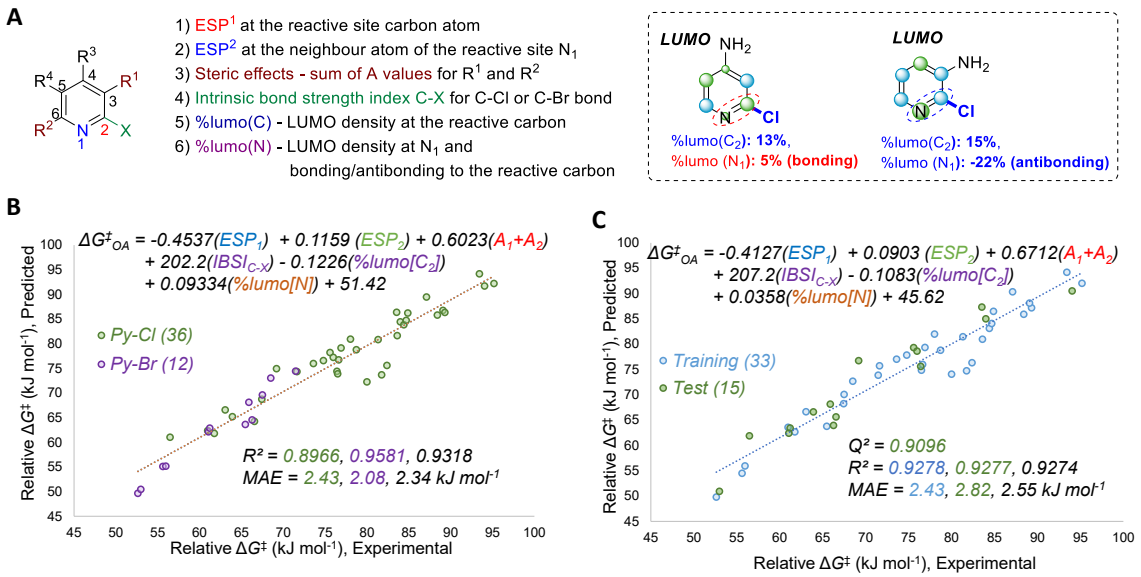


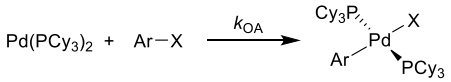
Figure 5.8. A) Molecular descriptors used to model oxidative addition reactivity as a function of substrate structure; examples of the LUMO-based descriptors for bonding (2-chloro-4-aminopyridine) and antibonding (2-chloro-3-aminopyridine) symmetries through the C=N bond. B). Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 48 bromo/chloro-pyridines in THF. C). Representative multivariate linear regression model generated using a 70/30 training/test split.

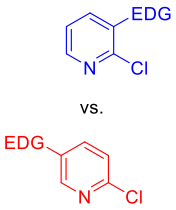
Table 5.2. Linear regression coefficients and % contribution to predicted $\Delta G^{\ddagger}_{OA}$ for min/max scaled descriptors.

Model	ESP_1	ESP_2	$A_1 + A_2$	IBSI	%lumo(C_2)	%lumo(N)
Fig. 4.9B	52.0%	13.0%	7.3%	20.8%	2.7%	4.1%

Next, I compared the prediction accuracy on relative rates for the three 2-chloro-3/5-EDG-pyridines pairs in Fig. 5.2 using the four-descriptor (*ESP₁, ESP₂, A₁+A₂, IBSI*) model in Fig. 5.2 with the six-descriptor (including the two LUMO descriptors) model in Fig. 5.8B. From the results shown in Table 5.3, the six-descriptor model still underestimates the relative rate ($\Delta\Delta G^{\ddagger}_{OA}$) between the 3/5-EDG pairs by 5.5 - 9.5 kJ mol^{-1}; however, I do see an improvement in prediction accuracy after accounting for the small contributions from the LUMO descriptors.

Table 5.3. Comparison of the prediction accuracy on relative rates ($\Delta\Delta G^{\ddagger}_{OA}$ = $\Delta G^{\ddagger}_{OA}$ (5-EDG) - $\Delta G^{\ddagger}_{OA}$ (3-EDG)) using the model in Fig. 5.2 with the model in Fig. 5.8B.

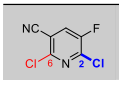
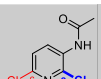
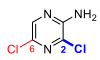
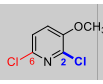


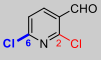


EDG	$\Delta\Delta G^{\ddagger}_{OA}$ Predicted (Fig. 5.2)	$\Delta\Delta G^{\ddagger}_{OA}$ Predicted (Fig. 5.8B)	$\Delta\Delta G^{\ddagger}_{OA}$ Experimental
NH_2	3.4 kJ mol^{-1}	5.8 kJ mol^{-1}	11.6 kJ mol^{-1}
OMe	2.6 kJ mol^{-1}	4.9 kJ mol^{-1}	10.4 kJ mol^{-1}
F	-0.3 kJ mol^{-1}	1.3 kJ mol^{-1}	10.8 kJ mol^{-1}

Then I compared the performance of the two models in making site-selectivity prediction for the 8 dichloropyridine derivatives in Fig. 5.6. Results of the 5 2,6-dichloro-3-EDG-pyridines are shown in Table 5.4 A, with the $\Delta\Delta G^{\ddagger}_{OA}$ being the relative rate between the major (C_2) and the minor (C_6) site ($\Delta\Delta G^{\ddagger}_{OA}$ = $\Delta G^{\ddagger}_{OA}$ (C_6) - $\Delta G^{\ddagger}_{OA}$ (C_2)). Prediction using the six-descriptor model gives $\Delta\Delta G^{\ddagger}_{OA}$ closer to the experimental $\Delta\Delta G^{\ddagger}_{OA}$ in all cases. The three 2,6-dichloro-3-EWG-pyridines in Table 5.4B have C_6 as the major site ($\Delta\Delta G^{\ddagger}_{OA}$ = $\Delta G^{\ddagger}_{OA}$ (C_2) - $\Delta G^{\ddagger}_{OA}$ (C_6)), and once again, the six-descriptor model including LUMO features makes better prediction for the $\Delta\Delta G^{\ddagger}_{OA}$.

Table 5.4. Comparison of the prediction accuracy on site-selectivity using the models before and after the LUMO features are included.

A. Substrate	$\Delta\Delta G^{\ddagger}_{OA}$ Predicted (Fig. 5.2)	$\Delta\Delta G^{\ddagger}_{OA}$ Predicted (Fig. 5.8B)	$\Delta\Delta G^{\ddagger}_{OA}$ Experimental
	0.4 kJ mol^{-1}	4.2 kJ mol^{-1}	>11.5 kJ mol^{-1} (>99:1)
	3.3 kJ mol^{-1}	5.8 kJ mol^{-1}	5.5 kJ mol^{-1} (9:1)
	1.8 kJ mol^{-1}	3.5 kJ mol^{-1}	5.2 kJ mol^{-1} (8:1)
	8.8 kJ mol^{-1}	11.8 kJ mol^{-1}	>11.5 kJ mol^{-1} (>99:1)
	6.8 kJ mol^{-1}	9.5 kJ mol^{-1}	>11.5 kJ mol^{-1} (>99:1)
$\Delta\Delta G^{\ddagger}_{OA}$ = $\Delta\Delta G^{\ddagger}_{OA}$ (C_6) - $\Delta\Delta G^{\ddagger}_{OA}$ (C_2)			

B. Substrate	$\Delta\Delta G^{\ddagger}_{OA}$ Predicted (Fig. 5.2)	$\Delta\Delta G^{\ddagger}_{OA}$ Predicted (Fig. 5.8B)	$\Delta\Delta G^{\ddagger}_{OA}$ Experimental
	0.0 kJ mol^{-1}	3.2 kJ mol^{-1}	6.0 kJ mol^{-1} (11:1)
	3.4 kJ mol^{-1}	4.0 kJ mol^{-1}	4.9 kJ mol^{-1} (7:1)
	6.8 kJ mol^{-1}	8.1 kJ mol^{-1}	>11.5 kJ mol^{-1} (>99:1)
$\Delta\Delta G^{\ddagger}_{OA}$ = $\Delta\Delta G^{\ddagger}_{OA}$ (C_2) - $\Delta\Delta G^{\ddagger}_{OA}$ (C_6)			

I took a few site-selectivity examples of 2,6-dihalopyridine derivatives in palladium-catalyzed reactions from literature for which the site-selectivity can be explained by frontier orbital symmetry (Fig. 5.9). Case A is a highly selective Stille coupling from the work by Nakamura and coworkers,[38] they reported that equimolar stannane and 3,5-dibromopyrazine under the given reaction condition yielded mostly the mono-substituted pyrazine at C_3 position. LUMO configuration of this substrate shows that the C_3 site, with high LUMO density and antibonding symmetry through the C_3=N bond, is preferable to coordinate to $L_2Pd(0)$ complex. Prediction using the model in Fig. 5.8B is consistent with the observed high selectivity at C_3.

The Suzuki-Miyaura coupling in Case B with the methyl 2,6-dichloronicotinate is reported to predominantly give the substituted product at C_6 site.[39] Based on the LUMO properties, C_6 is the favoured site when the reaction is catalyzed by $L_2Pd(0)$ catalyst. I predicted moderate selectivity between the two electronically identical sites, which is in excellent agreement with the observed 5:1 (C_6:C_2) selectivity. Another site-selective Suzuki-Miyaura coupling example shown in Case C was reported by Chung and coworkers,[40] they observed a ~10:1 ratio between the C_7 and C_5-subsituted products when the ligand PPh_3 is used. The LUMO properties at the two sites indicate that C_7 is preferred when $L_2Pd(0)$ is present, and prediction gives the correct major site (C_7), with a slightly lower C_7/C_5 ratio compared to the observed selectivity.

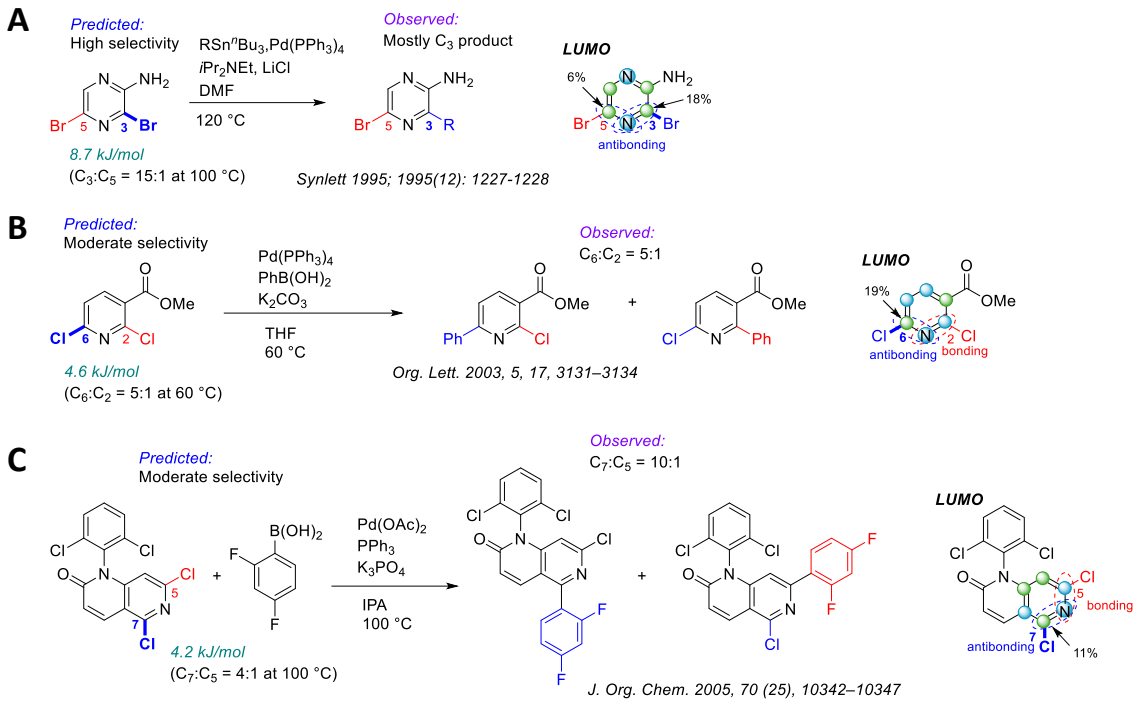
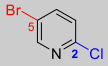
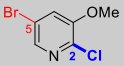
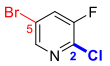


Figure 5.9. Quantitative selectivity predictions for dihalogenated heterocycles reported in literature. Coloured labels on the heterocycles correspond to predicted major site (blue), predicted minor site (red). The magnitude of the predicted $\Delta\Delta G^{\ddagger}_{OA}$ between the two sites is given in green.

Another major class of substrates in my database is the halobenzene derivatives, because of the less polarized C=C bond and weaker capability for the benzene ring to stabilize large amount of negative charges like the pyridyl ring does, the oxidative addition for this type of substrates follows the classic 3-centered mechanism according to the published work[4] and my computational study. I evaluated the model performance incorporating the frontier orbital features for prediction accuracy for the halobenzene derivatives, that I constructed a five-descriptor (*ESP*$_1$, *ESP*$_2$, A_1+A_2, *IBSI* and *%lumo(C)*) MLR model using a subset of 33 halobenzenes from my database. MLR using the normalized descriptors by min-max rescaling shows that contribution from the LUMO coefficient is trivial, only 0.06% to the prediction accuracy. This nearly zero contribution indicates that the LUMO properties is not relevant in controlling the oxidative addition outcome for this type of substrates. As a result, I use the four-descriptor model in Fig. 5.2 to make predictions for halobenzenes.

Table 5.5 shows the prediction comparison for the 4 2-chloro-5-bromopyridine derivatives. For the $\Delta\Delta G^{\ddagger}_{OA}$ values in column 3, the $\Delta G^{\ddagger}_{OA}$ of the C_2 site was obtained using the six-descriptor model in Fig. 5.8B; the C_5 is a halobenzene type site, and the $\Delta G^{\ddagger}_{OA}$ of this site is obtained using the four-descriptor model from Fig. 5.2. Before introducing the LUMO features, the model from Fig. 5.2 predicts C_5 as the preferred site, which is incorrect based on my experimental observation. After the LUMO features are included, the model makes better prediction by pointing out that the C_2 is the major site for the 3-NH_2 and 3-OMe derivatives, though the predicted $\Delta\Delta G^{\ddagger}_{OA}$ just qualitatively match with the experimental values. Prediction of the major site for the 3-F derivative is still incorrect; however, using the model incorporating LUMO features brings the $\Delta\Delta G^{\ddagger}_{OA}$ 1.3 kJ mol^{-1} closer to the experimental value.

Table 5.5. Comparison of the prediction accuracy on site-selectivity using the models before and after the LUMO features are included.

Substrate	$\Delta\Delta G^{\ddagger}_{OA}$ Predicted (Fig. 5.2)	$\Delta\Delta G^{\ddagger}_{OA}$ Predicted (Fig. 5.8B)	$\Delta\Delta G^{\ddagger}_{OA}$ Experimental
	-0.1 kJ mol^{-1}	0.2 kJ mol^{-1}	0 kJ mol^{-1} (1:1)
	-0.9 kJ mol^{-1}	0.3 kJ mol^{-1}	1.7 kJ mol^{-1} (2:1)
	-1.0 kJ mol^{-1}	1.0 kJ mol^{-1}	5.2 kJ mol^{-1} (8:1)
	-4.3 kJ mol^{-1}	-3.0 kJ mol^{-1}	4.9 kJ mol^{-1} (7:1)
$\Delta\Delta G^{\ddagger}_{OA} = \Delta\Delta G^{\ddagger}_{OA}(C_5) - \Delta\Delta G^{\ddagger}_{OA}(C_2)$			

5.4 Conclusions

In summary, from a combined experimental and computational study on reaction mechanisms and the resulting relative rate/site-selectivity for a series 2-chloropyridine derivatives oxidative addition to $L_2Pd(0)$, I have demonstrated that frontier molecular symmetry is a key feature leading to different mechanisms and reaction outcomes. The faster oxidative addition rate of the 2-chloro-3-EDG-pyridines than their 5-EDG analogues is based on the differences in frontier orbital features of these substrates. The LUMO properties of the 3-EDG derivatives – high LUMO density at C_2 and antibonding symmetry through the C=N bond – make this type of substrate capable of adding to $L_2Pd(0)$ via a polarized nucleophilic displacement $TS^{‡}$, which is lower in activation energy compared to the 3-centered $TS^{‡}$ found for the 5-EDG derivatives.

In addition, site-selectivity of specific types of substrates is controlled by frontier orbital properties. For the 2,6-dichloro-3-EDG/EWG-pyridines, the reactive site with LUMO properties matched to a nucleophilic displacement $TS^{‡}$ is the preferred site. Unconventional site-selectivity for the 2-chloro-5-bromo-3-EDG-pyridines, where the chlorine site is more reactive than the bromine site in intramolecular competition, is because these matched LUMO properties exist at C_2.

Finally, this work was inspired by the prediction outliers from the ESP-based oxidative addition model. This in turn helps to refine the model by incorporating newly discovered mechanistic aspects. Noting the importance of the frontier orbital properties in determining the reaction outcomes, I introduced two simple molecular descriptors to account for the LUMO density at the reactive carbon and antibonding/bonding symmetry through the C=N bond into the model. With the orbital based descriptors incorporated, there is improved prediction accuracy for both the relative rates and site-selectivity for 2-chloropyridine derivatives. I anticipate this work will help broaden our understanding of the mechanism of oxidative addition, a fundamental step in many catalytic reactions, and inspire more mechanistic studies by others in the field.

5.5 References

(1) Campeau, L.-C.; Hazari, N. Cross-Coupling and Related Reactions: Connecting Past Success to the Development of New Reactions for the Future. *Organometallics* **2019**, *38* (1), 3–35. https://doi.org/10.1021/acs.organomet.8b00720.

(2) Magano, J.; Dunetz, J. R. Large-Scale Applications of Transition Metal-Catalyzed Couplings for the Synthesis of Pharmaceuticals. *Chem. Rev.* **2011**, *111* (3), 2177–2250. https://doi.org/10.1021/cr100346g.

(3) Reeves, E. K.; Entz, E. D.; Neufeldt, S. R. Chemodivergence between Electrophiles in Cross-Coupling Reactions. *Chem. – Eur. J.* **2021**, *27* (20), 6161–6177. https://doi.org/10.1002/chem.202004437.

(4) Portnoy, M.; Milstein, D. Mechanism of Aryl Chloride Oxidative Addition to Chelated Palladium(0) Complexes. *Organometallics* **1993**, *12* (5), 1665–1673. https://doi.org/10.1021/om00029a026.

(5) Jutand, A.; Mosleh, A. Rate and Mechanism of Oxidative Addition of Aryl Triflates to Zerovalent Palladium Complexes. Evidence for the Formation of Cationic (.Sigma.-Aryl)Palladium Complexes. *Organometallics* **1995**, *14* (4), 1810–1817. https://doi.org/10.1021/om00004a038.

(6) Maes, B. U. W.; Verbeeck, S.; Verhelst, T.; Ekomié, A.; von Wolff, N.; Lefèvre, G.; Mitchell, E. A.; Jutand, A. Oxidative Addition of Haloheteroarenes to Palladium(0): Concerted versus SNAr-Type Mechanism. *Chem. – Eur. J.* **2015**, *21* (21), 7858–7865. https://doi.org/10.1002/chem.201406210.

(7) Mitchell, E. A.; Jessop, P. G.; Baird, M. C. A Kinetics Study of the Oxidative Addition of Bromobenzene to Pd(PCy3)2 (Cy = Cyclohexyl) in a Nonpolar Medium: The Influence on Rates of Added PCy3 and Bromide Ion. *Organometallics* **2009**, *28* (23), 6732–6738. https://doi.org/10.1021/om900679w.

(8) Palani, V.; Perea, M. A.; Sarpong, R. Site-Selective Cross-Coupling of Polyhalogenated Arenes and Heteroarenes with Identical Halogen Groups. *Chem. Rev.* **2022**, *122* (11), 10126–10169. https://doi.org/10.1021/acs.chemrev.1c00513.

(9) Hikawa, H.; Yokoyama, Y. Cross-Coupling Reaction on N-(3,5-Dibromo-2-Pyridyl)Piperazines: Regioselective Synthesis of 3,5-Disubstituted Pyridylpiperazines. *Tetrahedron* **2010**, *66* (49), 9552–9559. https://doi.org/10.1016/j.tet.2010.09.100.

(10) Houpis, I. N.; Huang, C.; Nettekoven, U.; Chen, J. G.; Liu, R.; Canters, M. Carboxylate Directed Cross-Coupling Reactions in the Synthesis of Trisubstituted Benzoic Acids. *Org. Lett.* **2008**, *10* (24), 5601–5604. https://doi.org/10.1021/ol802349u.

(11) Wang, B.; Shen, C.; Yao, J.; Yin, H.; Zhang, Y. Palladium(II)-Catalyzed Ortho-Olefination of Arenes Applying Sulfoxides as Remote Directing Groups. *Org. Lett.* **2014**, *16* (1), 46–49. https://doi.org/10.1021/ol402921w.

(12) Saito, H.; Yamamoto, K.; Sumiya, Y.; Liu, L.-J.; Nogi, K.; Maeda, S.; Yorimitsu, H. Palladium-Catalyzed C−H Iodination of Arenes by Means of Sulfinyl Directing Groups. *Chem. – Asian J.* **2020**, *15* (16), 2442–2446. https://doi.org/10.1002/asia.202000591.

(13) Cargill, M. R.; Sandford, G.; Tadeusiak, A. J.; Yufit, D. S.; Howard, J. A. K.; Kilickiran, P.; Nelles, G. Palladium-Catalyzed C−F Activation of Polyfluoronitrobenzene Derivatives in Suzuki−Miyaura Coupling Reactions. *J. Org. Chem.* **2010**, *75* (17), 5860–5866. https://doi.org/10.1021/jo100877j.

(14) Fukui, K.; Yonezawa, T.; Shingu, H. A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *J. Chem. Phys.* **2004**, *20* (4), 722–725. https://doi.org/10.1063/1.1700523.

(15) LEWIS, D. F. V. Frontier Orbitals in Chemical and Biological Activity: Quantitative Relationships and Mechanistic Implications*. *Drug Metab. Rev.* **1999**, *31* (3), 755–816. https://doi.org/10.1081/DMR-100101943.

(16) Houk, K. N. Frontier Molecular Orbital Theory of Cycloaddition Reactions. *Acc. Chem. Res.* **1975**, *8* (11), 361–369. https://doi.org/10.1021/ar50095a001.

(17) Vervoort, J.; Rietjens, I. M. C. M.; van BERKEL, W. J. H.; Veeger, C. Frontier Orbital Study on the 4-Hydroxybenzoate-3-Hydroxylase-Dependent Activity with Benzoate Derivatives. *Eur. J. Biochem.* **1992**, *206* (2), 479–484. https://doi.org/10.1111/j.1432-1033.1992.tb16950.x.

(18) Brinck, T.; Carlqvist, P.; Stenlid, J. H. Local Electron Attachment Energy and Its Use for Predicting Nucleophilic Reactions and Halogen Bonding. *J. Phys. Chem. A* **2016**, *120* (50), 10023–10032. https://doi.org/10.1021/acs.jpca.6b10142.

(19) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003**, *103* (5), 1793–1874. https://doi.org/10.1021/cr990029p.

(20) Koopmans, T. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. *Physica* **1934**, *1* (1), 104–113. https://doi.org/10.1016/S0031-8914(34)90011-2.

(21) Fukui, K.; Yonezawa, T.; Shingu, H. A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *J. Chem. Phys.* **2004**, *20* (4), 722–725. https://doi.org/10.1063/1.1700523.

(22) Wang, B.; Rong, C.; Chattaraj, P. K.; Liu, S. A Comparative Study to Predict Regioselectivity, Electrophilicity and Nucleophilicity with Fukui Function and Hirshfeld Charge. *Theor. Chem. Acc.* **2019**, *138* (12), 124. https://doi.org/10.1007/s00214-019-2515-1.

(23) Langenaeker, W.; Demel, K.; Geerlings, P. Quantum-Chemical Study of the Fukui Function as a Reactivity Index: Part 3. Nucleophilic Addition to α,β-Unsaturated compounds11For Part 2, See Ref. 6. *J. Mol. Struct. THEOCHEM* **1992**, *259*, 317–330. https://doi.org/10.1016/0166-1280(92)87022-R.

(24) Langenaeker, W.; Demel, K.; Geerlings', P. Quantum-Chemical Study of the Fukui Function as a Reactivity Index1 Part 2. Electrophilic Substitution on Mono-Substituted Benzenes.

(25) Li, Y.; Evans, J. N. S. The Fukui Function: A Key Concept Linking Frontier Molecular Orbital Theory and the Hard-Soft-Acid-Base Principle. *J. Am. Chem. Soc.* **1995**, *117* (29), 7756–7759. https://doi.org/10.1021/ja00134a021.

(26) Fuentealba, P.; Pérez, P.; Contreras, R. On the Condensed Fukui Function. *J. Chem. Phys.* **2000**, *113* (7), 2544–2551. https://doi.org/10.1063/1.1305879.

(27) Scales, S.; Johnson, S.; Hu, Q.; Do, Q.-Q.; Richardson, P.; Wang, F.; Braganza, J.; Ren, S.; Wan, Y.; Zheng, B.; Faizi, D.; McAlpine, I. Studies on the Regioselective Nucleophilic Aromatic Substitution (SNAr) Reaction of 2-Substituted 3,5-Dichloropyrazines. *Org. Lett.* **2013**, *15* (9), 2156–2159. https://doi.org/10.1021/ol4006695.

(28) O'Donnell, M.; Charrier, J.-D. A Complementary Route to Diaminopyrimidines through Regioselective SNAr Amination Reactions. *Tetrahedron* **2015**, *71* (10), 1515–1522. https://doi.org/10.1016/j.tet.2015.01.043.

(29) Martin, J. G.; Hill, R. K. Stereochemistry of the Diels-Alder Reaction. *Chem. Rev.* **1961**, *61* (6), 537–562. https://doi.org/10.1021/cr60214a001.

(30) Lu, J.; Paci, I.; Leitch, D. C. A Broadly Applicable Quantitative Relative Reactivity Model for Nucleophilic Aromatic Substitution (SNAr) Using Simple Descriptors. *Chem. Sci.* **2022**, *13* (43), 12681–12695. https://doi.org/10.1039/D2SC04041G.

(31) Norman, J. P.; Larson, N. G.; Neufeldt, S. R. Different Oxidative Addition Mechanisms for 12- and 14-Electron Palladium(0) Explain Ligand-Controlled Divergent Site Selectivity. *ACS Catal.* **2022**, *12* (15), 8822–8828. https://doi.org/10.1021/acscatal.2c01698.

(32) J. A. Hirsch *Topics in Stereochemistry,* John Wiley & Sons, Ltd, 1967, Pp. 199–222.

(33) Klein, J.; Khartabil, H.; Boisson, J.-C.; Contreras-García, J.; Piquemal, J.-P.; Hénon, E. New Way for Probing Bond Strength. *J. Phys. Chem. A* **2020**, *124* (9), 1850–1860. https://doi.org/10.1021/acs.jpca.9b09845.

(34) Large, J. M.; Clarke, M.; Williamson, D. M.; McDonald, E.; Collins, I. Synthesis of Trisubstituted Pyrimidines by Regioselective SNAr and Suzuki Reactions of Polyhalopyrimidines. *Synlett* **2006**, *2006* (6), 861–864. https://doi.org/10.1055/s-2006-939050.

(35) Montoir, D.; Tonnerre, A.; Duflos, M.; Bazin, M.-A. Efficient One-Pot Synthesis of 3,7-Disubstituted 1,6-Naphthyridin-2(1H)-Ones through Regioselective Palladium-Catalyzed Cross-Coupling and SNAr Reactions. *Tetrahedron* **2015**, *71* (21), 3303–3313. https://doi.org/10.1016/j.tet.2015.03.110.

(36) Kabri, Y.; Crozet, M. D.; Terme, T.; Vanelle, P. Efficient Access to 2,6,8-Trisubstituted 4-Aminoquinazolines through Microwave-Assisted One-Pot

Chemoselective Tris-Suzuki–Miyaura or SNAr/Bis-Suzuki–Miyaura Reactions in Water. *Eur. J. Org. Chem.* **2015**, *2015* (17), 3806–3817. https://doi.org/10.1002/ejoc.201500162.

(37) Keylor, M. H.; Niemeyer, Z. L.; Sigman, M. S.; Tan, K. L. Inverting Conventional Chemoselectivity in Pd-Catalyzed Amine Arylations with Multiply Halogenated Pyridines. *J. Am. Chem. Soc.* **2017**, *139* (31), 10613–10616. https://doi.org/10.1021/jacs.7b05409.

(38) Nakamura, H.; Takeuchi, D.; Murai, A. Synthesis of 5- and 3,5-Substituted 2-Aminopyrazines by Pd Mediated Stille Coupling. *Synlett* **1995**, *1995* (12), 1227–1228. https://doi.org/10.1055/s-1995-5249.

(39) Yang, W.; Wang, Y.; Corte, J. R. Efficient Synthesis of 2-Aryl-6-Chloronicotinamides via PXPd2-Catalyzed Regioselective Suzuki Coupling. *Org. Lett.* **2003**, *5* (17), 3131–3134. https://doi.org/10.1021/ol035188g.

(40) Chung, J. Y. L.; Cai, C.; McWilliams, J. C.; Reamer, R. A.; Dormer, P. G.; Cvetovich, R. J. Efficient Synthesis of a Trisubstituted 1,6-Naphthyridone from Acetonedicarboxylate and Regioselective Suzuki Arylation. *J. Org. Chem.* **2005**, *70* (25), 10342–10347. https://doi.org/10.1021/jo0514927.

Chapter 6 – Conclusions and Future Work

6.1 Summary and Conclusions

The objective of this thesis was to develop quantitative reactivity predictive models to assist reaction optimization and mechanistic elucidation in organic synthesis. Starting from a minireview in Chapter 1, I summarized the recent progress in data-rich approaches to reaction development and quantitative predictions for palladium-catalyzed reaction systems. High-quality reaction data from reliable sources are the basis to the success of any data-driven methodology; for palladium-catalyzed reactions, there are three major sources to obtain reaction data: literature reported reaction outcomes, computationally derived reaction-based parameters, and experimentally determined reaction data. I discussed those three approaches to reaction data collection with recently published studies, as well as the advantages and challenges of each approach in chemical space exploration and accurate reaction predictions.

Chapters 2 and 3 contain quantitative predictive models constructed for two pharmaceutically important reaction systems: nucleophilic aromatic substitution (S_NAr) and palladium oxidative addition, a fundamental and usually the rate/selectivity determining step in palladium-catalyzed cross-coupling reactions. Both models focused on the structure-reactivity relationship of the electrophiles. I developed both models by quantitatively mapping the molecular structures to the experimental reaction rates following the same general process. First, to obtain a training set consisting of high-quality reaction data, I used the high-throughput competition experimentation approach which allows us to collect reliable and diverse reaction rate datasets in a timely manner. Second, I selected a group of mechanistically meaningful molecular descriptors by performing univariate correlation to the reaction rates; instead of using the calculated activation energies from the computationally expensive transition state analysis, I only use simple and ground state descriptors which can be easily obtained from DFT calculations. Next, I constructed multivariate linear regression models by correlating the molecular

descriptors to the reaction rates and validated the prediction accuracy via a series of random train-test splits. In addition, to examine the robustness of the predictive models, I challenged the model performance by using them to make predictions on a wide variety of external reaction data. Using this approach combining high-throughput experimentation, DFT calculations and MLR analysis, I successfully achieved generally applicable models to make quantitative predictions on both the reaction rates and site-selectivity for both reaction systems (S_NAr and palladium oxidative addition).

Chapter 4 extends the application of the quantitative prediction model to the study of solvent effects in palladium oxidative addition. Using quantitative reactivity scales in different solvents, I identified cases exhibiting a significant solvent effect on the reaction outcomes. By investigating the role of solvent case by case, I discovered how solvent hydrogen bond basicity affects oxidative addition outcomes via its influence on the electronic properties of the electrophiles. Solvent polarity also influences site-selectivity of dichloropyridines via one type of mechanism being favoured over another. I also investigated the oxidative addition mechanism of 2-pyridyl triflate, and propose a concerted S_NAr-like mechanism with noticeable differences to the nucleophilic displacement and 3-centered mechanisms reported for (hetero)aryl halides.

In addition to making quantitative reaction predictions, a quantitative model constructed using high-quality data and mechanistically meaningful descriptors is very useful in gaining mechanistic insights because of the rich chemical information it incorporates. The mechanistic study described in Chapter 5 represents a good example of this feature. I identified a series of systematic prediction outliers from the ESP-based reactivity model for palladium oxidative addition. By investigating the underlying mechanistic causes, I found that in addition to the electronic features, the frontier orbital symmetry of the substrates is also an important factor in controlling the reaction mechanism and determining the reaction outcomes. This mechanistic information revealed by the model in return helps to improve its performance for both the predictive accuracy and applicable scope. I introduced

orbital-based descriptors into the oxidative addition model, leading to improved predictive accuracy for both the reaction rates and site-selectivity predictions for substrates which are previously outliers using the ESP-based model.

This thesis demonstrates the capabilities of the mechanistically driven quantitative structure-reactivity models in making accurate reaction predictions and in providing mechanistic insights. Future work includes extending the predictive power of the current models to additional reaction parameters, such as the nucleophiles in S_NAr and ligand effects in palladium oxidative addition, as well as constructing models for new types of chemical transformations, such as amide coupling and reductive elimination in palladium-catalysis.

6.2 Future Work

6.2.1 Quantitative Models for Amine Reactivity in Amide Coupling

Amide coupling is an extremely useful strategy for synthesizing a wide variety of active pharmaceutical ingredients.[1,2] According to a statistic given by Brown and Boström[3], amide coupling is top ranked among the most frequently occurring reactions in medicinal chemistry in 2014.

A multivariate linear regression model to predict amide coupling has been previously reported by the Sigman group and chemists at Pfizer.[4] Our proposed work is distinct from their prior work in two major aspects: the experimental approach for rate measurements, and the structural scope of amines investigated. In Sigman/Pfizer's work, the amide coupling rates were measured by traditional reaction progress kinetic analysis via monitoring the consumption of amine. We will collect the reaction rate data in a much more efficient way, via the high-throughput competition experimentation approach described in Chapter 2. Sigman/Pfizer have studied the structure and reactivity relationship of primary amines in amide coupling, and we will expand the investigated amine scope to include sterically bulky secondary amines and less reactive aniline derivatives.

As the reaction scheme in Fig. 6.1 shows, competition experiments will be conducted between two amines/anilines in excess but equal amounts with the limiting reagent being an acyl chloride. Acyl chlorides are a more reactive coupling partner compared to many activated carboxylic acid derivatives (e.g. the CDI-activated acids used in the Sigman/Pfizer work). Following the same procedures for the S_NAr reactivity model development, the reaction mixture will be analyzed by UPLC and a multivariate linear regression model for amine reactivity will be constructed by correlating the reaction rates to a group of simple ground state molecular descriptors. An initially proposed list of potentially useful amine descriptors contains electrostatic potential at the amine nitrogen, basicity of the amine, steric of the R group, HOMO energy of the amine and stretching frequency of the N-H bond.

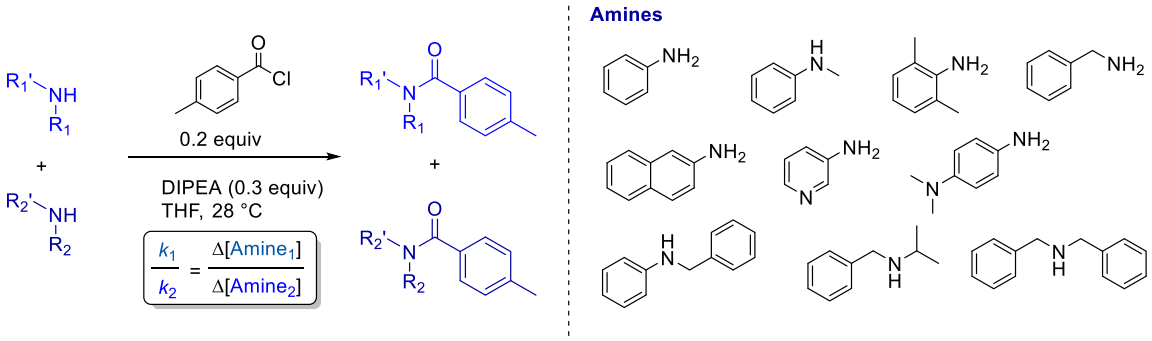


Figure 6.1. left: competition experiments for amide coupling to establish relative rates across amine library. right: representative amine structures from amine library.

6.2.2 Quantitative Model for Amine Reactivity in Buchwald-Hartwig Amination

Other strategies for C–N bond formation include S_NAr and Buchwald−Hartwig amination. The S_NAr reaction suffers from poor generality, since it requires aromatic compounds carrying strong electron-withdrawing groups. From preliminary screening results I of S_NAr reactions I obtained using cyclic amines as the nucleophile, only the reactions with extremely electron deficient heteroarenes achieved full conversion at 80 °C overnight (Fig. 6.2). The Buchwald-Hartwig

reaction, employing palladium catalysis, has dramatically expanded the substrate scope in C–N coupling, and it is now one of the most widely used transformations in medicinal chemistry.[2,5,6]

The rate-limiting step for Buchwald-Hartwig coupling is the C–N bond forming reductive elimination; therefore, constructing a quantitative amine reactivity scale would allow us to extend our QSAR study to another fundamental step in palladium catalysis. The competition experimental design is shown in Fig. 6.2; for ease of the reaction component analysis by UPLC, we plan to first isolate the oxidative addition product then have two amines competing with it. Synthesis and isolation of this oxidative addition product has been performed by our group,[7] with an isolated yield of 80% under the given condition. The DAB-Pd-$(CH_2TMS)_2$ used for this is an excellent Pd(II) precursor for oxidative addition that our group has synthesized recently.[7]

Diverse and internally-consistent datasets for amine reactivity in amide coupling and Buchwald-Hartwig amination would allow us to build quantitative predictive models for two of the most widely used C–N coupling strategies in medicinal chemistry. Also, comparison of the two amine reactivity scales may reveal interesting and unexpected mechanistic aspects for us to explore.

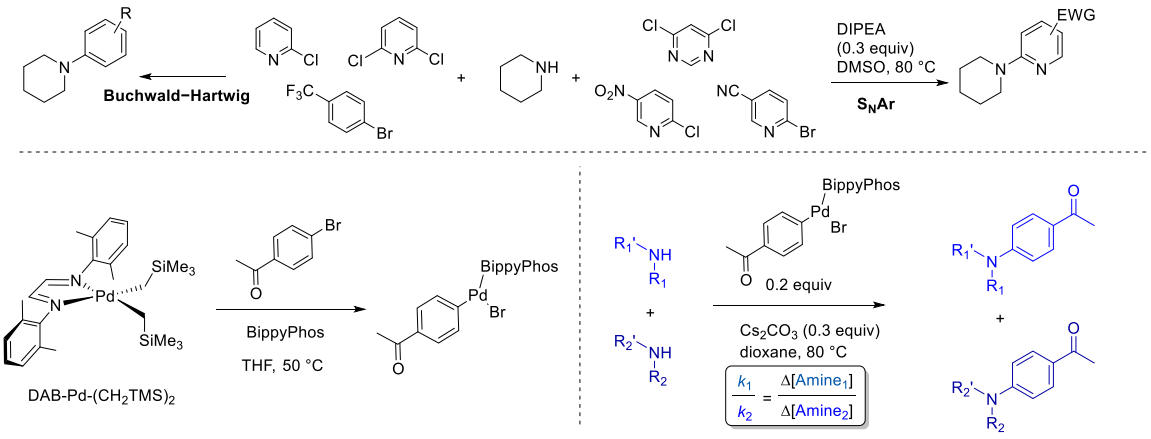


Figure 6.2. top: limited substrate scope for S_NAr: only the extremely electron deficient heteroarenes form C-N bond with amines via S_NAr. bottom left: synthesis of the oxidative addition product from the $Pd^{(II)}$ precursor DAB-Pd-$(CH_2TMS)_2$. Bottom right: competition experiments for Buchwald-Hartwig coupling to establish relative rates across amine library.

6.2.3 Ligand Effects in Palladium-Catalyzed Reactions

Ligand effects play a central role in homogeneous organometallic catalysis, since they usually determine the catalyst stability and efficiency.[8,9] In addition, ligand identity is reported to control the regio/stereoselectivity in many catalyzed reaction systems.[10,11] In an effort to study ligand effects in palladium oxidative addition beyond PCy_3, I made initial attempts using the two catalyst systems discussed below.

$Pd[P(t\text{Bu}_3)]_2$ is a frequently used catalyst among the bisligated palladium class.[12,13] Due to the steric bulk of $P(t\text{Bu}_3)$, this complex tends to undergo dissociation of one $P(t\text{Bu}_3)$ ligand and participate in oxidative addition in the form of monoligated $Pd[P(t\text{Bu}_3)]$.[14] $Pd[P(t\text{Bu}_3)]_2$ has been reported to enable oxidative addition of aryl chlorides under mild reaction conditions via the monoligated palladium oxidative addition pathway, and to achieve high C–Cl selectivity in Suzuki coupling of chlorophenol triflates.[15,16] For my preliminary trial, I conducted oxidative addition using $Pd(t\text{Bu}_3)_2$ and a series of 2-halopyridines. As the ^{31}P NMR spectra of the oxidative addition reaction mixtures in Fig. 6.3 show, one challenge to spectroscopic analysis is that the chemical shifts of different oxidative addition products are in a very narrow range. As a result, there is a high risk for peak overlap when conducting competition reactions between two substrates. In addition, I observe poor reactivity under the current reaction conditions, as indicated by the large peak at 85.3 ppm, which corresponds to unreacted $Pd[P(t\text{Bu}_3)]_2$.[14]

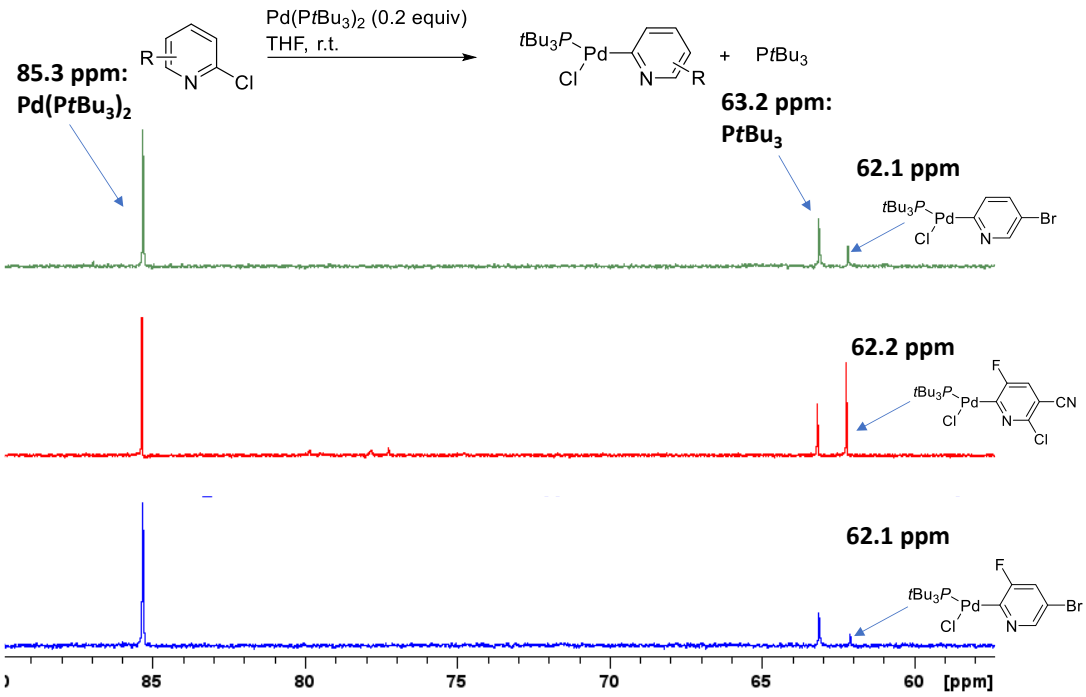


Figure 6.3. ^{31}P NMR spectra for oxidative addition of 2-chloro-5-bromopyridine (green), 2,6-dichloro-3-fluoro-5-cyanopyridine (red), 2-chloro-5-bromo-3-fluoropyridine (blue) to Pd(tBu$_3$)$_2$ in THF at room temperature.

I also investigated the reactivity of a group of frequently used phosphine ligands (*t*BuXPhos, XPhos, SPhos, RuPhos, *t*BuBippyPhos) by premixing the ligands with Pd_2dba_3•$CHCl_3$ to generate the active Pd(0) complexes. However, I observed poor reactivity for oxidative addition and unclear ^{31}P NMR spectra due to the appearance of multiple peaks.

Overall, investigating ligand effects remains challenging under the current reaction conditions. Additional effort is needed both in optimizing the reaction conditions and improving suitability of the analytical methods. The aforementioned DAB-Pd-$(CH_2TMS)_2$ precursor may be useful for these investigations.

6.3 References

(1) Montalbetti, C. A. G. N.; Falque, V. Amide Bond Formation and Peptide Coupling. *Tetrahedron* **2005**, *61* (46), 10827–10852. https://doi.org/10.1016/j.tet.2005.08.031.

(2) Dunetz, J. R.; Magano, J.; Weisenburger, G. A. Large-Scale Applications of Amide Coupling Reagents for the Synthesis of Pharmaceuticals. *Org. Process Res. Dev.* **2016**, *20* (2), 140–177. https://doi.org/10.1021/op500305s.

(3) Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **2016**, *59* (10), 4443–4458. https://doi.org/10.1021/acs.jmedchem.5b01409.

(4) Haas, B. C.; Goetz, A. E.; Bahamonde, A.; McWilliams, J. C.; Sigman, M. S. Predicting Relative Efficiency of Amide Bond Formation Using Multivariate Linear Regression. *Proc. Natl. Acad. Sci.* **2022**, *119* (16), e2118451119. https://doi.org/10.1073/pnas.2118451119.

(5) Torborg, C.; Beller, M. Recent Applications of Palladium-Catalyzed Coupling Reactions in the Pharmaceutical, Agrochemical, and Fine Chemical Industries. *Adv. Synth. Catal.* **2009**, *351* (18), 3027–3043. https://doi.org/10.1002/adsc.200900587.

(6) Ruiz-Castillo, P.; Buchwald, S. L. Applications of Palladium-Catalyzed C–N Cross-Coupling Reactions. *Chem. Rev.* **2016**, *116* (19), 12564–12649. https://doi.org/10.1021/acs.chemrev.6b00512.

(7) Huang, J. Development of Stable, Active, and High-Throughput Experimentation Compatible Palladium Precatalysts. Thesis, 2023. https://doi.org/10.26434/chemrxiv-2023-0xbh5.

(8) Popeney, C. S.; Guan, Z. Effect of Ligand Electronics on the Stability and Chain Transfer Rates of Substituted Pd(II) α-Diimine Catalysts. *Macromolecules* **2010**, *43* (9), 4091–4097. https://doi.org/10.1021/ma100220n.

(9) Marion, N.; Nolan, S. P. Well-Defined N-Heterocyclic Carbenes−Palladium(II) Precatalysts for Cross-Coupling Reactions. *Acc. Chem. Res.* **2008**, *41* (11), 1440–1449. https://doi.org/10.1021/ar800020y.

(10) Schoenebeck, F.; Houk, K. N. Ligand-Controlled Regioselectivity in Palladium-Catalyzed Cross Coupling Reactions. *J. Am. Chem. Soc.* **2010**, *132* (8), 2496–2497. https://doi.org/10.1021/ja9077528.

(11) Dai, X.; Chen, Y.; Garrell, S.; Liu, H.; Zhang, L.-K.; Palani, A.; Hughes, G.; Nargund, R. Ligand-Dependent Site-Selective Suzuki Cross-Coupling of 3,5-Dichloropyridazines. *J. Org. Chem.* **2013**, *78* (15), 7758–7763. https://doi.org/10.1021/jo401096u.

(12) Littke, A. F.; Schwarz, L.; Fu, G. C. Pd/P(t-Bu)3: A Mild and General Catalyst for Stille Reactions of Aryl Chlorides and Aryl Bromides. *J. Am. Chem. Soc.* **2002**, *124* (22), 6343–6348. https://doi.org/10.1021/ja020012f.

(13) Dai, C.; Fu, G. C. The First General Method for Palladium-Catalyzed Negishi Cross-Coupling of Aryl and Vinyl Chlorides: Use of Commercially Available Pd(P(t-Bu)3)2 as a Catalyst. *J. Am. Chem. Soc.* **2001**, *123* (12), 2719–2724. https://doi.org/10.1021/ja003954y.

(14) Bonney, K. J.; Proutiere, F.; Schoenebeck, F. Dinuclear Pd(I) Complexes—Solely Precatalysts? Demonstration of Direct Reactivity of a Pd(I) Dimer with an Aryl Iodide. *Chem. Sci.* **2013**, *4* (12), 4434–4439. https://doi.org/10.1039/C3SC52054D.

(15) Proutiere, F.; Schoenebeck, F. Solvent Effect on Palladium-Catalyzed Cross-Coupling Reactions and Implications on the Active Catalytic Species. *Angew. Chem. Int. Ed.* **2011**, *50* (35), 8192–8195. https://doi.org/10.1002/anie.201101746.

(16) Elias, E. K.; Rehbein, S. M.; Neufeldt, S. R. Solvent Coordination to Palladium Can Invert the Selectivity of Oxidative Addition. *Chem. Sci.* **2022**, *13* (6), 1618–1628. https://doi.org/10.1039/D1SC05862B.

Appendices

Appendix A: Supporting Information for Chapter 2

Other Supporting information for this chapter include the following:

Lu-PhD-thesis_Ch.2 SI tables. (separate file; *.xlsx, electronically available at *https://doi.org/10.20383/103.0795*)
Extended tables in Microsoft Excel format (.xlsx) that contain list of competition reactions, measured $\Delta G^{\ddagger}_{SNAr}$ and molecular descriptors.

Lu-PhD-thesis_Ch.2_MLR models. (separate file; *.xlsx, electronically available at *https://doi.org/10.20383/103.0795*)
Extended tables in Microsoft Excel format (.xlsx) that contain predicted $\Delta G^{\ddagger}_{SNAr}$ calculations, and statistical analysis for multivariate linear regression models.

General Considerations

Materials

All solvents, reagents, and organic substrates were used as purchased from commercial suppliers without further purification with the following exceptions. 4-(6-chloropyrimidin-4-yl)morpholine[1], 4-chloro-6-(pyrrolidin-1-yl)pyrimidine[1], 2-chloro-5-(1,3-dioxolan-2-yl)pyridine[2], 2-chloro-6-isopropoxypyridine[3] and 2-chloro-6-(tert-butoxy)-pyridine[4] were prepared using published procedures; 2-(benzyloxy)-4-chloropyridine and 4-(benzyloxy)-2-chloropyridine were prepared using the general procedures in the *Preparative Scale Synthesis of SNAr product* section. 4-Chloro-2-methylpyrimidine was purified by dissolving the commercial material in chloroform, followed by filtration to remove insoluble impurities; purity of this material was confirmed by NMR spectroscopy. Potassium *tert*-butoxide was purified by dissolving the commercial material in anhydrous THF, followed by filtration to remove insoluble impurities. The solvent was then evaporated *in vacuo.* This purification was performed inside an MBraun glovebox under an N_2 atmosphere.

Analysis and Spectroscopy

All ^{1}H and ^{13}C NMR spectra were recorded on either a Bruker AVANCE 300 MHz spectrometer or a Bruker AVANCE NEO 500 MHz spectrometer.

High-resolution electrospray ionization mass spectrometric analysis was performed using a Thermo Scientific Ultimate 3000 ESI-Orbitrap Exactive Plus.

LC analysis was performed on either a Shimadzu UPLC or a Waters LCMS. The Shimadzu Nexera X2 UPLC system is equipped with an autosampler, binary pump system, degassing unit, column oven and a diode-array UV/Vis detector. The chromatogram was recorded using a Raptor ARC-18 column (2.1 × 100 mm, particle size 1.8 µm) with 0.05% trifluoroacetic acid (TFA) in water as the mobile phase A and 0.05% TFA in acetonitrile

as the mobile phase B. The data was processed using LabSolutions. The Waters Acquity class H UPLC system is equipped with an autosampler, quaternary pump system, column oven, a photodiode array detector and a QDa Mass Spectrometer. The chromatogram was recorded using an ACQUITY UPLC® BEH C18 column (2.1 × 50 mm, particle size 1.7 μm) with 0.1% formic acid (FA) in water as the mobile phase A and 0.1% FA in acetonitrile as the mobile phase B. The data was processed using Masslynx.

Flash column purification of all the S_NAr products in *Preparative Scale Synthesis of S_NAr products* section was performed using a Biotage Selekt system. The system is equipped with a built-in QR reader for Biotage Sfar columns, RFID reader for Selekt collection racks, UV detector, pump, fraction collector, and touch screen. The column information is as follows: Biotage Sfar Silica 60 μm, 25 g, part No. FSRS-0445-0025. All the products were eluted at 70/30 hexanes/EtOAc except for **S4**, which was eluted at 100% methanol.

Experimental details for rate measurement by competition experimentation

All S_NAr reactions were performed inside an MBraun glovebox under an N_2 atmosphere. Benzyl alcohol was used as the nucleophile, and potassium *tert*-butoxide (*t*BuOK) was used as the base. Naphthalene or 1,3,5-trimethoxybenzene was used as the internal standard. A library of 74 (hetero)aryl fluorides, chlorides and bromides was used as the substrates. DMSO was used as the reaction solvent.

Stock solutions in DMSO were prepared for each component: the concentration of the substrate stock solutions was 0.5 M, the concentration of the base stock solution was 0.3 M, the concentration of the benzyl alcohol stock solution was 0.2 M, the concentration of the internal standard naphthalene stock solution was 0.03 M and 1,3,5-trimethoxybenzene was 0.2 M. Molecular sieves (4 Å, pellets) were added into the stock solutions of substrates and internal standards to remove possible water content from the commercial materials, which were allowed to sit at least overnight before use.

UPLC or LCMS was used as the analytical technique. To assess each individual (hetero)aryl halide for S_NAr reactivity with benzyl alcohol and determine the retention times of both the substrate and its S_NAr product, individual S_NAr reactions were carried out at room temperature on 1 mL reaction volume scale. In a 4 mL vial containing a stirbar, an aliquot of the substrate stock solution (200 μL, 0.1 mmol) was diluted with additional reaction solvent (DMSO, 600 μL), followed by addition of an aliquot of benzyl alcohol stock solution (100 μL, 0.02 mmol) and an aliquot of *t*BuOK stock solution (100 μL, 0.03 mmol). The resulting solution was mixed for 1-5 hours, then a 200 μL sample was transferred into a 1.5 mL UPLC vial, followed by dilution with 800 μL acetonitrile. The reaction mixture was analyzed by UPLC to confirm the S_NAr reaction occurred and went to completion (from disappearance of the benzyl alcohol peak), and to identify the substrate and its S_NAr product peaks by their retention times.

Competition experiments were performed to obtain the relative activation energies ($\Delta\Delta G^{\ddagger}_{SNAr}$) between two substrates in S_NAr. All reactions were conducted at room temperature on 1 mL reaction volume scale under *pseudo* first-order conditions by adding two substrates in excess but equal amount (0.1 mmol, 1.0 equivalent) to compete with benzyl alcohol (0.02 mmol, 0.2 equivalents). The competition reactions were prepared with 200 μL of each substrate stock solution, 200 μL of an internal standard stock solution (the choice of the internal standard is according to the retention times that the peak of the

internal standard will not overlap with either the substrate peaks or the product peaks), 200 µL of the reaction solvent DMSO, 100 µL of benzyl alcohol stock solution and 100 µL of *t*BuOK stock solution. The initial concentration of each substrate was 0.1 M, the initial concentration of *t*BuOK was 0.03 M and the initial concentration of benzyl alcohol was 0.02 M. The resulting solution was mixed for 1-5 hours, then a LC sample was prepared by transferring a 200 µL sample and 800 µL acetonitrile into a 1.5 mL UPLC vial (t_{end}). A solution with the initial concentration of the substrates was also prepared by adding 20 µL of each substrate stock solution and 20 µL of the internal standard stock solution into a 1.5 mL UPLC vial then diluting with 940 µL acetonitrile (t_{start}). For each competition experiment, LC chromatograms were recorded for both the t_{start} and t_{end} reaction solutions. The ratio between the two S_NAr rates were obtained from the relative concentrations of the two remaining substrates at t_{end}. This method of quantification avoids the need to obtain relative response factors between all 74 new S_NAr products and the internal standards.

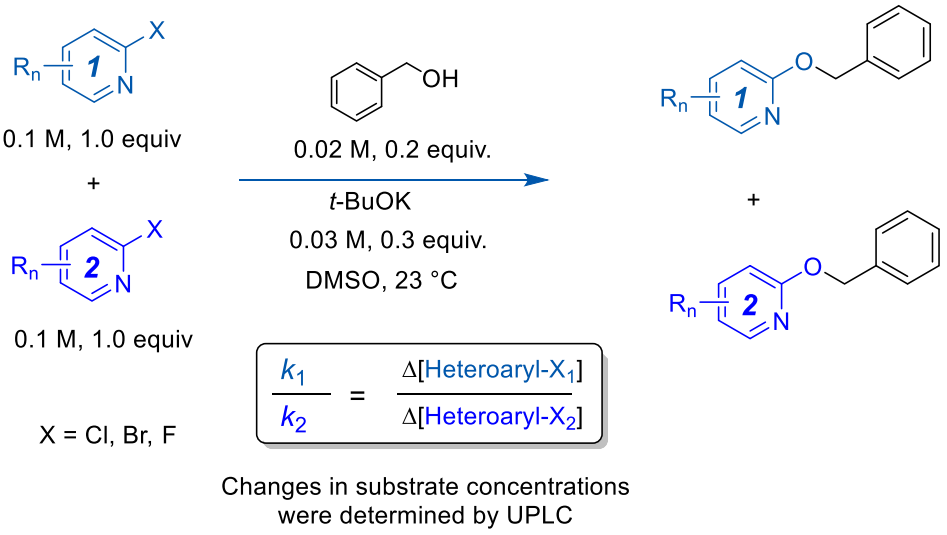


Fig. A1. General competition experiment design. Two electrophiles ***1*** and ***2*** in excess but equimolar amounts were reacted with benzyl alcohol as the limiting reagent, using potassium *tert*-butoxide as the base. Outcomes were quantified by changes to electrophile concentration, giving a ratio of *pseudo* first order rate constants k_1/k_2.

Three typical sets of LC method parameters are shown in Tables A1 to A3 (the 3 min and 10 min ones are on the UPLC and the 6 min one is on the LCMS). The LC inlet methods were adjusted as needed for different pairings of substrates, including the mobile phase gradient, type of columns, runtime and flowrate. To ensure reliable quantitative analysis for every single competition reaction, the inlet method was set up to satisfy the following criteria: the resolution between adjacent peaks should be no less than 1.5, the tailing factor of the peaks of interest should be within 0.8-2.0, and the peak height of the peaks of interest should be above the LOQ (limit of quantification) and below the risk of overloading (usually below 1800 mAu). The UV data was collected at all wavelength range (190 – 800 nm), then the spectra at either 220 nm or 254 nm were extracted to be processed for peak area integration. A representative set of LC chromatograms for a competition experiment is shown in Fig. A2.

Table A1. A typical set of LC method parameters (3 min method) on the Shimadzu Nexera X2 UPLC (equipped with a PDA detector).

LC method parameters
Instrument: Shimadzu UPLC (equipped with a PDA detector)

Instrument Runtime: 3.00 min			
Binary Pump			
Solvent A Name: 0.05% Trifluoroacetic acid in water			
Solvent B Name: 0.05% Trifluoroacetic acid in Acetonitrile			
Gradient Table:			
Time (min)	Flowrate (mL/min)	%A	%B
0.0	0.7	95	5
0.2	0.7	95	5
2.2	0.7	5	95
2.6	0.7	5	95
2.61	0.7	95	5
3.0	0.7	95	5
Column Manager			
Column: Waters Acquity UPLC BEH C18 (2.1 × 30 mm, particle size 1.7 µm)			
Target Column Temperature: 40.0 °C			
AutoSampler			
Needle Wash Solvent: 90:10 Water: Acetonitrile			
Target Sample Temperature: 15.0 °C			
Injection Volume (uL): 1.0			
PDA detector			
Wavelength range: 190 - 800 nm			

Table A2. A typical set of LC method parameters (6 min method) on the Waters Acquity class H UPLC (equipped with a PDA detector a QDa Mass Spectrometer).

LC method parameters
Instrument: Waters Acquity class H UPLC (equipped with a PDA detector a QDa Mass Spectrometer)

Instrument Runtime: 6.00 min					
Quaternary Pump					
Solvent A Name: 0.1% Formic acid in water					
Solvent B Name: 0.1% Formic acid in Acetonitrile					
Gradient Table:					
Time (min)	Flowrate (mL/min)	%A	%B	%C	%D
0.00	0.5	90	10	0	0
4.00	0.5	10	90	0	0
5.00	0.5	10	90	0	0
5.10	0.5	90	10	0	0
6.00	0.5	90	10	0	0
Waters Acquity Column Manager					
Column: ACQUITY UPLC® BEH C18 column (2.1 × 50 mm, particle size 1.7 µm)					
Target Column Temperature: 40.0 °C					
Waters ACQUITY FTN AutoSampler					
Needle Wash Solvent: 90:10 Water: Acetonitrile					
Target Sample Temperature: 10.0 °C					
Injection Volume (uL) - 1.00					

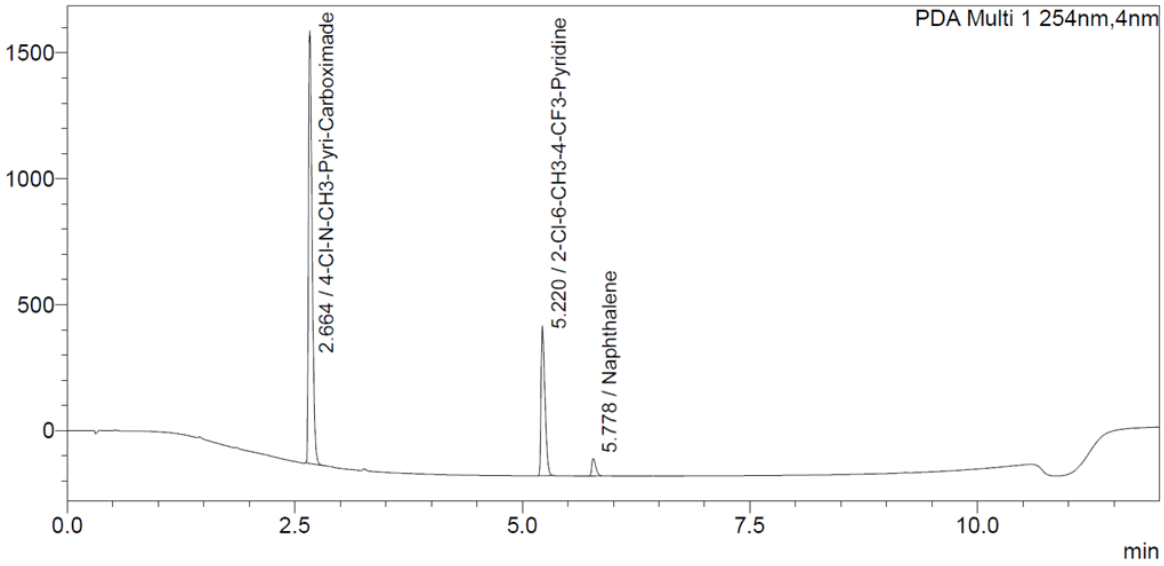
Waters Acquity PDA detector
PDA Detector Type: UPLC eLambda 800 nm
Wavelength range: 190 - 800 nm
QDa Mass Spectrometer
Mass Scanning range: 30 Da to 1250 Da

Table A3. A typical set of LC method parameters (10 min method) on the Shimadzu Nexera X2 UPLC (equipped with a PDA detector).

LC method parameters
Instrument: Shimadzu UPLC (equipped with a PDA detector)

Instrument Runtime: 10.00 min			
Binary Pump			
Solvent A Name: 0.05% Trifluoroacetic acid in water			
Solvent B Name: 0.05% Trifluoroacetic acid in Acetonitrile			
Gradient Table:			
Time (min)	Flowrate (mL/min)	%A	%B
0.0	0.6	90	10
7.5	0.6	5	95
8.5	0.6	5	95
8.51	0.6	90	10
10	0.6	90	10
Column Manager			
Column: Raptor ARC-18 (2.1 × 100 mm, particle size 1.8 µm)			
Target Column Temperature: 40.0 °C			
AutoSampler			
Needle Wash Solvent: 90:10 Water: Acetonitrile			
Target Sample Temperature: 15.0 °C			
Injection Volume (uL): 1.0			
PDA detector			
Wavelength range: 190 - 800 nm			

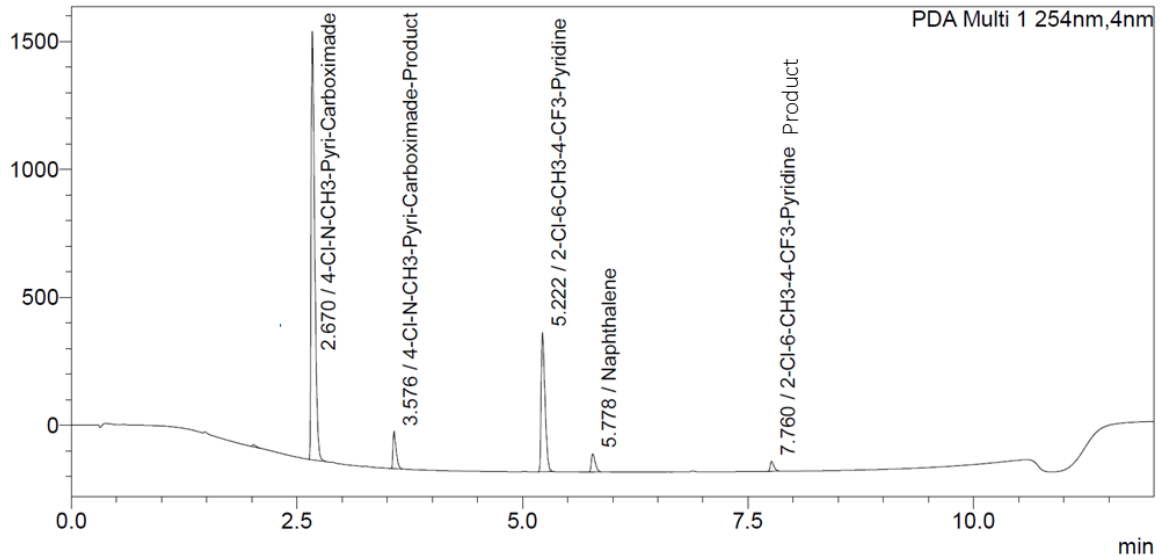
A



PDA Ch1 254 nm

Peak#	Name	Ret. Time	Height	Area	Tailing factor	Resolution
1	4-Cl-N-CH3-Pyri-Carboximade	2.664	1718310	4767682	1.674	--
2	2-Cl-6-CH3-4-CF3-Pyridine	5.220	593995	1751829	1,634	31.146
3	Naphthalene	5.778	69512	207583	1.589	6.509

B



PDA Ch1 254 nm

Peak#	Name	Ret. Time	Height	Area	Tailing factor	Resolution
1	4-Cl-N-CH3-Pyri-Carboximade	2.670	1675108	4615358	1.828	--
2	4-Cl-N-CH3-Pyri-Carboximade-Product	3.576	146336	346262	1.830	12.200
3	2-Cl-6-CH3-4-CF3-Pyridine	5.222	543570	1600488	1.668	21.293
4	Naphthalene	5.778	70360	207849	1.603	6.523
5	2-Cl-6-CH3-4-CF3-Pyridine Product	7.760	38942	106810	1.631	23.780

Fig. A2. UPLC chromatograms recorded by Shimadzu Nexera X2 UPLC system at 254 nm for competition reaction between 4-chloro-N-methylpyridine-2-carboxamide and 2-chloro-6-methyl-4 (trifluoromethyl)pyridine: A. LC chromatogram at t_{start}; B. LC chromatograms at t_{end}.

Experimental Details of kinetic analysis for the touchstone reactions

Three of the S_NAr reactions were chosen as the touchstone reactions and their absolute reaction rates were measured by kinetic analysis. The $\Delta G^{\ddagger}_{SNAr}$ (obtained from the *pseudo* first order rate constant) determined for one of those touchstone reactions (2-chloropyridine and benzyl alcohol in DMSO) was used to calibrate all the relative rates determined by competition reactions, giving the absolute $\Delta G^{\ddagger}_{SNAr}$ values for the entire array of substrates. The other two touchstone reactions were used to validate the accuracy of the rate constants determined from this competition experimentation approach, by comparing the reaction rates determined from kinetic analysis to those obtained from competition experiments.

All of the following procedures were performed inside an MBraun glovebox under an N_2 atmosphere. The reactions were conducted with the same concentrations of each species as for the competition reactions (0.1 M of substrate, 0.02 M of benzyl alcohol, 0.03 M of *t*BuOK and 0.003 M of naphthalene or 0.02 M 1,3,5-trimethoxybenzene as internal standard in 10 mL of DMSO). All the components except for *t*BuOK were added into a 20 mL glass vial charged with a stir bar and mixed well by stirring, then a small amount (200 μL) of this solution was transferred into an UPLC vial followed by dilution with 800 μL of acetonitrile as the time zero (t_0) point. A stopwatch was used to keep track of the reaction time immediately after the *t*BuOK stock solution was added. After a certain time period, a portion of the reaction mixture (500 μL) was transferred into a 2 mL UPLC vial containing 500 μL of 0.01 M hydrochloric acid in acetonitrile; the excess amount of acid was added to quench the reaction by consuming the base in the reaction solution. Reaction mixture aliquots were withdrawn at 10 time points to construct the reaction progress curves.

These reaction mixture aliquots were analyzed by UPLC and the peak areas of benzyl alcohol and the internal standard were determined. A [BnOH] versus time plot was constructed, and an exponential fit applied. Using a *pseudo* first-order rate law assumption, the rate constant (k) was calculated using Eq (A1). This k value was substituted into the Eyring equation to obtain the Gibbs free energy of this reaction ($\Delta G^{\ddagger}_{SNAr}$) according to Eq (A2). The results of the three reaction rate measurement experiments are summarized in Fig. A3.

According to the *pseudo* first-order rate law assumption, the rate constant (k) can be determined by Eq (S1):

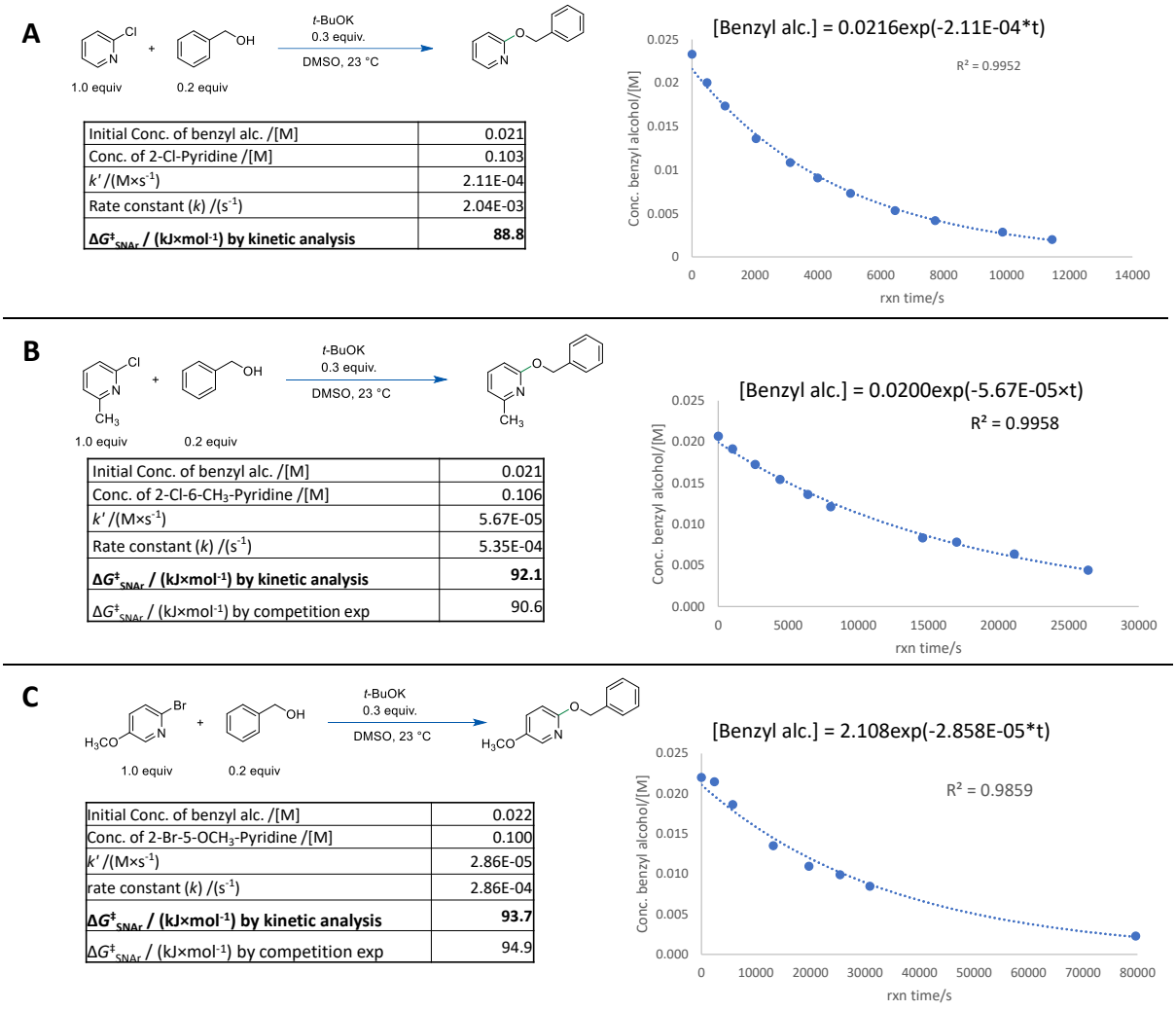
$$[benzyl\ alc.]_t = [benzyl\ alc.]_{t=0} e^{-k't}, where\ k' = k[substrate] \qquad Eq(A1)$$

Then the Gibbs free energy ($\Delta G^{\ddagger}_{SNAr}$) is determined by the Eyring equation Eq (A2) (transmission coefficient κ assumed to be 1):

$$k = \frac{k_B T}{h} \exp\left(-\frac{\Delta G^{\ddagger}}{RT}\right) \Rightarrow \Delta G^{\ddagger} = -RT ln \frac{kh}{k_B T} \qquad Eq(A2)$$

$$where\ R\ is\ the\ ideal\ gas\ constant: 8.314\ J \times (K \times mol)^{-1};$$
$$T\ is\ the\ reaction\ temperature: 299.55\ K;$$

$$k_B \; is \; the \; Boltzmann \; constant: 1.38 \times 10^{-23} \; J \times k^{-1};$$
$$h \; is \; Planck \; constant: 6.63 \times 10^{-34} \; J \times s$$



Initial Conc. of benzyl alc. /[M]	0.021
Conc. of 2-Cl-Pyridine /[M]	0.103
k' /(M×s^{-1})	2.11E-04
Rate constant (k) /(s^{-1})	2.04E-03
$\Delta G^{\ddagger}_{SNAr}$ / (kJ×mol^{-1}) by kinetic analysis	**88.8**

Initial Conc. of benzyl alc. /[M]	0.021
Conc. of 2-Cl-6-CH$_3$-Pyridine /[M]	0.106
k' /(M×s^{-1})	5.67E-05
Rate constant (k) /(s^{-1})	5.35E-04
$\Delta G^{\ddagger}_{SNAr}$ / (kJ×mol^{-1}) by kinetic analysis	**92.1**
$\Delta G^{\ddagger}_{SNAr}$ / (kJ×mol^{-1}) by competition exp	90.6

Initial Conc. of benzyl alc. /[M]	0.022
Conc. of 2-Br-5-OCH$_3$-Pyridine /[M]	0.100
k' /(M×s^{-1})	2.86E-05
rate constant (k) /(s^{-1})	2.86E-04
$\Delta G^{\ddagger}_{SNAr}$ / (kJ×mol^{-1}) by kinetic analysis	**93.7**
$\Delta G^{\ddagger}_{SNAr}$ / (kJ×mol^{-1}) by competition exp	94.9

Fig. A3. Determination of the rate constants for touchstone reactions: **A**. 2-chloropyridine and benzyl alcohol in DMSO; **B**. 2-chloro-6-methylpyridine and benzyl alcohol in DMSO; **C**. 2-bromo-5-methoxypyridine and benzyl alcohol in DMSO.

Calculating $\Delta G^{\ddagger}_{SNAr}$ for the entire array of Substrates

All of the competition reactions were conducted under *pseudo* first-order conditions, with the two electrophiles in excess but equal amounts ([substrate]:[nucleophile] = 5:1). The ratio of the reaction rates can be expressed as Eq (A3):

$$\frac{r_1}{r_2} = \frac{k_1[Nucleophile][Substrate_1]}{k_2[Nucleophile][Substrate_2]} = \frac{k_1[Substrate_1]}{k_2[Substrate_2]} \qquad Eq(A3)$$

The concentrations of the two substrates are assumed to stay constant throughout the reaction; then, Eq (A1) can be simplified to Eq (A4):

$$\frac{r_1}{r_2} = \frac{k_1}{k_2} = \frac{\frac{d[Substrate_1]}{dt}}{\frac{d[Substrate_2]}{dt}} = \frac{\Delta[Substrate_1]}{\Delta[Substrate_2]} \qquad Eq(A4)$$

From UPLC chromatography analysis, $\frac{\Delta[Substrate_1]}{\Delta[Substrate_2]}$ is equal to the percent change of the peak area of the two substrates from the reaction mixture at t_{start} to that at t_{end}. By substituting the Eyring equation (Eq (A5)):

$$k = \frac{k_B T}{h} \exp\left(-\frac{\Delta G^{\ddagger}}{RT}\right) \qquad Eq(A5)$$

into Eq (A4), the relative activation energy ($\Delta\Delta G^{\ddagger}_{SNAr}$) of the two S_NAr reactions from the competition can be calculated by Eq (A6):

$$\frac{k_1}{k_2} = \exp\left(\frac{\Delta G^{\ddagger}{}_2 - \Delta G^{\ddagger}{}_1}{RT}\right) \Rightarrow \Delta\Delta G^{\ddagger}_{SNAr} = \ln\left(\frac{k_1}{k_2}\right) RT \Rightarrow$$

$$\Delta\Delta G^{\ddagger}_{SNAr} = \ln\left(\frac{\%\Delta Peak\ area_{Substrate1}}{\%\Delta Peak\ area_{Substrate2}}\right) RT \qquad Eq(A6)$$

$$where\ \Delta\Delta G^{\ddagger}_{SNAr} = \Delta G^{\ddagger}_2 - \Delta G^{\ddagger}_1$$

The $\Delta\Delta G^{\ddagger}_{SNAr}$ values determined from the competition experiments were related to the absolute $\Delta G^{\ddagger}_{SNAr}$ determined from kinetic analysis of the touchstone reaction between 2-chloropyridine and benzyl alcohol in DMSO; therefore, all other $\Delta G^{\ddagger}_{SNAr}$ values are given relative to 2-chloropyridine.

The S_NAr rate ratio of 94 competition reactions is shown in Table A4, and the experimental $\Delta G^{\ddagger}_{SNAr}$ for the 74 substrates determined from the 94 competition reactions in DMSO is summarized in Table A5. Substrates highlighted are those with $\Delta G^{\ddagger}_{SNAr}$ determined by multiple competition experiments with different substrate pairings. The given $\Delta G^{\ddagger}_{SNAr}$ values for these substrates are averages of those determined by at least two different competition experiments, and the relative standard deviation (RSD) as well as the $\Delta G^{\ddagger}_{SNAr}$ from each individual measurement for these examples are also given.

Table A4 –S_NAr rate ratios (r_1/r_2) calculated using Eq (A3) for 94 competition reactions in DMSO (Lu-PhD-thesis_Ch.2 SI tables).

Table A5 –Observed $\Delta G^{\ddagger}_{SNAr}$ for 74 substrates determined from competition reactions in DMSO (Lu-PhD-thesis_Ch.2 SI tables).

Hammett Analyses of *para- and meta*-Substituted Substrates

To further validate the kinetic parameters obtained by competition experiments, we have obtained reaction constants (ρ) through construction of Hammett plots – $\log(k_Z/k_H)$ versus substituent σ values – for two sets of *para*-substituted substrates and two set of *meta*-substituted substrates undergoing S_NAr. These include: 5-substituted-2-chloropyridines and 5-substituted-2-bromopyridines (σ_{para}); 4-substituted-2-chloropyridines and 6-substituted-2-chloropyridines (σ_{meta}). Substituent σ values were obtained from published tables.[5]

We obtain reaction constants of ρ = 5.1 in DMSO for S_NAr of the 5-Z-2-Cl-pyridines (Fig. A4), ρ = 4.4 in DMSO for S_NAr of the 5-Z-2-Br-pyridines (Fig. A5), ρ = 5.1 in DMSO for S_NAr of the 4-Z-2-Cl-pyridines (Fig. A6), and ρ = 4.8 in DMSO for S_NAr of the 6-Z-2-Cl-pyridines (Fig. A7).

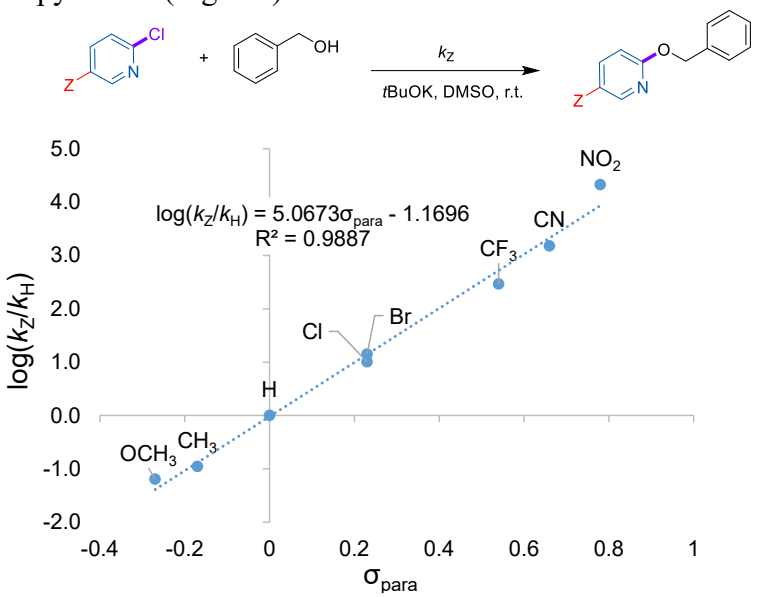


Fig. A4. Hammett plot of $\log(k_Z/k_H)$ versus σ_{para} for S_NAr with benzyl alcohol of a group of 5-Z-2-chloropyridines in DMSO.

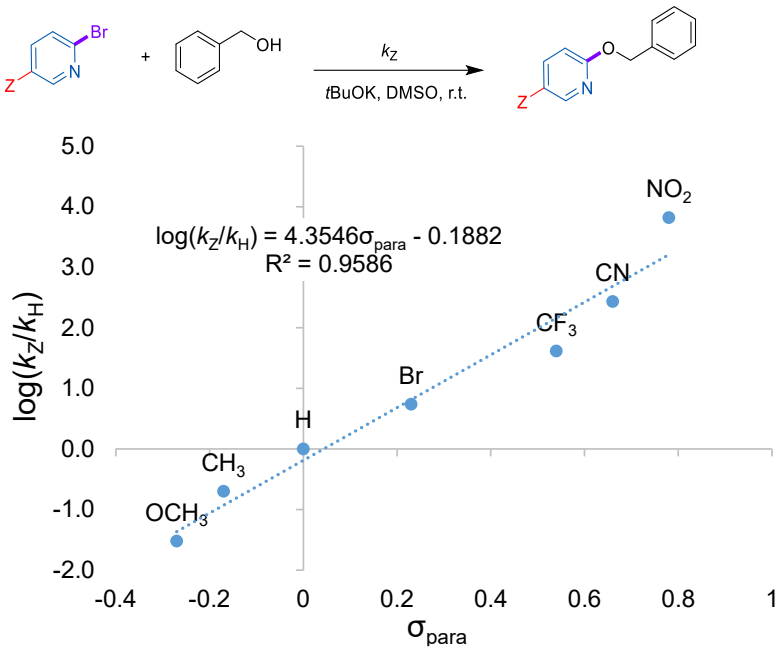


Fig. A5. Hammett plot of $\log(k_Z/k_H)$ versus σ_{para} for S_NAr with benzyl alcohol of a group of 5-Z-2-bromopyridines in DMSO.

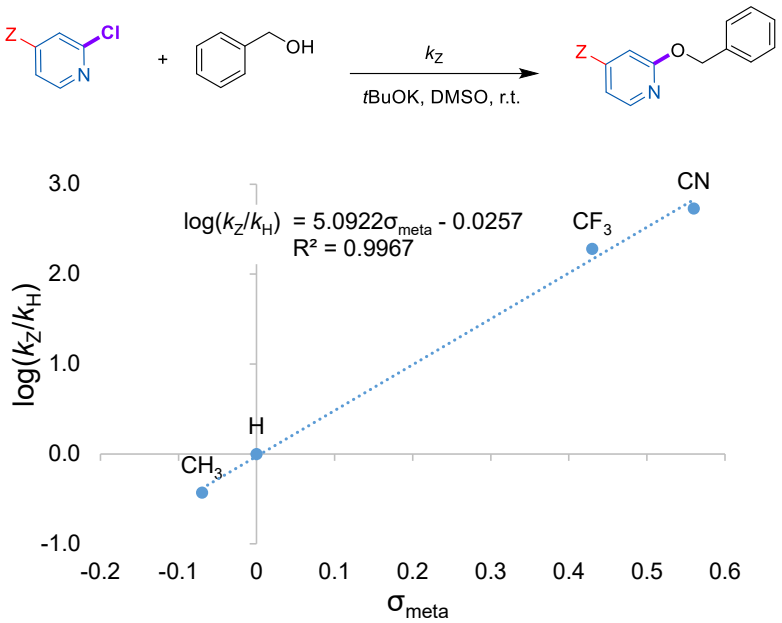


Fig. A6. Hammett plot of $\log(k_Z/k_H)$ versus σ_{para} for S_NAr with benzyl alcohol of a group of 4-Z-2-chloropyridines in DMSO.

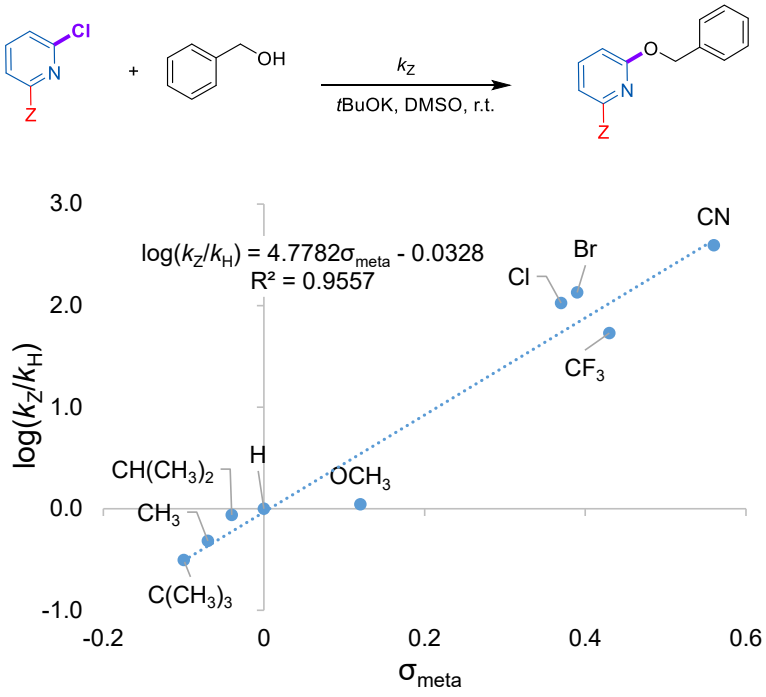


Fig. A7. Hammett plot of $\log(k_Z/k_H)$ versus σ_{para} for S_NAr with benzyl alcohol of a group of 6-Z-2-chloropyridines in DMSO.

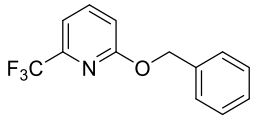
Preparative Scale Synthesis of S_NAr products

Five representative products from S_NAr reactions were isolated, purified, and characterized using the following general procedure.

In the glovebox, a 50 mL round bottom flask containing a stir bar was charged with 3 mmol of the substrate (1.0 equiv.), 3.9 mmol of benzyl alcohol (1.3 mmol) and 15 mL DMSO. 4.5 mmol *t*BuOK was dissolved into 5 mL DMSO and the solution was added dropwise into the round bottom flask. The reaction mixture was stirred at room temperature overnight. On the next day, the reaction mixture was transferred outside the glovebox for workup and isolation.

The reaction mixture was poured into 30 mL water, then extracted with ethyl acetate (3×20 mL). The combined ethyl acetate phase was washed using saturated NaCl solution (30 mL) then dried over Mg_2SO_4, filtered and evaporated under vacuum. Then the crude product was purified by automated flash chromatography (Biotage Selekt) on silica gel using hexanes/ethyl acetate to obtain the purified product.

Characterization of Isolated S_NAr products



S1

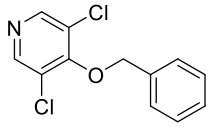
The S_NAr product **S1** was prepared by the general procedure using 2-chloro-6-(trifluoromethyl)pyridine. 307.9 mg of colorless liquid was obtained (31% yield)

^1H NMR: (500 MHz, $CDCl_3$): δ 5.3 (s, 2H, 2 x CH_2), 6.8 (d, 1H, 1 x Py-H), 7.1 (d, 1H, 1 x Py-H), 7.2-7.4 (m, 5H, 5 x Ph-H), 7.5-7.6 (t, 1H, 1 x Py-H).

$^{13}C\{^1H\}$ NMR: (126 MHz, $CDCl_3$): δ 68.3 (1 x CH_2), 113.4 (1 x Py), 114.9 (1 x Py), 118.2-124.8 (q, J=273.6 Hz, 1 x CF3), 128.2-128.6 (5 x Ph), 136.7 (1 x Ph), 139.5 (1 x Py), 145.0-145.9 (q, J=34.8 Hz, 1 x Py), 163.6 (1 x Py).

$^{19}F\{^1H\}$ NMR: (471 MHz, $CDCl_3$): δ -68.4.

HRMS (ESI): $[C_{13}H_{10}F_3NO{\cdot}H]^+$ (major isotopomer, protonated): 254.07873 (calc'd), 254.07871 (found).



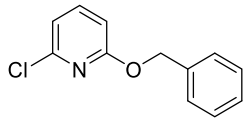
S2

The S_NAr product **S2** was prepared by the general procedure using 3,4,5-trichloropyridine. 172.5 mg of white powder was obtained (23% yield)

^1H NMR: (500 MHz, $CDCl_3$): δ 5.1 (s, 2H, 2 x CH_2), 7.5-7.3 (m, 5H, 5 x Ph-H), 8.4 (s, 2H, 2 x Py-H).

$^{13}C\{^1H\}$ NMR: (126 MHz, $CDCl_3$): δ 75.4 (1 x CH_2), 126.7 (2 x Py), 128.6-128.8 (5 x Ph), 135.3 (1 x Ph), 149.2 (2 xPy), 157.5 (1 x Py).

HRMS (ESI): $[C_{12}H_9Cl_2NO{\cdot}H]^+$ (major isotopomer, protonated): 254.01340 (calc'd), 254.01339 (found).



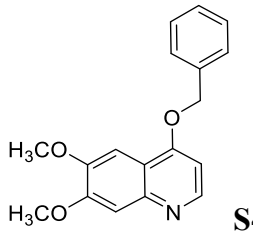
S3

The S_NAr product **S3** was prepared by the general procedure using 2,6-dichloropyridine. 320.4 mg of colorless liquid was obtained (49% yield)

^1H NMR: (500 MHz, $CDCl_3$): δ 5.3 (s, 2H, 2 x CH_2), 6.6 (dd, 1H, 1 x Py-H), 6.8 (dd, 1H, 1 x Py-H), 7.2-7.4 (m, 5H, 5 x Ph-H), 7.4-7.5 (1H 1 x Py-H).

$^{13}C\{^1H\}$ NMR: (126 MHz, $CDCl_3$): δ 68.3 (1 x CH_2), 109.4 (1 x Py), 116.5 (1 x Py), 128.1-128.5 (5 x Ph), 136.6 (1 x Ph), 140.7 (1 x Py), 148.3 (1 x Py), 163.2 (1 x Py).

HRMS (ESI): $[C_{12}H_{10}ClNO{\cdot}H]^+$ (major isotopomer, protonated): 220.05237 (calc'd), 220.05239 (found).



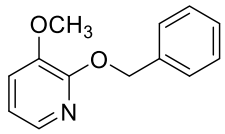
S4

The S_NAr product **S4** was prepared by the general procedure using 4-chloro-6,7-dimethoxy-quinoline. During the flash chromatography purification, the targeted product could not be eluted with hexanes/ethyl acetate gradient flow, and the flash column was flushed with 100% methanol to collect the targeted product. 212.3 mg of bright yellow crystal was obtained (24% yield).

^1H NMR: (500 MHz, $CDCl_3$): δ 3.9 (s, 3H, 3 x OCH_3), 3.9 (s, 3H, 3 x OCH_3), 5.2 (s, 2H, 2 x CH_2), 6,6 (d, 1H, 1 x Qu-H), 7.3-7.4 (m, 7H, 5 x Ph-H and 2 x Qu-H), 8.5 (d, 1H, 1 x Qu-H).

$^{13}C\{^1H\}$ NMR: (126 MHz, $CDCl_3$): δ 54.1 (2 x OCH_3), 68.3 (1 x CH_2), 97.9 (1 x Qu), 98.5 (1 x Qu), 105.8 (1 x Qu), 114.0 (1 x Qu), 125.4-126.8 (5 x Ph), 134.0 (1 x Ph), 144.2 (1 x Qu), 147.1 (1 x Qu), 147.2 (1 x Qu), 150.6 (1 x Qu), 158.2 (1 x Qu).

HRMS (ESI): $[C_{18}H_{17}NO_3{\cdot}H]^+$ (major isotopomer, protonated): 296.12812 (calc'd), 296.12813 (found).



S5

The S_NAr product **S5** was prepared by the general procedure using 2-chloro-3-methoxypyridine. 168.5 mg of colorless liquid was obtained (26% yield).

^1H NMR: (500 MHz, $CDCl_3$): δ 3.8 (s, 3H, 3 x OCH_3), 5.4 (2, 2H, 2 x CH_2), 6.7-6.8 (dd, 1H, 1 x Py-H), 7.0 (d, 1H, 1 x Py-H), 7.2-7.4 (m, 5H, 5 x Ph-H), 7.6-7.7 (d, 1H, 1 x Py-H).

$^{13}C\{^1H\}$ NMR: (126 MHz, $CDCl_3$): δ 55.7 (1 x OCH_3), 67.5 (1 x CH_2), 117.0 (1 x Py), 117.5 (1 x Py), 127.7-128.4 (5 x Ph), 137.0 (1 x Ph), 137.4 (1 x Py), 144.2 (1 x Py), 153.9 (1 x Py).

HRMS (ESI): $[C_{13}H_{13}NO_2{\cdot}H]^+$ (major isotopomer, protonated): 216.10191 (calc'd), 216.10193 (found).

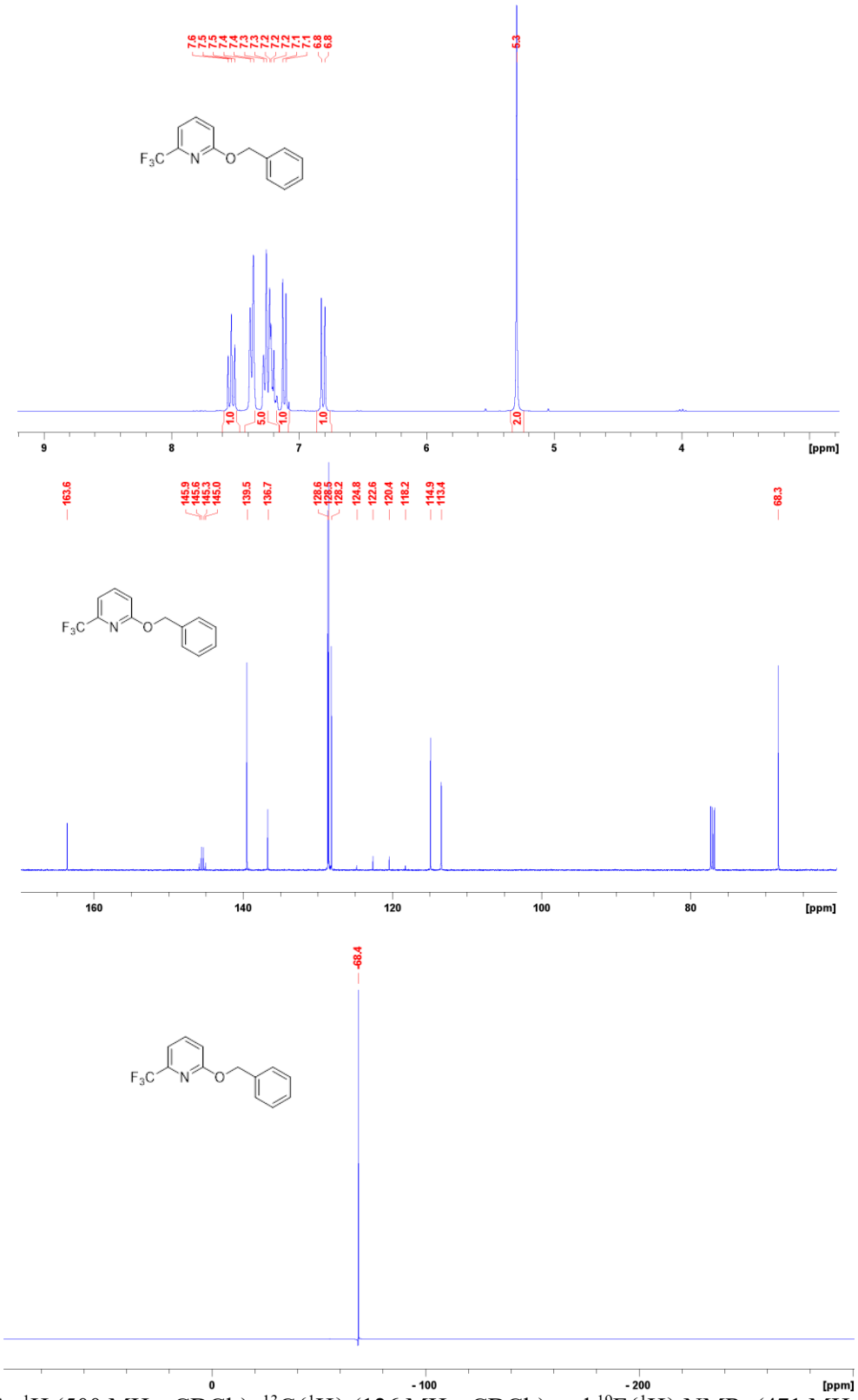


Fig. SA. ^{1}H (500 MHz, $CDCl_3$), $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) and $^{19}F\{^{1}H\}$ NMR: (471 MHz, $CDCl_3$) NMR spectra of **S1**.

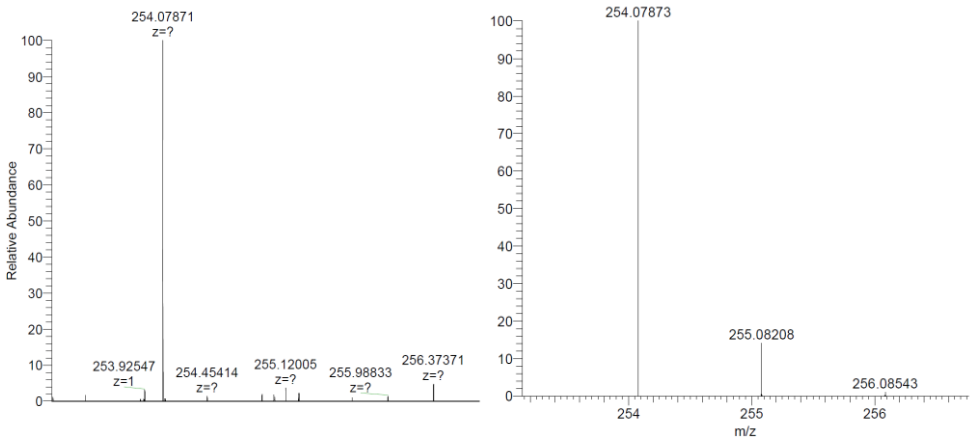


Fig. SA. Left: Experimental HRMS-ESI spectrum of [**S1**+H]$^{+}$. Right: Calculated HRMS isotope pattern for [**S1**+H]$^{+}$.

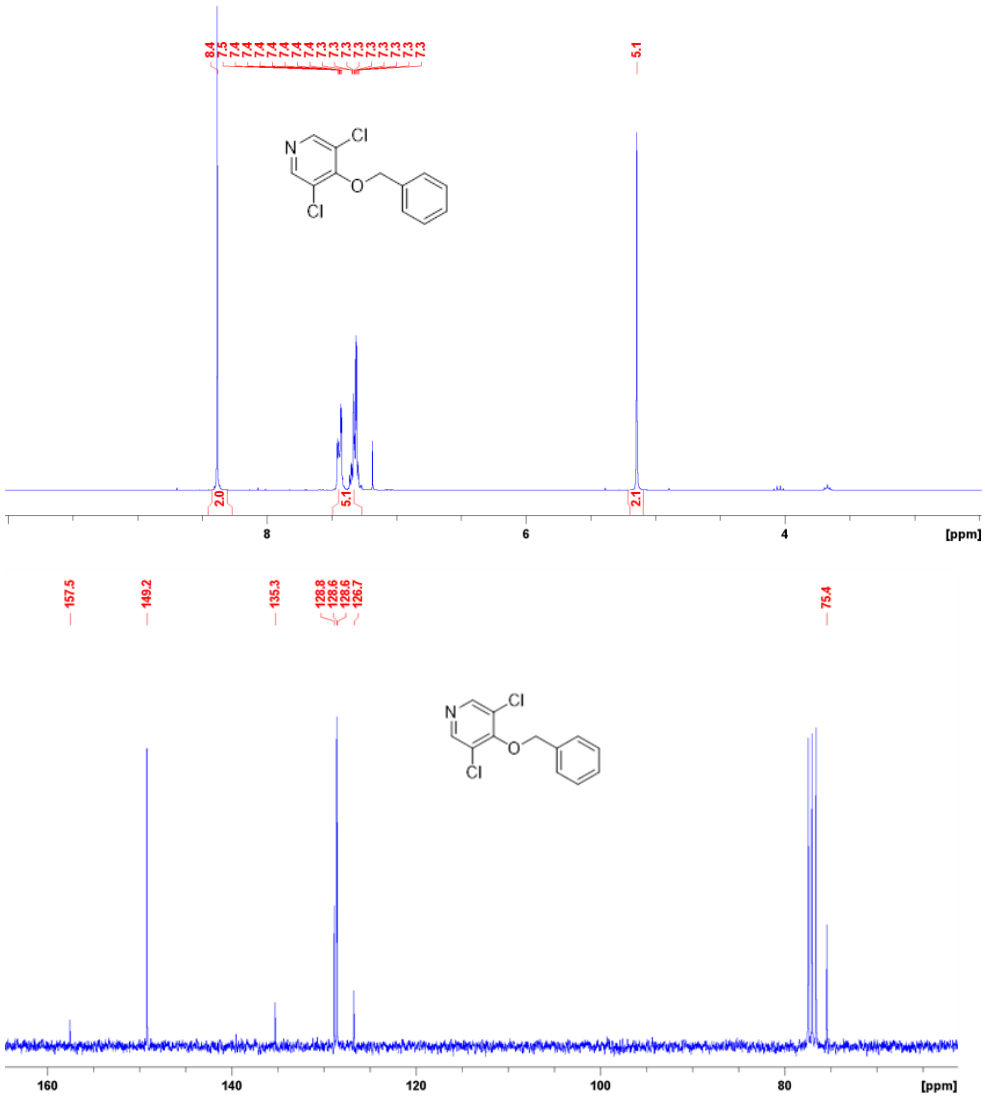


Fig. A10. ^{1}H (500 MHz, $CDCl_3$), and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **S2**.

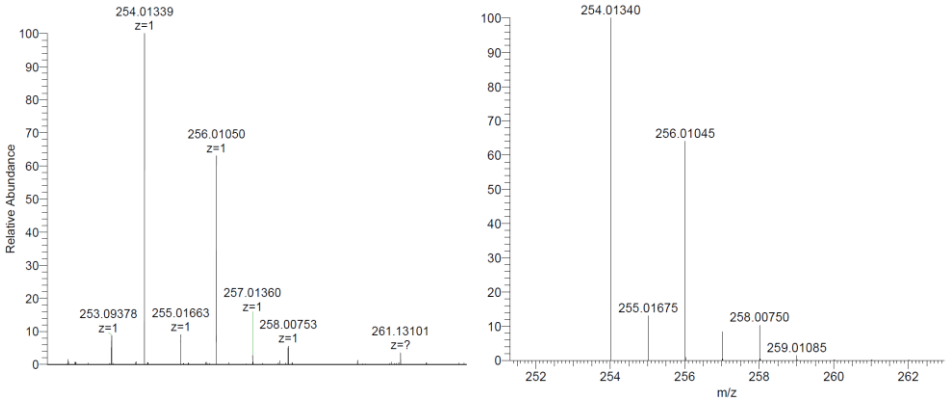


Fig. A11. Left: Experimental HRMS-ESI spectrum of [**S2**+H]$^+$. Right: Calculated HRMS isotope pattern for [**S2**+H]$^+$.

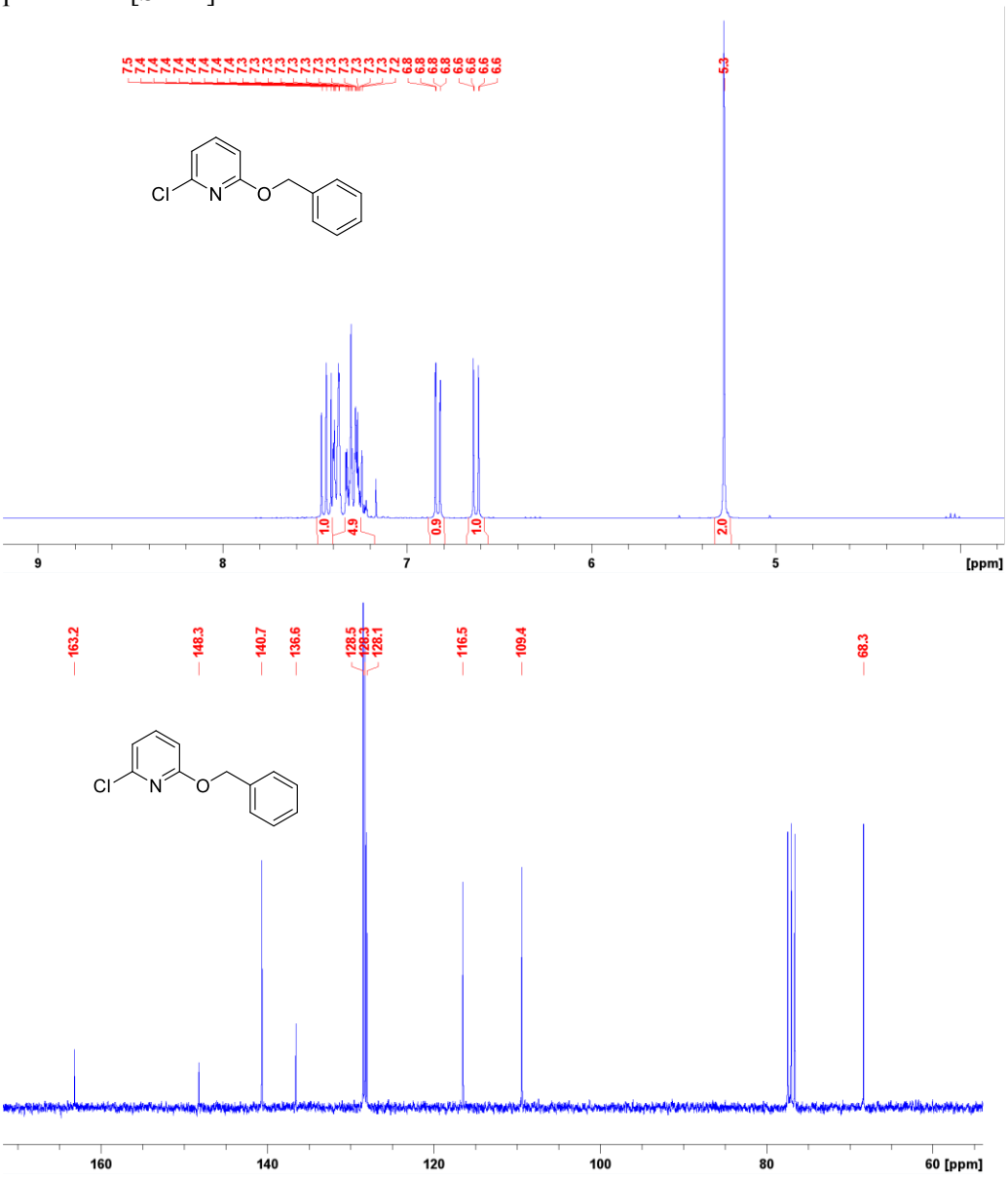


Fig. A12. ^{1}H (500 MHz, $CDCl_3$), and $^{13}C\{^1H\}$ (126 MHz, $CDCl_3$) NMR spectra of **S3**.

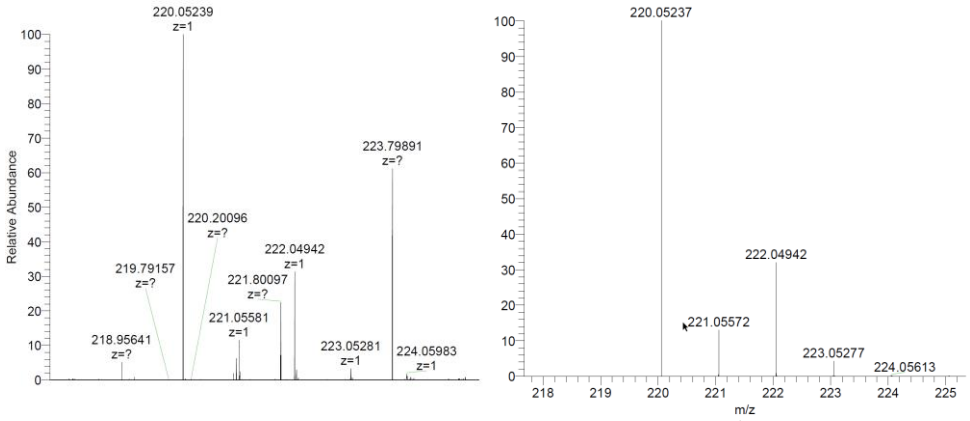


Fig. A13. Left: Experimental HRMS-ESI spectrum of [**S3**+H]$^+$. Right: Calculated HRMS isotope pattern for [**S3**+H]$^+$.

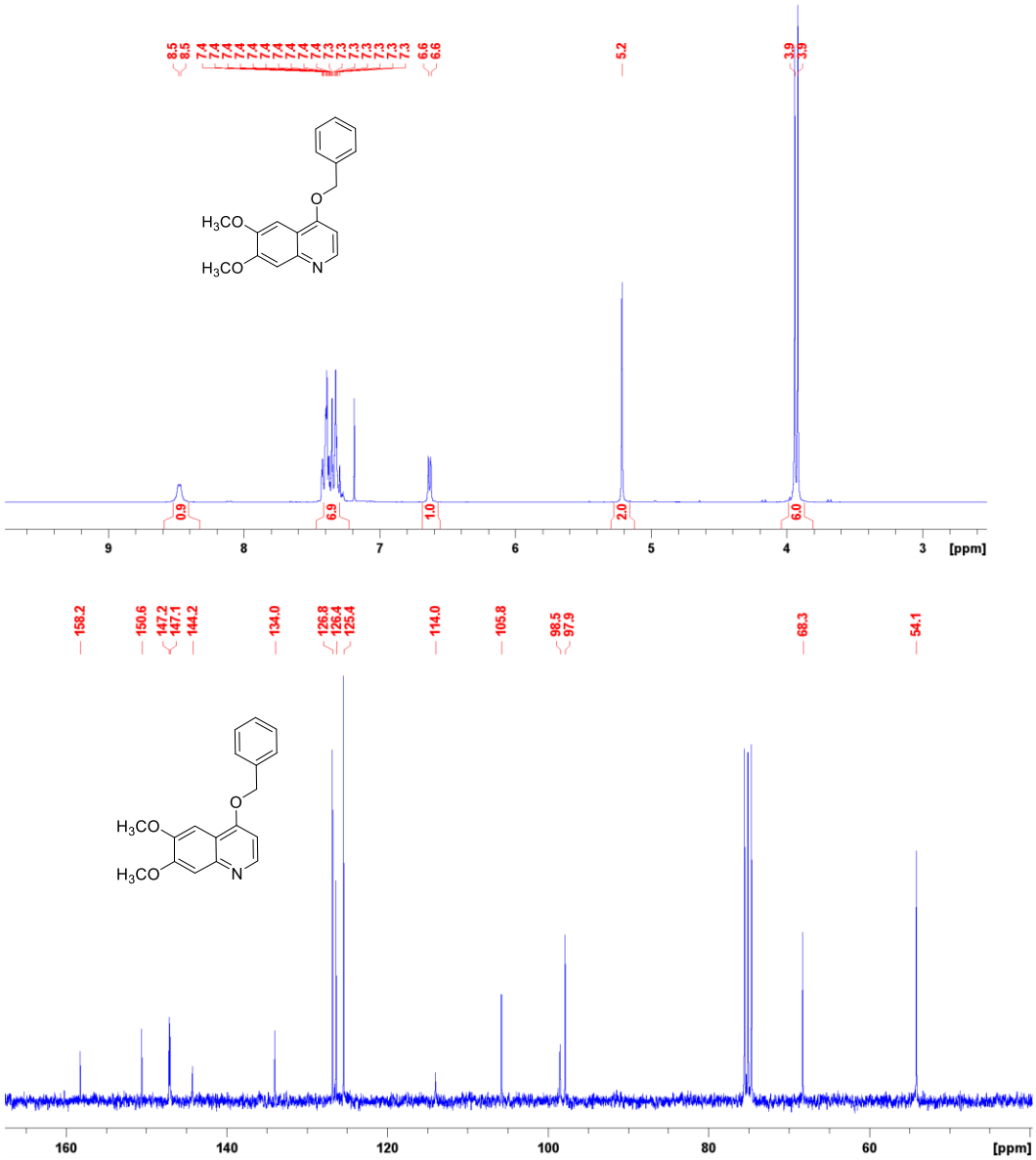


Fig. A14. ^{1}H (500 MHz, $CDCl_3$), and $^{13}C\{^1H\}$ (126 MHz, $CDCl_3$) NMR spectra of **S4**.

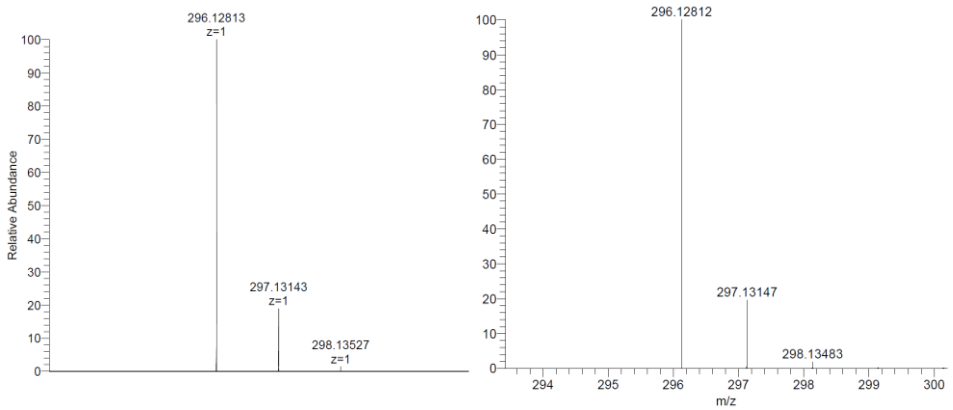


Fig. A15. Left: Experimental HRMS-ESI spectrum of [**S4**+H]$^{+}$. Right: Calculated HRMS isotope pattern for [**S4**+H]$^{+}$.

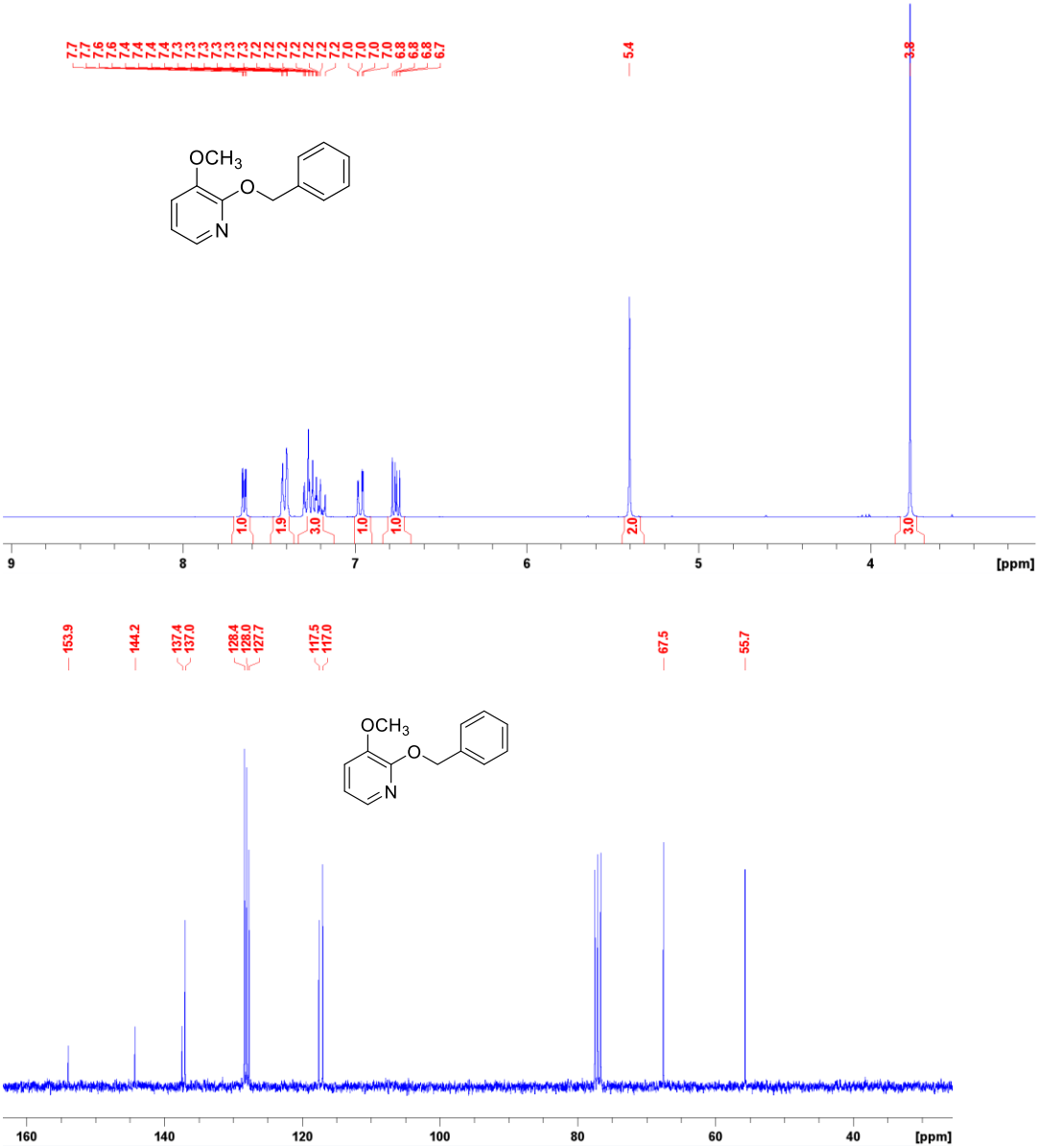


Fig. A16. ^{1}H (500 MHz, $CDCl_3$), and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **S5**.

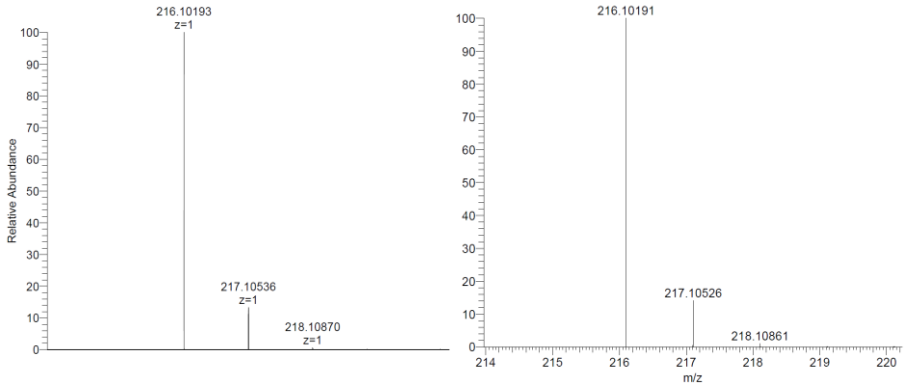


Fig. A17. Left: Experimental HRMS-ESI spectrum of [**S5**+H]$^+$. Right: Calculated HRMS isotope pattern for [**S5**+H]$^+$.

Computational Determination of Molecular Descriptors

All molecular geometry optimizations were performed using ORCA version 4.0.1.2[6] and all electrostatic potential (*ESP*) calculations were performed using Multiwfn version 3.7[7,8].

Initial substrate structures were either downloaded from the ChemSpider database[9] or generated using Avogadro version 4.1.[10] Geometry optimizations were performed for the uncharged species (N-electron system) using a B3LYP/def2-TVZPD approach. The energy of the anionic species (N+1-electron system) was calculated at the optimized geometry of the uncharged species.

Using the conceptual link between *LUMO* energy and electron affinity (*EA*) related to Koopmans's theorem[11], we calculated *EA* for each substrate according to:

$$EA = E(N) - E(N+1) \sim -\mathcal{E}(LUMO) \qquad Eq(A7)$$

The corrections to the *EA* values for substrates in the main text Fig. 5 that proceed through overlap of the *LUMO*+1 orbitals were conducted by adding the orbital energy difference between the *LUMO*+1 and *LUMO*, as output by the DFT calculations, to the *EA* values.

A *.gbw* file that contains a binary summary of the calculation was generated automatically during the execution of ORCA, then transformed to a *.wfn* file by the utility program ORCA_2aim. The *.wfn* file contains electron density information and can be used for molecular surface analysis by the Multiwfn program. The utility program ORCA_2mkl was used to generate a *.fch* file, containing information on the full range of calculated MOs from the *.gbw* file.

The *.wfn* file was imported into Multiwfn for *ESP* calculations. The molecular *ESP* was calculated at a 0.004 au isosurface of electron density. By selecting *Output surface properties of each atom,* the surface area occupied by each atom, as well as the maximal, minimal and average *ESP* values at that surface were displayed. The average *ESP* values at the reactive center (ESP_1) and sum of the *ESP* values at the ortho and para atoms in relation to the reactive center (ESP_2) were used as the electronic descriptors in construction of the S_NAr predictive model.

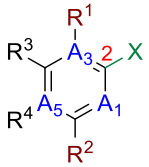
The electrostatic potential maps and molecular orbitals were plotted and rendered in VMD, a molecular visualization program.[12] To do so, the electron density cube file and the *ESP* cube file were generated in Multiwfn from the *.wfn* file. The two cube files can be

read by VMD and the *ESP* map can be plotted, by projecting the *ESP* values at each grid point to a defined electron density isosurface. To obtain molecular orbital pictures, orbital cube files were generated in Multiwfn using *.fch* file as input, then plotted in VMD at a defined electron density isosurface.

Table A6 –Calculated molecular descriptors for the 74 (hetero)aryl halide substrates used in S_NAr experiments (Lu-PhD-thesis_Ch.2 SI tables).

Construction of the Multivariate Linear Regression Model

The selection of the molecular descriptors used to correlate $\Delta G^{\ddagger}_{SNAr}$ with substrate structures was guided by the mechanistic features of nucleophilic aromatic substitution,[13,14] DFT calculations on the transition states, and iterative refinement of the included descriptors based on our experimental observations. As summarized in Fig. A18, there are three descriptors that lead to accurate predictions of S_NAr reactivity: *Electron affinity (EA)* of the substrate (a measurable molecular property to approximate the *LUMO* energy), average electrostatic potential (ESP_1) at the reactive carbon, sum of the average electrostatic potential at the *ortho* and *para* atoms in relation to the reactive carbon (ESP_2). Linear regression analysis with the normalized descriptors has been performed to obtain the contribution of the individual descriptor to the reactivity of the substrate, and the results are also summarized in Fig. A19.



1) *EA* of the substrate
2) ESP at the reactive site carbon atom (ESP_1)
3) sum of ESP at A_1, A_3 and A_5 (ESP_2)
(A = Nitrogen or Carbon)

Contribution of each descriptor:
EA (15.3%)
ESP_1 (50.2%)
ESP_2 (34.5%)

Fig. A18. Substrate molecular descriptors to construct the nucleophilic aromatic substitution (S_NAr) predictive model.

The initial multivariate linear regression model constructed from these descriptors and the $\Delta G^{\ddagger}_{SNAr}$ of 74 (hetero)aryl halides in DMSO achieved excellent linear correlation with a squared correlation coefficient (R^2) of 0.92, and a mean absolute error (MAE) of 1.87 kJ/mol, and has an expected random distribution of residuals as summarized in Fig. S19.

To validate the inclusion of these descriptors and evaluate simpler potential models, we have compared the three-descriptor model to a series of alternatives (Table A7).

First, we evaluated three univariate linear regression models constructed by each individual descriptor. The linear regression with ESP_1 gives fairly good performance with an R^2 of 0.77 and an MAE of 3.0 kJ/mol, as shown in entry 2. In contrast, the univariate models with the other two descriptors only give a weak linear relationship (*EA*, entry 4) and almost no linear relationship (ESP_2, entry 3).

ESP_1 is the descriptor that contributes the most to predicting S_NAr reactivity among the three, as indicated by the performance of the univariate models. Then we evaluated two bivariate linear regression models as shown in entries 5 and 6. Compared to the univariate model in entry 2, the inclusion of one more descriptor, either *EA* or ESP_2 improves the prediction accuracy, but still in a slightly lower level compared to the prediction accuracy of the multivariate model with all the three descriptors included.

The inclusion of ESP_2 (sum of the *ESP* at the *ortho* and *para* atoms) in the model was explored based on our DFT transition state calculations and the experimental observations. From our iterative analysis of the linear regression model, we discovered that the two *ESP* values – *ESP* at the ortho atoms and the *ESP* at the para atom – have approximately equal contributions when linear regression fitting is done with ESP_{ortho} and ESP_{para} as separate descriptors. This alternate, four-descriptor model is shown in entry 7. An R^2 of 0.92 and a MAE of 1.9 kJ/mol were obtained from the four-descriptor linear regression model, which are almost identical to those obtained from the original three-variable regression model (entry 1). The two models give the same level of prediction accuracy, and thus the ESP_{ortho} and ESP_{para} are treated as one summed value for simplification.

The model with steric effect as an additional descriptor (steric A value[15]) has also been evaluated. The statistic measures in the entry 8 show that the prediction accuracy is almost identical compared to the initial model shown in Fig. A19. Linear regression analysis with normalized descriptors shows that the contribution of the steric A value is only 2%, which is much lower than any other descriptor. As a result, the steric descriptor is not included because of its insignificant impact to the reactivity of the substrates in our training dataset.

Overall, the alternatives presented in Table A7 demonstrate that all of the three descriptors are necessary and sufficient to predict $\Delta G^{\ddagger}_{SNAr}$ for this diverse set of substrates.

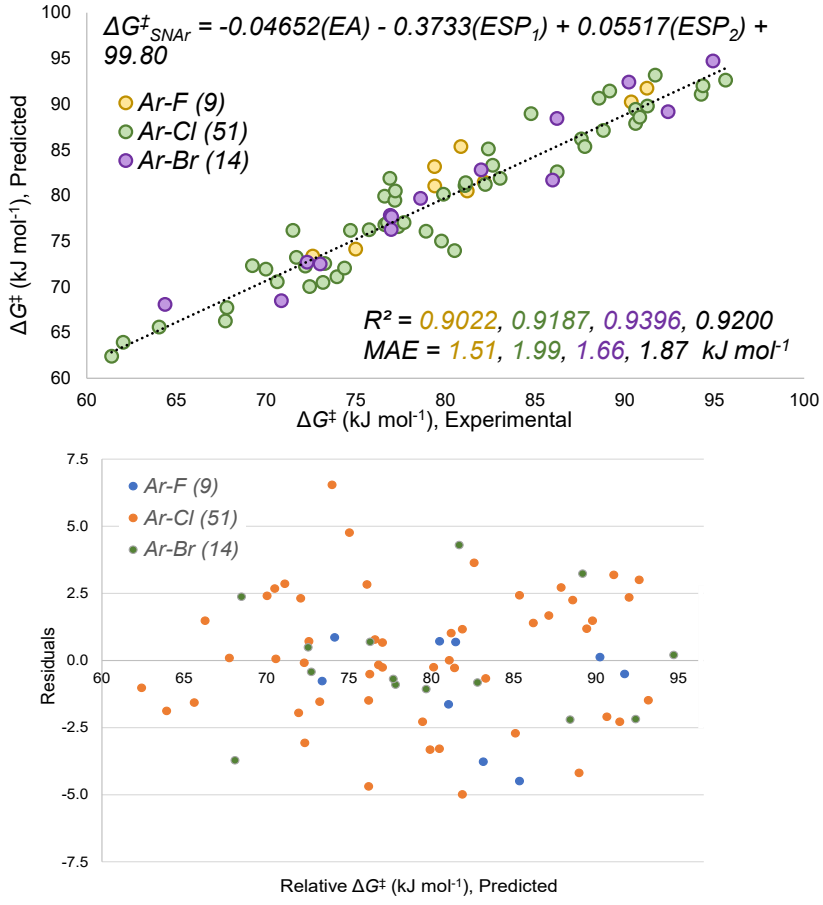
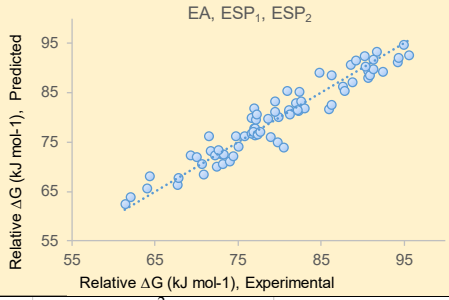
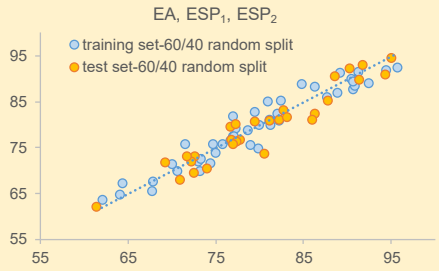
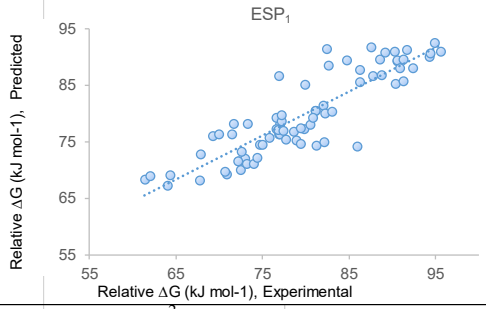
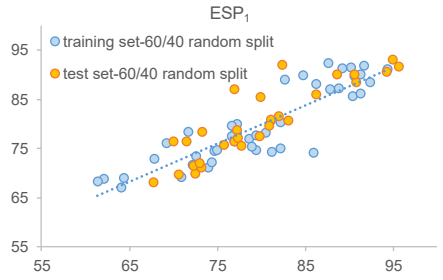
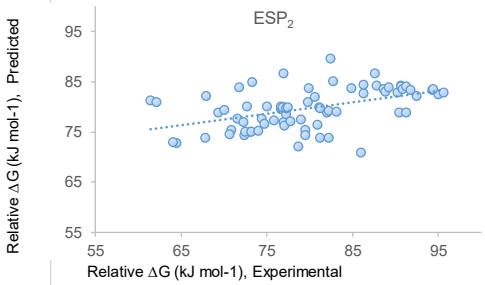
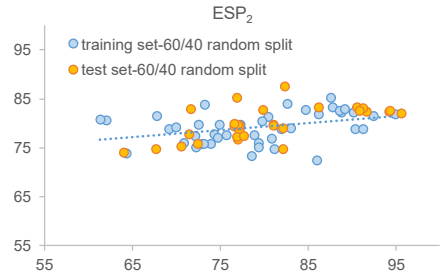
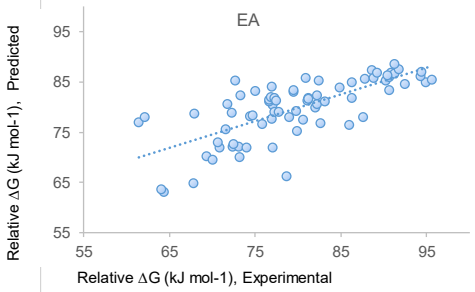
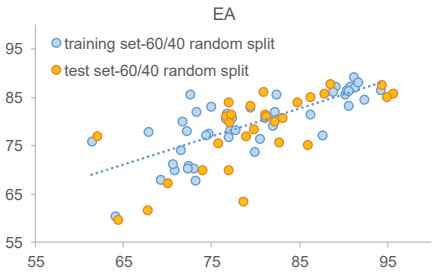
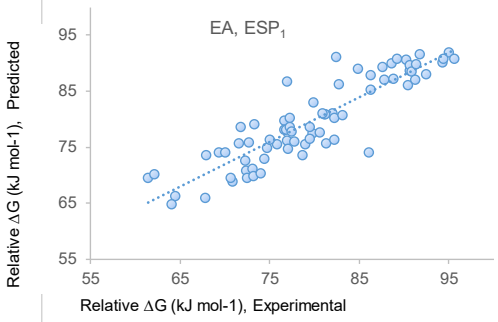


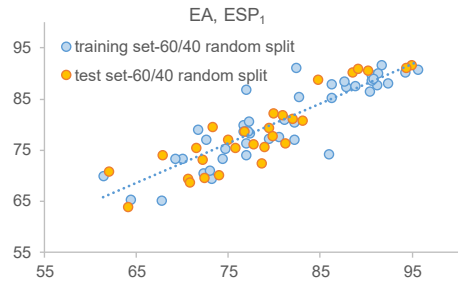
Fig. A19. Multivariate linear regression model of $\Delta G^{\ddagger}_{SNAr}$ for 74 substrates, including experimental versus predicted plot (top) and predicted versus residuals plot (bottom).

Table A7. Comparison of model performance for different combinations of molecular descriptors.

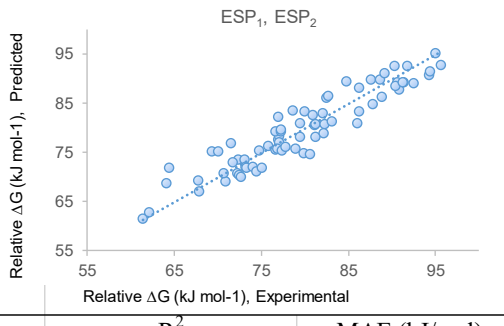
Comparison of Model Performance				
Variables	**Multivariate linear regression model performance**		**60/40 random split performance for test set**	
	R^2	MAE (kJ/mol)	Q^2	MAE (kJ/mol)
EA, ESP_1, ESP_2	0.92	1.90	0.91	2.0
				
ESP_1	R^2	MAE (kJ/mol)	Q^2	MAE (kJ/mol)
	0.77	3.0	0.82	2.6
				
ESP_2	R^2	MAE (kJ/mol)	Q^2	MAE (kJ/mol)
	0.23	5.8	0.36	5.50
				
EA	R^2	MAE (kJ/mol)	Q^2	MAE (kJ/mol)
	0.53	4.4	0.45	4.8
				

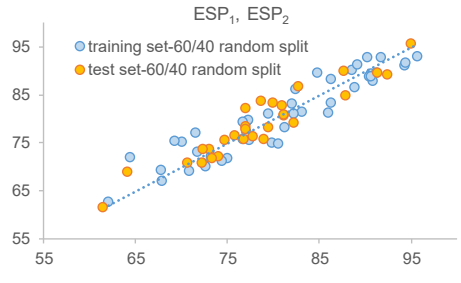
	R^2	MAE (kJ/mol)	Q^2	MAE (kJ/mol)
EA, ESP_1	0.80	2.9	0.82	2.7



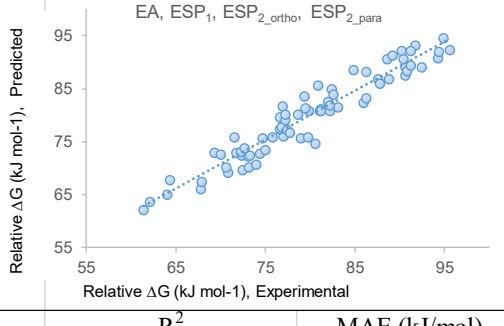


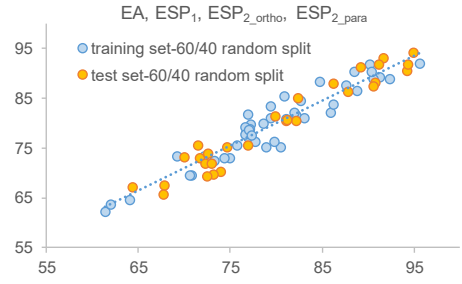
	R^2	MAE (kJ/mol)	Q^2	MAE (kJ/mol)
ESP_1, ESP_2	0.88	2.4	0.92	2.0



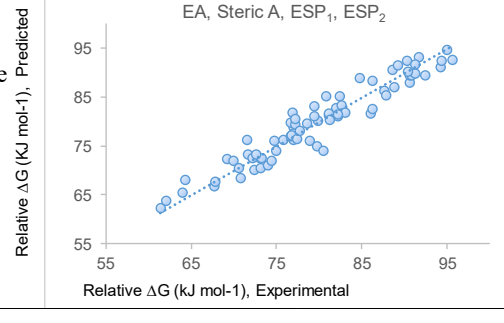


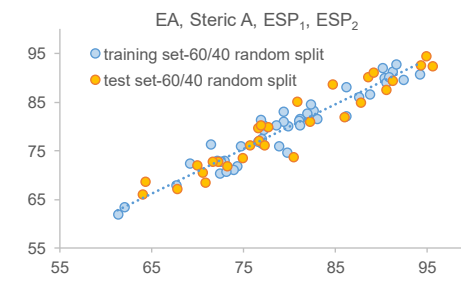
	R^2	MAE (kJ/mol)	Q^2	MAE (kJ/mol)
EA, ESP_1, ESP_{2_ortho}, ESP_{2_para}	0.92	1.9	0.91	1.9





	R^2	MAE (kJ/mol)	Q^2	MAE (kJ/mol)
EA, Steric A value, ESP_1, ESP_2	0.92	1.9	0.88	2.2





Cross Validation and Out-of-Sample Prediction

To further evaluate the linear regression model, we performed cross-validation by doing five random 60/40 training/test data splits (Figs A20-A24). Excellent linear correlation was achieved between the observed and predicted $\Delta G^{\ddagger}_{SNAr}$, as indicated by the range of R^2 from 0.89 to 0.93 for training set, Q^2 from 0.86 to 0.93, and MAE from 1.63 kJ/mol to 2.30 kJ/mol for test set. The good agreement between the observed and predicted $\Delta G^{\ddagger}_{SNAr}$ obtained from this random split cross-validation has indicated that our multivariate linear regression model is appropriately fitted, with no overfitting issues.

We also calculated the 95% prediction intervals of the predicted $\Delta G^{\ddagger}_{SNAr}$ for the test set for one of the 60/40 random split model (Fig. A20). The narrow prediction intervals (±5.1 to ±5.5 kJ/mol) indicate that the model makes reliable prediction of the S_NAr rates.

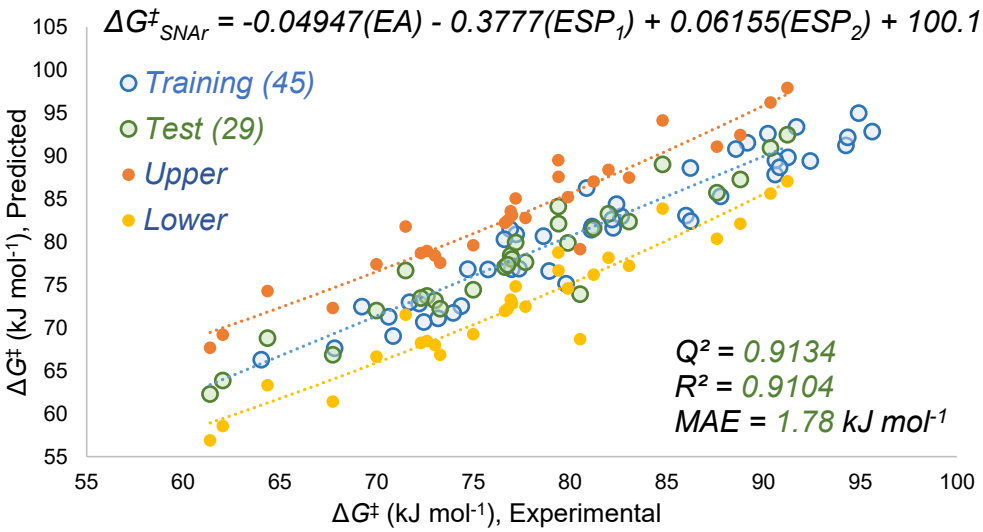


Fig. A20. Multivariate linear regression model from one of the five 60/40 random split divisions (1/5), and the 95% prediction intervals of the 29 predicted for the test set. The prediction intervals are in the range of ±5.1 to ±5.5 kJ/mol.

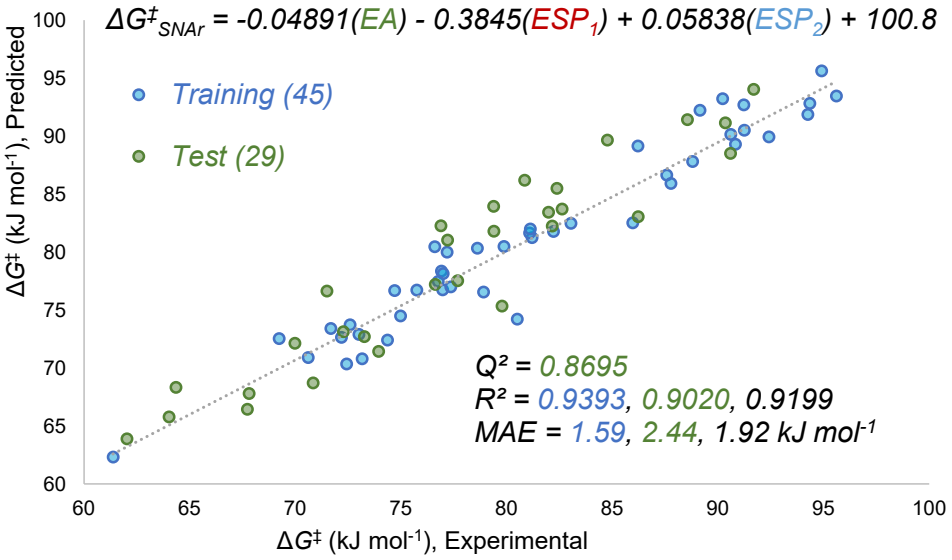


Fig. A21. Multivariate linear regression (MLR) model from one of the five 60/40 random split divisions (2/5).

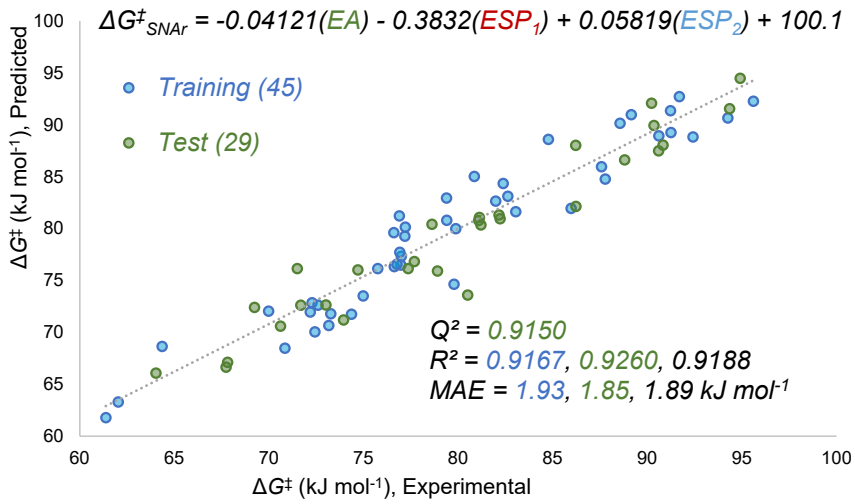


Fig. A22. MLR model from one of the five 60/40 random split divisions (3/5).

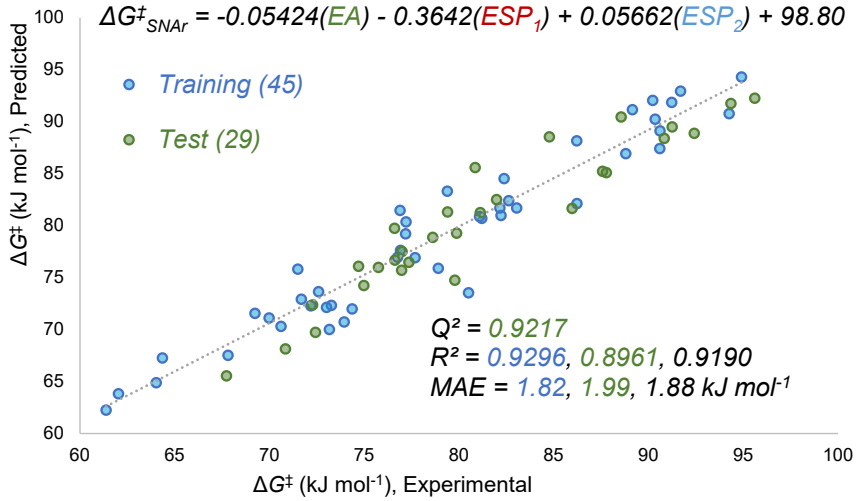


Fig. A23. MLR model from one of the five 60/40 random split divisions (4/5).

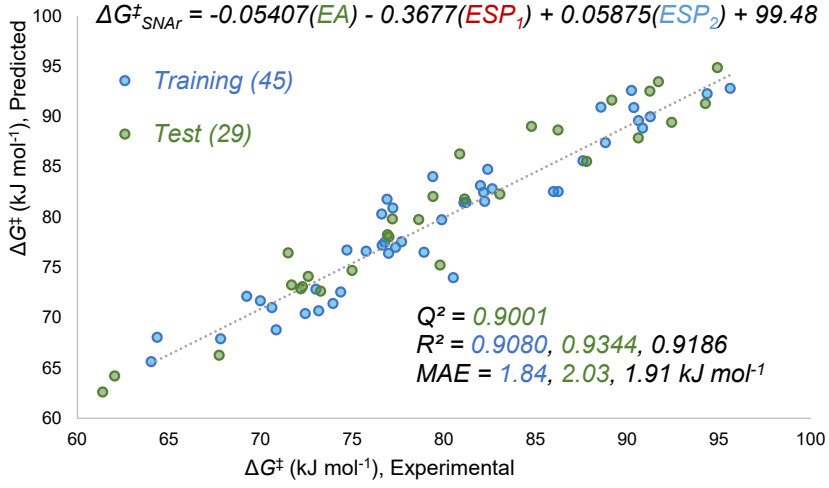


Fig. A24. MLR model from one of the five 60/40 random split divisions (5/5).

The model performance was further evaluated by an out-of-sample prediction. To test if the model can give reliable predictions for molecules with a variety of structural features, we split the data set into a training set containing only the chloropyridine substrates, and a test set containing all other substrates. The model has achieved an excellent performance with a R^2 of 0.92 for the training set, and a MAE of 1.81 kJ/mol and a Q^2 of 0.93 for the test set (Fig. A25).

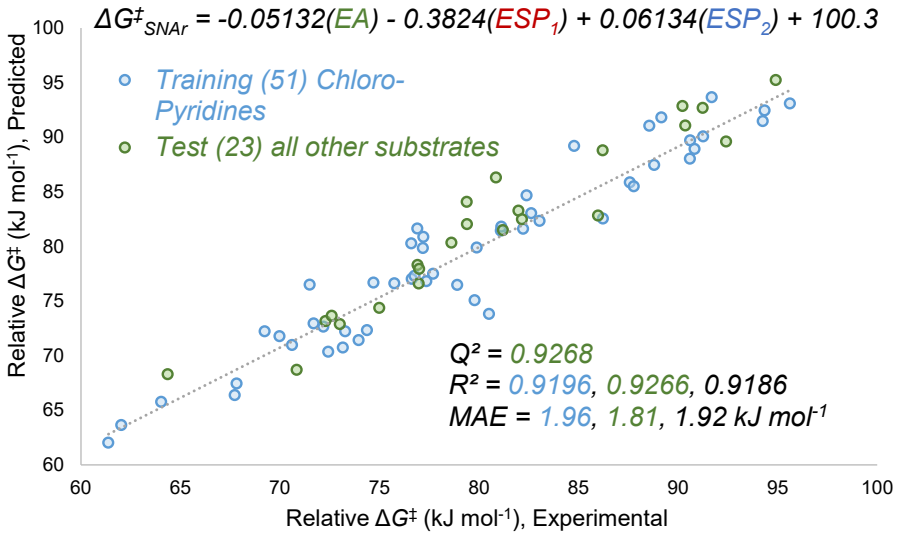


Fig. A25. Multivariate linear regression model obtained using a training set containing all chloropyridines, and a test set containing all other substrates.

LUMO energy as the Electron Affinity alternatives

LUMO energy at B3LYP/def2-TVZPD level by Orca

We use the electron affinity of a substrate as an approximation to its LUMO energy. Electron affinity is a measurable physical property, it is defined as the amount of energy released after an electron is attached to a neutral molecule. Electron affinity is calculated as the energy difference between a neutral molecule and its anionic radical by DFT. The orbital energy calculation requires less computational effort since only the DFT calculation on the neutral molecule is needed. While LUMO is a virtual orbital, and the calculated virtual orbital energy can introduce uncertainties since it is sensitive to the selection of the basis set.

We compared the LUMO energies to the electron affinity calculated at the same level of theory and basis set (B3LYP/def2-TVZPD), obtaining a linear correlation as shown in Fig. A26. The strong correlation (R^2=0.94) indicates that the LUMO energies may be used as an alternative to the electron affinity to reduce computational cost. Then we constructed a multivariate linear model using LUMO energies and ESP values as shown in Fig. A27, and the model performance (R^2=0.92 and MAE=1.91 kJ/mol) is equivalent to the model using electron affinity (Fig. A19).

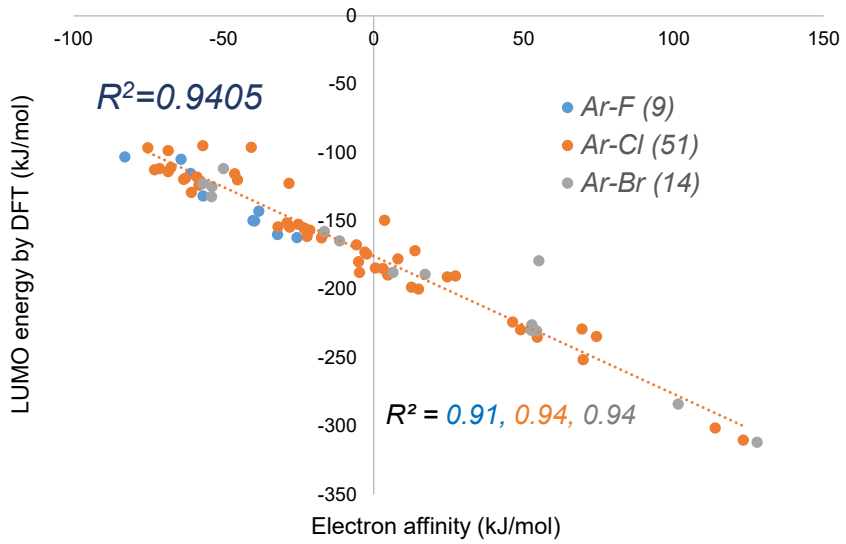


Fig. A26. Linear correlation of electron affinity (*EA*) vs. DFT-calculated *LUMO* energies at B3LYP/def2-TVZPD of the 74 substrates in the training dataset.

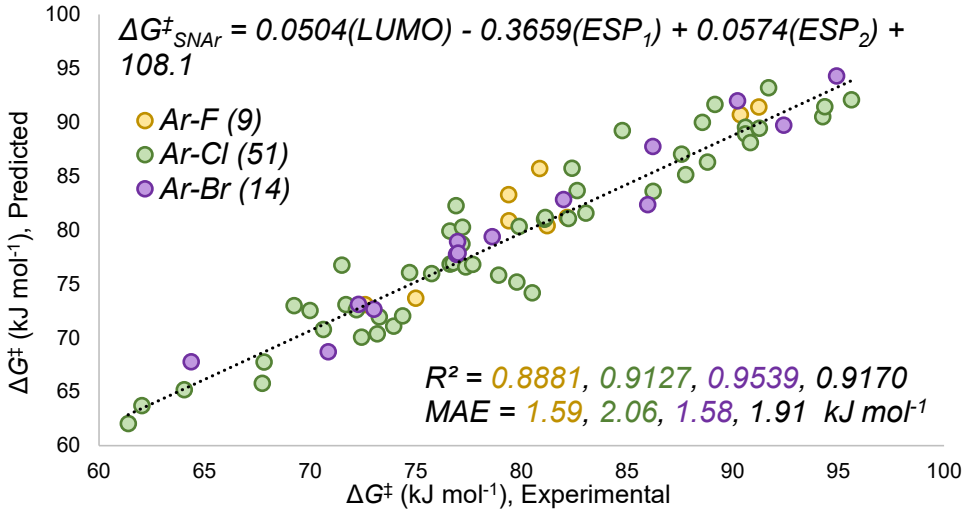


Fig. A27. Multivariate linear regression model constructed using DFT-calculated *LUMO* energies and *ESP* values at B3LYP/def2-TVZPD, and the linear regression analysis for experimental versus predicted $\Delta G^{\ddagger}_{SNAr}$.

LUMO energy at GFN1-xTB level by Entos Envision

To assess whether a rapid method at a lower level of theory could be suitable for generating *LUMO* energies that give accurate predictions, we evaluated Entos Envision[16], an interactive web-based platform for molecular simulation and visualization developed by Entos, Inc. It requires the molecular structure as input to run computational calculations using GFN1-xTB[17], a semi-empirical tight-binding method, and outputs a series of properties including molecular orbital energies. Calculation at this theory level only takes a few seconds for the size of a typical substrate in S_NAr reaction, much faster than hybrid functional DFT calculations, which may run for hours. We compared the *LUMO* energies computed in Entos to electron affinity at B3LYP/dev2-TVZPD level, obtaining a linear

correlation as shown in Fig. A28. This good correlation, with an $R^2 = 0.88$, indicates that these semi-empirical calculations give reasonable *LUMO* energy values. We used the *LUMO* energies obtained from Entos in our multivariate regression model, and compared its performance with the model built using DFT-calculated electron affinity (Fig. A19). The two models are almost identical in their performance, in terms of R^2 and MAE values (Fig. A29).

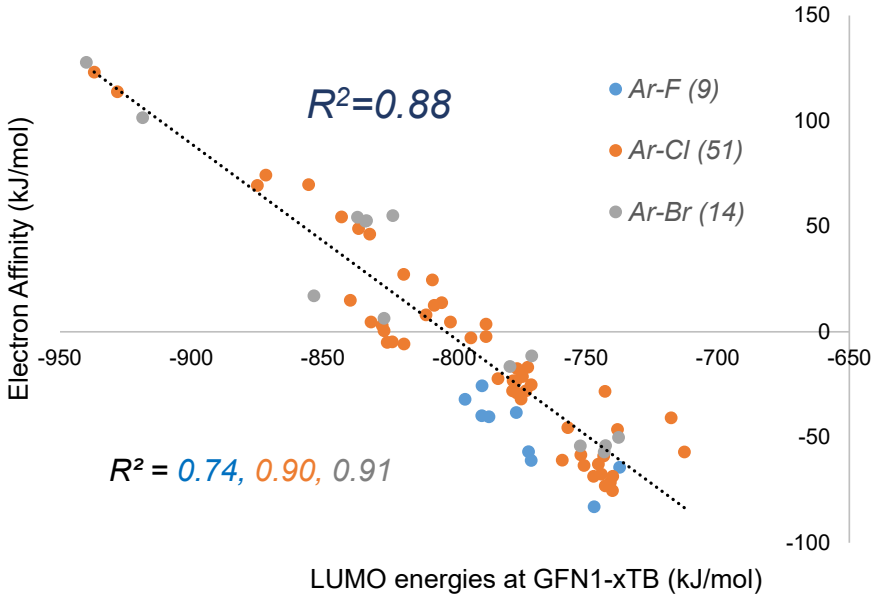


Fig. A28. Linear correlation of the electron affinity at B3LYP/def2-TVZPD vs. the LUMO energies at GFN1-xTB of the 74 substrates in the training dataset.

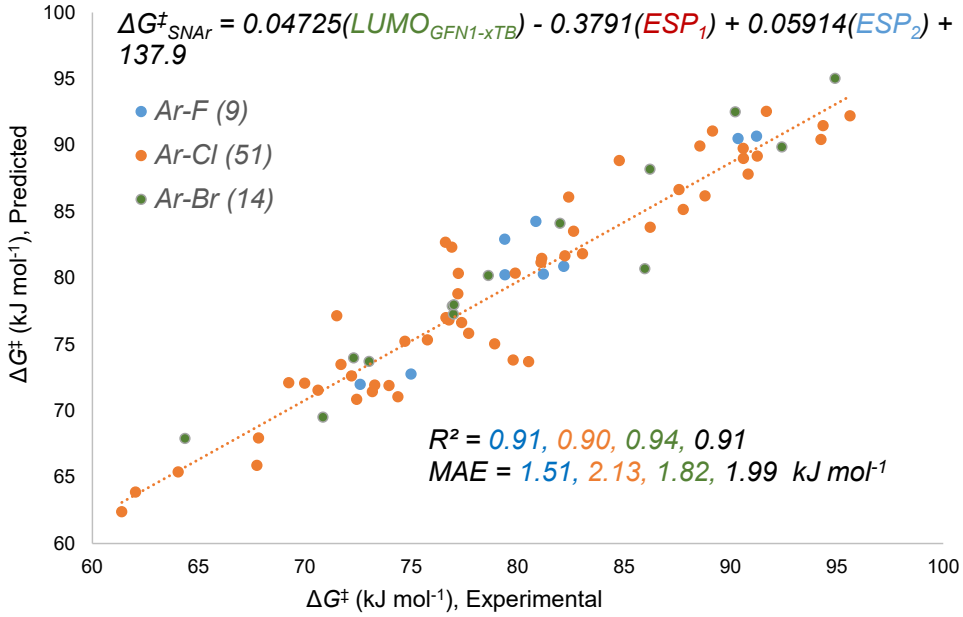


Fig. A29. Multivariate linear regression model constructed using *LUMO* energies at GFN1-xTB by Entos and *ESP* values at B3LYP/def2-TVZPD by ORCA/Multiwfn, and the linear regression analysis for experimental versus predicted $\Delta G^{\ddagger}_{SNAr}$.

Transition State Calculations of S_NAr by DFT

All transition state calculations were performed in ORCA 4.0.1.2. Geometry optimization and frequency calculations for starting materials, intermediates, transition states and products were performed at B3LYP/def2-SVPD level with D3BJ dispersion for all atoms and implicit solvation model CPCM when solvent effect is included. Additional single point calculations were performed at B3LYP/def2-TVZPD level with D3BJ dispersion and CPCM when needed.

Transition state geometries were sought through relaxed scans along a simplified reaction coordinate (the breaking or forming bond length), at a B3LYP/def2-SVPD/CPCM (when needed) level with D3BJ dispersion. The geometry at the maximum energy point of the reaction coordinate scan was used as input geometry for continuing the transition state search via saddle-point optimization.

The optimization to a saddle point was performed using a B3LYP/def2-SVPD/CPCM (when needed) approach with D3BJ dispersion. The transition state structure was confirmed by vibrational analysis. The vibrational mode corresponding to the single imaginary frequency was visualized to ensure that it lied along the reaction coordinate bond stretch.

The free energies of the starting materials, intermediates and transition states were calculated as the sum of the electronic energies (single point energy at B3LYP/def2-TVZPD level) and the Gibbs free energy correction (obtained from frequency calculations at B3LYP/def2-SVPD level). Gibbs free energies of activation ($\Delta G^{\ddagger}_{SNAr}$) for the reactions were calculated as the difference in free energies between the transition state and the starting materials.

These labor-intensive transition state calculations were pursued for 6 di-halogenated heterocycles at both reactive sites except for 2,4-dichloro-1,3,5-triazine, for which the two sites are identical. In implicit solvation model CPCM(DMSO), concerted pathways were found at all reactive centers, and no sigma-complex intermediates could be located; the starting materials transformed into the products *via* only one intrinsic reaction coordinate (the bond length between the reactive carbon at the substrate and the negatively charged oxygen at deprotonated benzyl alcohol). Transition state structures and energy diagrams are shown in Figs. A30 – A39. All molecular images were generated in Vesta[18].

In gas phase, stepwise mechanisms were found for the carbon-fluorine sites at 2-chloro-3-fluoro-pyridine and 2-chloro-4-fluoro-pyridine (Figs. A38 – A39). The stable sigma-complex was located from geometry optimization and confirmed by frequency calculations, that no negative frequencies were found. The first step, addition of the nucleophile, is an exothermic reaction that involves a very early transition state. In these cases, the transition state could not be found as no maximum point was accessible from the potential energy surface scan. This computational challenge has been reported by Jacobsen and co-workers' in their S_NAr mechanistic study, where transition state searches failed in similar exothermic steps.[13] Transition states for the second step (the departure of the fluorine) were located and the structures were optimized and confirmed by vibrational analysis for both of the carbon-fluorine sites.

HOMO diagrams for the transition state were calculated at both the major and minor reactive sites for the following 3 substrates: 2,4-dichloropyridine (Fig. 6, main text), 2-chloro-4-bromopyridine (Fig. A32) and 2,4-dichloropyrimidine (Fig. A34). One motivation to conduct transition state calculations is to look for the mechanistic factors that

determine the intramolecular selectivity, and our orbital analysis shows that the frontier orbital interaction is certainly among those important factors. Transition state frontier orbital analysis shows that different sites of a substrate use different antibonding orbitals to interact with the nucleophile, where those orbitals are accessible through thermal fluctuations. In these cases, the major site undergoes S_NAr corresponding to the *LUMO* orbital, while the minor site corresponds to *LUMO*+1.

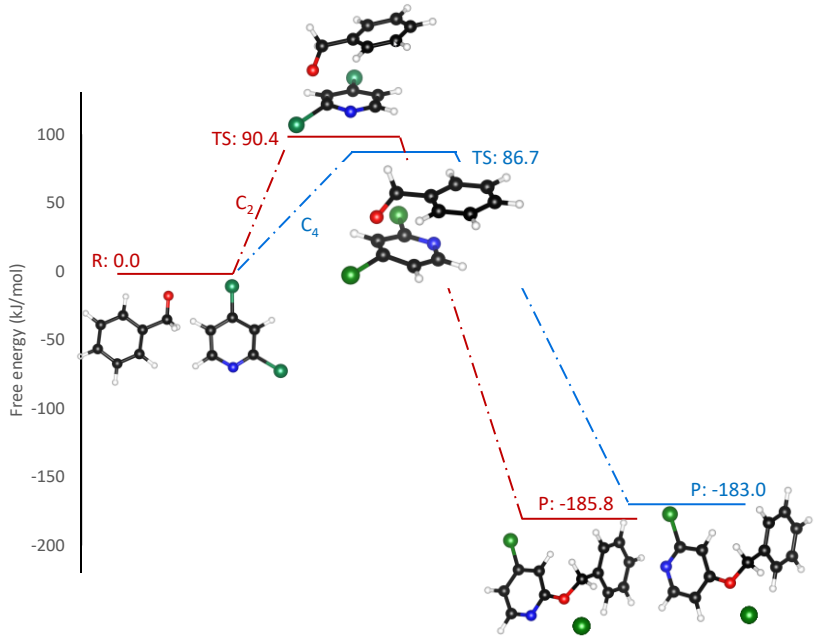


Fig. A30. Calculated reaction coordinates for S_NAr of 2,4-dichloropyridine and the anion of benzyl alcohol in DMSO. The blue pathway is the S_NAr coordinate at C_4, the major reactive site; and the red pathway is the S_NAr coordinate at C_2, the minor reactive site.

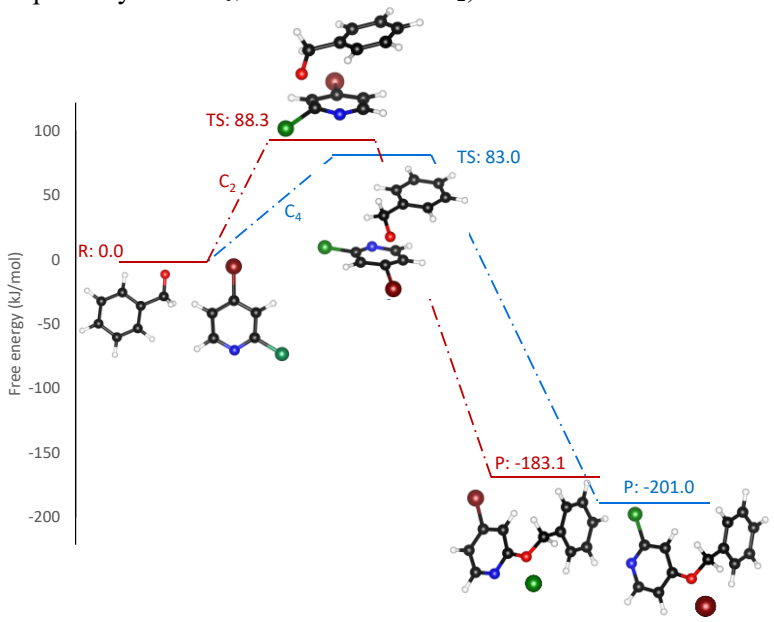


Fig. A31. Calculated reaction coordinates for S_NAr of 2-chloro-4-bromopyridine and the anion of benzyl alcohol in DMSO. The blue pathway is the S_NAr coordinate at C_4, the major reactive site; and the red pathway is the S_NAr coordinate at C_2, the minor reactive site.

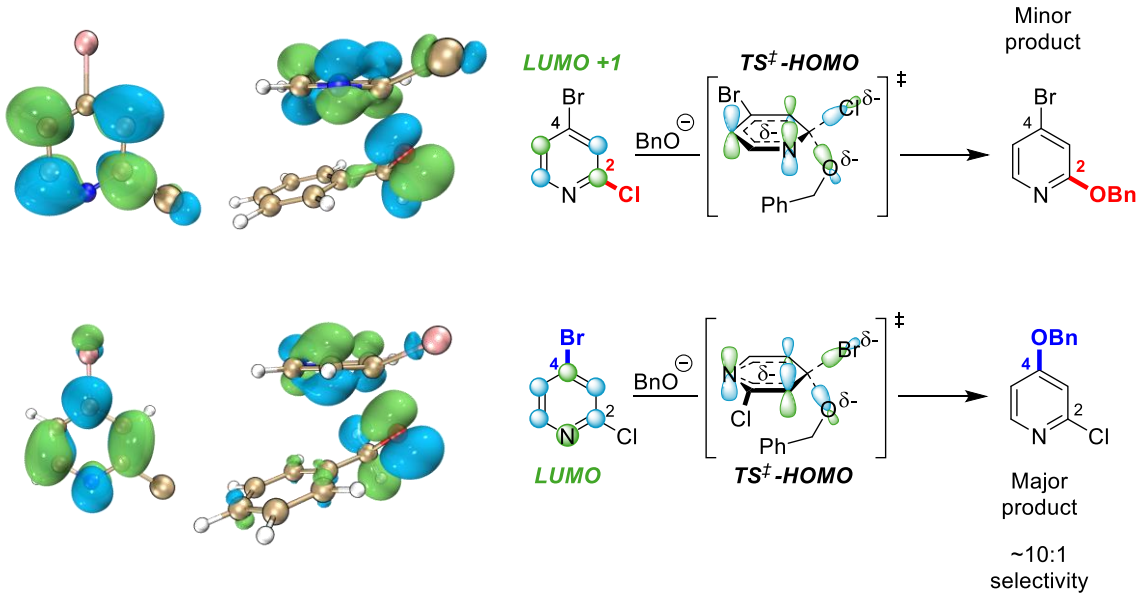


Fig. A32. Frontier orbitals of 2-chloro-4-bromopyridine and the anion of benzyl alcohol involved in S_NAr, and *HOMO* orbitals of the transition states at C_2 and C_4. The major reactive site C_4 undergoes S_NAr with the *LUMO* orbital, as shown as the blue reaction pathway in Fig. A31. The minor reactive site C_2 undergoes S_NAr with the *LUMO*+1 orbital, as shown as the red reaction pathway in Fig. A31.

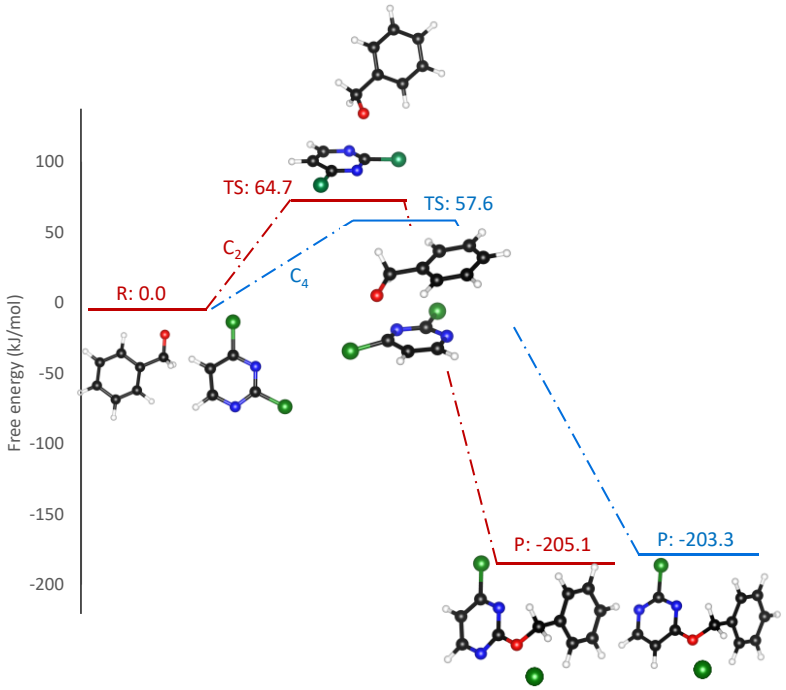


Fig. A33. Calculated reaction coordinates for S_NAr of 2,4-dichloro-pyrimidine and the anion of benzyl alcohol in DMSO. The blue pathway is the S_NAr coordinate at C_4, the major reactive site; and the red pathway is the S_NAr coordinate at C_2, the minor reactive site.

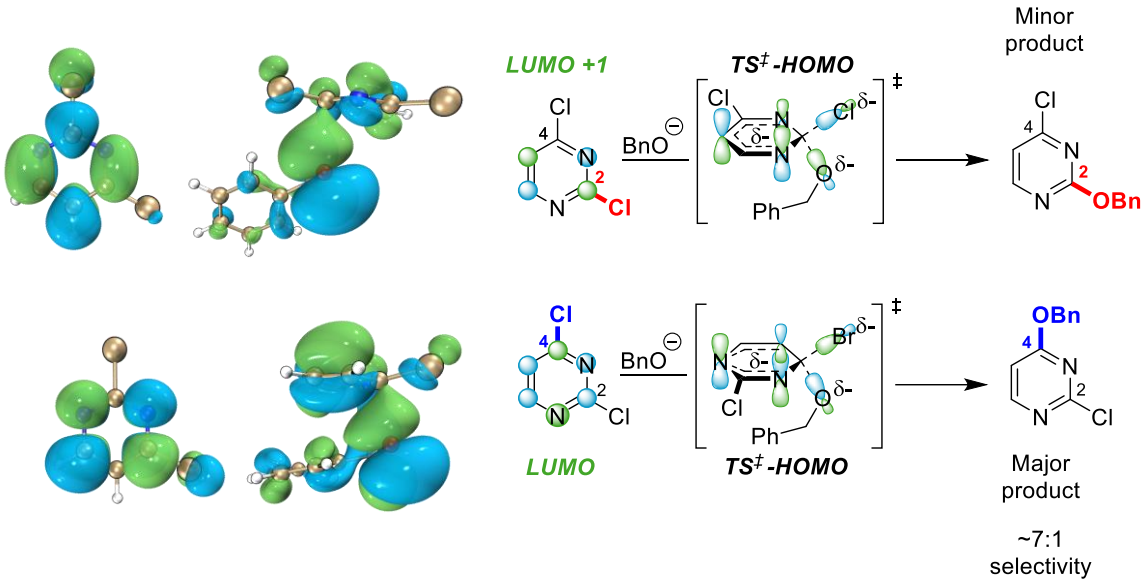


Fig. S34. Frontier orbitals of 2,4-dichloropyrimidine and the anion of benzyl alcohol involved in S_NAr, and *HOMO* orbitals of the transition states at C_2 and C_4. The major reactive site C_4 undergoes S_NAr with the *LUMO* orbital, as shown as the blue reaction pathway in Fig. A33. The minor reactive site C_2 undergoes S_NAr with the *LUMO*+1 orbital, as shown as the red reaction pathway in Fig. A33.

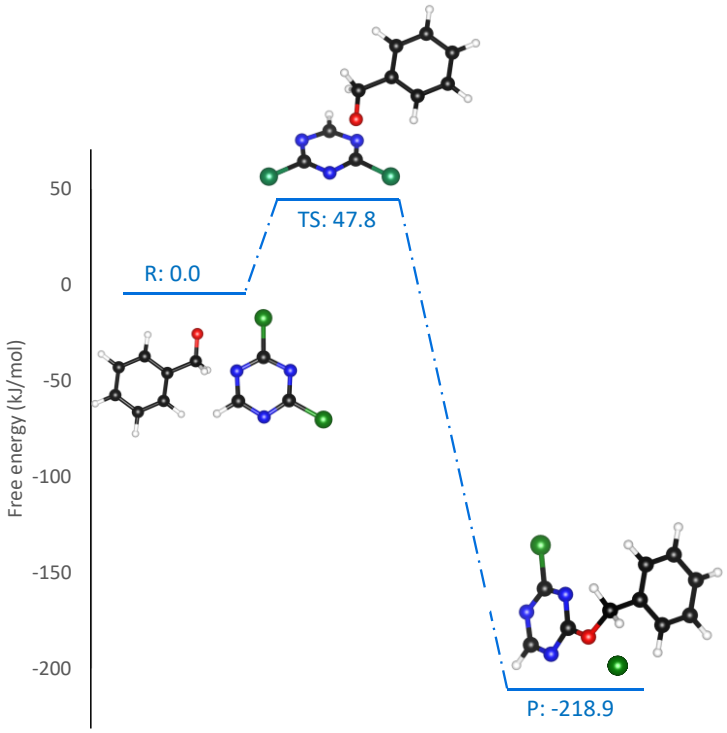


Fig. A35. Calculated reaction coordinates for S_NAr of 2,4-dichloro-1,3,5-triazine and the anion of benzyl alcohol in DMSO.

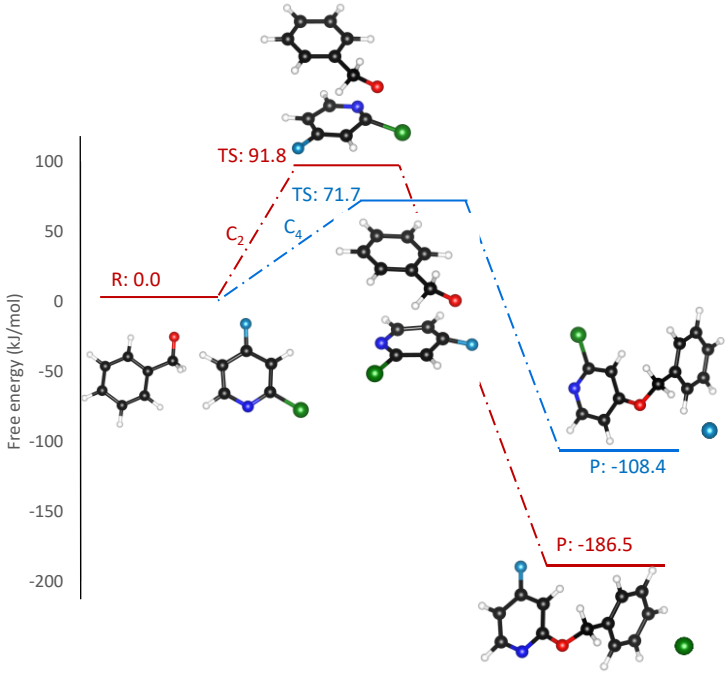


Fig. A36. Calculated reaction coordinates for S_NAr of 2-chloro-4-fluoropyridine and the anion of benzyl alcohol in DMSO. The blue pathway is the S_NAr coordinate at C_4, the major reactive site; and the red pathway is the S_NAr coordinate at C_2, the minor reactive site.

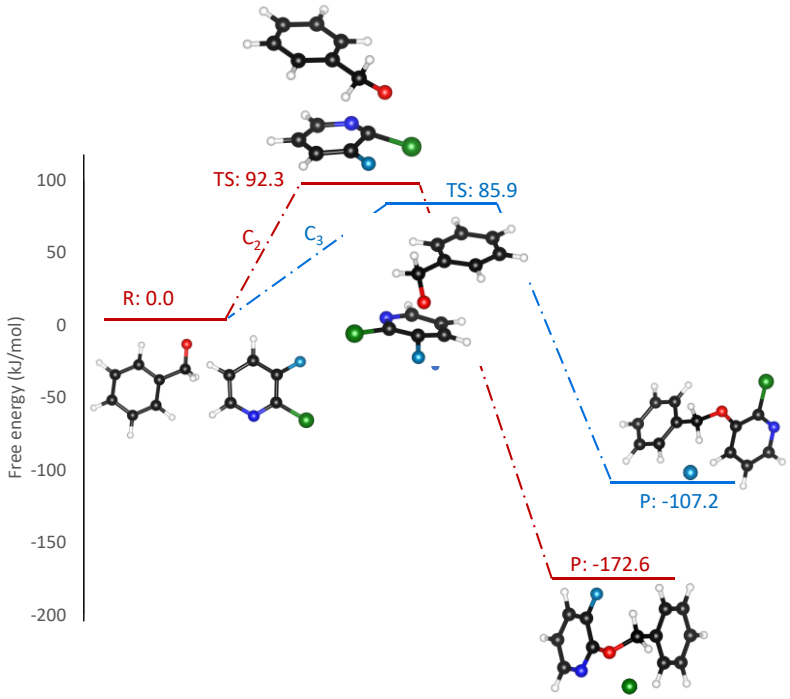


Fig. A37. Calculated reaction coordinates for S_NAr of 2-chloro-3-fluoropyridine and the anion of benzyl alcohol in DMSO. The blue pathway is the S_NAr coordinate at C_3, the major reactive site; and the red pathway is the S_NAr coordinate at C_2, the minor reactive site.

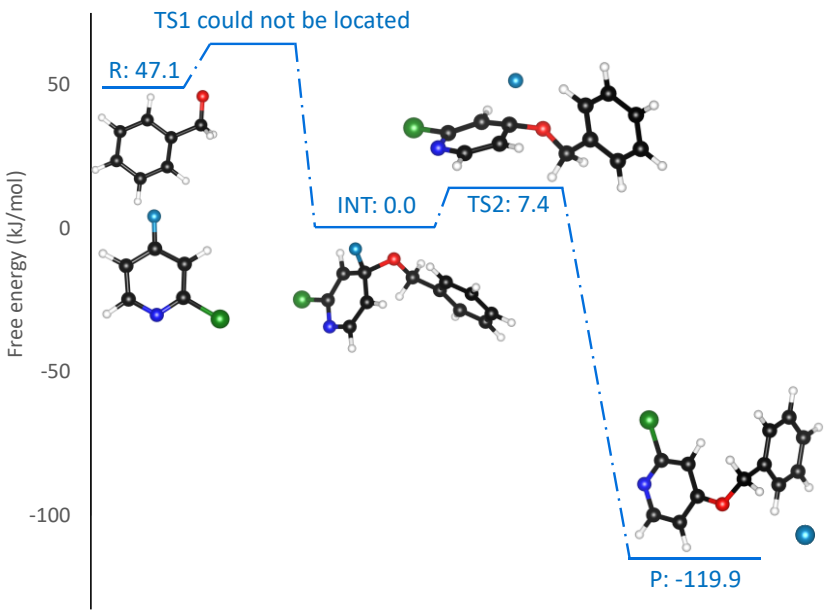


Fig. A38. Calculated reaction coordinate for S_NAr at the carbon-fluorine site (C_4) of 2-chloro-4-fluoropyridine and the anion of benzyl alcohol in the gas phase.

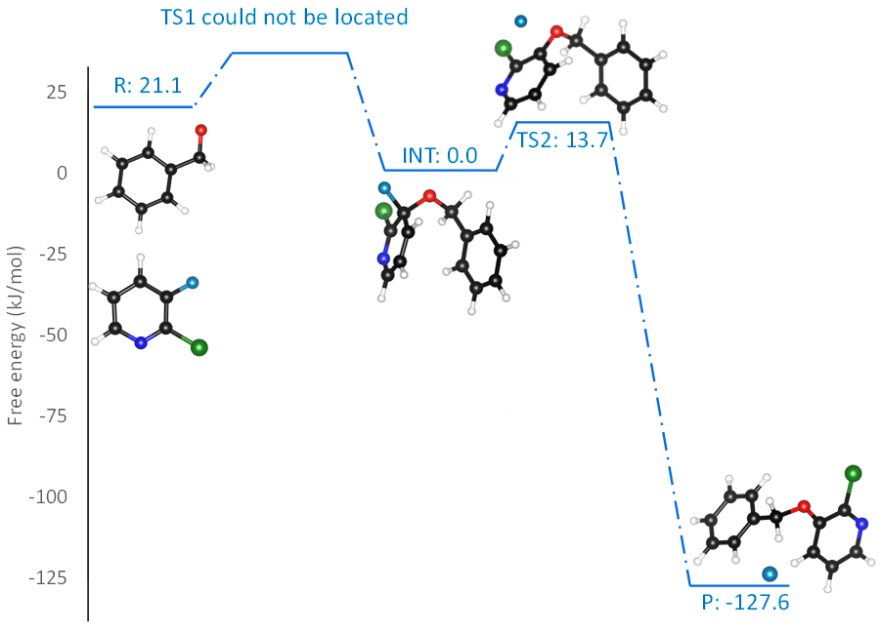


Fig. A39. Calculated reaction coordinates for S_NAr at the carbon-fluorine site (C_3) of 2-chloro-3-fluoropyridine and the anion of benzyl alcohol in the gas phase.

Experimental details for assessing outliers from external case studies (Fig. 2.9C)

To assess the model's applicability on new S_NAr reactions beyond our training dataset, we conducted a series of external validation on both S_NAr rate correlation and site-selectivity prediction. We also performed experimental checks on the one significant outlier that we observed from the S_NAr rate correlation test (Substrate 6C, Fig. 2.9C).

This substrate 6C (1-bromo-4-fluoro-2-nitrobenzene) has two potential S_NAr sites (Ar–F and Ar–Br) and is expected to form multiple products from reaction with piperidine.

The substrate was purchased from a commercial supplier and used as is. The reaction was conducted using the same reaction conditions from literature[19] (which is also shown in Fig. A40A) for two days, then the reaction solution was analyzed by LCMS. The molecular ion peaks from the LCMS analysis gives experimental evidence that the reaction of substrate 6C and piperidine leads to two S_NAr products. The product ratio of 1.5:1 was obtained from the peak area ratio from the UV spectrum at 254 nm, with the Ar-Br site slightly favoured.

Excellent linear correlation ($R^2 = 0.90$) was obtained for the dataset in Fig. 2.9C with all the substrates included, and exclusion of the outlier 6C gives rise to a stronger correlation, with an R^2 of 0.94.

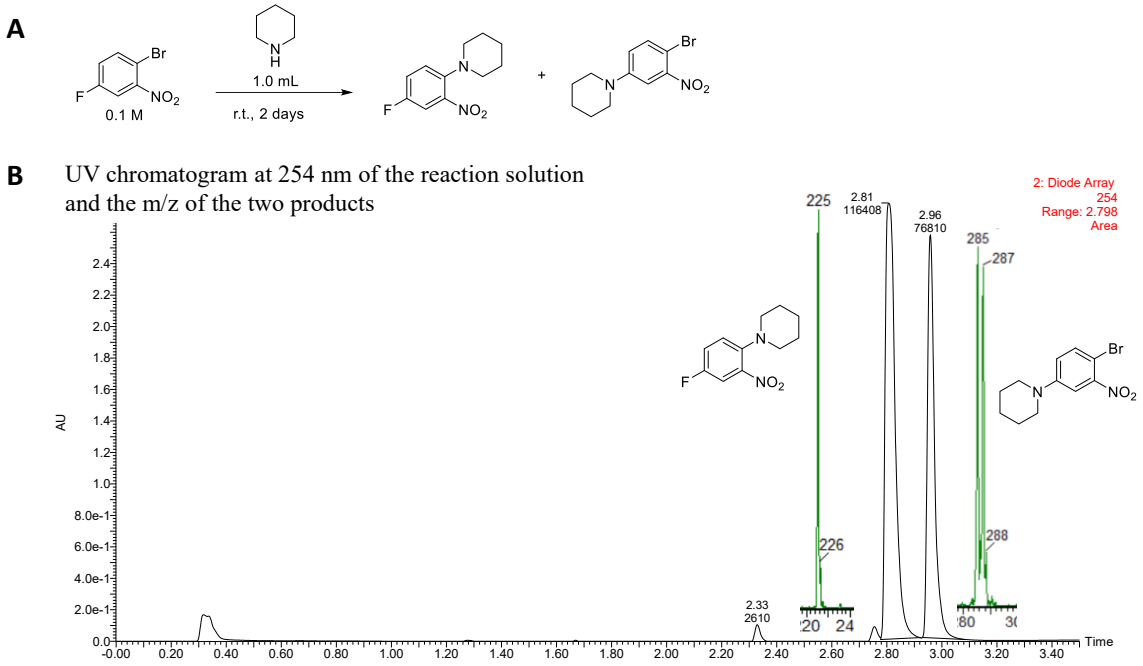


Fig. A40. A). Reaction condition of substrate 6 in Fig.7C reacting with piperidine. B). UV chromatogram at 254 nm extracted from LCMS analysis of the reaction solution, and the m/z of the two S_NAr products.

Reference

(1) Sengmany, S.; Lebre, J.; Le Gall, E.; Léonel, E. Selective Mono-Amination of Dichlorodiazines. *Tetrahedron* **2015**, *71* (29), 4859–4867. https://doi.org/10.1016/j.tet.2015.05.056.

(2) Pöller, S.; Schuhmann, W. A Miniaturized Voltammetric PH Sensor Based on Optimized Redox Polymers. *Electrochimica Acta* **2014**, *140*, 101–107. https://doi.org/10.1016/j.electacta.2014.03.116.
(3) Cash, B. M.; Prevost, N.; Wagner, F. F.; Comins, D. L. Studies toward the Total Synthesis of Dihydrolycolucine. Preparation of AB and CEF Ring Fragments. *J. Org. Chem.* **2014**, *79* (12), 5740–5745. https://doi.org/10.1021/jo500878v.
(4) Gellrich, U.; Huang, J.; Seiche, W.; Keller, M.; Meuwly, M.; Breit, B. Ligand Self-Assembling through Complementary Hydrogen-Bonding in the Coordination Sphere of a Transition Metal Center: The 6-Diphenylphosphanylpyridin-2(1H)-One System. *J. Am. Chem. Soc.* **2011**, *133* (4), 964–975. https://doi.org/10.1021/ja108639e.
(5) Hansch, Corwin.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91* (2), 165–195. https://doi.org/10.1021/cr00002a004.
(6) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA Quantum Chemistry Program Package. *J. Chem. Phys.* **2020**, *152* (22), 224108. https://doi.org/10.1063/5.0004608.
(7) Lu, T.; Chen, F. Quantitative Analysis of Molecular Surface Based on Improved Marching Tetrahedra Algorithm. *J. Mol. Graph. Model.* **2012**, *38*, 314–323. https://doi.org/10.1016/j.jmgm.2012.07.004.
(8) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33* (5), 580–592. https://doi.org/10.1002/jcc.22885.
(9) Pence, H. E.; Williams, A. ChemSpider: An Online Chemical Information Resource. *J. Chem. Educ.* **2010**, *87* (11), 1123–1124. https://doi.org/10.1021/ed100697w.
(10) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminformatics* **2012**, *4* (1), 17. https://doi.org/10.1186/1758-2946-4-17.
(11) Koopmans, T. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. *Physica* **1934**, *1* (1), 104–113. https://doi.org/10.1016/S0031-8914(34)90011-2.
(12) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.
(13) Rohrbach, S.; Smith, A. J.; Pang, J. H.; Poole, D. L.; Tuttle, T.; Chiba, S.; Murphy, J. A. Concerted Nucleophilic Aromatic Substitution Reactions. *Angew. Chem. Int. Ed.* **2019**, *58* (46), 16368–16388. https://doi.org/10.1002/anie.201902216.
(14) Terrier, F. Rate and Equilibrium Studies in Jackson-Meisenheimer Complexes. *Chem. Rev.* **1982**, *82* (2), 77–152. https://doi.org/10.1021/cr00048a001.
(15) *J. A. Hirsch Topics in Stereochemistry , John Wiley & Sons, Ltd, 1967, Pp. 199–222*.
(16) *Interactive Chemistry | Envision*. https://www.entos.ai/envision (accessed 2023-08-11).
(17) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All Spd-Block Elements (Z = 1–86). *J. Chem. Theory Comput.* **2017**, *13* (5), 1989–2009. https://doi.org/10.1021/acs.jctc.7b00118.
(18) Momma, K.; Izumi, F. VESTA: A Three-Dimensional Visualization System for Electronic and Structural Analysis. *J. Appl. Crystallogr.* **2008**, *41* (3), 653–658. https://doi.org/10.1107/S0021889808012016.
(19) Berliner, E.; Monack, L. C. Nucleophilic Displacement in the Benzene Series [1]. *J. Am. Chem. Soc.* **1952**, *74* (6), 1574–1579. https://doi.org/10.1021/ja01126a069.

Appendix B: Supporting Information for Chapter 3

Other Supporting information for this chapter include the following:

Lu-PhD-thesis_Ch.3 SI tables. (separate file; *.xlsx, electronically available at *https://doi.org/10.20383/103.0795*)
Extended tables in Microsoft Excel format (.xlsx) that contain list of competition reactions, measured $\Delta G^{\ddagger}_{OA}$ and molecular descriptors.

Lu-PhD-thesis_Ch.3_MLR models. (separate file; *.xlsx, electronically available at *https://doi.org/10.20383/103.0795*)
Extended tables in Microsoft Excel format (.xlsx) that contain predicted $\Delta G^{\ddagger}_{OA}$ calculations, and statistical analysis for multivariate linear regression models.

General Considerations

Materials

All solvents, reagents, and organic substrates were used as purchased from commercial suppliers without further purification with the following exceptions. 4-Chloro-6-(piperidin-1-yl)pyrimidine and 4-chloro-6-(pyrrolidin-1-yl)pyrimidine were prepared using published procedures.[1] 4-Chloro-2-methylpyrimidine and 2-Chloro-5-aminopyridine were purified by dissolving the commercial material in THF or chloroform, followed by filtration to remove insoluble impurities; purity of these materials was confirmed by NMR spectroscopy. The 9 aryl triflates were prepared using a published procedure,[2] and the NMR characterization data has been reported in published works.[2–8] Bis(tricyclohexylphosphine)palladium(0) was purchased from Strem Chemicals and used as received. All reactions were performed inside an MBraun glovebox under an N_2 atmosphere.

Analysis and Spectroscopy

All NMR spectra were recorded on either a Bruker AVANCE 300 MHz spectrometer or a Bruker AVANCE NEO 500 MHz spectrometer. All $^{31}P\{^{1}H\}$ qNMR (quantitative Nuclear Magnetic Resonance) spectra were recorded on the 500 MHZ spectrometer using a $^{31}P\{^{1}H\}$ NMR parameter set with a relaxation delay (D_1) of 20 s and 64 scans. This delay time was chosen based on the measured T_1 relaxation time for the oxidative addition products generated in a competition reaction between 2-chloro-pyridine and 2-chloro-5-methylpyridine reacting with $Pd(PCy_3)_2$ (Fig. B1). The longest T_1 observed peaks from 0 ppm to 50 ppm was 4.35 s. By setting the delay time D_1 to 20 s and the acquisition time to 1.7 s, there is 5 times the T_1 period between scans to ensure all nuclei giving peaks from 0 ppm to 50 ppm have fully relaxed after each scan. Note that the ^{31}P nuclei from the internal standard triphenylphosphine (dissolved in C_6D_6 in a sealed capillary added to the NMR tube) has a T_1 longer than 13 s; as a result, the signal of PPh_3 has not fully relaxed during a D_1 of 20 s. In our $^{31}P\{^{1}H\}$ NMR spectroscopy analysis, PPh_3 was only used as a chemical shift standard, and the product ratio in a competition reaction was calculated directly from the peak area ratio of the two oxidative addition products from the ^{31}P qNMR spectrum.

High-resolution electrospray ionization mass spectrometric (HRMS) analysis was performed using a Thermo Scientific Ultimate 3000 ESI-Orbitrap Exactive Plus.

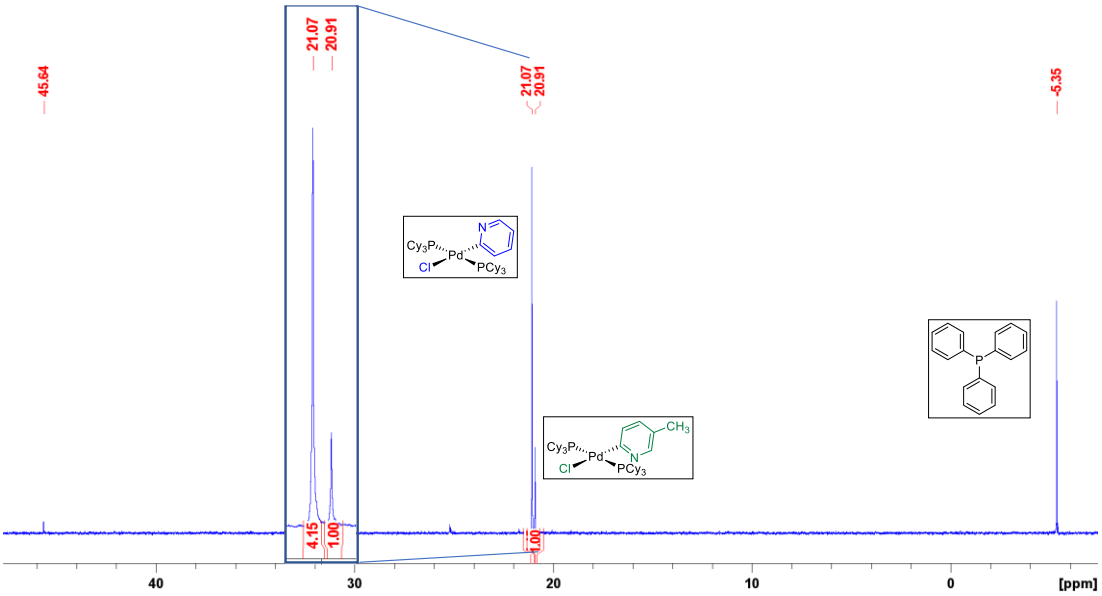


Fig. B1. $^{31}P\{^{1}H\}$ qNMR spectrum of the competition reaction of 2-chloro-pyridine and 2-chloro-5-methyl-pyridine with $Pd(PCy_3)_2$; T_1 relaxation values were determined for these two Pd species, which was used to inform the delay time setting of 20 s.

Experimental Details for Oxidative Addition Competition Studies

Bis(tricyclohexylphosphine)palladium(0) ($Pd(PCy_3)_2$) was used as the palladium source. A library of 70 (hetero)aryl chlorides and bromides was used as the substrates. THF was used as the reaction solvent. Stock solutions were prepared for each component: the concentration of the substrate stock solutions was 0.375 M, and the concentration of $Pd(PCy_3)_2$ stock solution was 0.0375 M.

To assess each individual (hetero)aryl halide for oxidative addition reactivity in combination with $Pd(PCy_3)_2$ and determine the ^{31}P NMR chemical shift of the resulting Pd(II) product, individual oxidative addition reactions were carried out at room temperature on 1 mL scale. In a 4 mL vial containing a stirbar, an aliquot of the substrate stock solution (400 µL, 0.150 mmol) was diluted with additional reaction solvent (THF, 400 µL), followed by addition of an aliquot of the $Pd(PCy_3)_2$ stock solution (200 µL, 0.00750 mmol). The resulting solution was mixed for 2-18 hours, then a 600 µL sample was transferred to an NMR tube containing a capillary filled with PPh_3 in C_6D_6. A $^{31}P\{^{1}H\}$ NMR spectrum was recorded to confirm the oxidative addition reaction occurred and went to completion, and to identify the ^{31}P chemical shift of the oxidative addition product. The reaction progress was determined after 2 hours by NMR analysis of an aliquot of the solution. If not complete, the remaining reaction solution was stirred overnight for a total reaction time of at least 18 hours before $^{31}P\{^{1}H\}$ NMR analysis.

Competition experiments were performed to obtain the relative activation energies ($\Delta\Delta G^{\ddagger}_{OA}$) between two substrates in palladium oxidative addition (Fig. B2). All reactions were conducted at room temperature on 1 mL scale under *pseudo* first-order conditions by

adding the two substrates in excess but equal amount (0.150 mmol, 1.0 equivalent) to compete with one palladium source (0.0750 mmol, 0.05 equivalent). In order to accurately measure the oxidative addition product ratio in a competition reaction, the ^{31}P NMR chemical shifts of the products must differ by at least 0.1 ppm; appropriate resolution was confirmed prior to competition experiment set-up by comparing the chemical shifts obtained in the aforementioned individual oxidative addition experiments. The competition reactions were prepared similarly to the individual oxidative addition experiments, with 400 µL of each substrate stock solution and 200 µL of $Pd(PCy_3)_2$ stock solution mixed in a small vial. The initial concentration of each substrate was 0.150 M, and the initial concentration of $Pd(PCy_3)_2$ was 0.00750 M. After mixing the reaction solution for 2-18 hours, ^{31}P qNMR spectra were recorded to measure the peak area ratio between two oxidative addition products.

The competition experiments listed in Table B1, entries 92-95 were performed between an aryl bromide and an aryl triflate. Stock solution of tetrabutylammonium was prepared and the concentration was 0.075 M. After mixing the reaction solution for 2 hours, 300 µL of the tetrabutylammonium bromide was added and the reaction solution was stirred for 20 minutes before NMR analysis. Excess amount of tetrabutylammonium bromide was added to convert the oxidative addition product of Pd(II)-triflate species into Pd(II)-bromide species (Fig. B2). Because the Pd(II) oxidative addition product of 4-cyano phenyl triflate is not soluble in THF, stock solution of tetrabutylammonium bromide was also added into the competition reaction mixture that involves 4-cyano phenyl triflate (Table B1, entry 96: competition reaction between 4-cyano phenyl triflate and 4-acetylphenyl triflate).

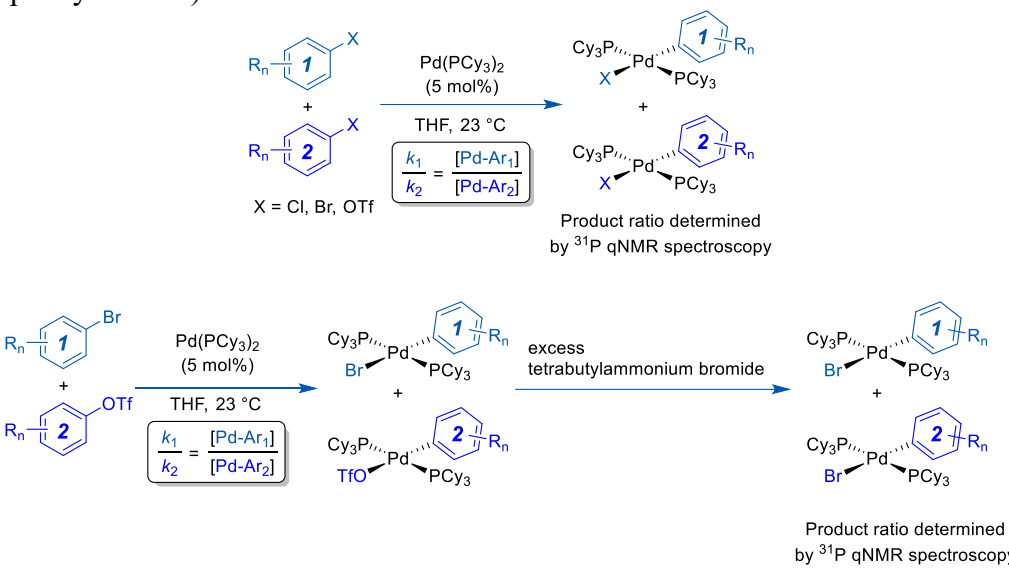


Fig. B2. *Top*: General competition experiment design. *Bottom*: Specific design of competition experiments between an aryl bromide and an aryl triflate.

Table S1 contains the ^{31}P NMR chemical shifts and peak area ratios of 98 competition reactions performed with 79 substrates in THF. All ^{31}P NMR chemical shifts are referenced to triphenylphosphine (-6.00 ppm) as an internal chemical shift standard.

Table B1 ^{31}P NMR chemical shifts and the peak area ratios of 98 competition reactions in THF (Lu-PhD-thesis_Ch.3 SI tables.xlsx).

Figures B3 to B5 contain $^{31}P\{^{1}H\}$ qNMR spectra for three representative competition reactions.

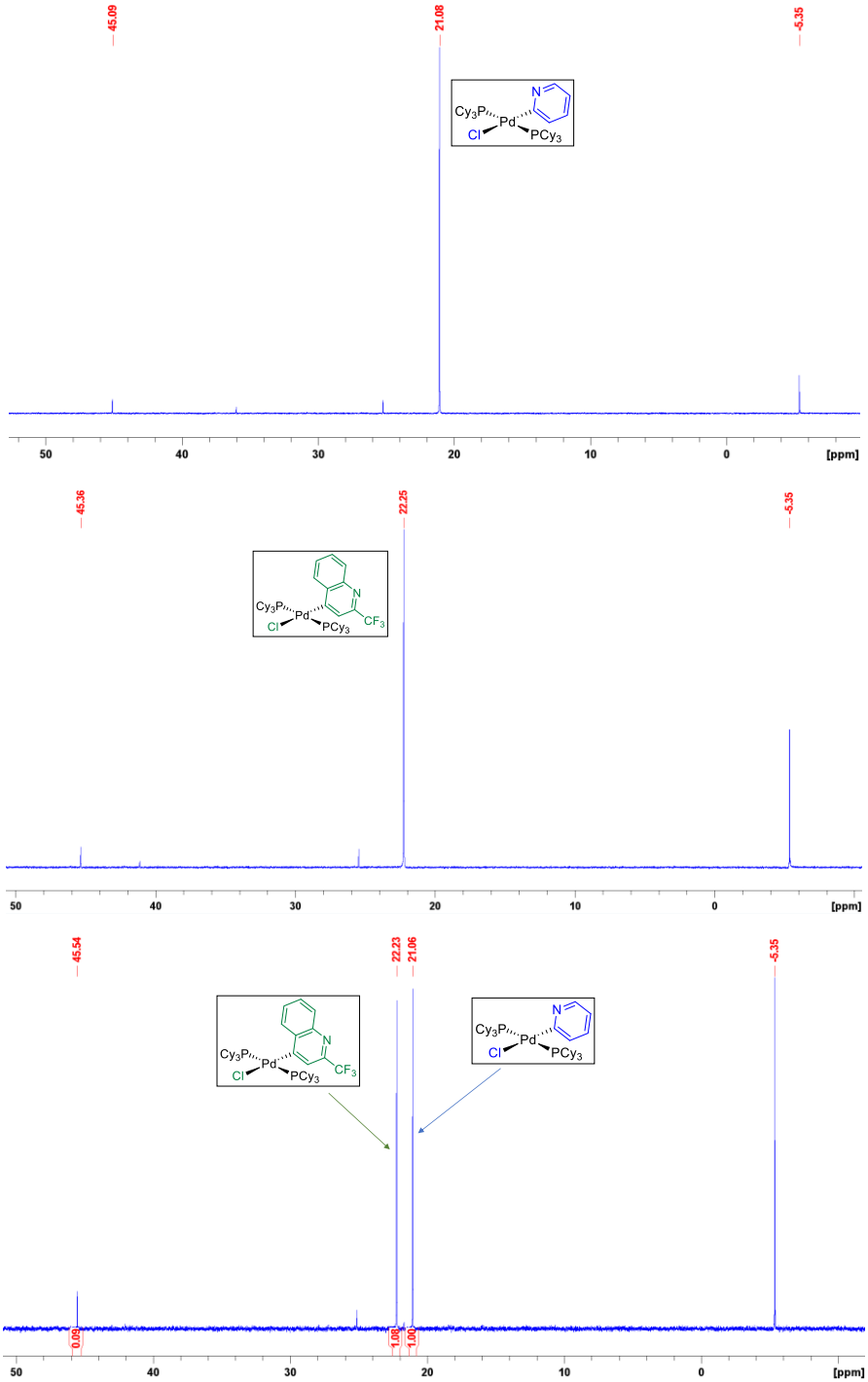


Fig. B3. Top: $^{31}P\{^{1}H\}$ NMR spectrum of the 2-chloro-pyridine and $Pd(PCy_3)_2$ oxidative addition product; Middle: $^{31}P\{^{1}H\}$ NMR spectrum of the 4-chloro-2-trifluoromethyl-quinoline and

$Pd(PCy_3)_2$ oxidative addition product; Bottom: $^{31}P\{^1H\}$ qNMR spectrum of the competition reaction of the two above substrates with $Pd(PCy_3)_2$.

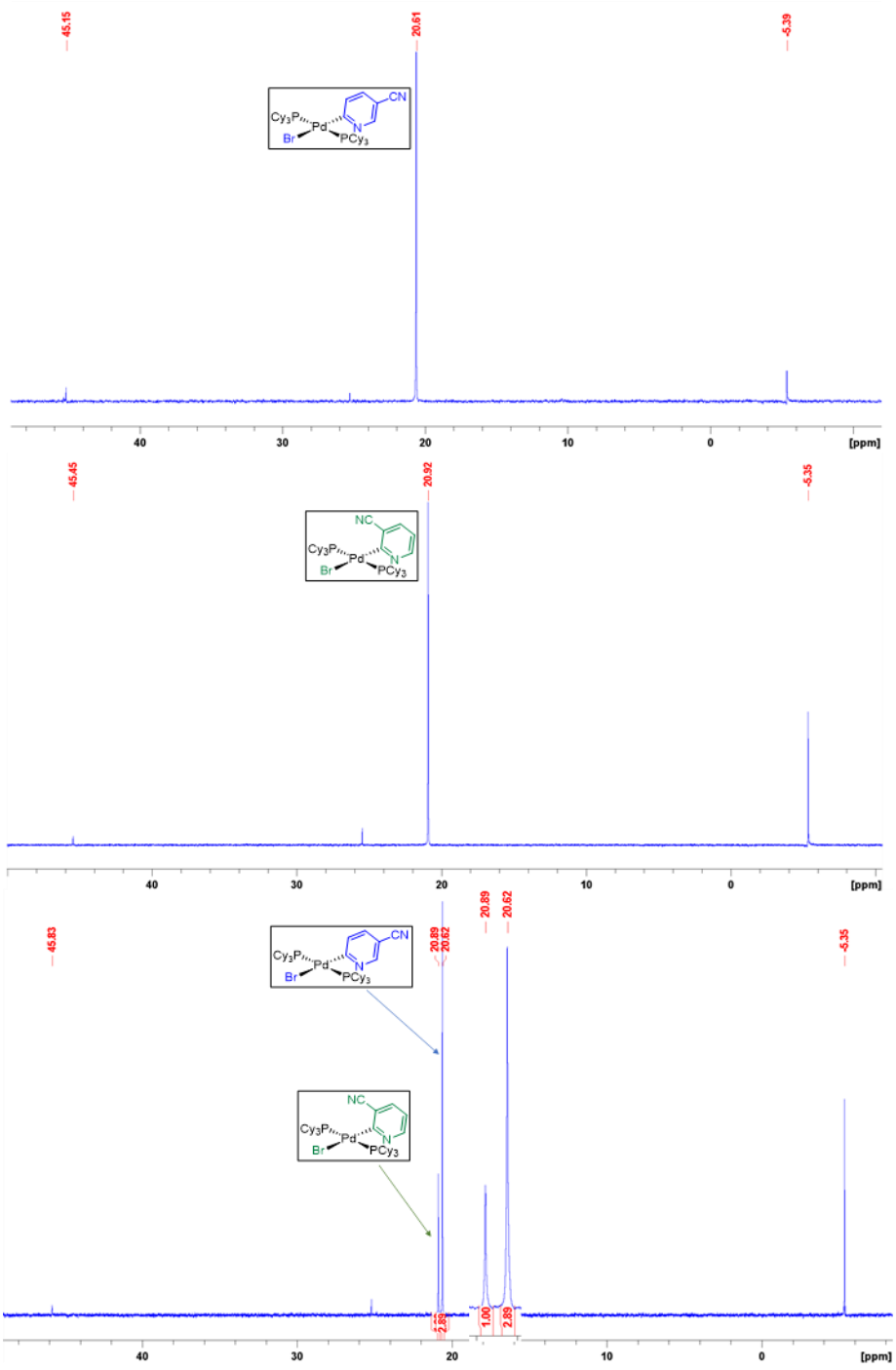


Fig. B4. Top: $^{31}P\{^1H\}$ NMR spectrum of the 2-bromo-5-cyano-pyridine and $Pd(PCy_3)_2$ oxidative addition product; middle: $^{31}P\{^1H\}$ NMR spectrum of the 2-bromo-3- cyano-pyridine and $Pd(PCy_3)_2$ oxidative addition product; Bottom: $^{31}P\{^1H\}$ qNMR spectrum of the competition reaction of the two above substrates with $Pd(PCy_3)_2$; product peak region expanded in inset.

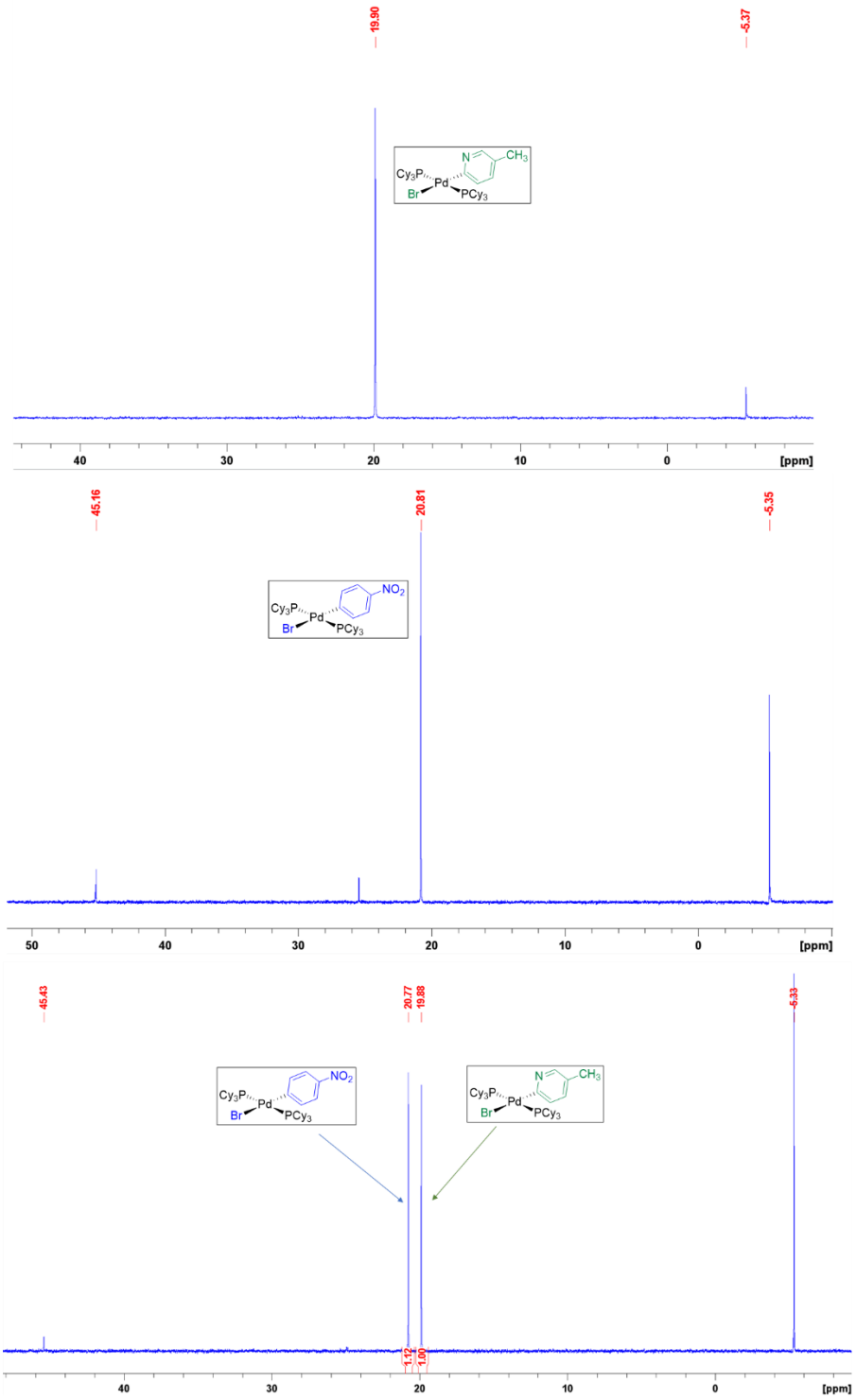


Fig. B5. Top: $^{31}P\{^{1}H\}$ NMR spectrum of the 2-bromo-5-methyl-pyridine and $Pd(PCy_3)_2$ oxidative addition product; middle: $^{31}P\{^{1}H\}$ NMR spectrum of the 4-bromo-1-nitro-benzene and $Pd(PCy_3)_2$ oxidative addition product; Bottom: $^{31}P\{^{1}H\}$ qNMR spectrum of the competition reaction of the two above substrates with $Pd(PCy_3)_2$.

Calculating Relative $\Delta G^{\ddagger}_{OA}$ for Each Substrate

All of the competition reactions were conducted under *pseudo* first-order conditions, with the two substrates being present in large excess ([substrate]:[Pd source] =20:1). The ratio of the reaction rates can be expressed as Eq (B1):

$$\frac{r_1}{r_2} = \frac{k_1[Pd][Substrate_1]}{k_2[Pd][Substrate_2]} = \frac{k_1[Substrate_1]}{k_2[Substrate_2]} \qquad Eq(B1)$$

The concentrations of the two substrates are assumed to stay constant throughout the reaction; then, Eq (S1) can be simplified to Eq (S2):

$$\frac{r_1}{r_2} = \frac{k_1}{k_2} = \frac{\frac{d[Product_1]}{dt}}{\frac{d[Product_2]}{dt}} = \frac{[Product_1]}{[Product_2]} \qquad Eq(B2)$$

In the $^{31}P\{^1H\}$ qNMR spectroscopy analysis, $\frac{[Product_1]}{[Product_2]}$ is equal to the peak area ratio of the two oxidative addition products from the ^{31}P qNMR spectrum. By substituting the Eyring equation (Eq (B3)):

$$k = \frac{k_B T}{h} \exp\left(-\frac{\Delta G^{\ddagger}}{RT}\right) \qquad Eq(B3)$$

into Eq (B2), the relative activation energy ($\Delta\Delta G^{\ddagger}_{OA}$) of the two oxidative addition reactions from the competition can be calculated by Eq (B4):

$$\frac{k_1}{k_2} = \exp\left(\frac{\Delta G^{\ddagger}{}_2 - \Delta G^{\ddagger}{}_1}{RT}\right) \Rightarrow \Delta\Delta G^{\ddagger}_{OA} = \ln\left(\frac{k_1}{k_2}\right) RT \Rightarrow$$

$$\Delta\Delta G^{\ddagger}_{OA} = \ln\left(\frac{Peak\ area_{Product1}}{Peak\ area_{Product2}}\right) RT \qquad Eq(B4)$$

$$where\ \Delta\Delta G^{\ddagger}_{OA} = \Delta G^{\ddagger}_2 - \Delta G^{\ddagger}_1$$

The experimental $\Delta G^{\ddagger}_{OA}$ for the 79 substrates determined from 98 competition reactions in THF is summarized in Table B2. The most reactive substrate – 2-bromo-5-nitropyridine – is used as the zero point for this scale ($\Delta G^{\ddagger}_{OA}$ set to 0 kJ mol^{-1}), and all other $\Delta G^{\ddagger}_{OA}$ values are given relative to this substrate. Substrates highlighted are those with $\Delta G^{\ddagger}_{OA}$ determined by multiple competition experiments with different substrate pairings. The given $\Delta G^{\ddagger}_{OA}$ values for these substrates are averages of those determined by at least two different competition experiments, and the relative standard deviation (RSD) for these examples is also given.

Table B2 – Observed $\Delta G^{\ddagger}_{OA}$ for 79 substrates determined from competition reactions in THF. The $\Delta G^{\ddagger}_{OA}$ for 2-bromo-5-nitropyridine (first entry) is set to 0.00 kJ mol^{-1} (Lu-PhD-thesis_Ch.3 SI tables.xlsx).

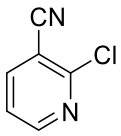
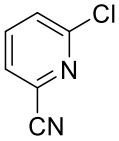
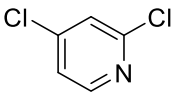
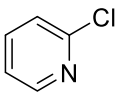
Assessing Possible Reversibility of Oxidative Addition

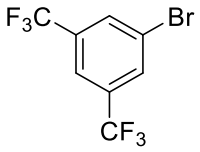
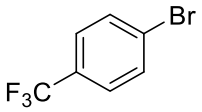
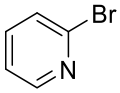
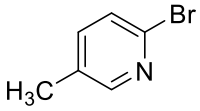
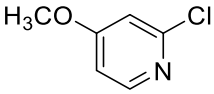
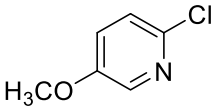
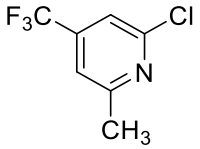
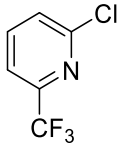
Our experimental approach of determining the rate constant ratio from the oxidative addition product ratio is only valid if the oxidative addition reactions are under kinetic control at room temperature; therefore, we conducted a series of experiments to confirm that kinetically controlled product mixtures are formed, with no reversible oxidative addition. We selected 6 pairs of substrates to test for reversibility, and the experimental details and test results are summarized in Table B3.

For each pair of substrates, we first conducted individual oxidative addition reactions for the two substrates separately using the general procedure outlined previously. The reaction solutions were analyzed by $^{31}P\{^{1}H\}$ NMR to confirm that $Pd(PCy_3)_2$ reacted to completion before the next step. Once confirmed complete, the individual reaction solutions were mixed in a 1:1 v/v ratio, resulting in a 1:1 molar ratio of Pd(II) products. Half of the resulting mixture was immediately analyzed by $^{31}P\{^{1}H\}$ qNMR spectroscopy to confirm this 1:1 ratio of the two Pd oxidative addition complexes, while the other half of the mixture was left stirring overnight. After 24 hours, a $^{31}P\{^{1}H\}$ qNMR spectrum of the stirred reaction mixture was recorded, revealing no change to the product ratio. If the product ratio in our competition experiments was thermodynamically controlled through reversible oxidative addition, the 1:1 product mixture would have changed to reflect that ratio. Comparing the results from these 6 reversibility tests to the results from the corresponding competition reactions reveals that the competition product ratio is kinetically controlled.

$^{31}P\{^{1}H\}$ qNMR spectra from one representative reversibility test as well as the $^{31}P\{^{1}H\}$ qNMR spectrum of the corresponding competition reaction are shown in Figure B6.

Table B3. Test reactions for possible reversibility in oxidative addition competition experiments

Exp No.	Substrate A	Substrate B	Initial Product ratio	Product ratio after 24 h	Competition reaction product ratio
1	 2-chloro-3-cyanopyridine	 2-chloro-6-cyanopyridine	1.03	1.03	6.16
2	 2,4-dichloropyridine	 2-chloropyridine	0.97	0.97	22.17

3	 1-bromo-3,5-bis(trifluoromethyl) benzene	 1-bromo-4-trifluoromethyl benzene	1.04	1.04	32.25
4	 2-bromopyridine	 2-bromo-5-methylpyridine	1.00	1.00	3.02
5	 2-chloro-4-methoxypyridine	 2-chloro-5-methoxypyridine	0.87	0.87	4.25
6	 2-chloro-4-trifluoromethyl -6-methylpyridine	 2-chloro-6-trifluoromethyl pyridine	1.04	1.04	5.02

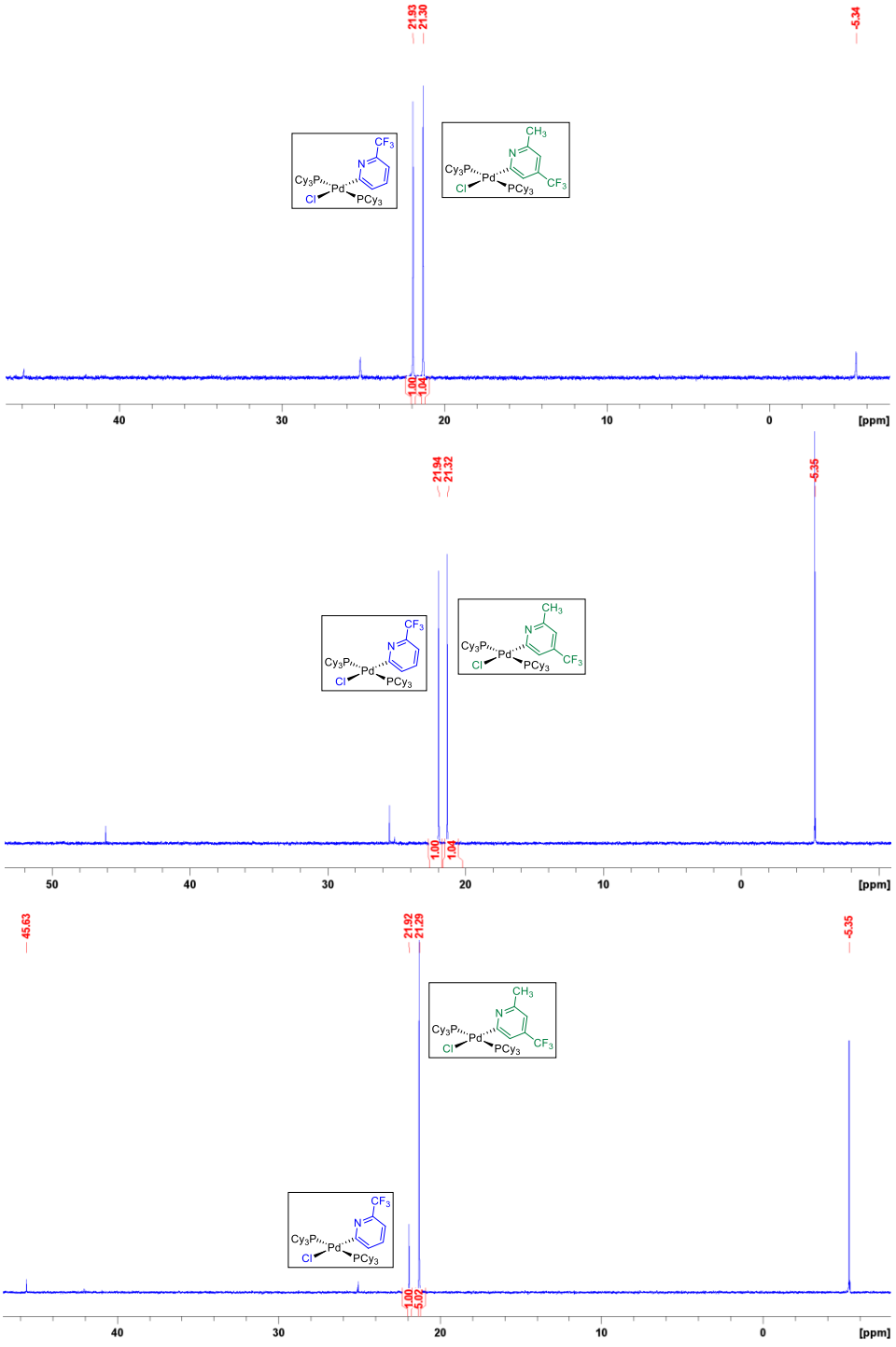


Fig. B6. $^{31}P\{^{1}H\}$ NMR spectra for the reversibility test of 2-chloro-4-trifluoromethyl-6-methyl-pyridine and 2-chloro-6-trifluoromethyl-pyridine oxidative addition products. Top: $^{31}P\{^{1}H\}$ NMR spectrum of the 1:1 product mixture right after mixing. Middle: $^{31}P\{^{1}H\}$ NMR spectrum of the 1:1 product mixture after 24 hours. Bottom: $^{31}P\{^{1}H\}$ NMR spectrum of the competition reaction outcome.

Preparative Scale Synthesis of Oxidative Addition Complexes

Six representative Pd(II) oxidative addition complexes derived from 2-halopyridine substrates were isolated, purified, and characterized using the following general procedure.

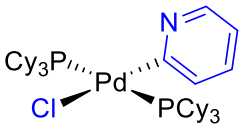
All stock solutions were prepared in the following concentrations in THF: substrate (0.375 M), and $Pd(PCy_3)_2$ (0.0375 M).

In the glovebox, a 20 mL vial containing a stir bar was charged with 300 µL of the substrate stock solution (0.1125 mmol, 1.5 equiv.), 2 mL of the $Pd(PCy_3)_2$ stock solution (0.075 mmol, 1.0 equiv.), and 5 mL of THF. The reaction mixture was stirred at room temperature overnight. A $^{31}P\{^1H\}$ NMR spectrum was recorded to confirm the $Pd(PCy_3)_2$ had reacted to completion. The reaction mixture was then transferred outside glovebox for further purification.

The solvent was evaporated under vacuum using a rotary evaporator. The solid left in the vial was washed with either pentane or diethyl ether (2 x 7 mL) to remove unreacted substrate. The solid was dried under vacuum to give the oxidative addition product, and the isolated yield was recorded.

Pentane was used to wash the products of 2-chloropyridine, 2-chloro-6-trifluoropyridine and 2-bromo-5-chloropyridine with $Pd(PCy_3)_2$. Diethyl ether was used to wash the products of 2-chloro-3-cyanopyridine, 2-chloro-4-nitropyridine and 2-chloro-3-aminopyridine with $Pd(PCy_3)_2$.

Characterization of Isolated Oxidative Addition Complexes



P1

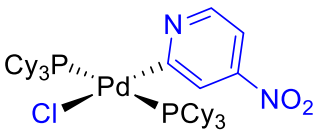
The oxidative addition product **P1** was prepared by the general procedure using 2-chloropyridine. 43.2 mg of a white solid was obtained (73.8% yield)

^1H NMR: (500 MHz, $CDCl_3$): δ 1.01-1.98 (m, 66H, 66 x Cy-H), 6.70 (t, 1H, 1 x Py-H), 7.06 (t, 1H, 1 x Py-H), 7.32 (d, 1H, 1 x Py-H), 8.36 (dd, 1H, 1 x Py-H).

$^{13}C\{^1H\}$ NMR: (126 MHz, $CDCl_3$): δ 25.61-33.71 (36 x Cy), 116.9 (1 x Py), 131.90 (1 x Py), 134.90 (1 x Py), 147.96 (1 x Py), 178.89 (1 x Py).

$^{31}P\{^1H\}$ NMR: (203 MHz, $CDCl_3$): δ 20.26.

HRMS (ESI): $[C_{41}H_{70}P_2PdN]^+$ (target compound minus the chlorine): 744.40128 (calc'd), 744.40204 (found); $[C_{41}H_{70}P_2PdNCl{\cdot}H]^+$ (hydrogen adduct): 780.37796 (calc'd), 780.37888 (found).



P2

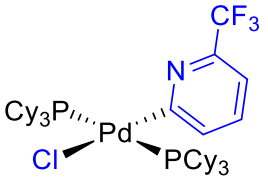
The oxidative addition product **P2** was prepared by the general procedure using 2-chloro-4-nitropyridine. 19.5 mg of a light-yellow solid was obtained (31.5% yield) as the isolated product.

^1H NMR: (300 MHz, $CDCl_3$): δ 0.92-1.82 (m, 66H, 66 x Cy-H), 7.31 (d, 1H, 1 x Py-H), 7.97 (s, 1H, 1 x Py-H), 8.63 (d, 1H, 1 x Py-H).

$^{13}C\{^{1}H\}$ NMR: (126 MHz, $CDCl_3$): δ 26.49-33.84 (36 x Cy), 108.19 (1 x Py), 124.47 (1 x Py), 149.34 (1 x Py), 149.73 (1 x Py), 186.34 (1 x Py).

$^{31}P\{^{1}H\}$ NMR: (203 MHz, $CDCl_3$): δ 21.56.

HRMS (ESI): $[C_{41}H_{69}P_2PdN_2O_2]^+$ (target compound minus the chlorine): 789.38636 (calc'd), 789.38722 (found); $[C_{41}H_{69}P_2PdN_2O_2Cl{\cdot}H]^+$ (hydrogen adduct): 825.36304 (calc'd), 825.36395 (found).



P3

The oxidative addition product was prepared by the general procedure using 2-chloro-6-trifluoromethylpyridine. 29.7 mg of a white solid was obtained (57.3%) as the isolated product.

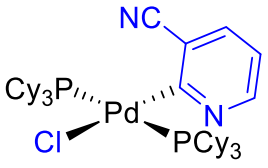
^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.02-1.96 (m, 66H, 66 x Cy-H), 7.10 (d, 1H, 1 x Py-H), 7.16 (t, 1H, 1 x Py-H), 7,57 (d, 1H, 1 x Py-H).

$^{13}C\{^{1}H\}$ NMR: (126 MHz, $CDCl_3$): δ 22.35-33.60 (36 x Cy), 131.46 (1 x Py), 182.33 (1 x Py).

$^{31}P\{^{1}H\}$ NMR: (203 MHz, $CDCl_3$): δ 21.05.

$^{19}F\{^{1}H\}$ NMR: (471 MHz, $CDCl_3$): δ 21.05.

HRMS (ESI): $[C_{42}H_{69}P_2PdNF_3]^+$ (target compound minus the chlorine): 812.38867 (calc'd), 812.38996 (found); $[C_{42}H_{69}P_2PdNF_3Cl{\cdot}H]^+$ (hydrogen adduct): 848.36535 (calc'd), 848.36632 (found).



P4

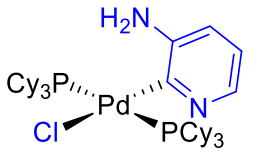
The oxidative addition product **P4** was prepared by the general procedure using 2-chloro-3-cyanopyridine. 42.1 mg of a white solid was obtained (69.7%) as the isolated product.

^{1}H NMR: (500 MHz, $CDCl_3$): δ 0.99-2.06 (m, 66H, 66 x Cy-H), 6.83 (dd, 1H, 1 x Py-H), 7.47 (d, 1H, 1 x Py-H), 8.60 (dd, 1H, 1 x Py-H).

$^{13}C\{^{1}H\}$ NMR: (126 MHz, $CDCl_3$): δ 26.51-34.17 (36 x Cy), 116.32 (1 x CN-Py), 120.14 (1 x Py), 121.11 (1 x Py), 137.36 (1 x Py), 150.52 (1 x Py), 186.91 (1 x Py).

$^{31}P\{^{1}H\}$ NMR: (203 MHz, $CDCl_3$): δ 21.76.

HRMS (ESI): $[C_{42}H_{69}P_2PdN_2]^+$ (major isotopomer, target compound minus the chlorine): 769.39653 (calc'd), 769.39698 (found); $[C_{42}H_{69}P_2PdN_2Cl{\cdot}H]^+$ (hydrogen adduct): 805.37321 (calc'd), 805.37288 (found); $[C_{42}H_{69}P_2PdN_2Cl{\cdot}Na]^+$ (sodium adduct): 827.35516 (calc'd), 827.35502 (found).



P5

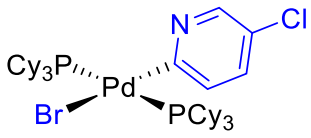
The oxidative addition product **P5** was prepared by the general procedure using 2-chloro-3-aminopyridine. 31.6 mg of a light brown solid was obtained (52.9%) as the isolated product.

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.04-2.03 (m, 66H, 66 x Cy-H), 4.26 (s, 2H, N*H2*-Py), 6.39 (d, 1H, 1 x Py-H), 6.60 (dd, 1H, 1 x Py-H), 7.96 (d, 1H, 1 x Py-H).

^{13}C{1H} NMR: (126 MHz, $CDCl_3$): δ 27.73-34.15 (36 x Cy), 115.80 (1 x Py), 117.53 (1 x Py), 138.91 (1 x Py), 145.20 (1 x Py), 164.21 (1 x Py).

^{31}P{^{1}H} NMR: (203 MHz, $CDCl_3$): δ 23.20.

HRMS (ESI): $[C_{41}H_{71}P_2PdN_2Cl{\cdot}H]^+$ (major isotopomer, hydrogen adduct): 795.38886 (calc'd), 795.38898 (found).



P6

The oxidative addition product **P6** was prepared by the general procedure from 2-bromo-5-chloropyridine. 35.3 mg of a light-yellow solid was obtained (54.8%) as the isolated product.

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.04-2.00 (m, 66H, 66 x Cy-H), 7.11 (dd, 1H, 1 x Py-H), 7.31 (d, 1H, 1 x Py-H), 8.42 (d, 1H, 1 x Py-H).

^{13}C{^{1}H} NMR: (126 MHz, $CDCl_3$): δ 26.37-35.15 (36 x Cy), 126.31 (1 x Py), 131.42 (1 x Py), 134.57 (1 x Py), 146.34 (1 x Py), 177.87 (1 x Py).

^{31}P{^{1}H} NMR: (203 MHz, $CDCl_3$): δ 19.89.

HRMS (ESI): $[C_{41}H_{69}P_2PdNCl]^+$ (major isotopomer, target compound minus the bromine): 778.36231 (calc'd), 778.36264 (found).

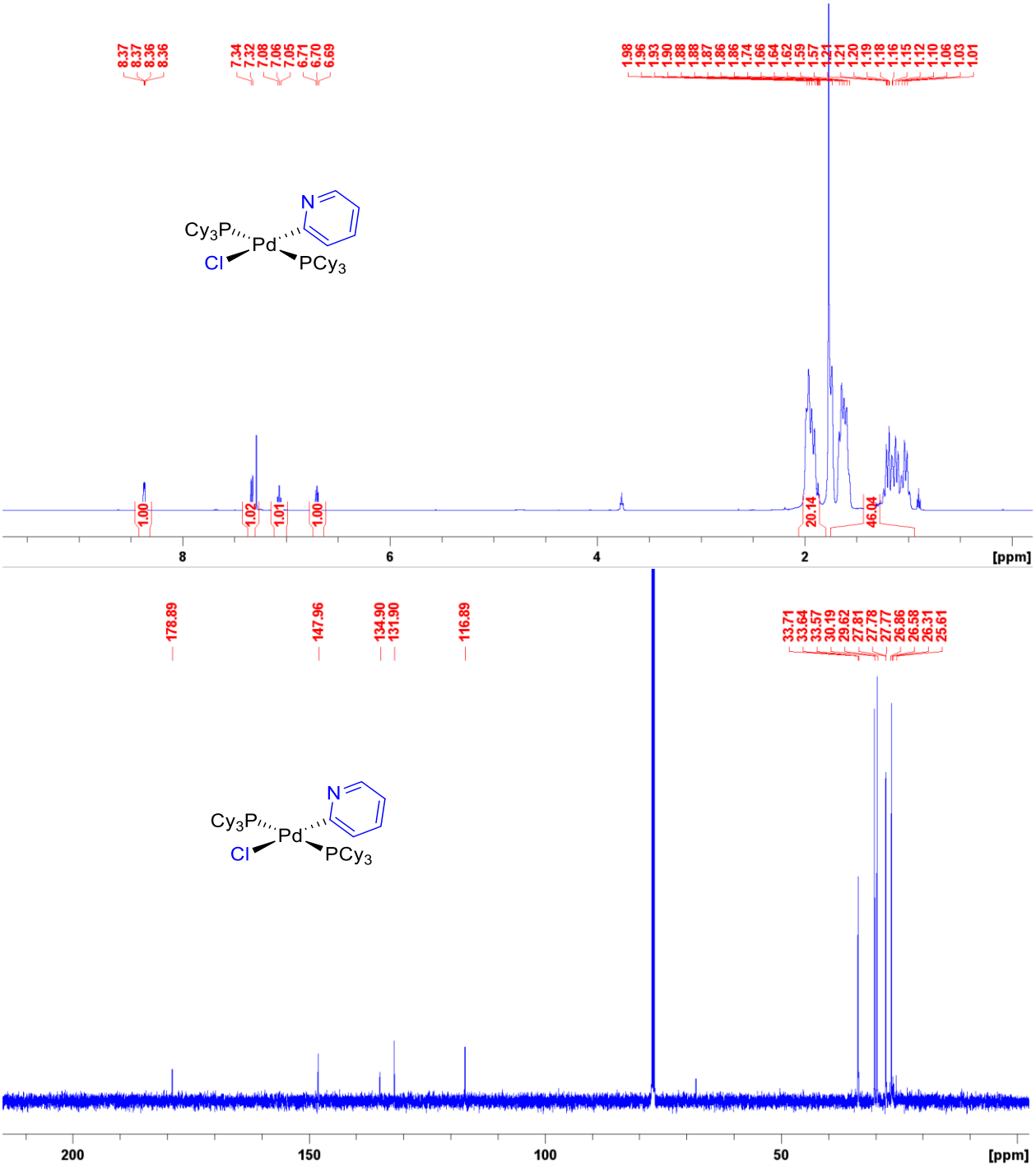


Fig. B7. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **P1**.

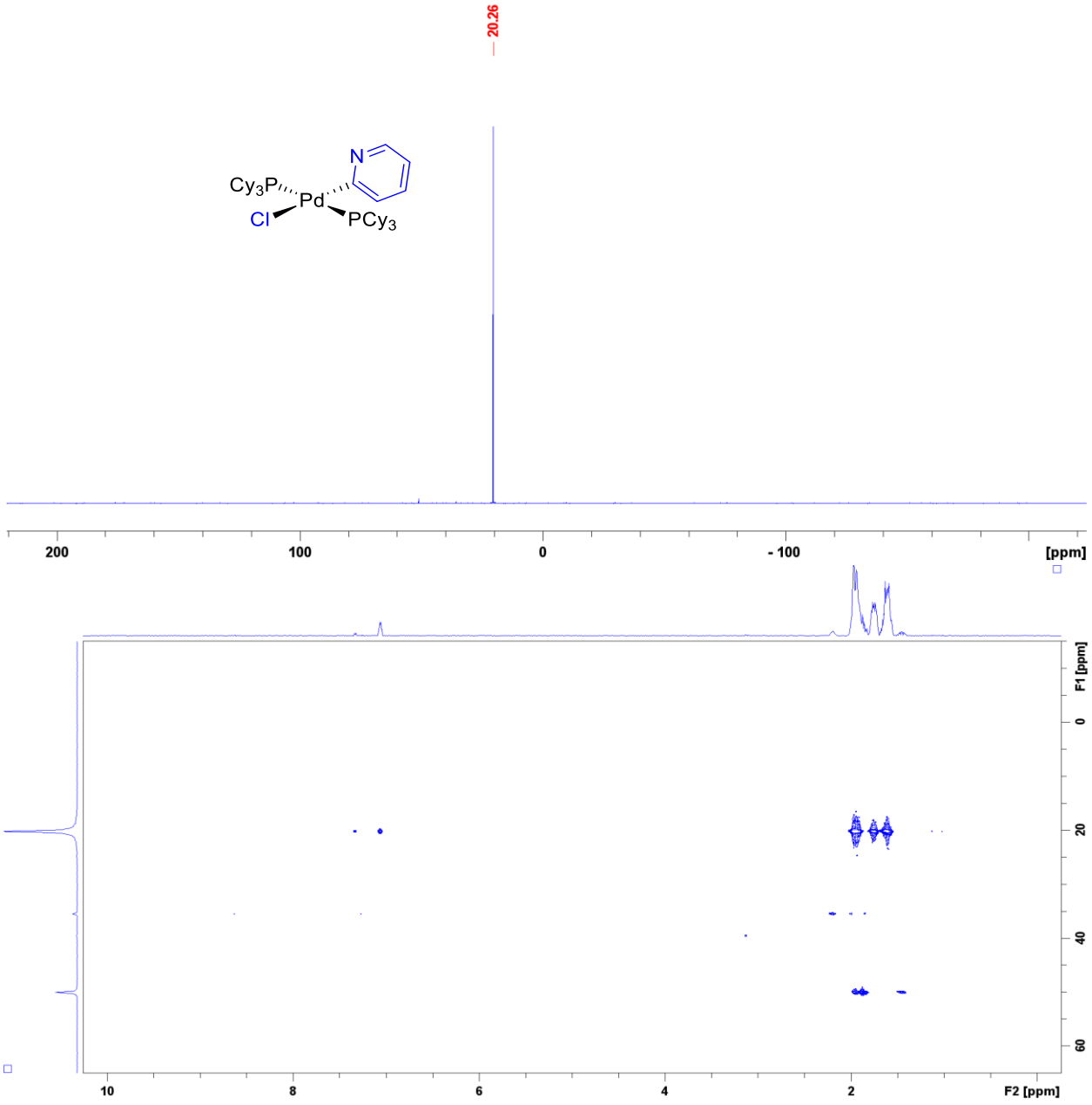


Fig. B8. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **P1**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

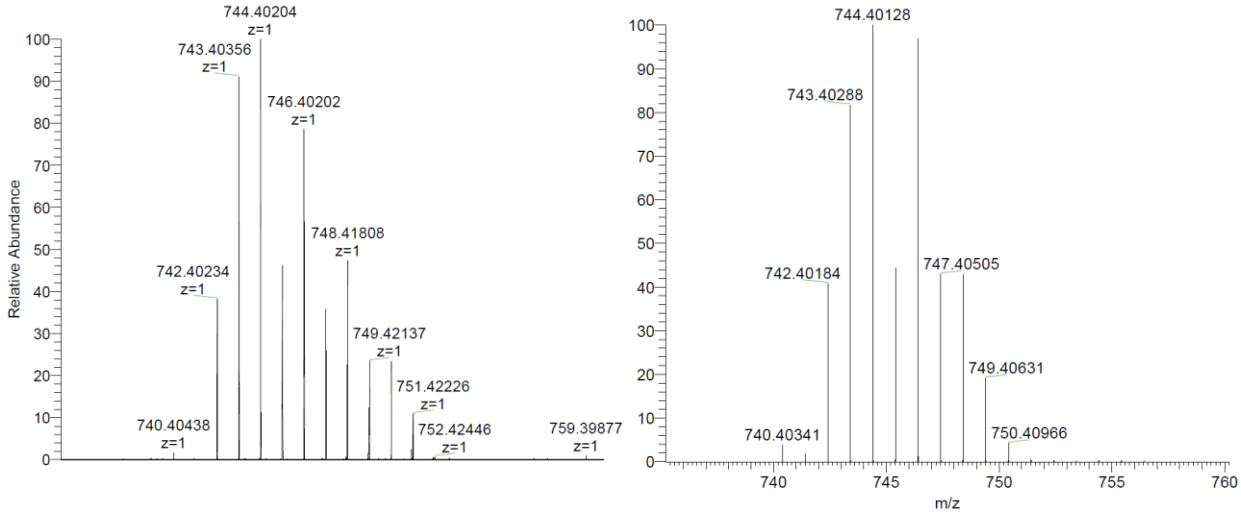


Fig. B9. Left: Experimental HRMS-ESI spectrum of [**P1**-Cl]$^+$. Right: Calculated HRMS isotope pattern for [**P1**-Cl] $^+$.

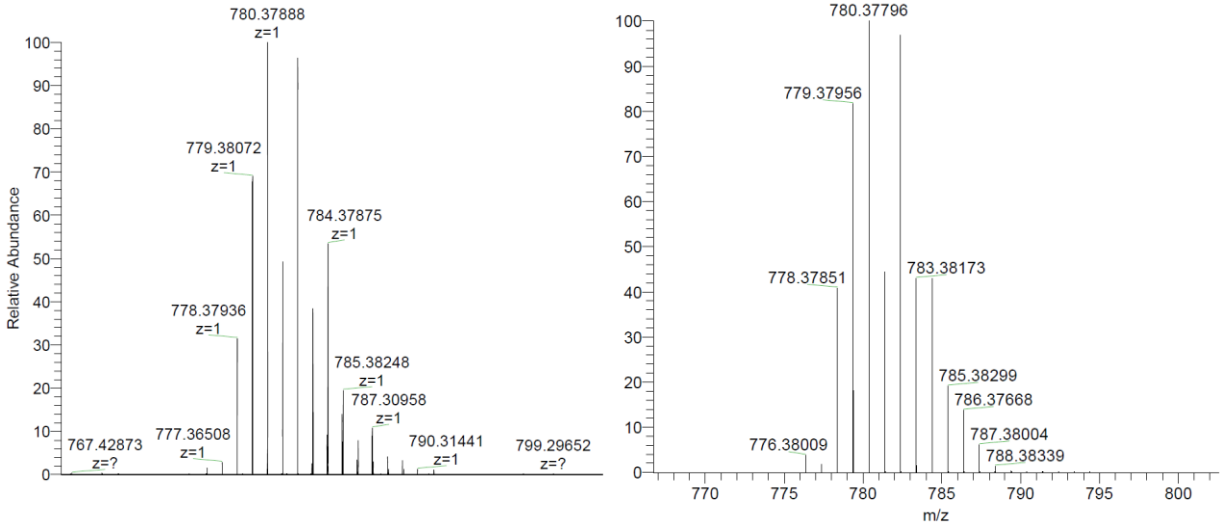


Fig. B10. Left: Experimental HRMS-ESI spectrum of [**P1**+H]$^+$. Right: Calculated HRMS isotope pattern for [**P1**+H]$^+$.

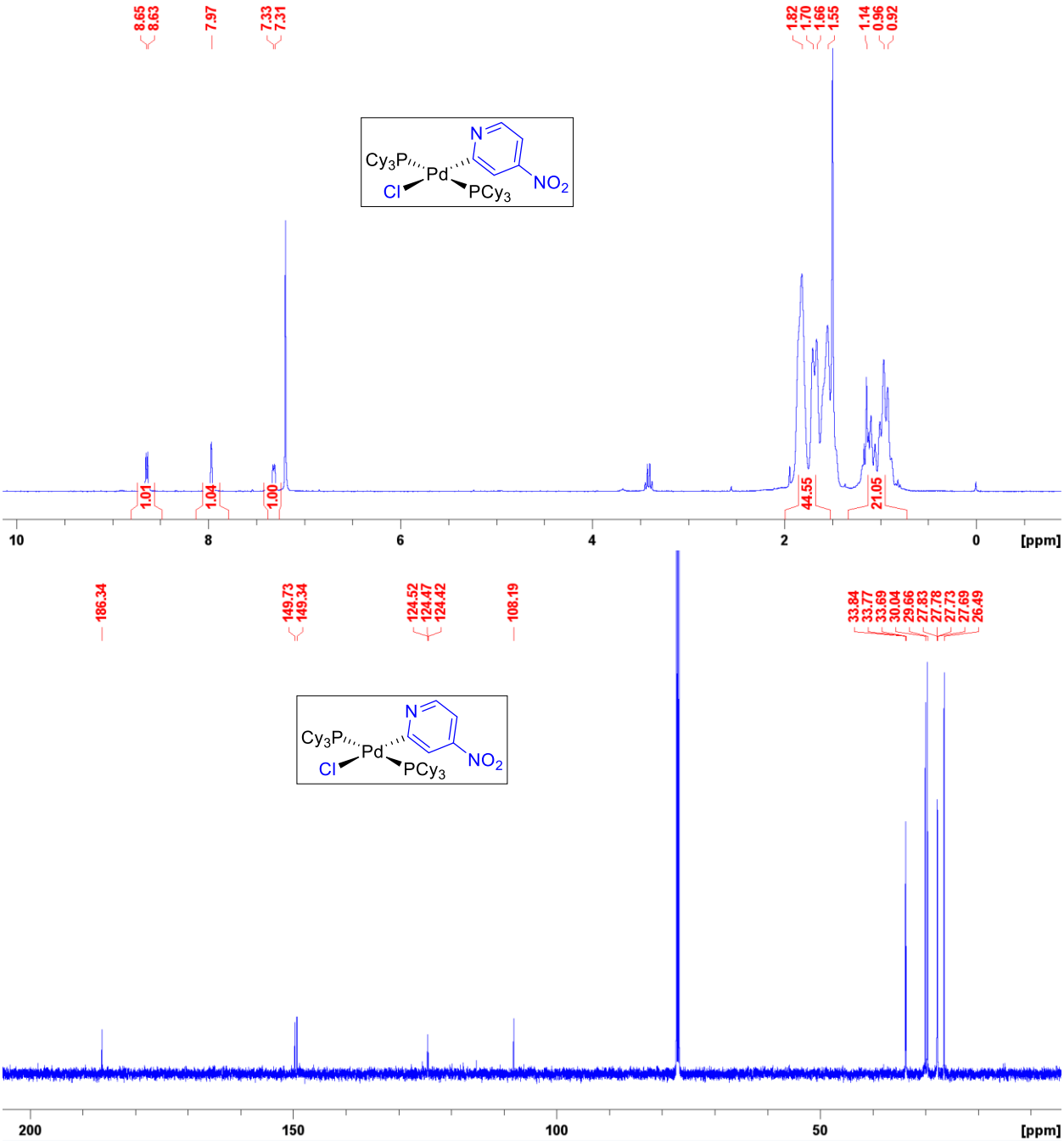


Fig. B11. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **P2**.

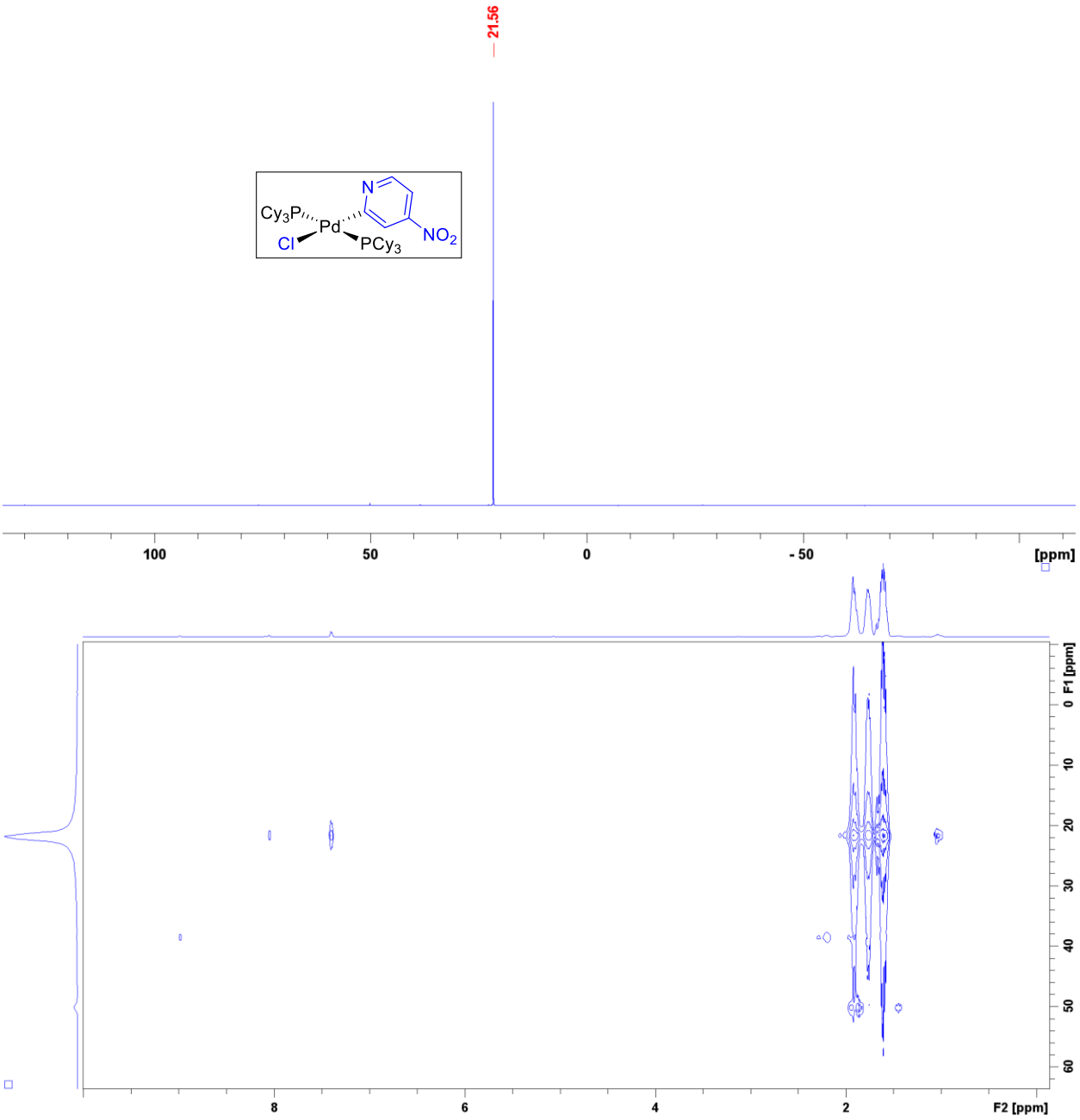


Fig. B12. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **P2**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

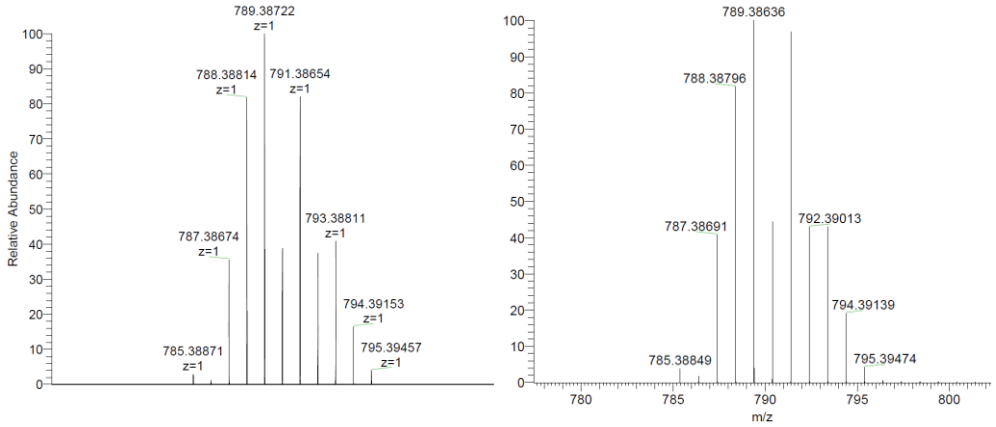


Fig. B13. Left: Experimental HRMS-ESI spectrum of [**P2**-Cl]$^+$. Right: Calculated HRMS isotope pattern for [**P2**-Cl]$^+$.

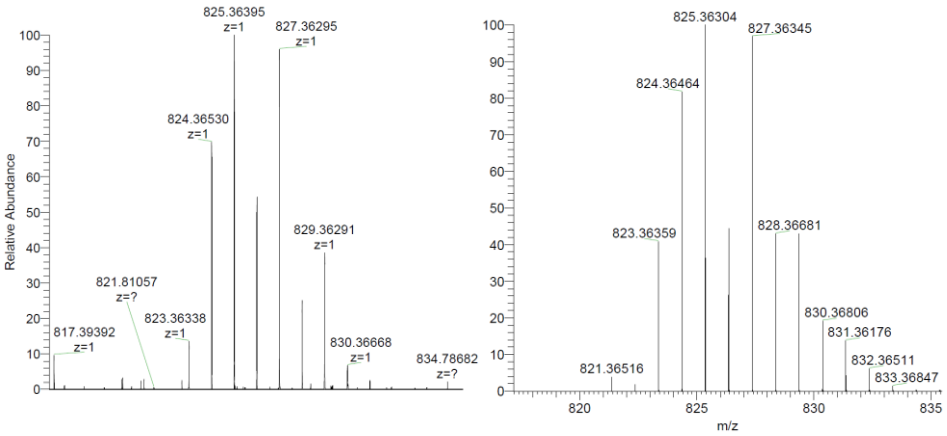


Fig. B14. Left: Experimental HRMS-ESI spectrum of [**P2**+H]$^+$. Right: Calculated HRMS isotope pattern for [**P2**+H]$^+$.

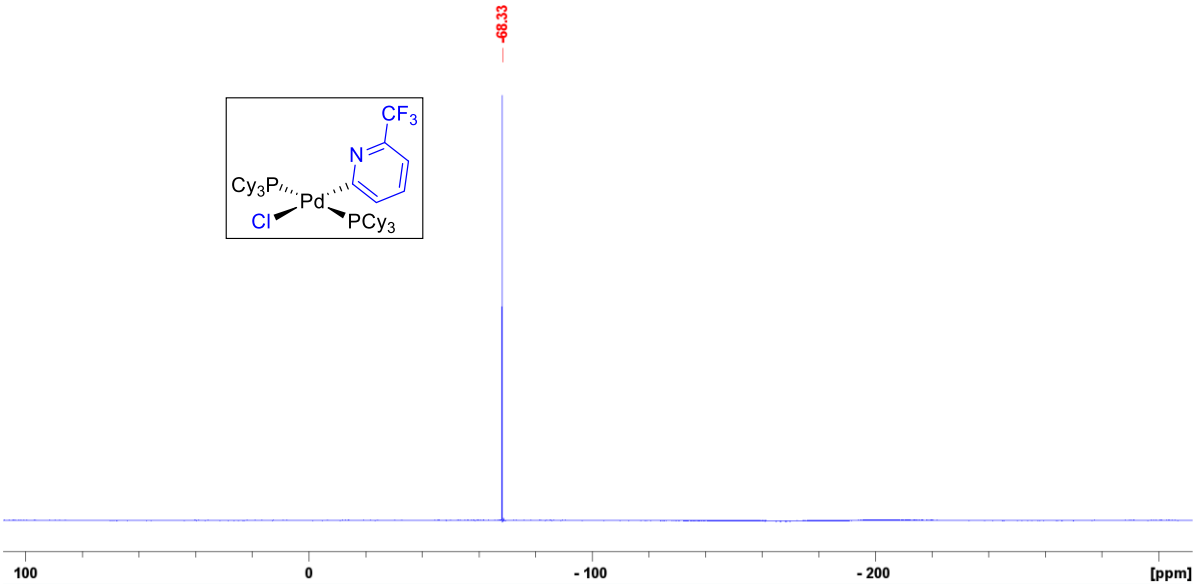


Fig. B15. $^{19}F\{^1H\}$ (471 MHz, $CDCl_3$) NMR spectrum of **P3**.

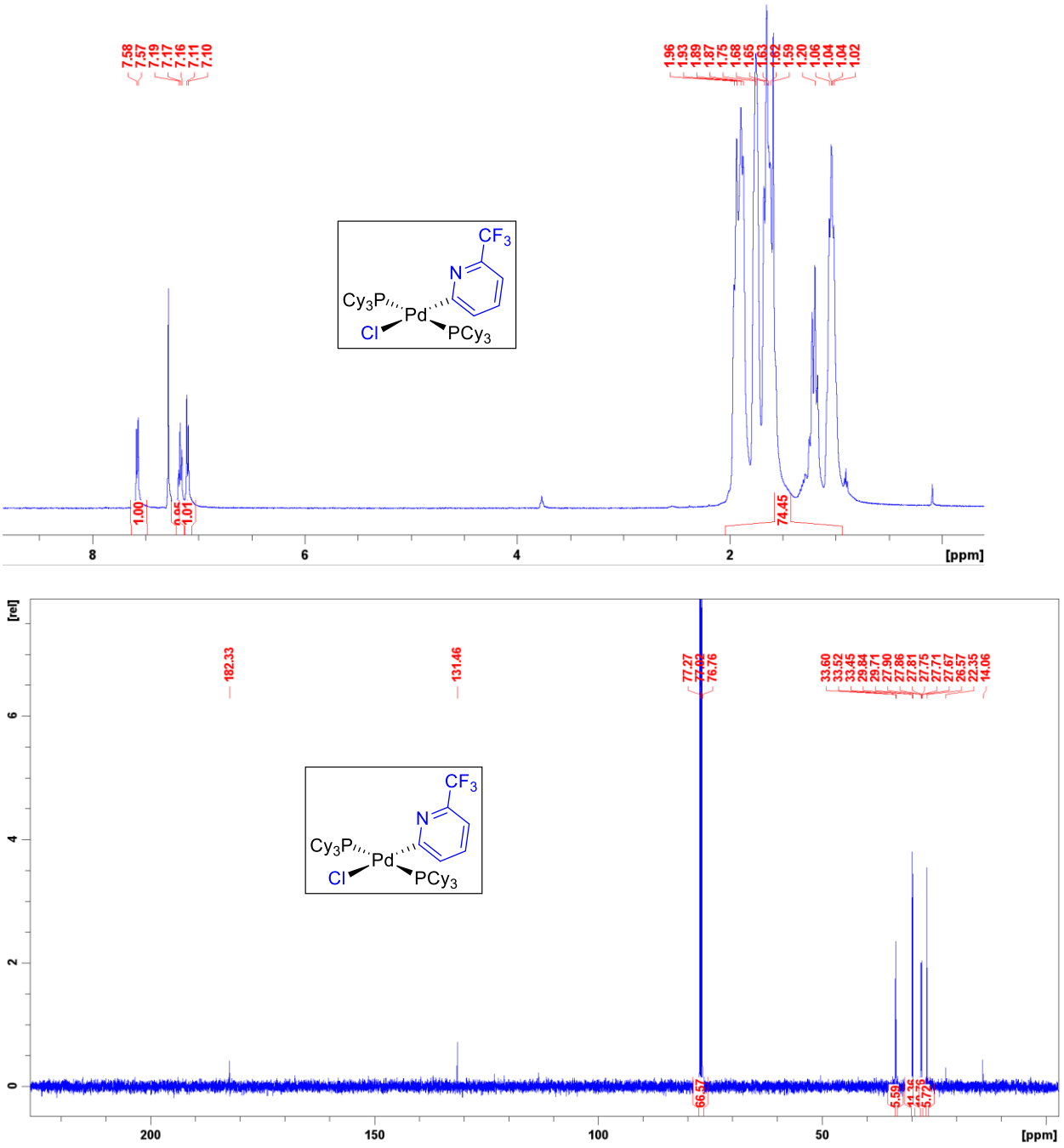


Fig. B16. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **P3**.

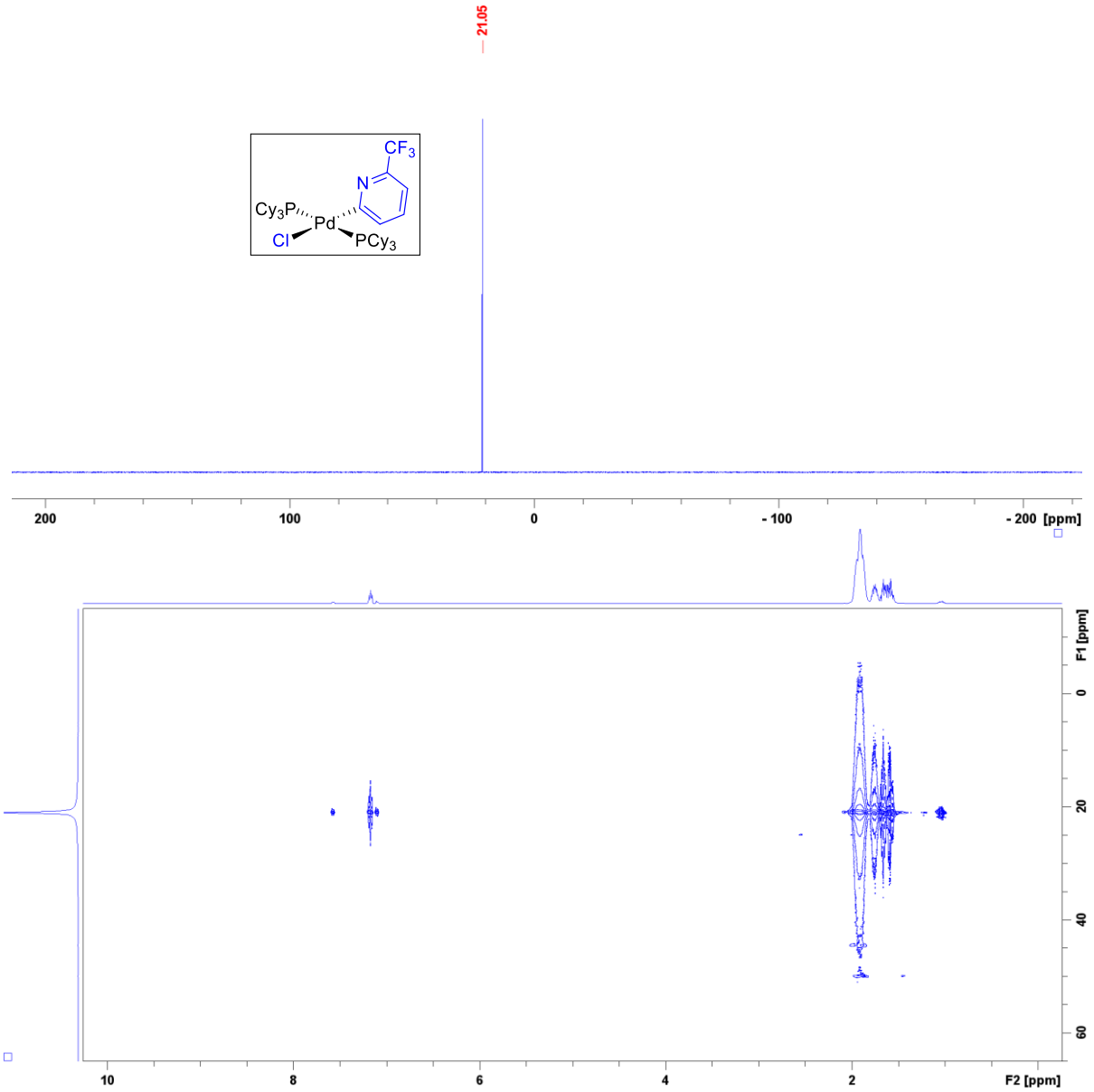


Fig. B17. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **P3**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

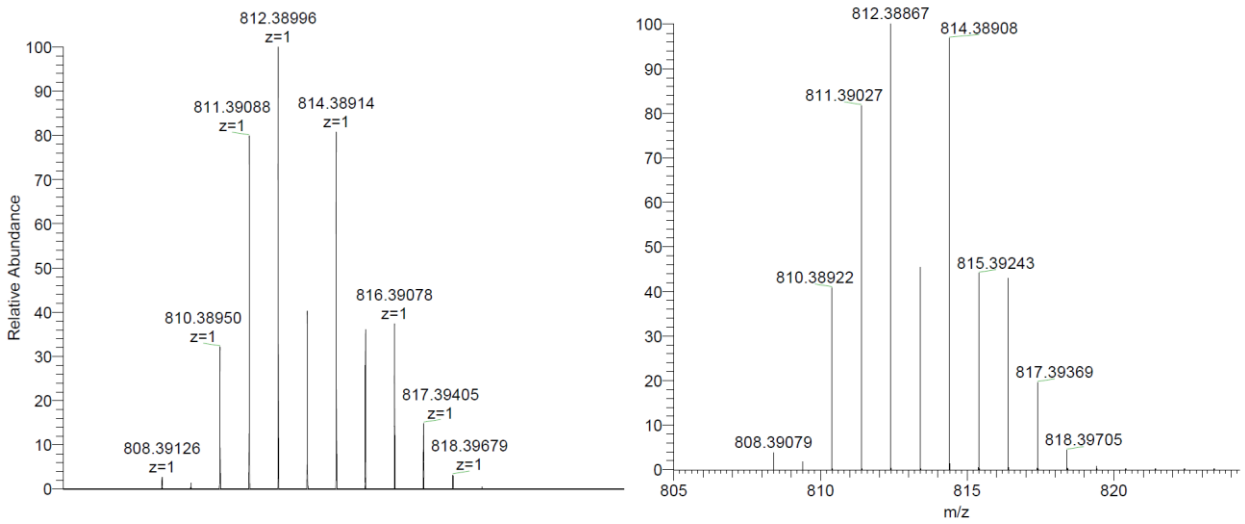


Fig. B18. Left: Experimental HRMS-ESI spectrum of [**P3**-Cl]$^+$. Right: Calculated HRMS isotope pattern for [**P3**-Cl]$^+$.

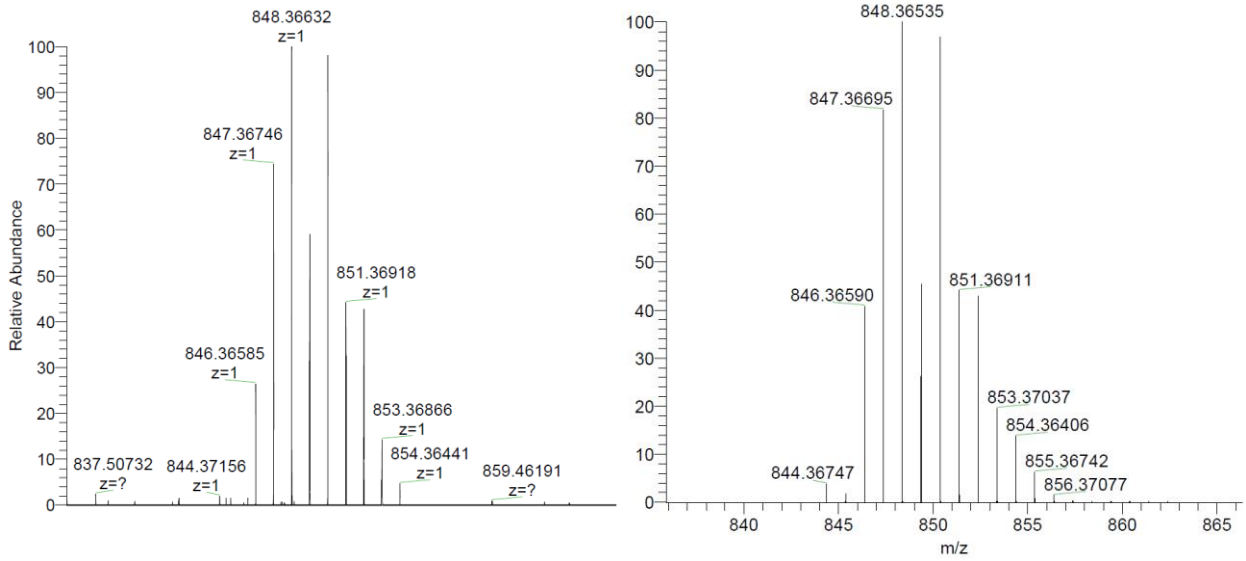


Fig. B19. Left: Experimental HRMS-ESI spectrum of [**P3**+H]$^+$. Right: Calculated HRMS isotope pattern for [**P3**+H]$^+$.

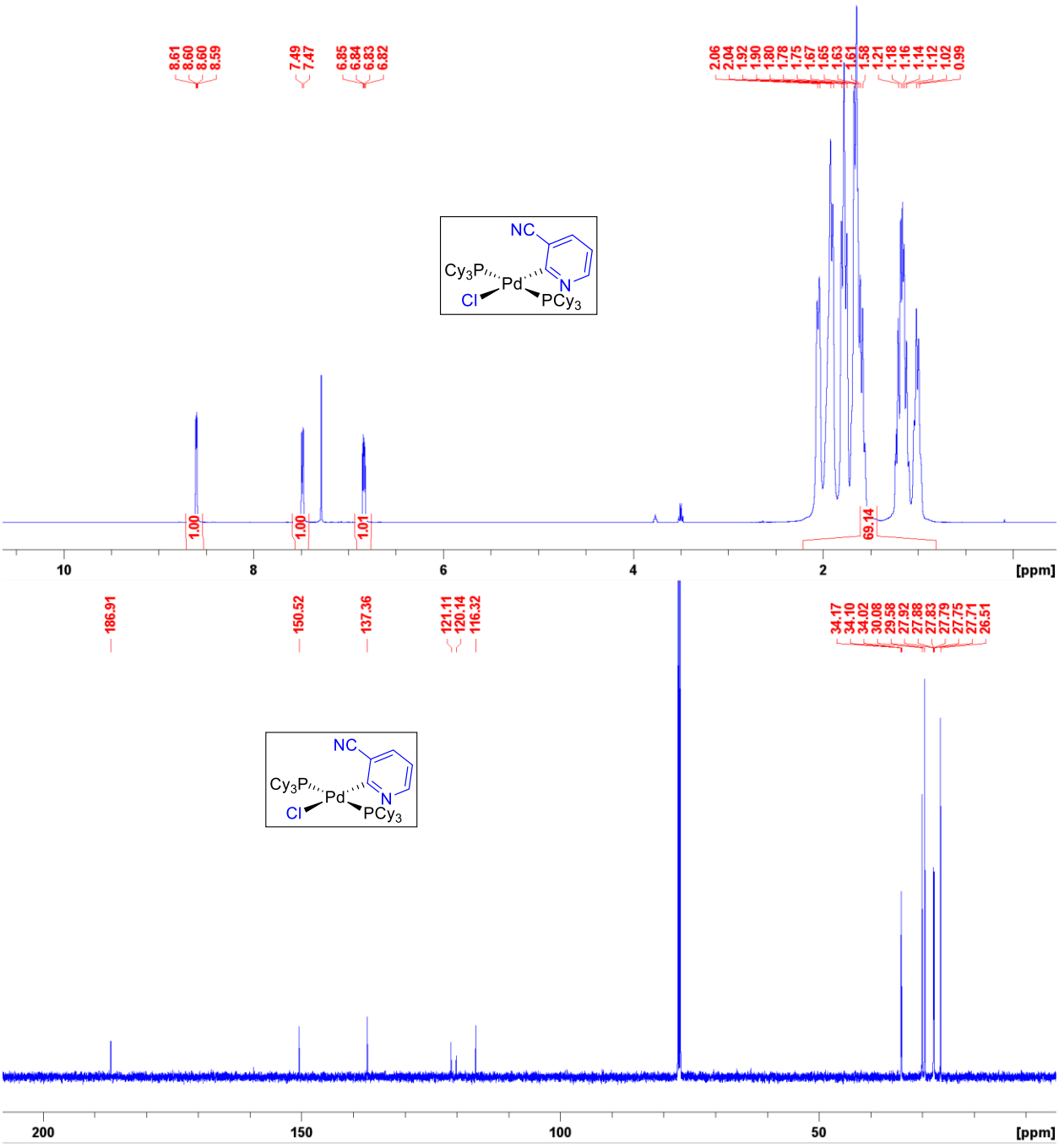


Fig. B20. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **P4**.

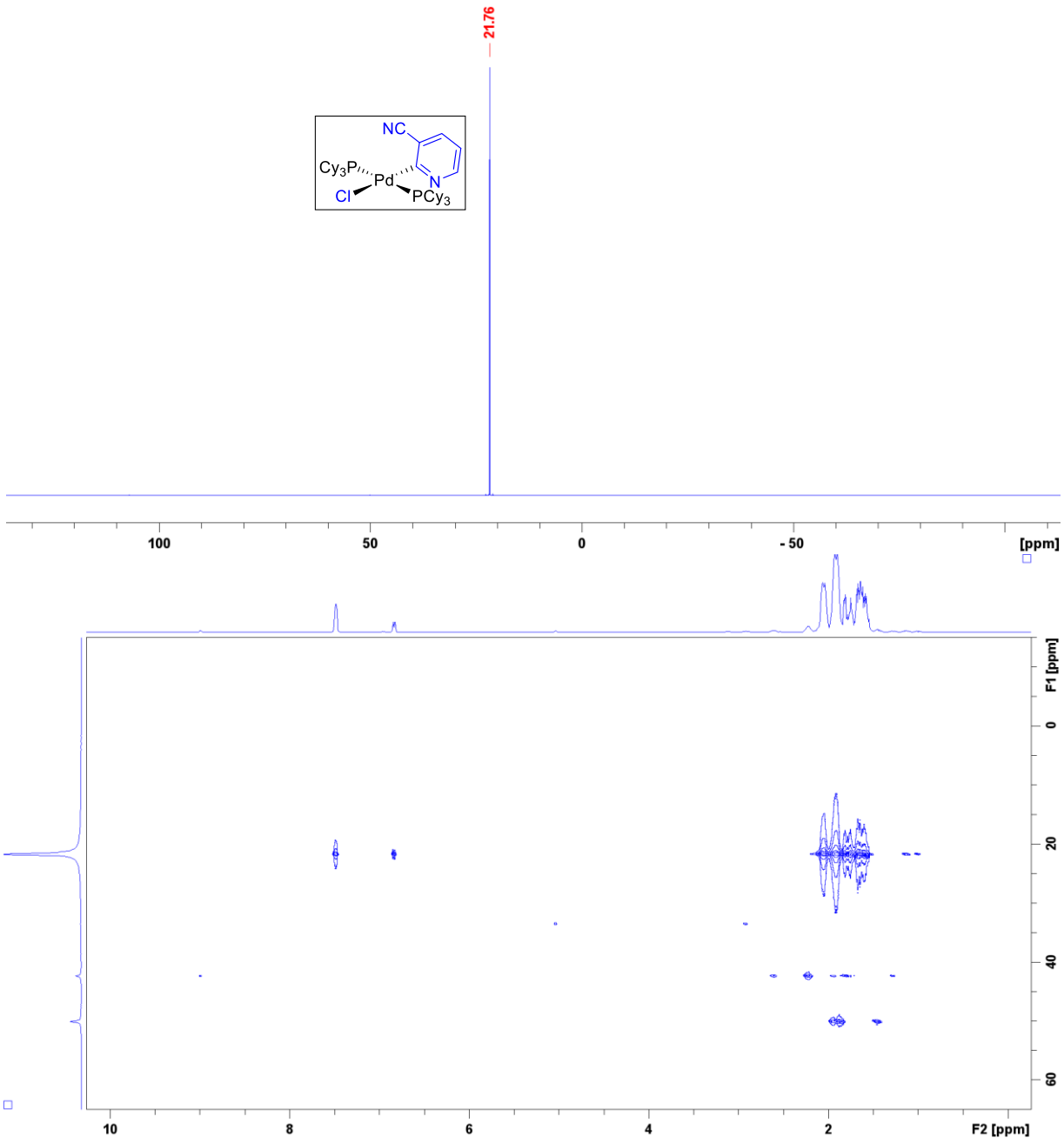


Fig. B21. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **P4**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

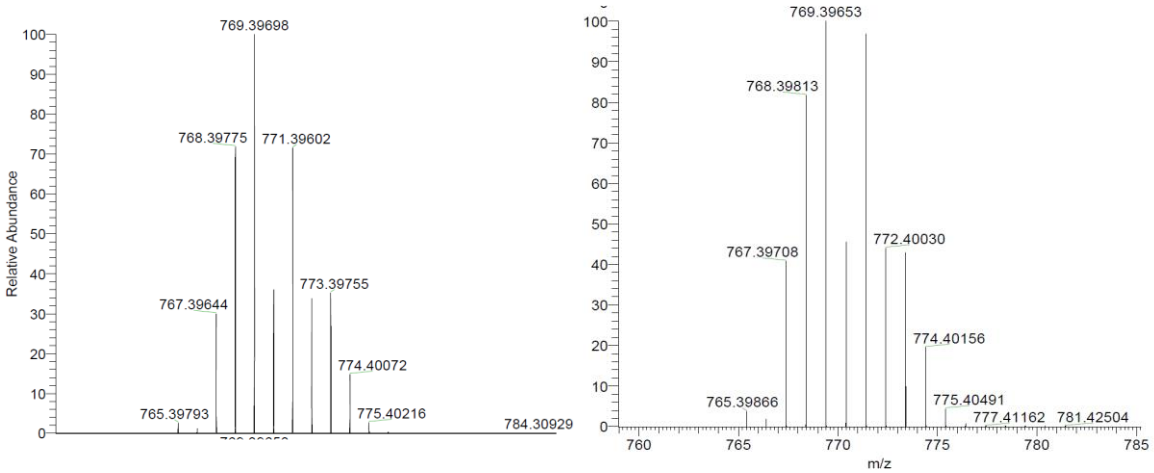


Fig. B22. Left: Experimental HRMS-ESI spectrum of [**P4**-Cl]$^+$. Right: Calculated HRMS isotope pattern for [**P4**-Cl]$^+$.

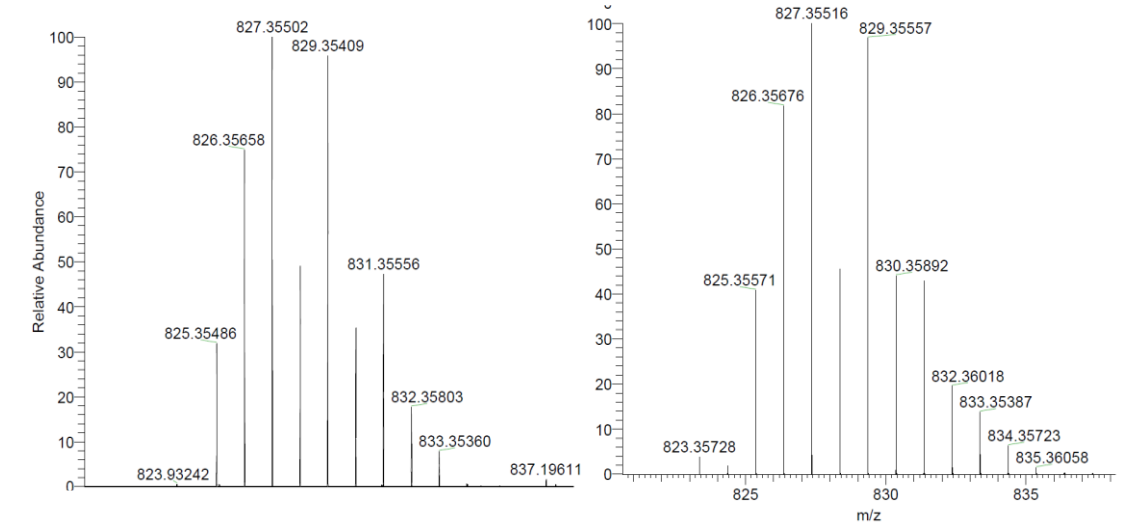


Fig. B23. Left: Experimental HRMS-ESI spectrum of [**P4**+Na]$^+$. Right: Calculated HRMS isotope pattern for [**P4**+Na]$^+$.

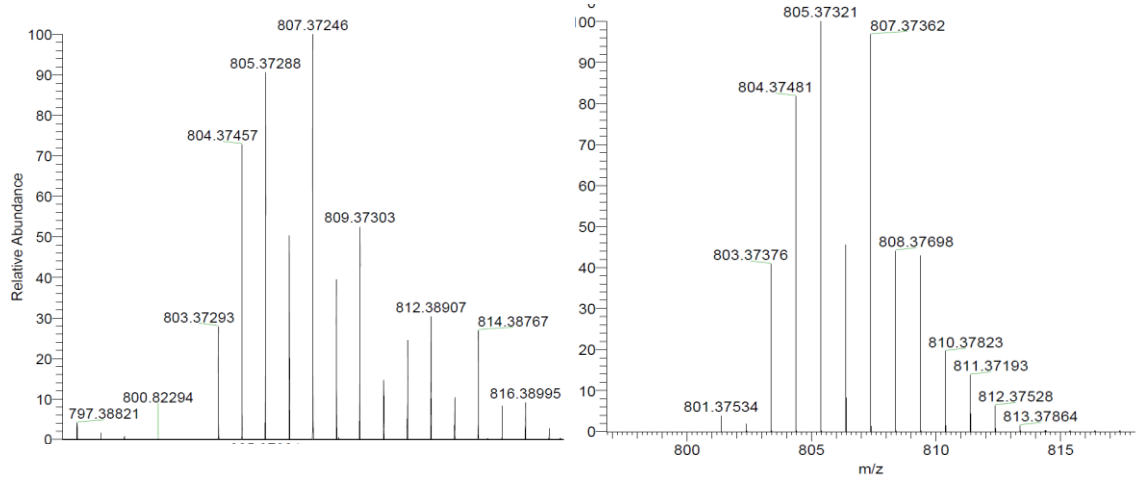


Fig. B24. Left: Experimental HRMS-ESI spectrum of [**P4**+H]$^+$. Right: Calculated HRMS isotope pattern for [**P4**+H]$^+$.

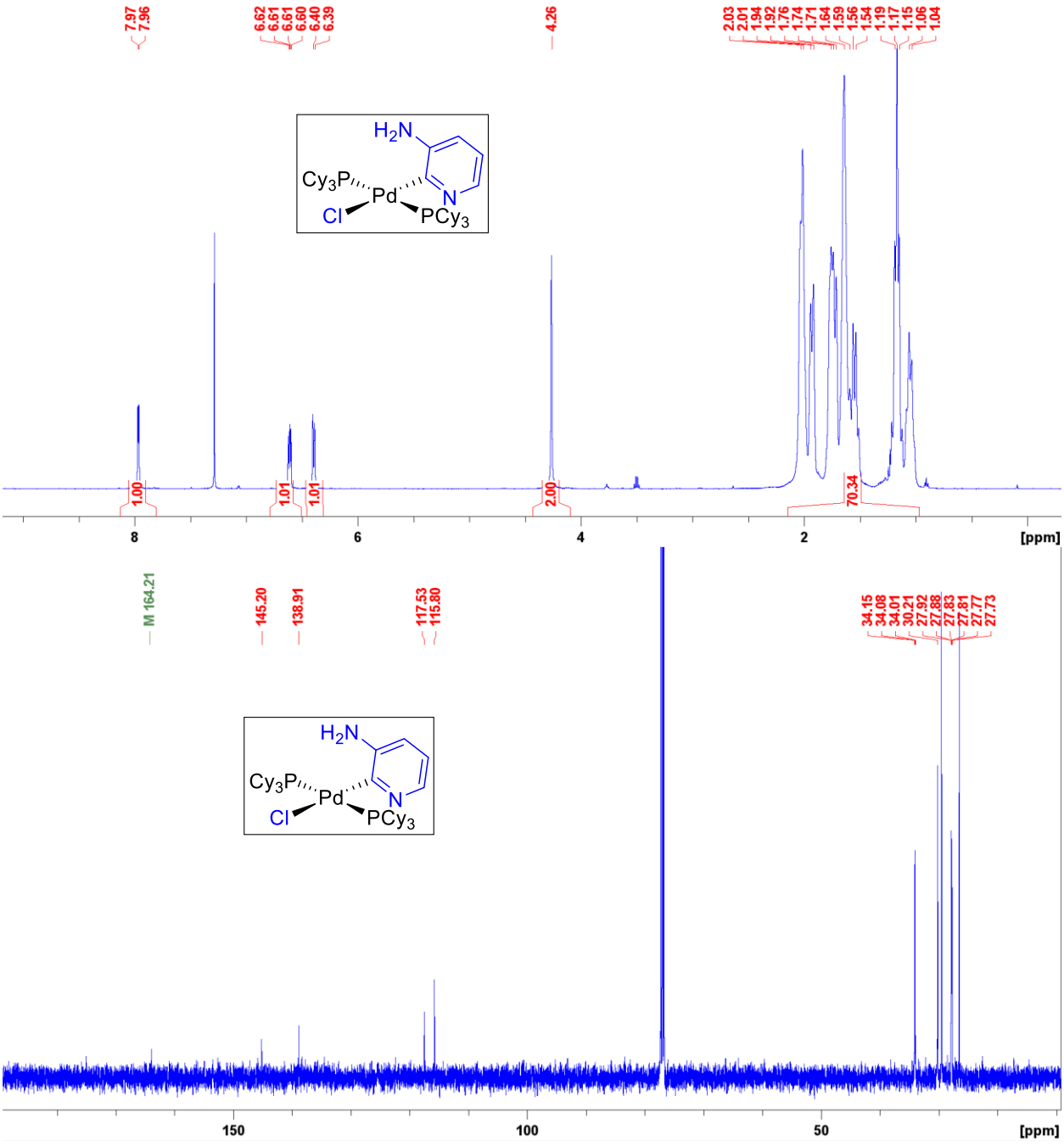


Fig. B25. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **P5**.

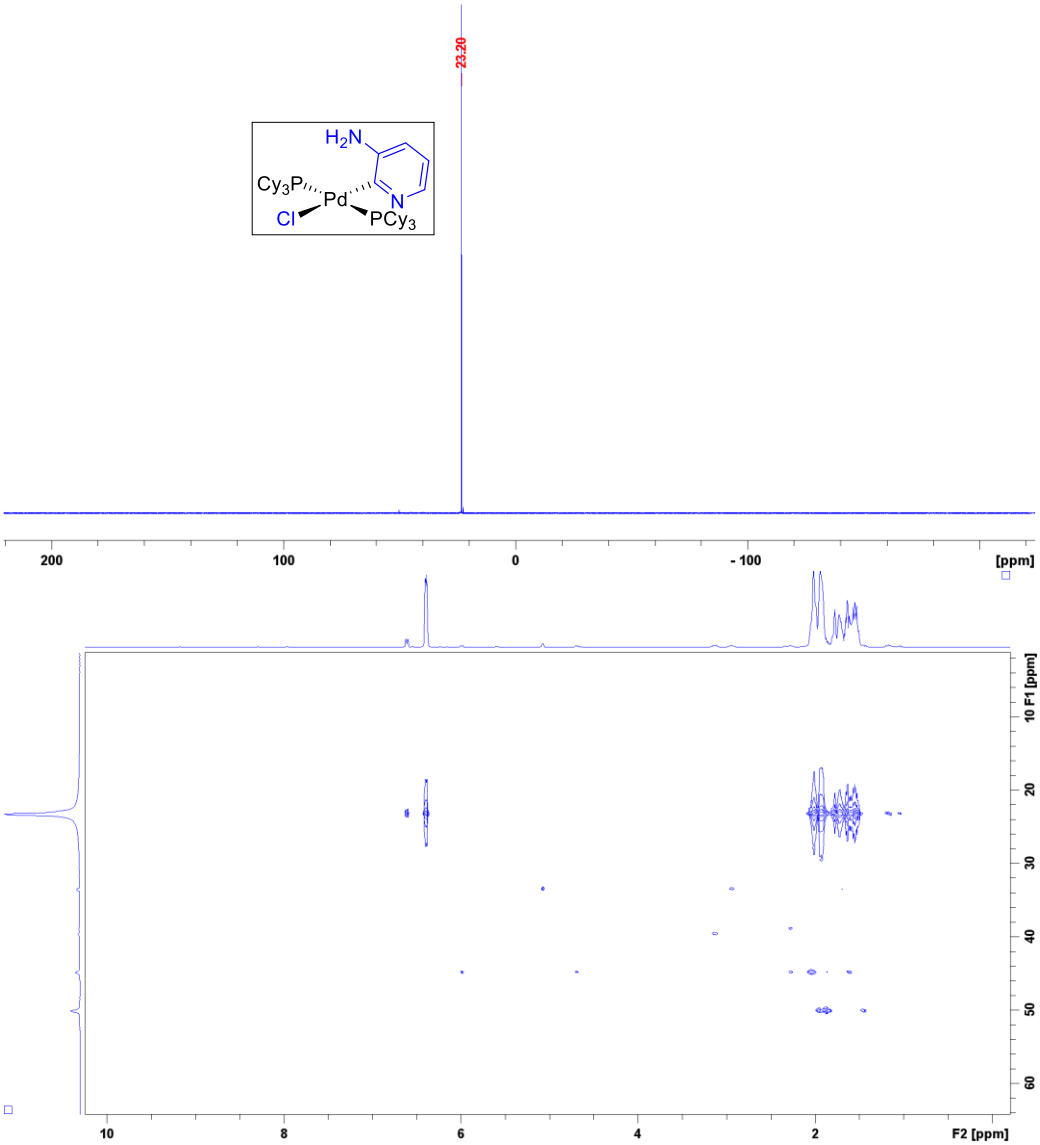


Fig. B26. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **P5**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

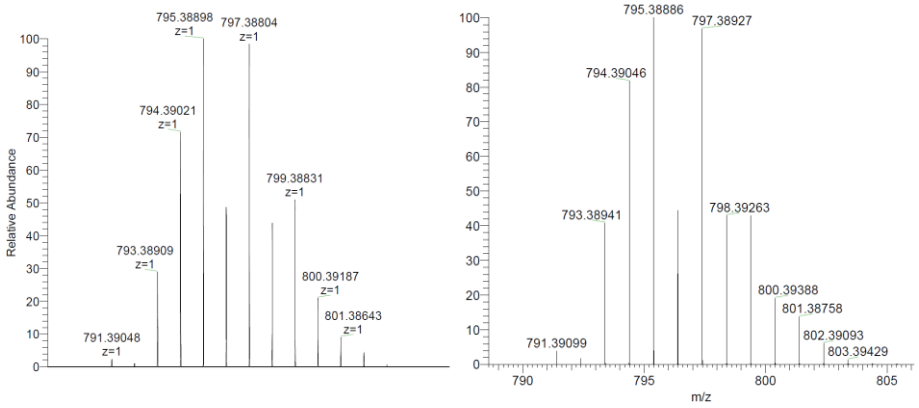


Fig. B27. Left: Experimental HRMS-ESI spectrum of [**P5**+H]$^{+}$. Right: Calculated HRMS isotope pattern for [**P5**+H]$^{+}$.

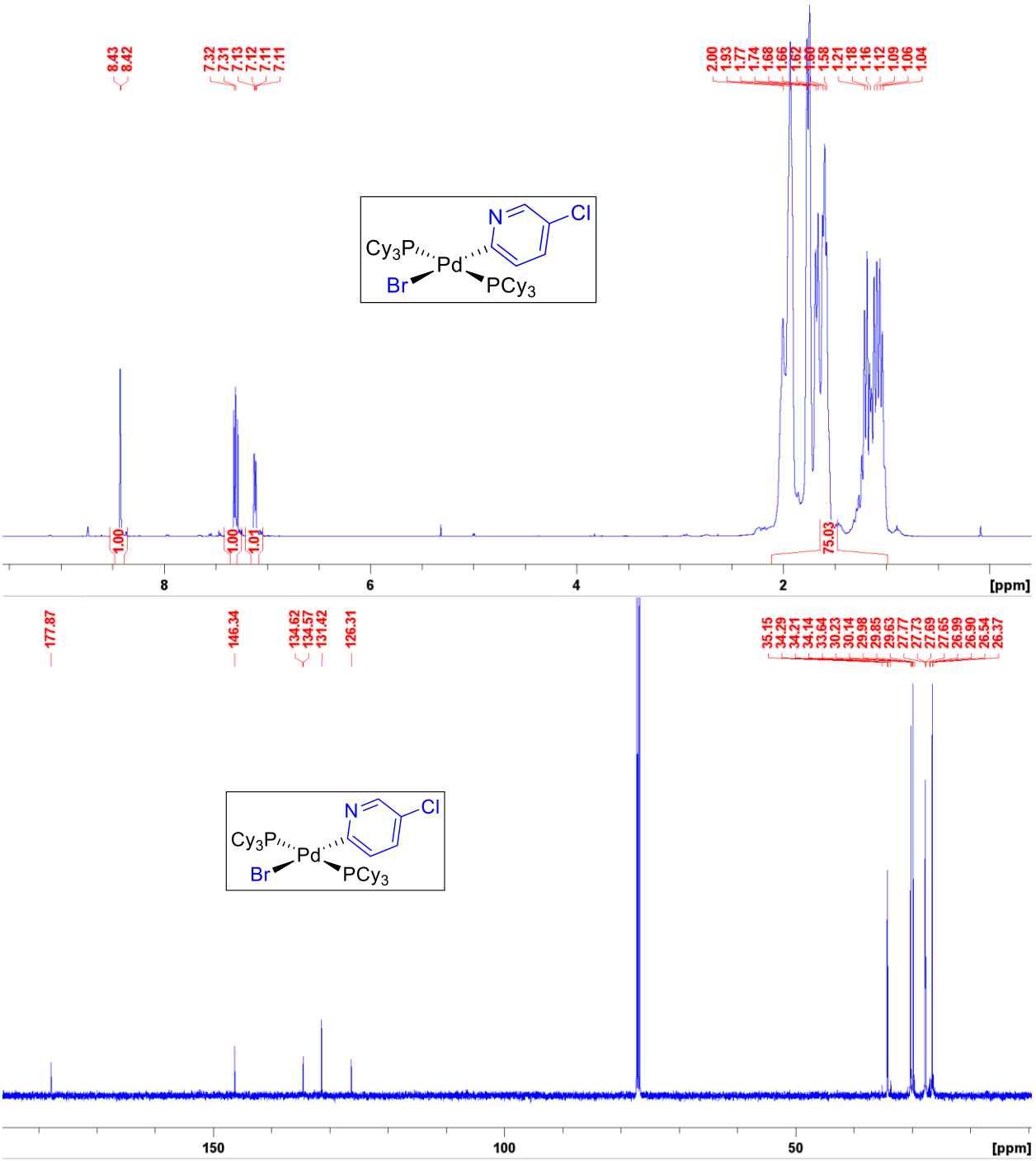


Fig. B28. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **P6**.

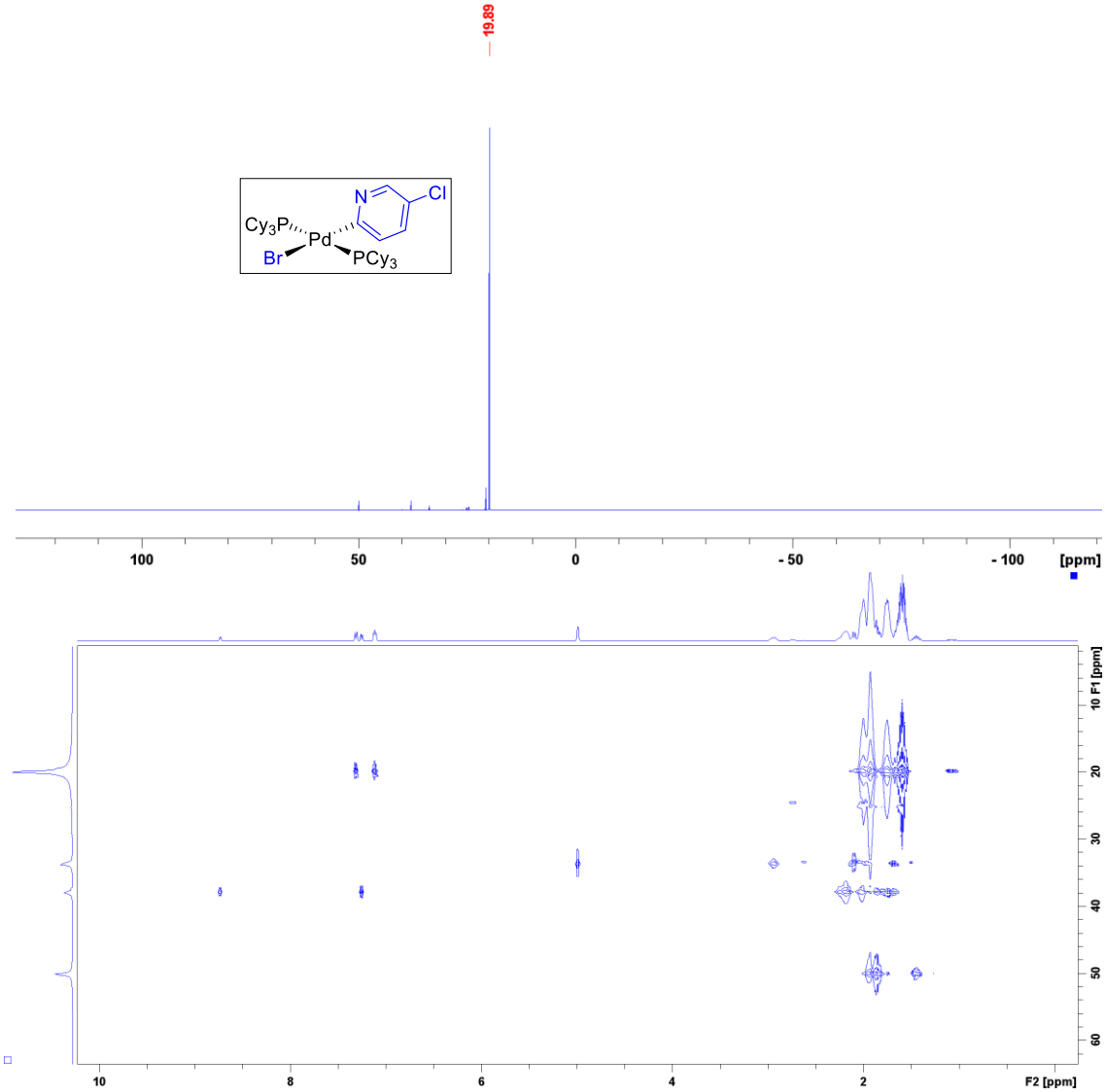


Fig. B29. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **P6**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

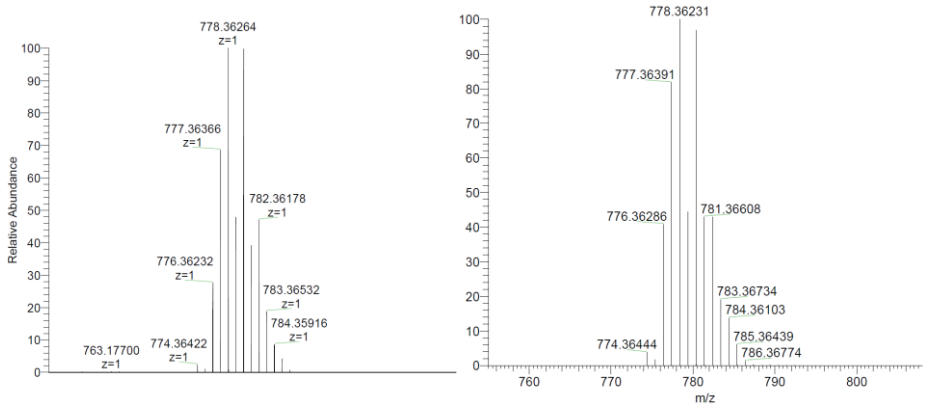


Fig. B30. Left: Experimental HRMS-ESI spectrum of [**P6**-Br]$^{+}$. Right: Calculated HRMS isotope pattern for [**P6**-Br]$^{+}$.

Computational Determination of Molecular Descriptors

All geometry optimizations were performed using ORCA 4.0.1.2[9]. Initial substrate structures were either downloaded from the ChemSpider database[10] or generated using Avogadro version 4.1.[11] Geometry optimizations were performed in ORCA with a B3LYP/def-TVZPD approach. Wavefunction *.wfn* files were generated with the ORCA_2aim utility and imported in the Multiwfn program version 3.7[12,13] for calculation of electrostatic potential (*ESP*) and intrinsic bond strength index (*IBSI*). A-values were obtained from published tables.[14]

The molecular *ESP* was calculated at a 0.004 au isosurface of electron density. Surface properties for individual atoms were selected, including the atom-based surface area and the maximal, minimal and average *ESP* values at that surface Average *ESP* values at the reactive center (carbon) and its neighboring atom (nitrogen or carbon) were used as the electronic descriptors in construction of the oxidative addition predictive model. Electrostatic potential maps were plotted in VMD,[15] using electron density and *ESP* cube files generated by Multiwfn.

The *IBSI* was used as the bond strength descriptor between the reactive center (carbon) and the leaving group (Br, Cl, or OTf). *IBSI* is a recently proposed interpretation to quantify the bond strength of the covalent bonds.[16] The *IBSI* of the carbon-halogen bonds for each substrate was calculated using the Hirshfeld independent gradient model (IGMH) at high quality in Multiwfn.

To obtain molecular descriptors for the catalysts in the Sonogashira case study, the structures of three bisligated palladium(0) complexes $Pd(PCy_3)_2$, $Pd(P\textit{i}Pr_3)_2$ and $Pd(P\textit{t}Bu_3)_2$ were obtained from the literature.[17] The structures of the other 14 bisligated Pd(0) complexes were initially edited in Avogadro using one of the available structures as the starting point. Geometry optimizations of the 17 bisligated, 17 monoligated Pd(0) catalysts and 17 phosphine ligands were carried out using the B3LYP/def-TVZPD approach for all atoms except Pd. The LANL2DZ basis set with effective core potentials was used for Pd. Since effective core potentials are not supported by ORCA_2aim, single-point energy calculations with the all-electron relativistic ZORA-def2-TZVP basis set were carried out on the optimized geometries of Pd-containing compounds. *ESP* calculations for bisligated catalysts used a 0.01 au isosurface of electron density, whereas for monoligated catalysts and phosphine ligands a 0.004 au isosurface was used. The larger isosurface value for bisligated catalysts was due to the inability of the code to calculate surface areas of the buried Pd center. The $\%V_{bur}$ descriptor is calculated using Sambvca.

A complete list of the molecular descriptors for the 79 oxidative addition substrates is given in Table B4. The descriptors for the 17 phosphine ligands and the corresponding PdL_2 and PdL complexes is given in Table B6.

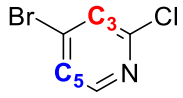
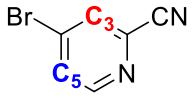
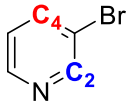
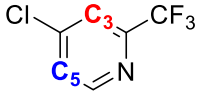
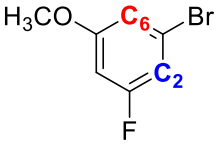
Table B4 –Calculated molecular descriptors for the 70 (hetero)aryl halide and 9 aryl triflate substrates used in oxidative addition experiments (Lu-PhD-thesis_Ch.3 SI tables.xlsx).
** pKa of HCl, HBr and HOTf is 0.2, -4.4, -11.3, respectively.[17]

For substrates where ESP_2 values at the two adjacent carbon atoms are different, the smaller ESP value is used as ESP_2 in constructing the linear regression model in order to give smaller $\Delta G^{\ddagger}_{OA}$.

The ESP values at the two adjacent atoms, the predicted $\Delta G^{\ddagger}_{OA}$ using both ESP values, and the absolute difference between the two $\Delta G^{\ddagger}_{OA}$ values are shown below in Table B5 for

comparison, confirming that the selection of one or the other ESP_2 value does not appreciably change the predicted $\Delta G^{\ddagger}_{OA}$.

Table B5.

Substrate	Average ESP_2 (kJ mol^{-1})	*Predicted* $\Delta G^{\ddagger}_{OA}$ (kJ mol^{-1})	*Absolute difference* (kJ mol^{-1})
 4-bromo-2-chloro-pyridine	C5: 52.924	16.868	0.372
	C3: 56.524	17.240	
 4-bromo-2-cyano-pyridine	C5: 86.644	9.566	0.464
	C3: 91.132	10.031	
 3-bromo-pyridine	C2: 34.990	26.579	0.546
	C4: 40.269	27.125	
 4-chloro-2-trifluoromethyl-pyridine	C5: 71.900	27.168	0.447
	C3: 76.2212	27.615	
 1-bromo-3-fluoro-5-methoxybenzene	C2: 19.917	30.492	0.271
	C6: 22.534	30.762	
	C5: 48.6364	32.838	0.932

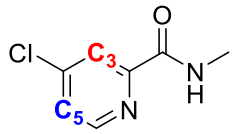
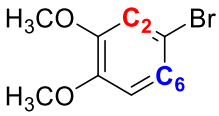
 N-Methyl-4-chloropyridine-2-carboxamide	C3: 57.6509	33.770	
 1-bromo-3,4-dimethoxybenzene	C6: -24.978	41.169	1.361
	C2: -11.822	42.530	

Table B6 –Calculated molecular descriptors for bisligated Pd(0) complexes (PdL_2), monoligated Pd(0) complexes (PdL) and phosphines (L) for the 17 ligands used in Sonogashira case study (Lu-PhD-thesis_Ch.3 SI tables.xlsx).

Hammett Analyses of para-Substituted Substrates

To further validate the kinetic parameters obtained by competition experiments, we have obtained reaction constants (ρ) through construction of Hammett plots – $\log(k_Z/k_H)$ versus substituent σ values – for five sets of *para*-substituted substrates undergoing oxidative addition. These include: 5-substituted-2-chloro-pyridines, 5-substituted-2-bromo-pyridines, 4-substituted-1-bromo-benzenes, 4-substituted-2-chloropyridines and 4-substituted phenyl triflates. Substituent σ values were obtained from published tables.[18]

For oxidative addition of the 5-Z-2-Cl-pyridines, we obtain reaction constants of ρ = 4.8 in THF (Fig. B31). This is similar to the published value of ρ = 4.3 obtained for an analogous set of reactions using $Pd(PPh_3)_4$ in THF.[19]

For oxidative addition of the 5-Z-2-Br-pyridines, we obtain a reaction constant of ρ = 3.2 in THF (Fig. B32), which is smaller than the published value of ρ = 4.4 obtained for an analogous set of reactions using $Pd(PPh_3)_4$ in THF.[19]

For oxidative addition of the 4-Z-1-Br-benzenes, we obtain reaction constants of ρ = 3.4 in THF (Fig. B33). We also noted slightly worse linear correlation when using standard σ_{para} values ($R^2 = 0.93$). Plotting $\log(k_Z/k_H)$ versus σ_{para}^- gives better linear correlation ($R^2 = 0.98$), and reaction constant of ρ = 2.3 in THF (Fig. B34).

For oxidative addition of the 4-Z-2-Cl-pyridines, we plotted $\log(k_Z/k_H)$ versus both σ_{meta} and σ_{para} to assess the inductive effect of Z on the C–X position, and the resonance effect on the adjacent pyridine nitrogen. Using σ_{meta} (Fig. B35), we obtain reaction constant of ρ = 5.9 in THF, though with relatively poor linear correlation ($R^2 = 0.92$ in THF). In contrast, the Hammett plots using σ_{para} give better linear correlation ($R^2 = 0.99$ in THF), and reaction constant of ρ = 4.0 in THF (Fig. B36). This is comparable to the published

value of ρ = 3.3 using $Pd(PPh_3)_4$ in THF (though this previous work only included 3 substrates).[19] The strong linear correlation for σ_{para} indicates that resonance stabilization of negative charge at N in the transition state accelerates the oxidative addition reaction. This is consistent with the proposed S_NAr-like mechanism, and further validates the inclusion of ESP_2 in our quantitative model.

For oxidative addition of the 4-Z-phenyl-triflates, we obtain reaction constant of ρ = 3.9 in THF (Fig. B37). We plotted $\log(k_Z/k_H)$ versus both σ_{para} and σ_{para}^- and noted slightly worse linear correlation when using standard σ_{para} values ($R^2 = 0.93$). Plotting $\log(k_Z/k_H)$ versus σ_{para}^- gives better linear correlation ($R^2 = 0.98$), and reaction constant of ρ = 2.3 in THF.

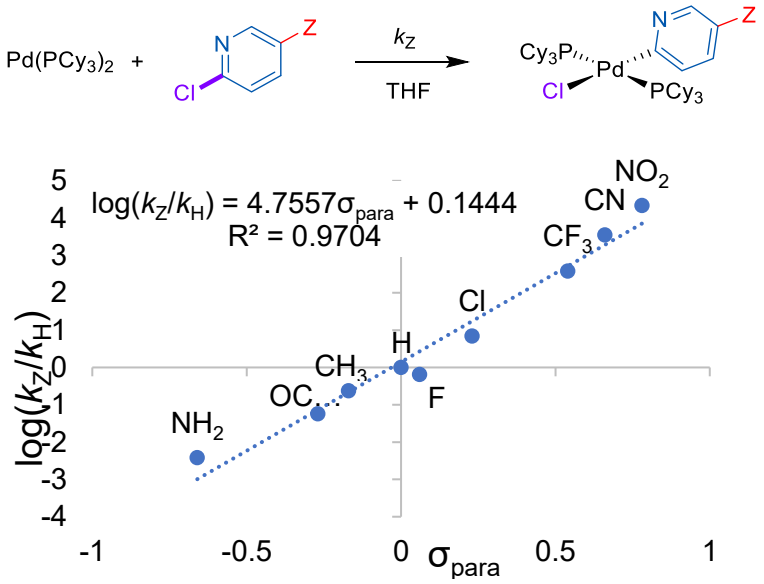


Fig. B31. Hammett plot of $\log(k_Z/k_H)$ versus σ_{para} for oxidative addition of a group of 5-Z-2-chloropyridines.

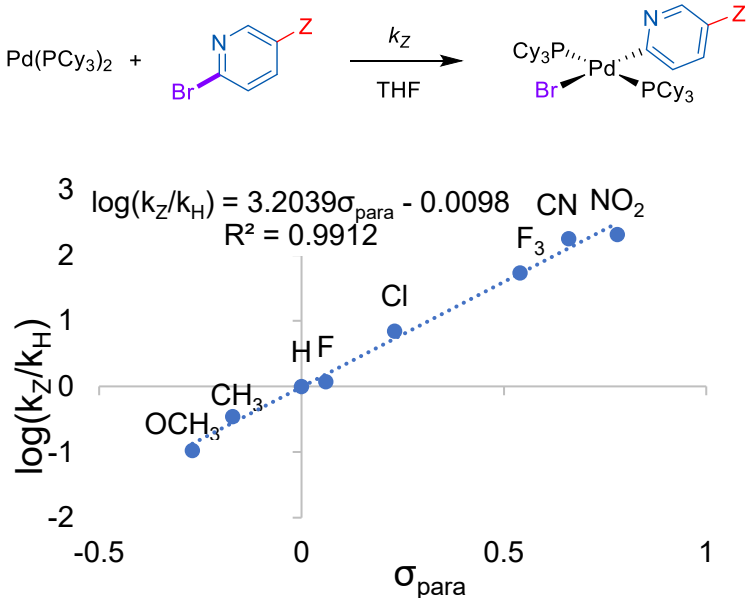


Fig. B32. Hammett plot of log(k_Z/k_H) versus σ_{para} for oxidative addition of a group of 5-Z-2-bromopyridines.

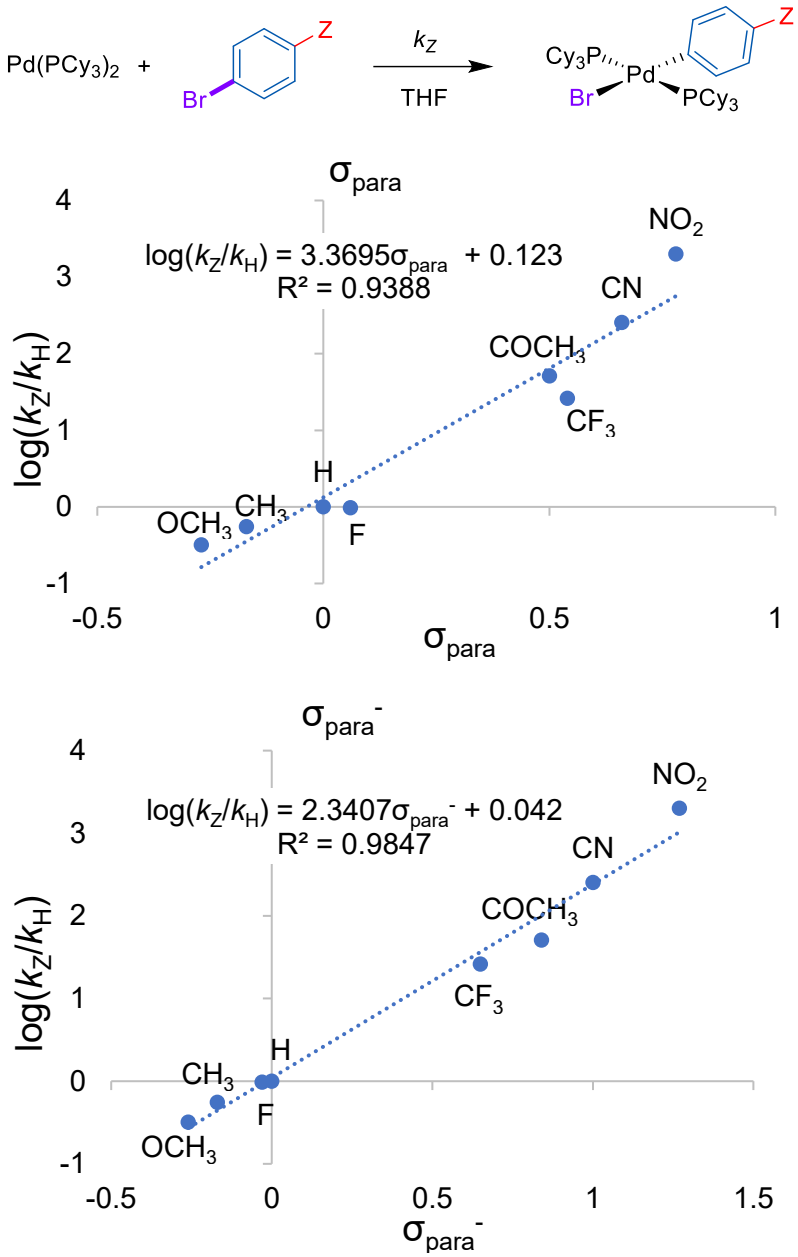


Fig. B33. Hammett plots of log(k_Z/k_H) versus σ_{para} and σ_{para}^- for oxidative addition of a group of 1-Z-4-bromobenzenes.

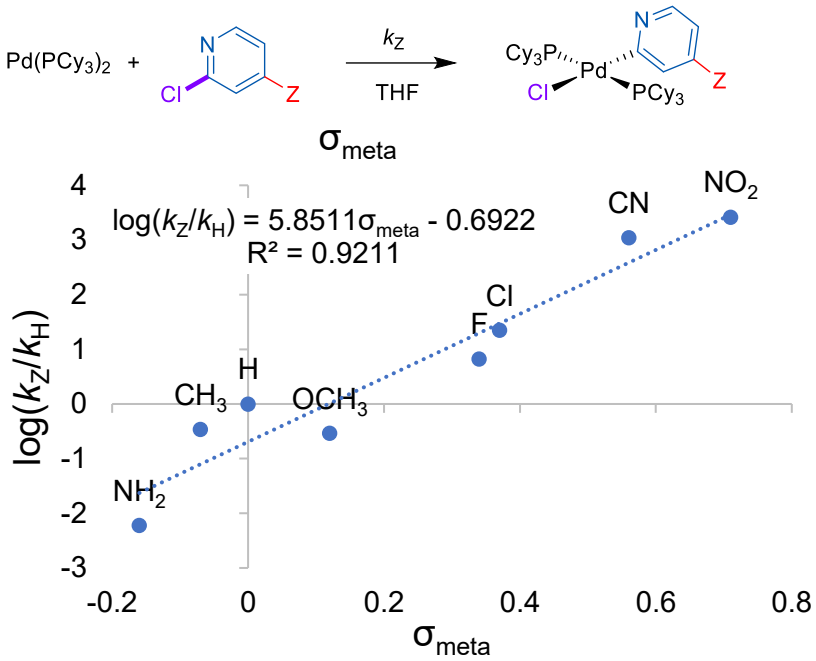


Fig. B34. Hammett plot of $\log(k_Z/k_H)$ versus σ_{meta} for oxidative addition of a group of 4-Z-2-chloropyridines.

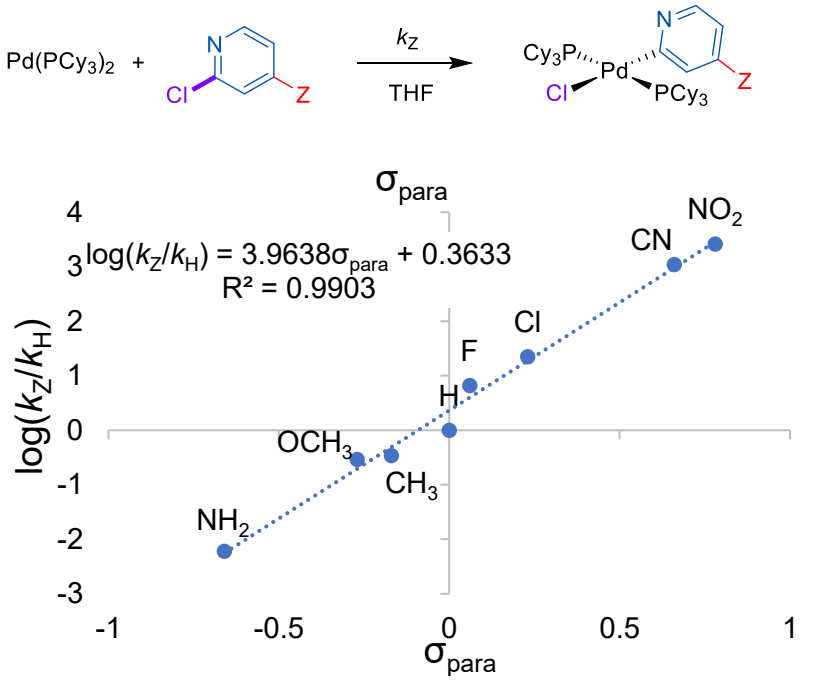


Fig. B35. Hammett plots of $\log(k_Z/k_H)$ versus σ_{para} for oxidative addition of a group of 4-Z-2-chloropyridines.

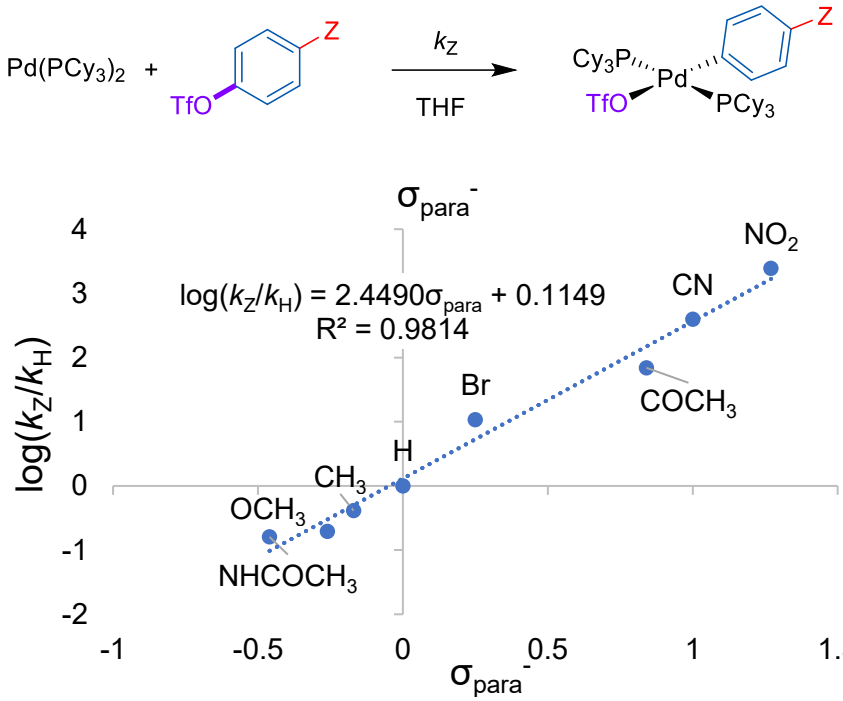
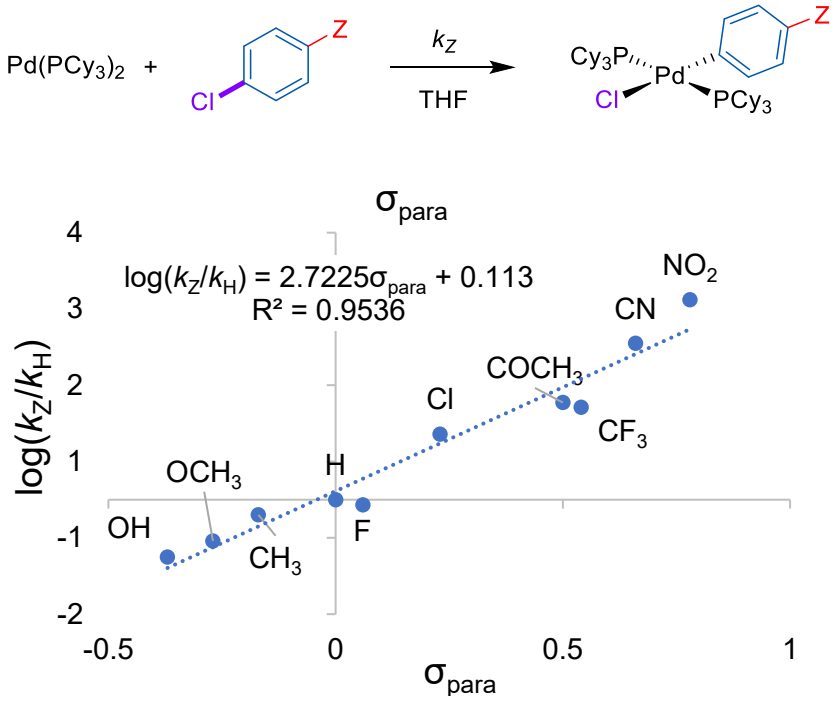


Fig. B36. Hammett plots of $\log(k_Z/k_H)$ versus σ_{para}^- for oxidative addition of a group of 4-Z-phenyl triflates.



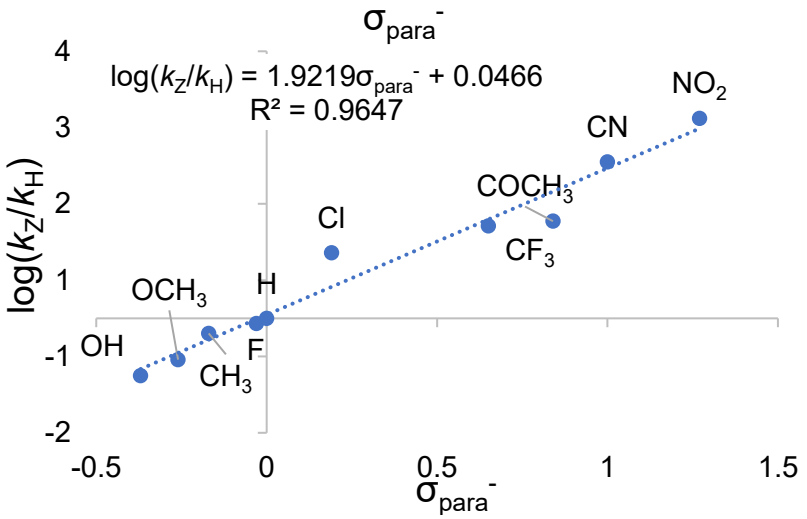


Fig. B37. Hammett plots of $\log(k_Z/k_H)$ versus σ_{para} and σ_{para}^- for oxidative addition of a group of 1-Z-4-chlorobenzenes.

π-Complex Intermediates and Transition States for Oxidative Addition

To assess the substrate-catalyst effects on the π-complex intermediate and oxidative addition transition states, we calculated the structures and *ESP* maps for 11 π-complexes and 6 transition states (structures for 7 of the 11 π-complexes are shown in Fig. 3.3A, main text). To focus our analysis, we did not calculate the transition state for the coordination of the substrate to form the π-complex intermediate.

Transition state structures were obtained in ORCA 4.0.1.2[9] using relaxed scans with an RI BP86/def2-SVP approach with D3BJ dispersion for all atoms except for Pd, for which a def2-TZVP basis set was used. Single point calculations were then performed with a RI-B2PLYP/def2-TZVP approach with D3 dispersion to calculate TS energies. These structures were used for subsequent calculations of energies at the TZP level and average *ESP* values. An example input file to compute a geometry is provided below.

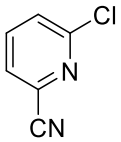
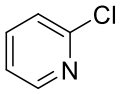
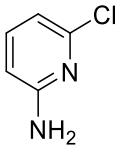
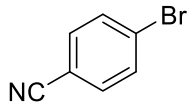
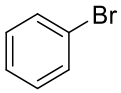
Geometry:
! RI BP86 def2-SVP def2/J D3BJ TIGHTSCF Opt Grid3 FinalGrid5
%basis newgto Pd "def2-TZVP" end end
Energy:
! RI-B2PLYP D3 def2-TZVP def2-TZVP/C TIGHTSCF

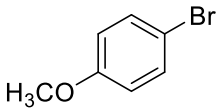
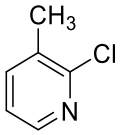
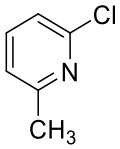
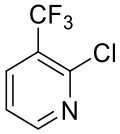
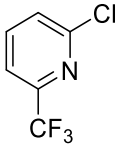
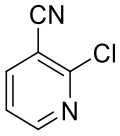
The bent geometry of the $Pd(PCy_3)_2$ starting structure (P–Pd–P bond angle of 160.3°) agrees with geometries previously reported in the literature.[20] Our π-complex intermediate structures and transition states are also consistent with previously reported geometries for related systems.[19]

To locate transition states for oxidative addition, different reaction coordinates were scanned depending on the nature of the aromatic substrate. For chlorinated pyridine substrates, the transition state of the oxidative addition step was found by stretching along the Cl-C2 bond axis to locate the "S_NAr-like" transition state. To locate the "3-centered" oxidative addition transition state for aryl bromides, we used a relaxed scan opening the

C1-Pd-Br angle from the π-complex intermediate to the *cis* oxidative addition product; this *cis* complex then isomerizes to the more stable *trans* geometry (which is the experimentally observed geometry). The energies of the six intermediates and their corresponding transition states and products are given in Table B7. The simplified reaction coordinates for oxidative addition of the 2-chloropyridines and bromobenzenes are given in Figures B38 and B39 respectively. Overall, the trends in reactivity mirror those observed experimentally. One minor discrepancy is the transition state energy for 4-bromoanisole is slightly lower than for bromobenzene (3 kJ mol^{-1} difference), whereas the experimental $\Delta\Delta G^{\ddagger}$ is ~1.5 kJ mol^{-1} with bromobenzene lower than 4-bromoanisole.

Table B7. Energies of key species in the oxidative addition pathway (kJ/mol), relative to the unbound substrate and $Pd(PCy_3)_2$.

	Energy (kJ/mol)			
Substrate	**π-complex Intermediate**	**Transition state**	**Product (*cis* geometry)**	**Product (*trans* geometry)**
 2-chloro-6-cyanopyridine	14	27	-103	-198
 2-chloropyridine	18	36	-125	-164
 2-chloro-6-aminopyridine	45	57	-101	-179
 4-bromobenzonitrile	-9	15	-128	-206
 bromobenzene	11	34	-110	-190

 4-bromoanisole	7	31	-108	-157
 2-chloro-3-methylpyridine	-3	n/d	n/d	n/d
 2-chloro-6-methylpyridine	9	n/d	n/d	n/d
 2-chloro-3-trifluoromethylpyridine	3	n/d	n/d	n/d
 2-chloro-6-trifluoromethylpyridine	2	n/d	n/d	n/d
 2-chloro-3-cyanopyridine	-41	n/d	n/d	n/d

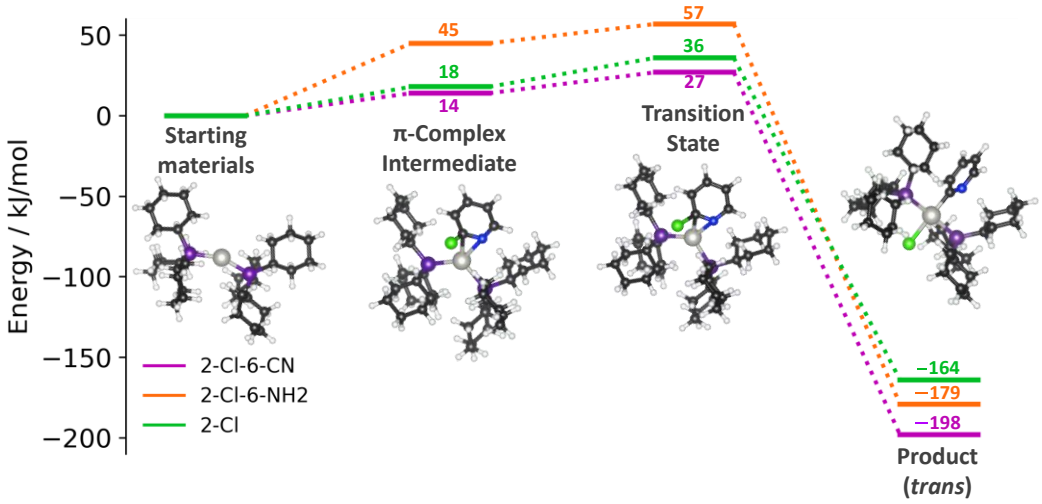


Fig. B38. Simplified calculated reaction coordinate for oxidative addition of 2-chloro-6-cyanopyridine, 2-chloro-6-aminopyridine, and 2-chloropyridine to $Pd(PCy_3)_2$. Calculated structures for the 2-chloropyridine derivatives are shown as representative. The *cis* oxidative addition product is not shown.

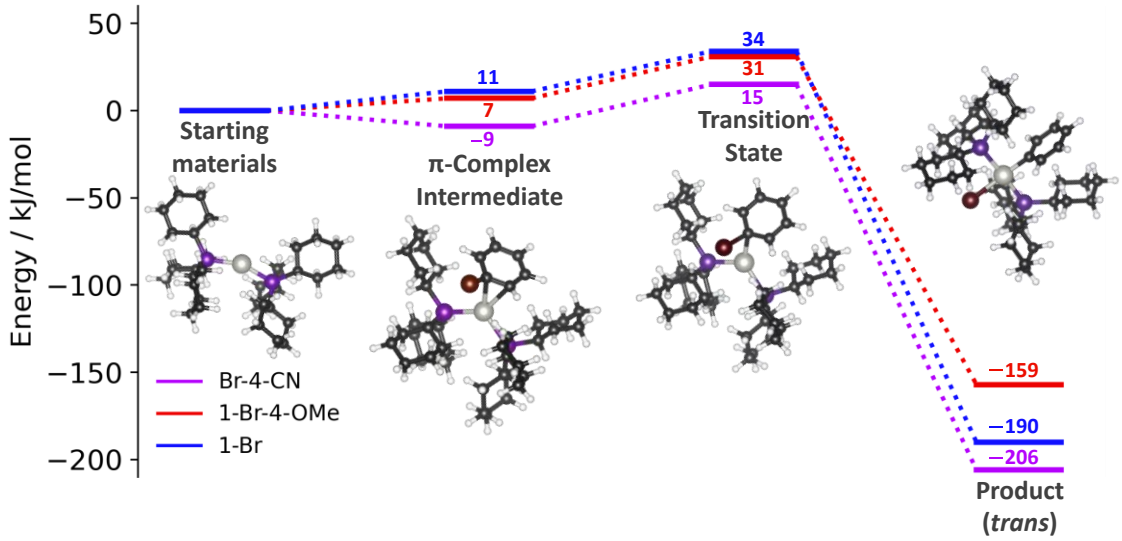


Fig. B39. Simplified calculated reaction coordinate for oxidative addition of 4-bromobenzonitrile, 4-bromoanisole, and bromobenzene to $Pd(PCy_3)_2$. Calculated structures for the bromobenzene derivatives are shown as representative. The *cis* oxidative addition product is not shown.

Figures B40 – B56 show the calculated structures of the oxidative addition transition states/intermediates, and *ESP* maps of the free substrate and the transition states/intermediates.

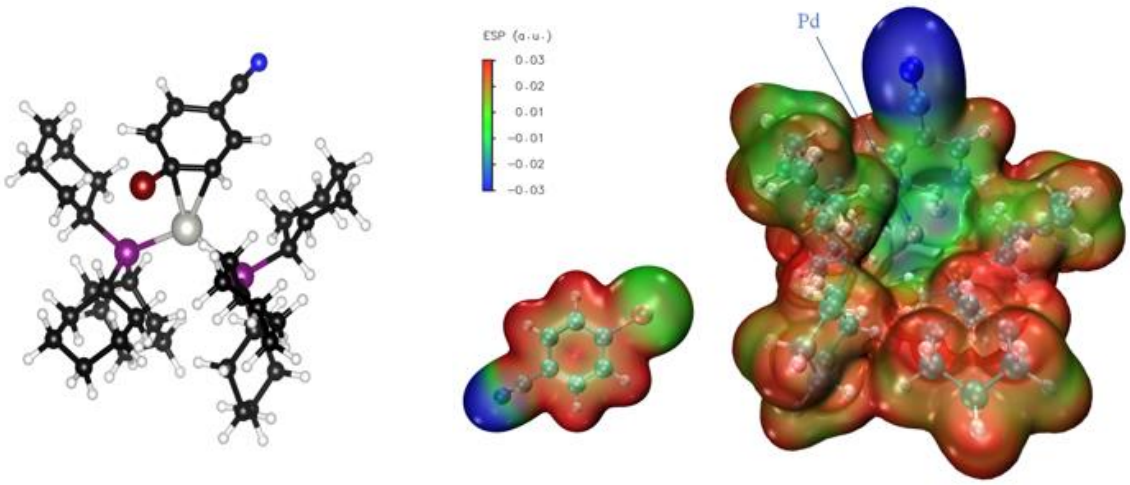


Fig. B40. Calculated structure of the π-complex intermediate for 4-bromobenzonitrile, and *ESP* maps of the free substrate and the π-complex intermediate. ESP_{Pd} = 12.4 kJ mol^{-1}.

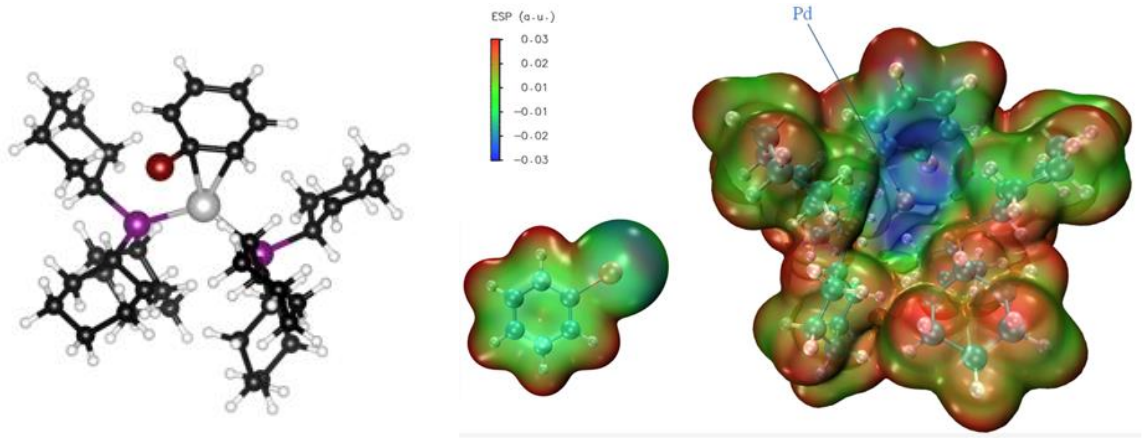


Fig. B41. Calculated structure of the π-complex intermediate for bromobenzene, and *ESP* maps of the free substrate and the π-complex intermediate. ESP_{Pd} = -23.2 kJ mol^{-1}.

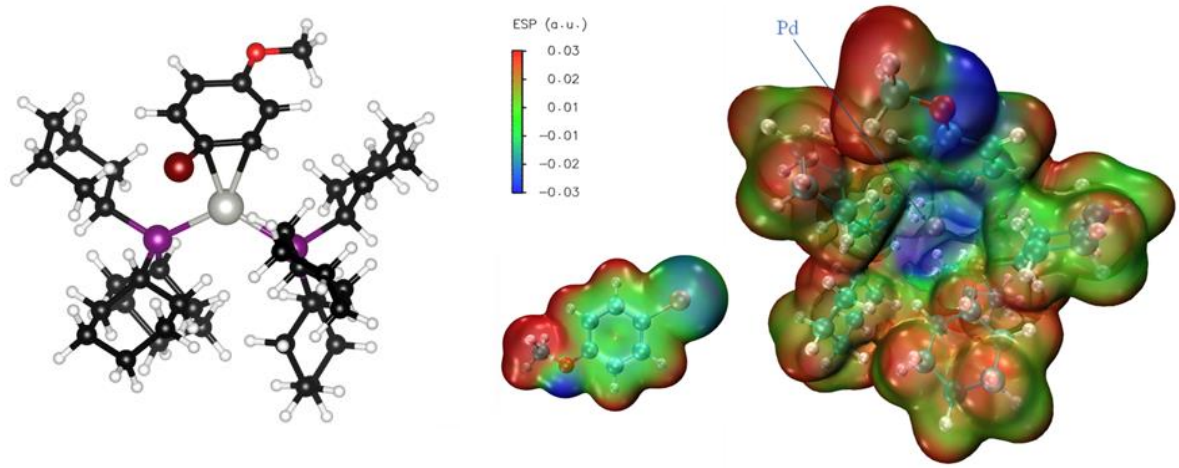


Fig. B42. Calculated structure of the π-complex intermediate for 4-bromoanisole, and *ESP* maps of the free substrate and the π-complex intermediate. ESP_{Pd} = -27.2 kJ mol^{-1}.

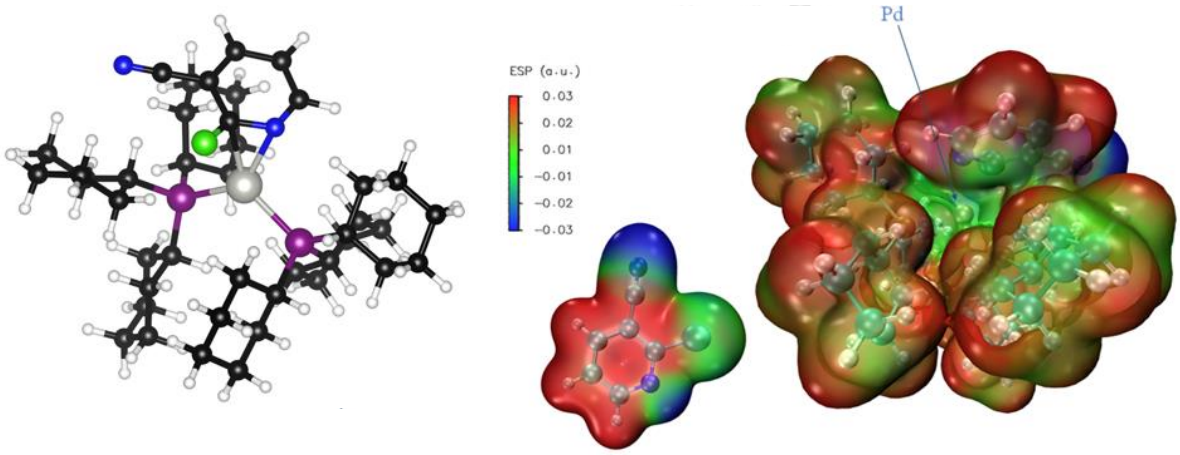


Fig. B43. Calculated structure of the π-complex intermediate for 2-chloro-3-cyanopyridine, and *ESP* maps of the free substrate and the π-complex intermediate. ESP_{Pd} = 26.1 kJ mol^{-1}.

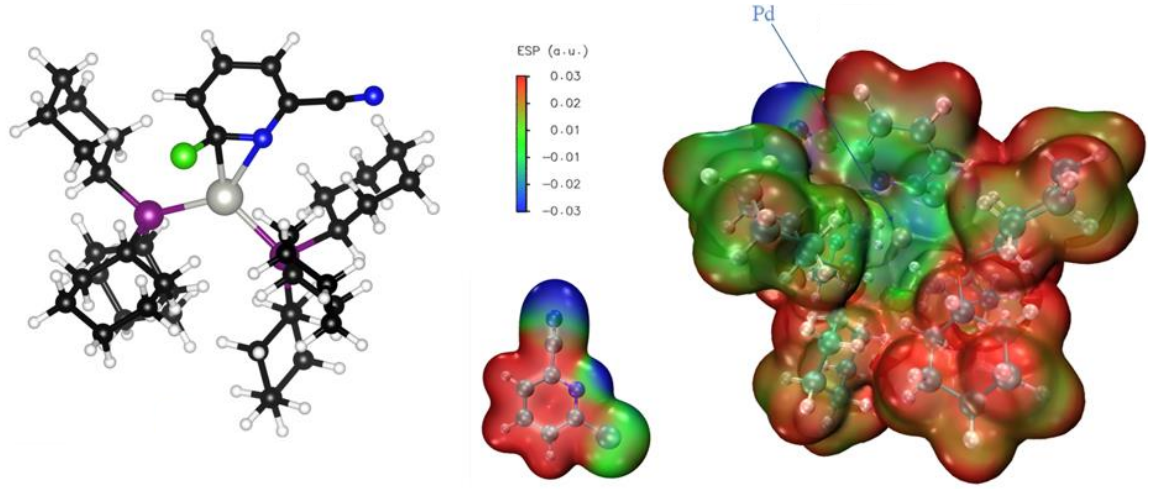


Fig. B44. Calculated structure of the π-complex intermediate for 2-chloro-6-cyanopyridine, and *ESP* maps of the free substrate and the π-complex intermediate. ESP_{Pd} = 25.3 kJ mol^{-1}.

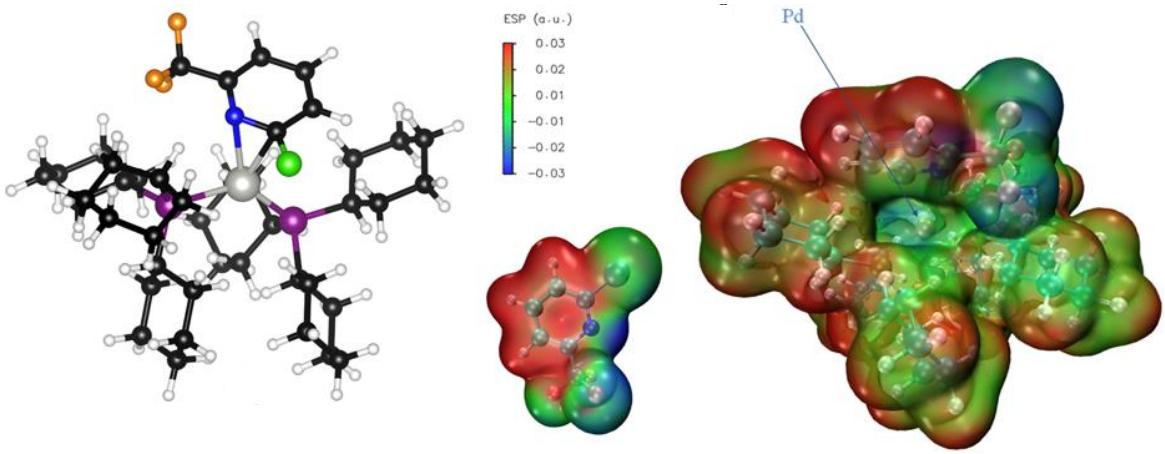


Fig. B45. Calculated structure of the π-complex intermediate for 2-chloro-6-trifluoromethylpyridine, and *ESP* maps of the free substrate and the π-complex intermediate. ESP_{Pd} = 1.7 kJ mol^{-1}.

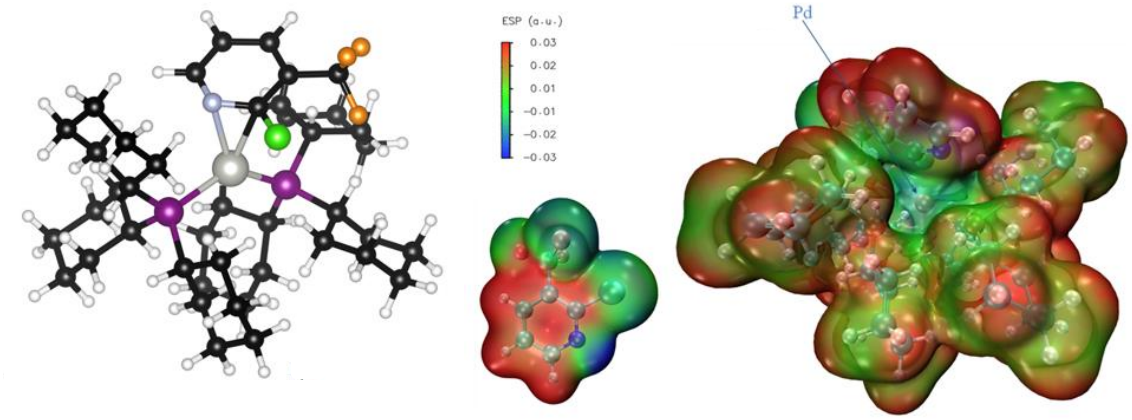


Fig. B46. Calculated structure of the π-complex intermediate for 2-chloro-3-trifluoromethylpyridine, and *ESP* maps of the free substrate and the π-complex intermediate. ESP_{Pd} = -3.6 kJ mol^{-1}.

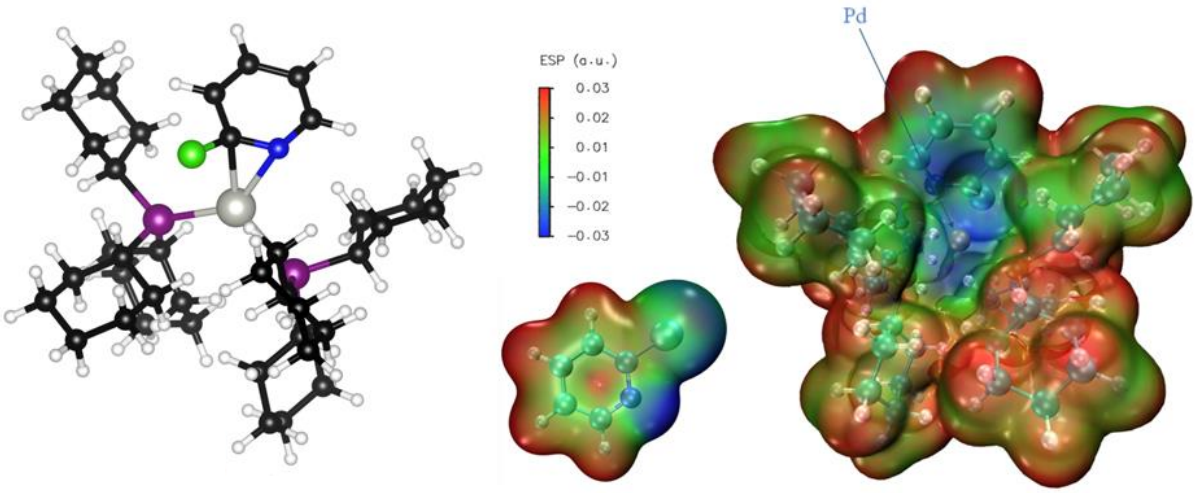


Fig. B47. Calculated structure of the π-complex intermediate for 2-chloropyridine, and *ESP* maps of the free substrate and the π-complex intermediate. ESP_{Pd} = -11.7 kJ mol^{-1}.

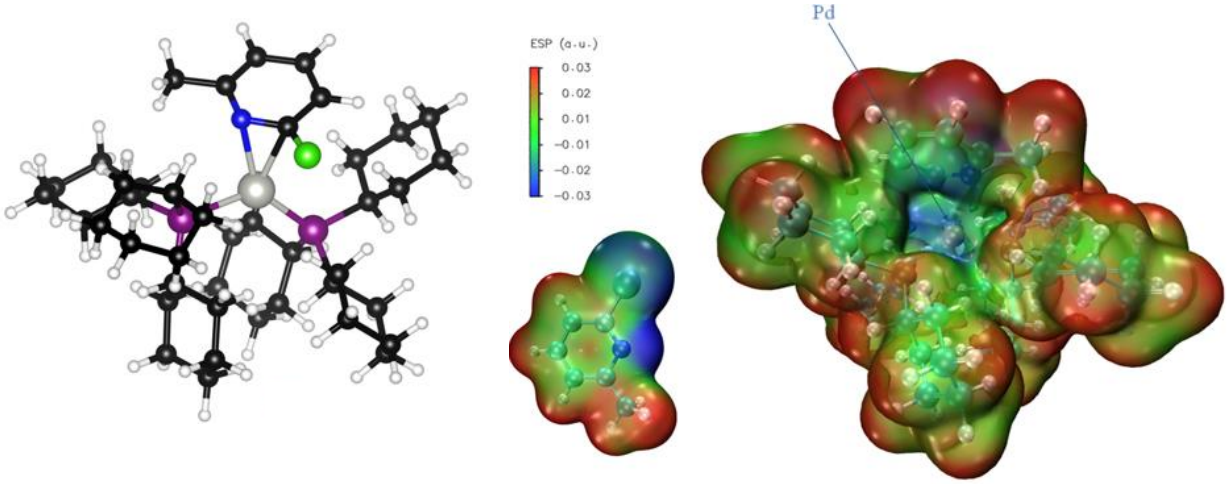


Fig. B48. Calculated structure of the π-complex intermediate for 2-chloro-6-methylpyridine, and *ESP* maps of the free substrate and the π-complex intermediate. ESP_{Pd} = -11.3 kJ mol^{-1}.

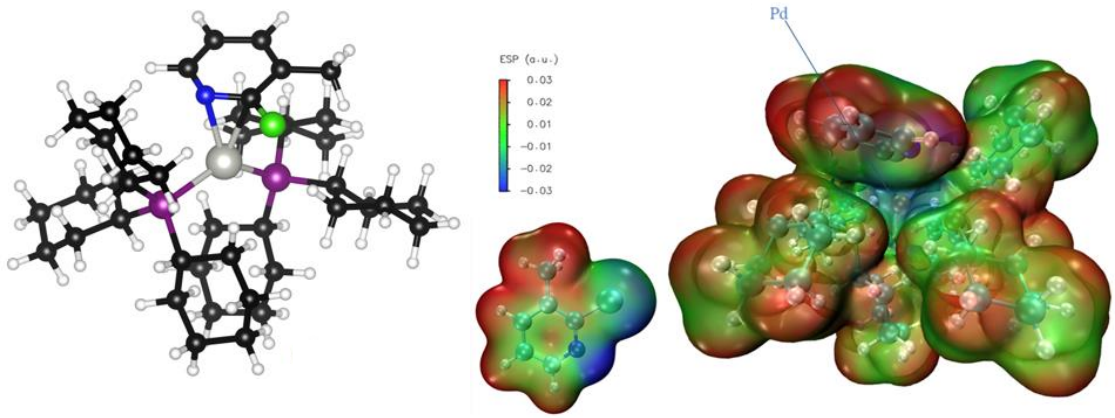


Fig. B49. Calculated structure of the π-complex intermediate for 2-chloro-3-methylpyridine, and *ESP* maps of the free substrate and the π-complex intermediate. ESP_{Pd} = -9.7 kJ mol^{-1}.

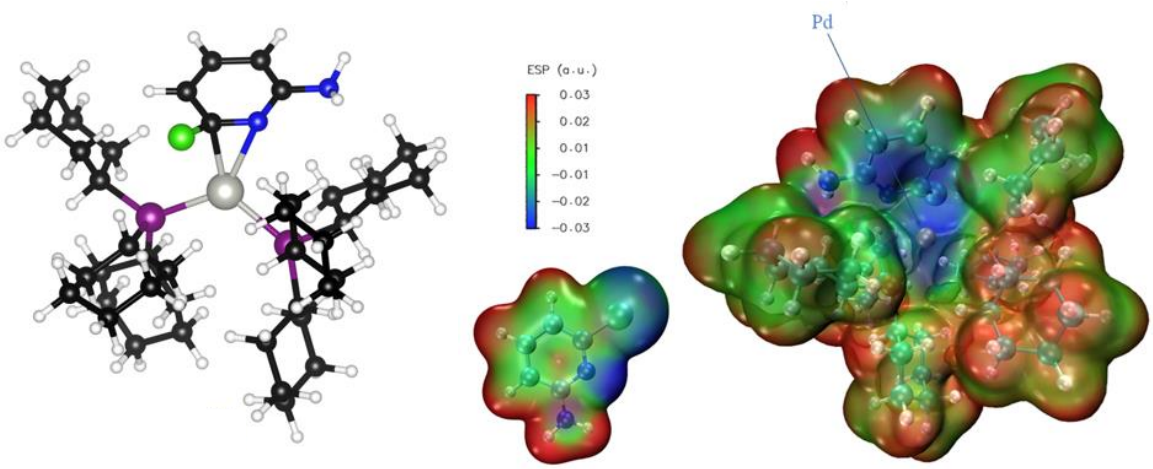


Fig. B50. Calculated structure of the π-complex intermediate for 2-chloro-6-aminopyridine, and *ESP* maps of the free substrate and the π-complex intermediate. ESP_{Pd} = -15.0 kJ mol^{-1}.

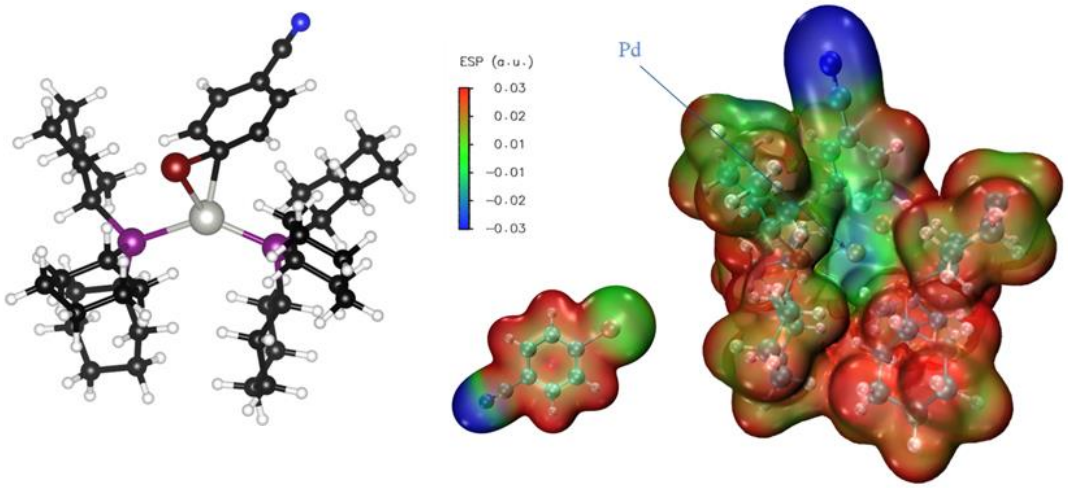


Fig. B51. Calculated structure of the oxidative addition transition state for 4-bromobenzonitrile, and *ESP* maps of the free substrate and the transition state. ESP_{Pd} = 22.9 kJ mol^{-1}.

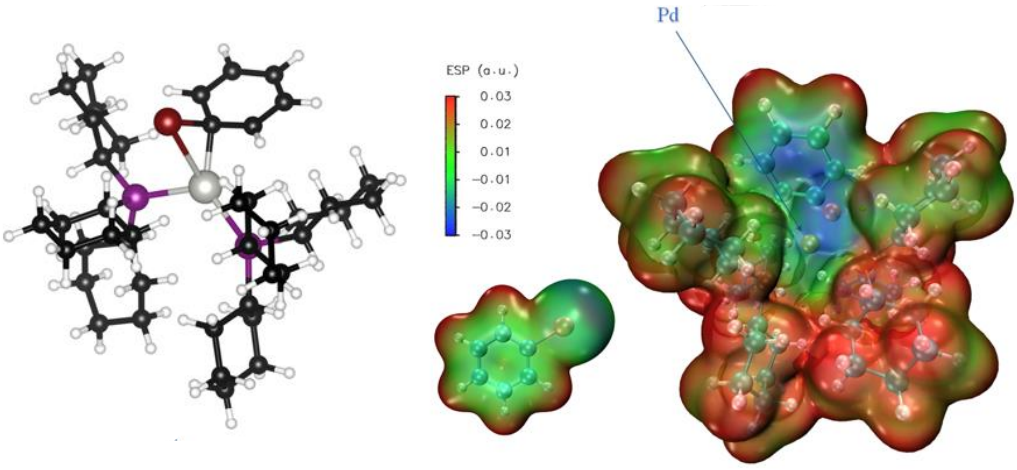


Fig. B52. Calculated structure of the oxidative addition transition state for bromobenzene, and *ESP* maps of the free substrate and the transition state. ESP_{Pd} = 13.5 kJ mol^{-1}.

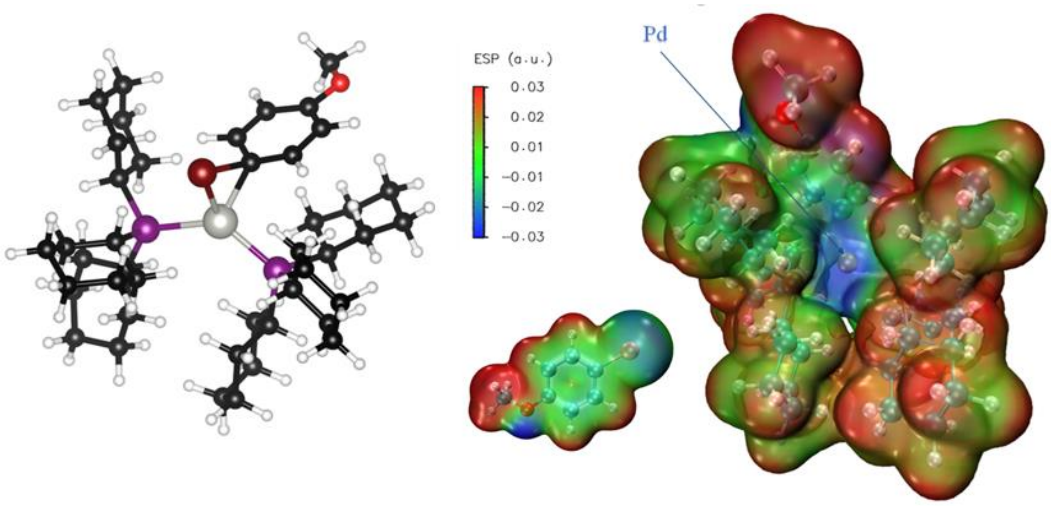


Fig. B53. Calculated structure of the oxidative addition transition state for 4-bromoanisole, and *ESP* maps of the free substrate and the transition state. ESP_{Pd} = -29.8 kJ mol^{-1}.

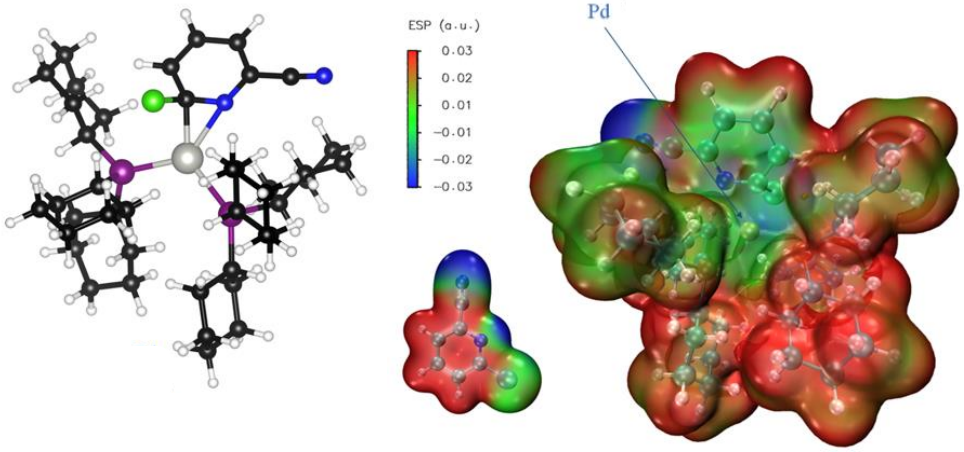


Fig. B54. Calculated structure of the oxidative addition transition state for 2-chloro-6-cyanopyridine, and *ESP* maps of the free substrate and the transition state. ESP_{Pd} = 42.2 kJ mol^{-1}.

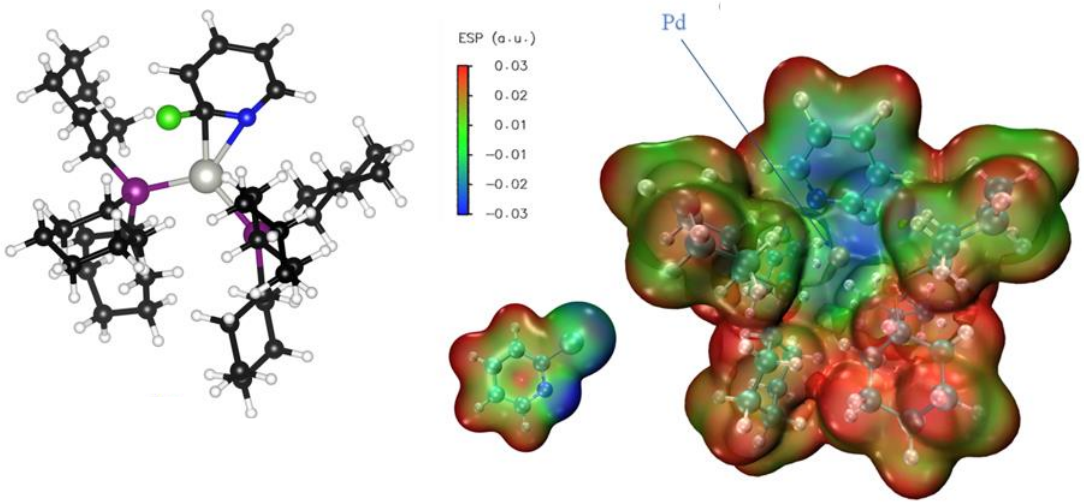


Fig. B55. Calculated structure of the oxidative addition transition state for 2-chloropyridine, and *ESP* maps of the free substrate and the transition state. ESP_{Pd} = 7.7 kJ mol^{-1}.

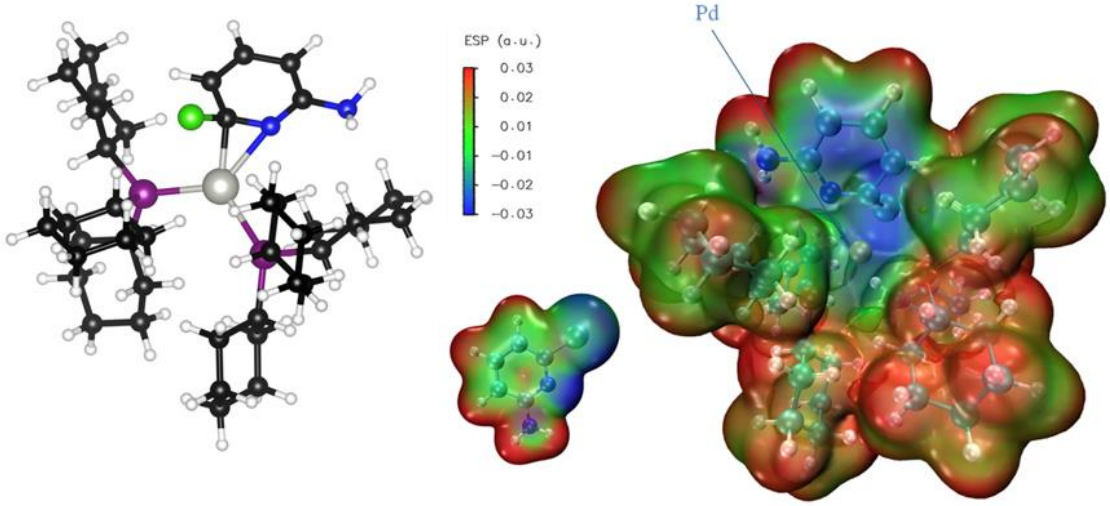


Fig. B56. Calculated structure of the oxidative addition transition state for 2-chloro-6-aminopyridine, and *ESP* maps of the free substrate and the transition state. ESP_{Pd} = 5.8 kJ mol^{-1}.

Analyzing how ESP_{Pd} relates to experimental $\Delta G^{\ddagger}_{OA}$ revealed the linear correlation from Fig. 3.3A (main text). A fully labelled version of this chart is shown in Fig. B57.

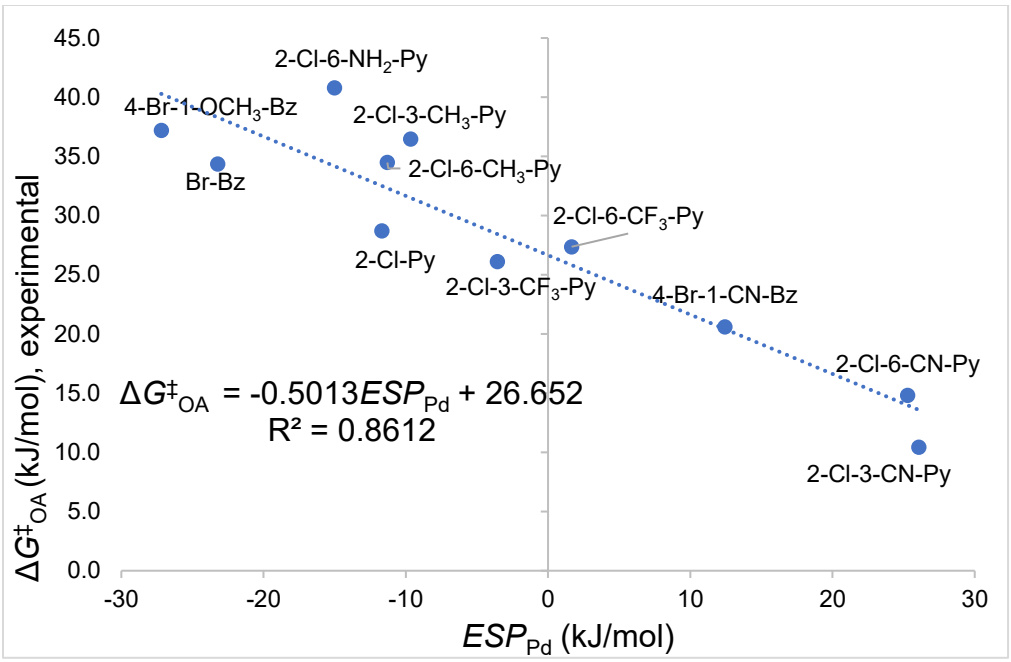
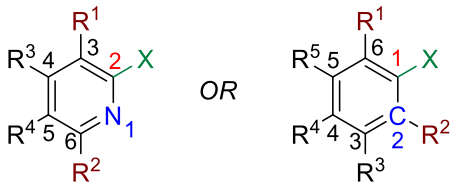


Fig. B57. Linear correlation between observed $\Delta G^{\ddagger}_{OA}$ and ESP_{Pd} for 11 π-complex intermediates with the indicated substrates bound to Pd.

Construction of the Multivariate Linear Regression Model

The selection of the molecular descriptors used to correlate $\Delta G^{\ddagger}_{OA}$ with substrate structure was guided by the mechanistic features of oxidative addition elucidated previously, computational calculations on the π-complex intermediates and transition states, and iterative refinement of the included descriptors based on our experimental observations. As summarized in Fig. B58, there are four descriptors that lead to accurate predictions of relative oxidative addition reactivity: average molecular electrostatic potential (ESP_1) at the reactive carbon, average molecular electrostatic potential (ESP_2) at the adjacent atom to the reactive center, sum of the A-values for substituents R_1 and R_2, and the intrinsic bond strength index (*IBSI*) of the C–X (X = Cl, Br, O) bond. To make interpretation of the model easier, we multiplied the unitless *IBSI* values (~0.23-0.40) by 1000 to give values with a similar order of magnitude to bond dissociation energies (~300-450 kJ mol^{-1}). In this way, the coefficients in the multivariate linear equation more accurately represent the relative contribution of each term to the overall $\Delta G^{\ddagger}_{OA}$.



1) ESP[1] at the reactive site carbon atom
2) ESP[2] at the neighbour atom of the reactive site N_1
3) Steric effects - sum of A values for R^1 and R^2
4) Intrinsic bond strength index C-X for C-Cl, C-Br, or C-O bond

Fig. B58. Substrate molecular descriptors to construct the oxidative addition prediction model.

The initial multivariate linear regression model constructed from these descriptors and the $\Delta G^{\ddagger}_{OA}$ of 70 (hetero)aryl halides in THF achieved excellent linear correlation with a squared correlation coefficient (R^2) of 0.92, and a mean absolute error (MAE) of 2.50 kJ mol^{-1}, and has an expected random distribution of residuals and no significant outliers as summarized in Figure B59.

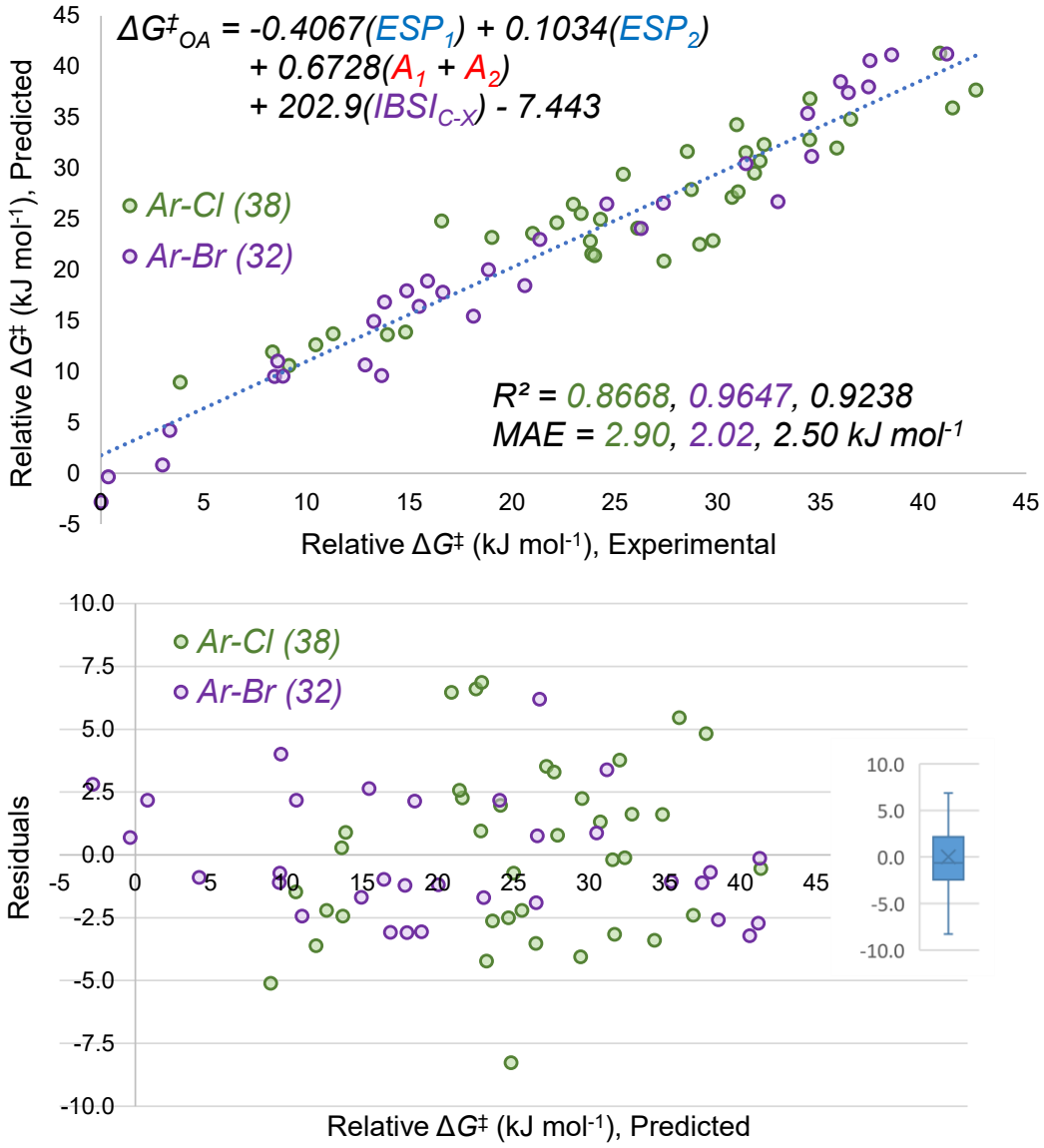


Fig. B59. Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 70 substrates, including experimental versus predicted plot (top) and predicted versus residuals plot with corresponding box plot inset (bottom).

To validate the inclusion of these descriptors and evaluate simpler potential models, we have compared the four-descriptor model to a series of alternatives (Table B8).

First, as stated in the main text and above, we discovered that the two steric A-values have approximately equal contributions when linear regression fitting is done with A_1 and A_2 as separate descriptors. This alternate, five descriptor model is shown in entry 2. An R^2 of 0.92 and a MAE of 2.46 kJ mol^{-1} were obtained from the five-variable linear regression model, which are almost identical to those obtained from the original four-variable regression model (entry 1). The two models give the same level of prediction accuracy, and thus the steric A-values of R_1 and R_2 are treated as one summed value for simplification.

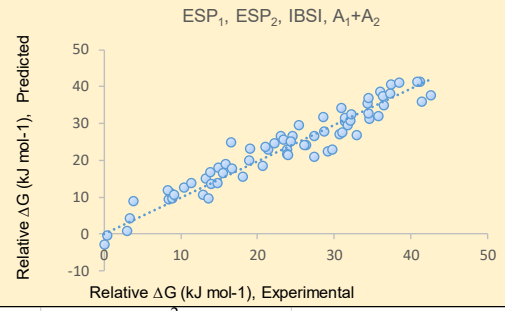
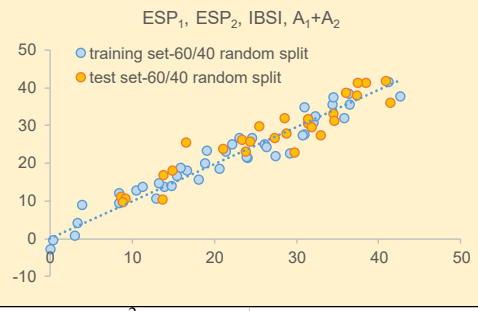
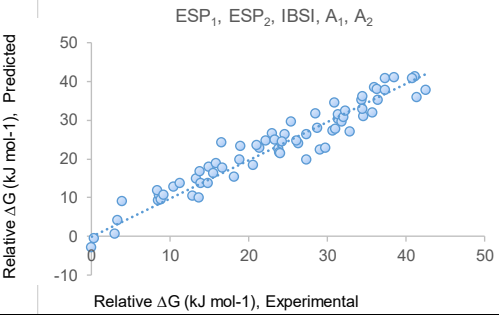
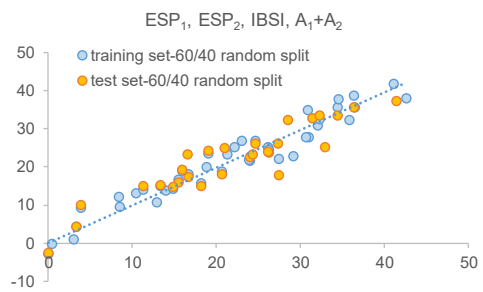
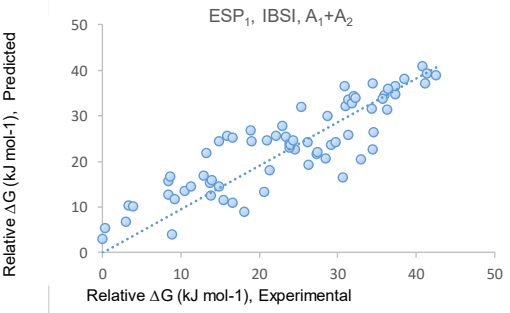
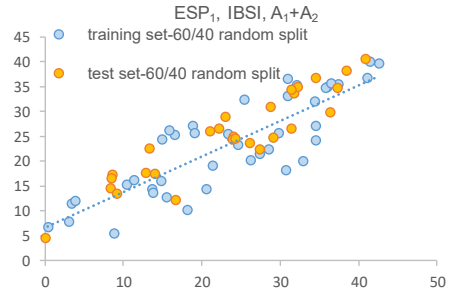
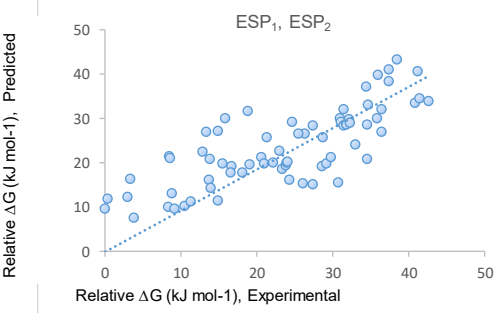
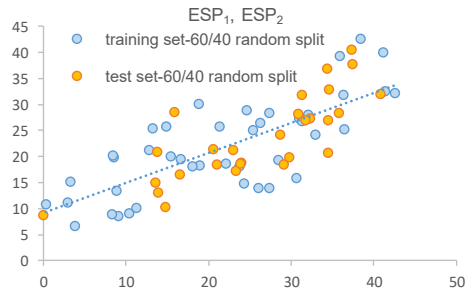
The inclusion of ESP_2 (corresponding to the adjacent atom) in the model was explored based on the nature of the palladium oxidative addition mechanism transition state, and the Hammett analysis on the 4-Z-2-chloropyridines (Figs. B35-B36). To confirm that both ESP_1 and ESP_2 are necessary to predict $\Delta G^{\ddagger}_{OA}$, a linear regression model was constructed with only 3 descriptors (ESP_1, A_1+A_2, *IBSI*). This model, shown in entry 3, gives poorer performance, with one 60/40 training/test split giving a Q^2 of 0.83 and a MAE of 3.91 kJ mol^{-1} for the test data.

The inclusion of the *IBSI* descriptor is necessary to create a single model that incorporates both Ar–Cl and Ar–Br substrates. Linear regression models that do not include this descriptor (entries 4-6) perform poorly with respect to linear correlation and predictive ability, with test set MAE values of 4.44-4.97 kJ mol^{-1}.

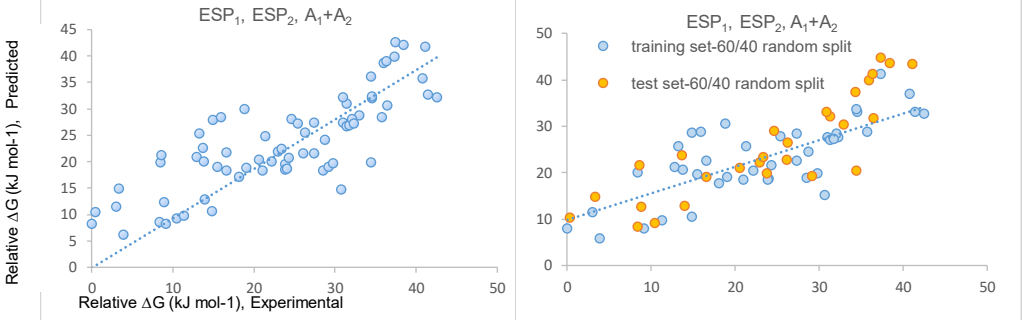
Finally, we evaluated a model that does not include steric effects (entry 7). While the descriptive statistics of this model appear to be fairly good, this is because many of the substrates in our library do not include substituents at R_1 or R_2. The all-data plot in entry 7 shows that many of the substrates with non-zero A_1+A_2 values (dark orange points) are significant systematic outliers, where the model underestimates $\Delta G^{\ddagger}_{OA}$.

Overall, the alternatives presented in Table B8 demonstrate that all of the four descriptors are necessary and sufficient to predict $\Delta G^{\ddagger}_{OA}$ for this diverse set of substrates.

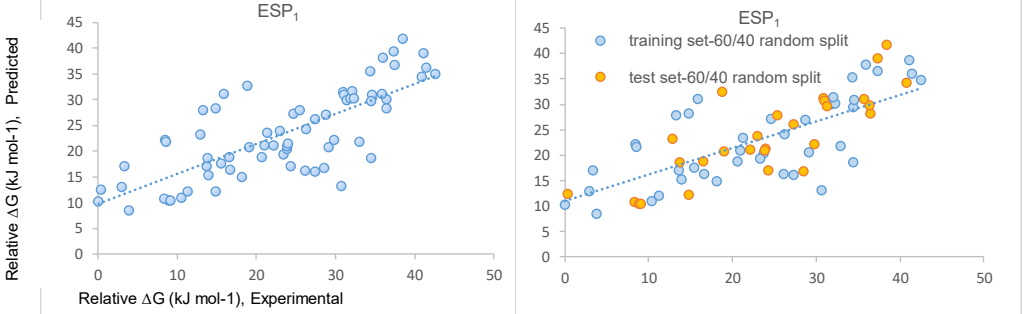
Table B8. Comparison of model performance for different combinations of molecular descriptors.

Comparison of Model Performance				
Variables	**Multivariate linear regression model performance**		**60/40 random split performance for test set**	
	R^2	MAE (KJ/mol)	Q^2	MAE (KJ/mol)
ESP_1, ESP_2, IBSI, A_1+A_2	0.92	2.50	0.91	2.77
				
ESP_1, ESP_2 IBSI, A_1, A_2	R^2	MAE (KJ/mol)	Q^2	MAE (KJ/mol)
	0.92	2.46	0.9	2.9
				
ESP_1, IBSI, A_1+A_2	R^2	MAE (KJ/mol)	Q^2	MAE (KJ/mol)
	0.76	4.36	0.83	3.91
				
ESP_1, ESP_2	R^2	MAE (KJ/mol)	Q^2	MAE (KJ/mol)
	0.60	5.40	0.71	4.97
				

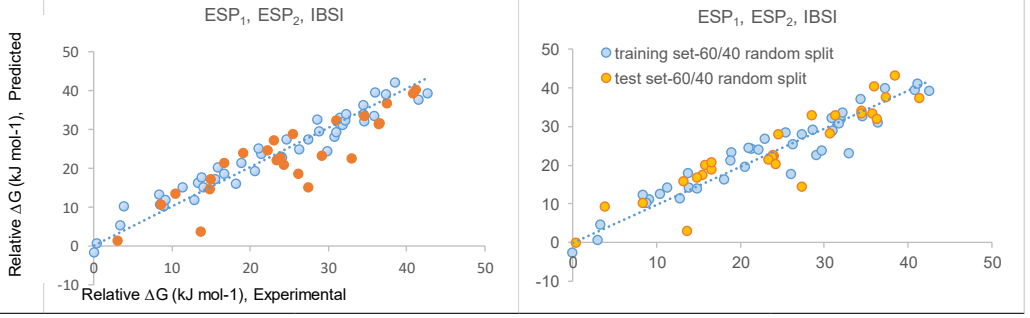
	R^2	MAE (KJ/mol)	Q^2	MAE (KJ/mol)
ESP_1, ESP_2, A_1+A_2	0.65	5.15	0.67	4.55



	R^2	MAE (KJ/mol)	Q^2	MAE (KJ/mol)
ESP_1	0.58	5.27	0.74	4.44



	R^2	MAE (KJ/mol)	Q^2	MAE (KJ/mol)
ESP_1, ESP_2, IBSI	0.89	2.86	0.85	3.28



Cross Validation and Out-of-Sample Prediction

To further evaluate the linear regression model, we performed cross-validation by doing five random 60/40 training/test data splits. Excellent linear correlation was achieved between the observed and predicted $\Delta G^{\ddagger}_{OA}$, as indicated by the range of R^2 from 0.90 to 0.95 for training set, Q^2 from 0.88 to 0.91, and MAE from 2.63 kJ mol^{-1} to 3.00 kJ mol^{-1} for test set. A plot for one such 60/40 split is shown in Fig. 3.2D, main text, and the plots for other four divisions are shown in Figures B60 to B63. The good agreement between the observed and predicted $\Delta G^{\ddagger}_{OA}$ obtained from this random split cross-validation has indicated that our multivariate linear regression model is appropriately fitted, without overfitting issues.

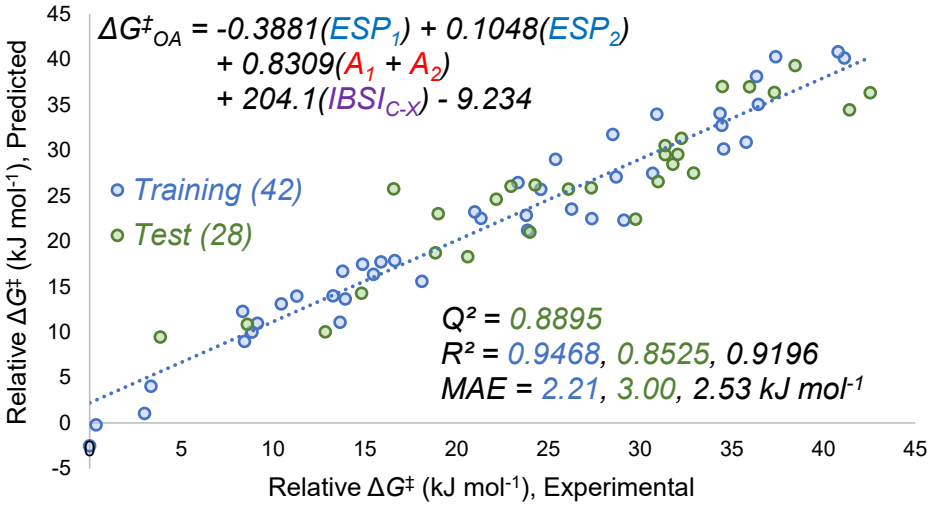


Fig. B60. Multivariate linear regression model from one of the five 60/40 random split divisions (2/5).

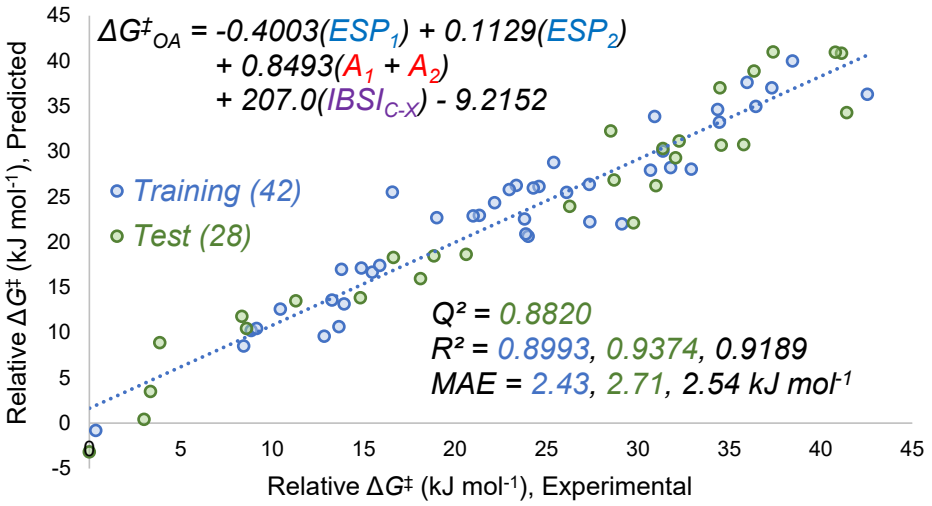


Fig. B61. Multivariate linear regression model from one of the five 60/40 random split divisions (3/5).

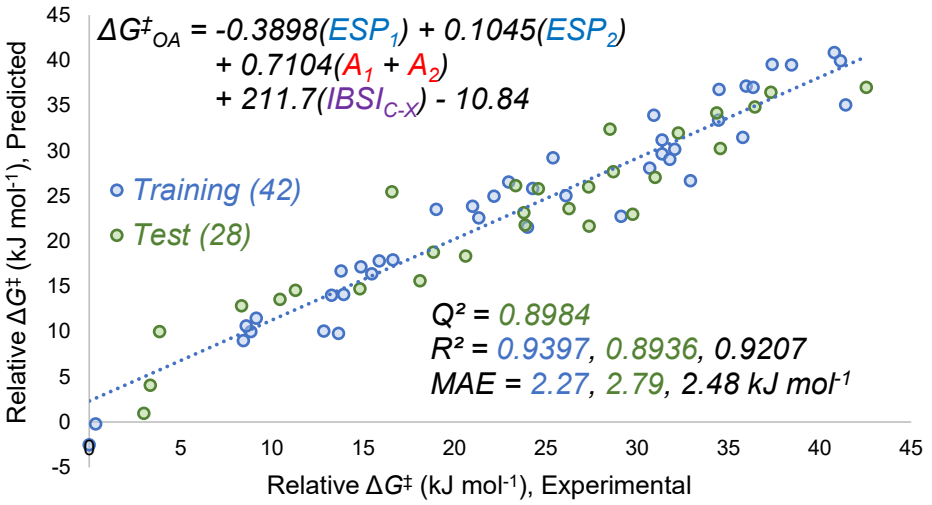


Fig. B62. Multivariate linear regression model from one of the five 60/40 random split divisions (4/5).

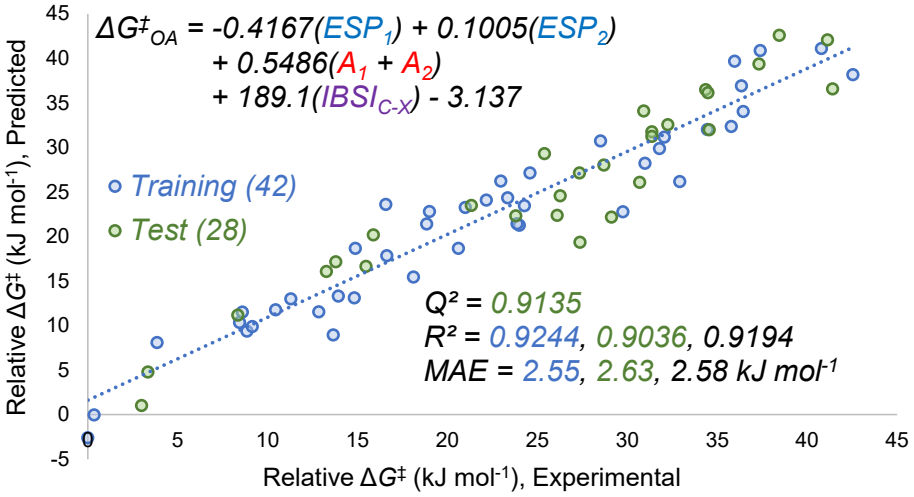


Fig. B63. Multivariate linear regression model from one of the five 60/40 random split divisions (5/5).

The model performance was further evaluated by out-of-sample predictions. To test if the model can derive reliable prediction for molecules with a variety of structural features, we split the data set into a training set containing only the 2-halo pyridines, and a test set containing all other substrates. The model has achieved an excellent performance with a R^2 of 0.92 for the training set; a MAE of 2.68 kJ mol^{-1} and a Q^2 of 0.92 for the test set (Fig. B64).

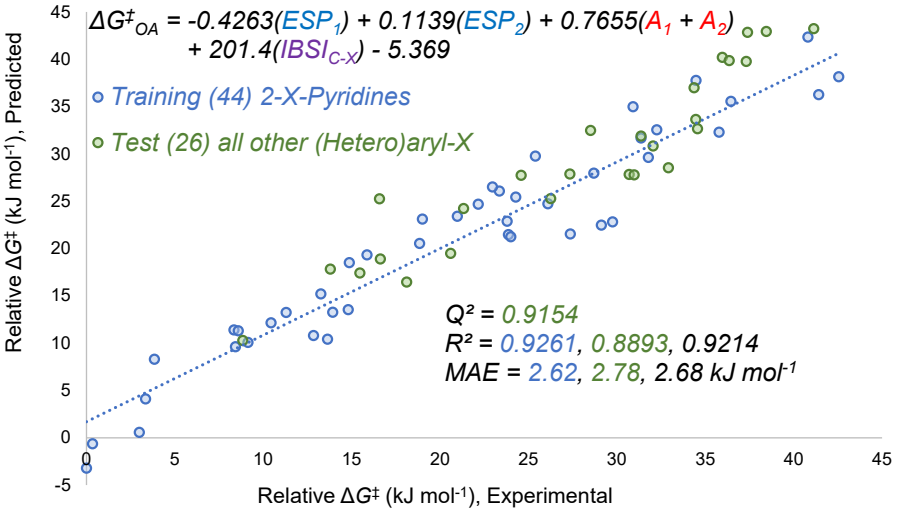


Fig. B64. Multivariate linear regression model obtained using a training set containing all 2-halopyridines, and a test set containing all other substrates.

We performed another sample split and trained the model with just the mono-substituted (hetero)aryl halides, reserving multisubstituted (hetero)aryl halides as a test set (Fig. B65). This out-of-sample prediction has also showed an excellent performance with a R^2 of 0.92 for the training set; a MAE of 2.49 kJ mol^{-1} and a Q^2 of 0.93 for the test set.

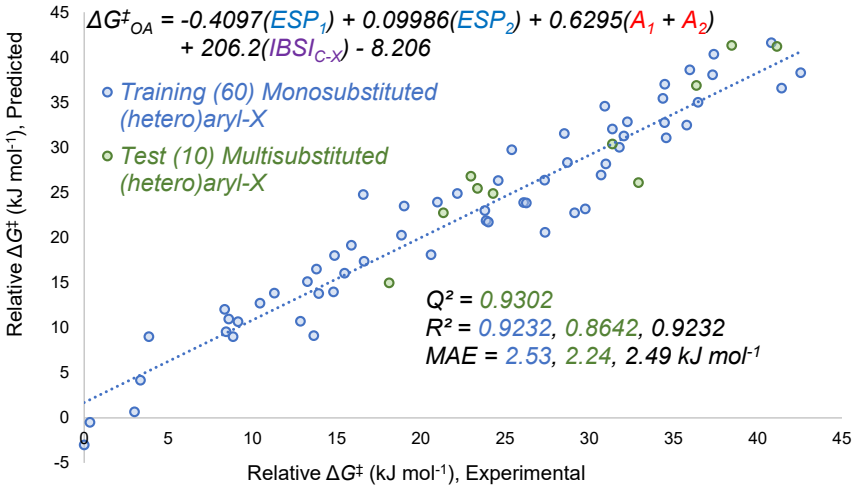


Fig. B65. Multivariate linear regression model obtained using a training set containing all the monosubstituted (hetero)aryl halides, and a test set containing all the multisubstituted (hetero)aryl halides.

We also tested the model performance in predicting reactivity that falls outside the range of that in the training data set. We sorted the substrates by their observed $\Delta G^{\ddagger}_{OA}$ and performed out-of-sample predictions with an 80/20 (fast/slow) split (Fig. B66) and an 80/20 (slow/fast) split (Fig. B67). Both models have good predictive ability on the test set (MAE=2.90 kJ mol^{-1}, Q^2=0.87 for 80/20 (fast/slow) split; MAE=2.56 kJ mol^{-1}, Q^2=0.86 for 80/20 (slow/fast) split).

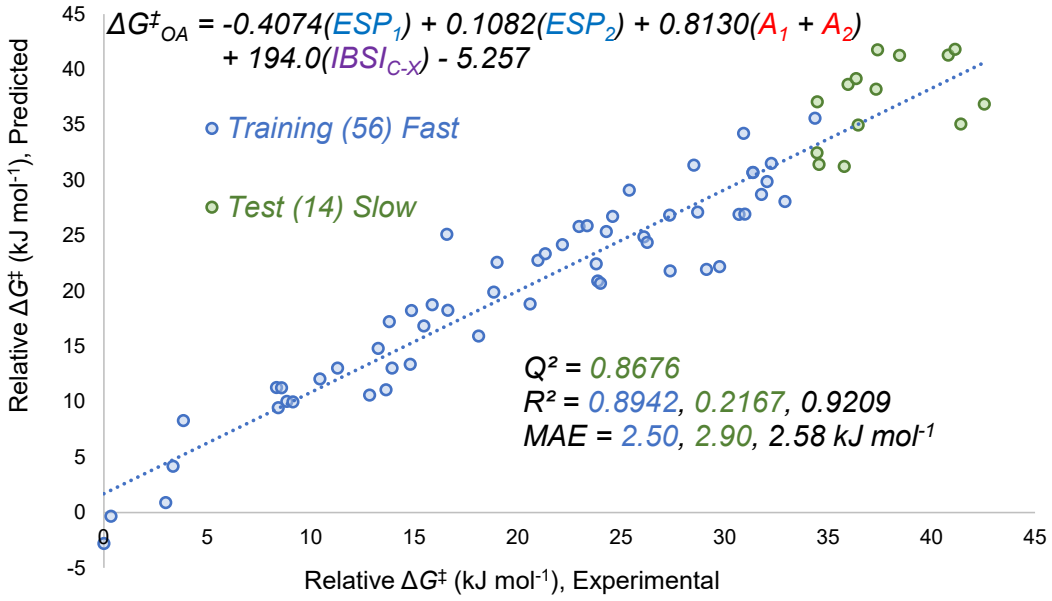


Fig. B66. Multivariate linear regression model from the 80/20 (fast/slow) split.

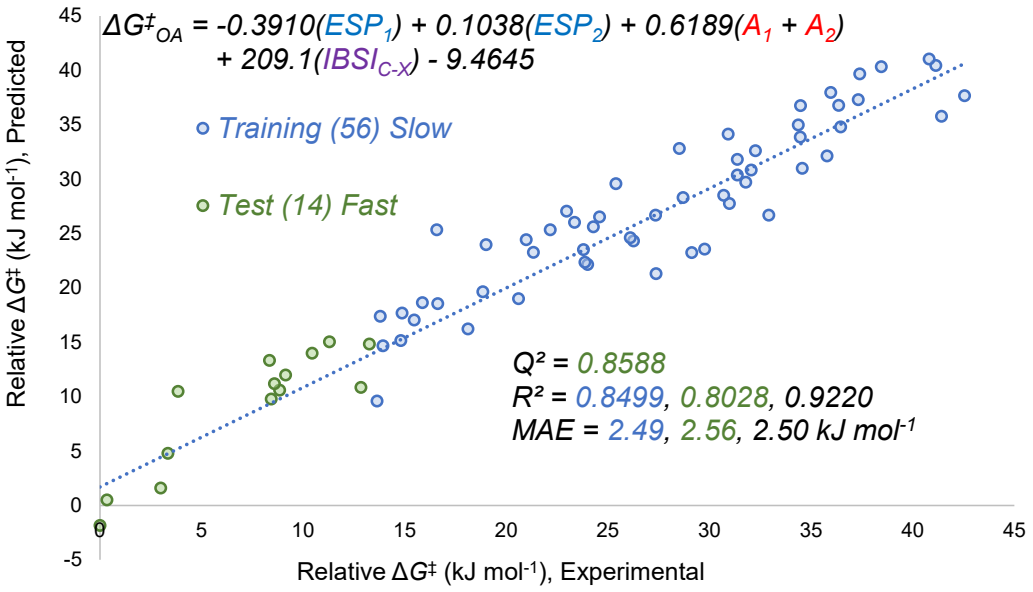
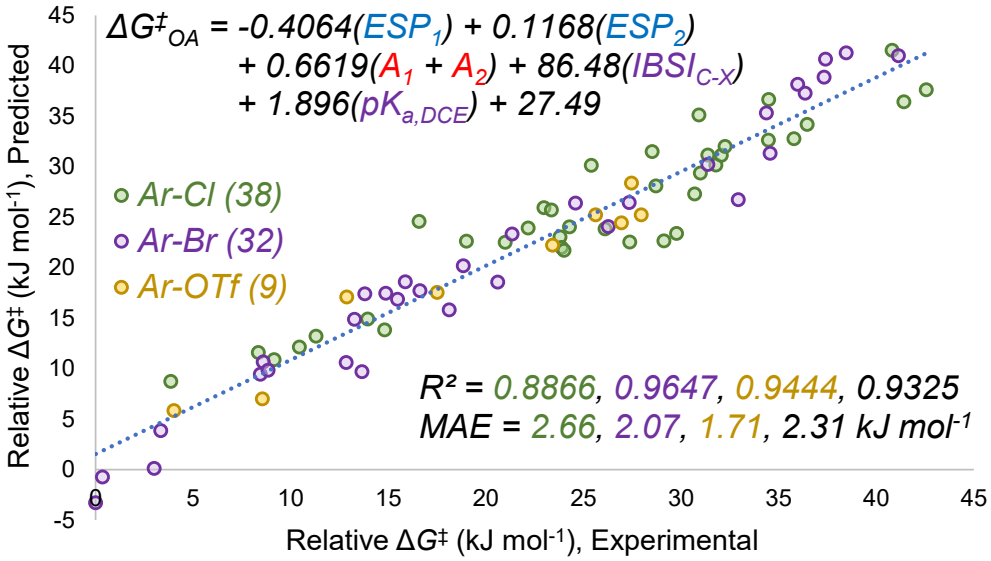


Fig. B67. Multivariate linear regression model from the 80/20 (slow/fast) split.

Including Ar-OTf substrates into the predictive model

The inclusion of pK_a of the conjugate acid of the leaving group creates a unified predictive model that incorporates Ar-Cl, Ar-Br and Ar-OTf. The multivariate linear regression model constructed from those five descriptors (ESP_1, ESP_2, A_1+A_2, IBSI and pK_a) and the $\Delta G^{\ddagger}_{OA}$ of 79 substrates achieved excellent linear correlation with an R^2 of 0.93, and an MAE of 2.31 kJ mol^{-1} and has an expected random distribution of residuals and only one significant outlier as summarized in Figure B68.



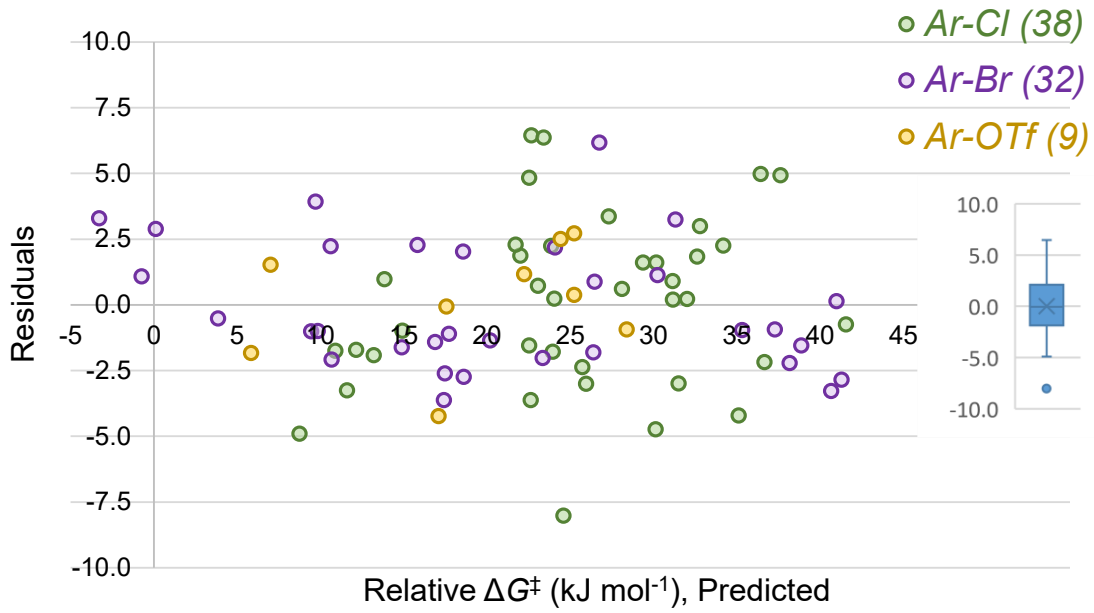


Fig. B68. Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ in THF for 79 substrates, including experimental versus predicted plot (top) and predicted versus residuals plot with corresponding box plot inset (bottom).

The model performance was evaluated by cross-validation with five random 60/40 training/test data splits. Excellent linear correlation was achieved between the observed and predicted $\Delta G^{\ddagger}_{OA}$, as indicated by the range of R^2 from 0.91 to 0.95 for training set, Q^2 from 0.90 to 0.93, and MAE from 2.18 kJ mol^{-1} to 2.91 kJ mol^{-1} for the test set (Figure B69 to Figure B73).

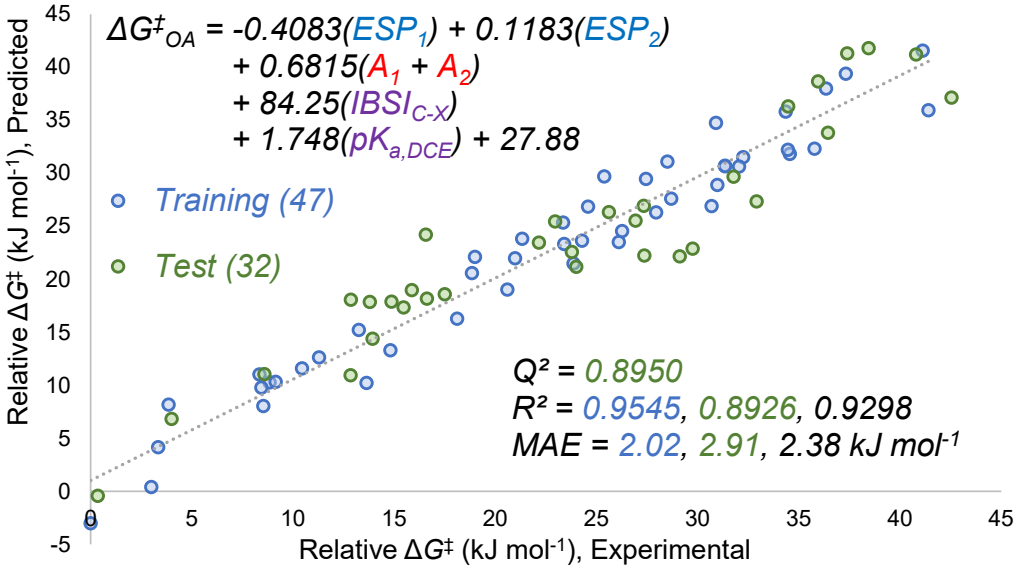


Fig. B69. Multivariate linear regression model from one of the five 60/40 random split divisions (1/5).

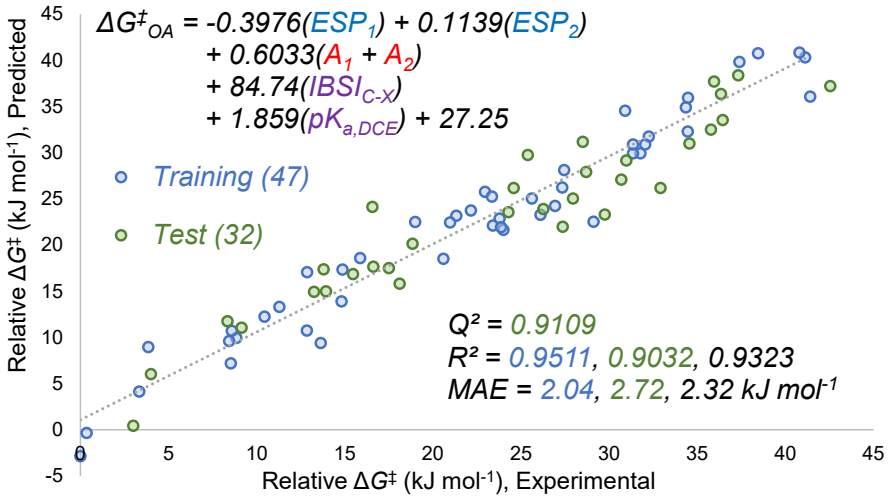


Fig. B70. Multivariate linear regression model from one of the five 60/40 random split divisions (2/5).

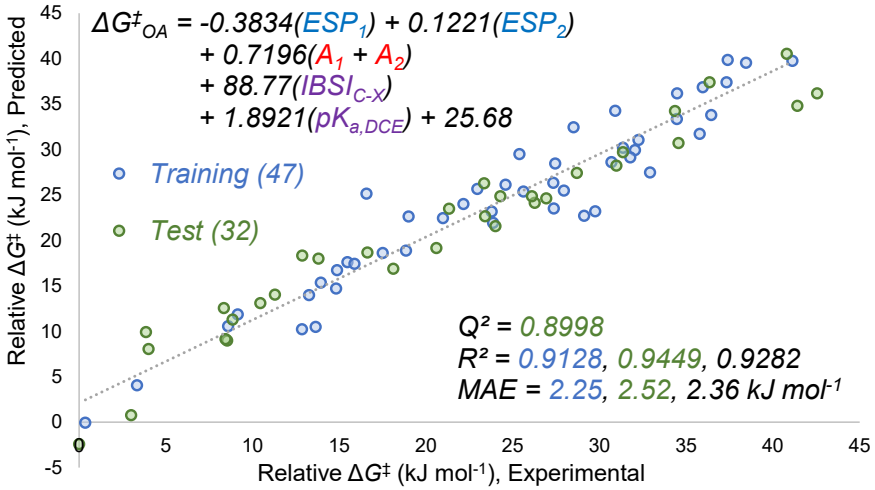


Fig. B71. Multivariate linear regression model from one of the five 60/40 random split divisions (3/5).

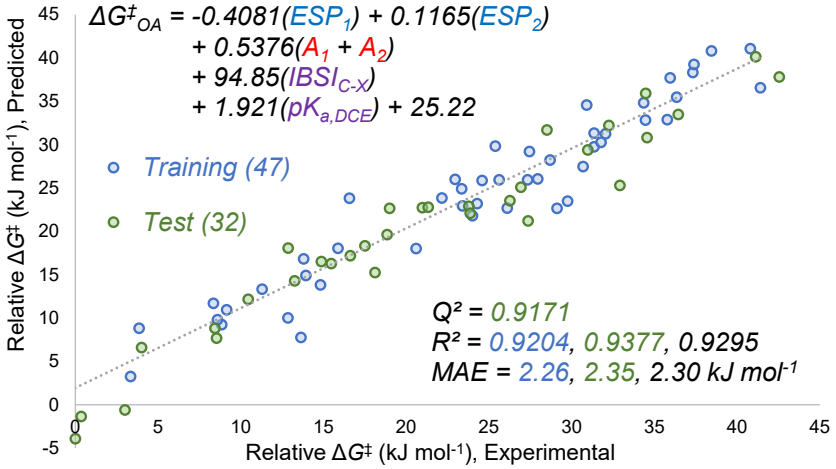


Fig. B72. Multivariate linear regression model from one of the five 60/40 random split divisions (4/5)

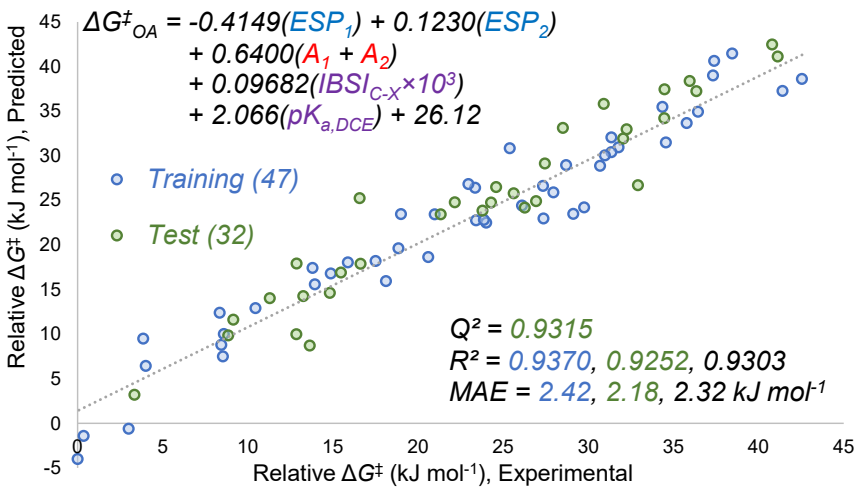


Fig. B73. Multivariate linear regression model from one of the five 60/40 random split divisions (5/5).

Sonogashira Initial Rate Prediction Modelling

To build the predictive model for Sonogashira initial rates, we calculated the key molecular descriptors and the predicted $\Delta G^{\ddagger}_{OA}$ for each of the 29 aryl bromides in substrate sets #1 and #2 (Table B9). Linear correlations of the predicted $\Delta G^{\ddagger}_{OA}$ with ln *k* for each Sonogashira reaction using a particular phosphine are shown in Figs. B74 and B75. For the 10 phosphines in Fig. B74, only substrate set #1 is used, while the 7 phosphines in Fig. B75 include data from both substrate sets. Good to excellent linear correlations are observed (R^2 = 0.79-0.92) across the data for all 17 phosphines; only one substrate appears to be a significant outlier: 1-bromo-2,4,6-triisopropylbenzene, where the Sonogashira reaction rate is much lower than expected based on the predicted $\Delta G^{\ddagger}_{OA}$.

Table B9 –Predicted $\Delta G^{\ddagger}_{OA}$ and molecular descriptors for the Sonogashira substrates (Lu-PhD-thesis_Ch.3 SI tables.xlsx).

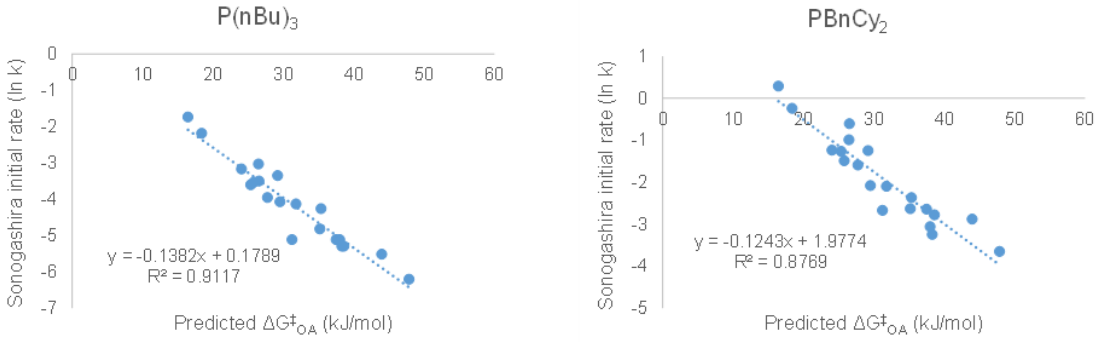


Fig. B74. Univariate correlations between ln k for Sonogashira reactions and the predicted $\Delta G^{\ddagger}_{OA}$ for each substrate across 10 ligands used only with substrate set #1.

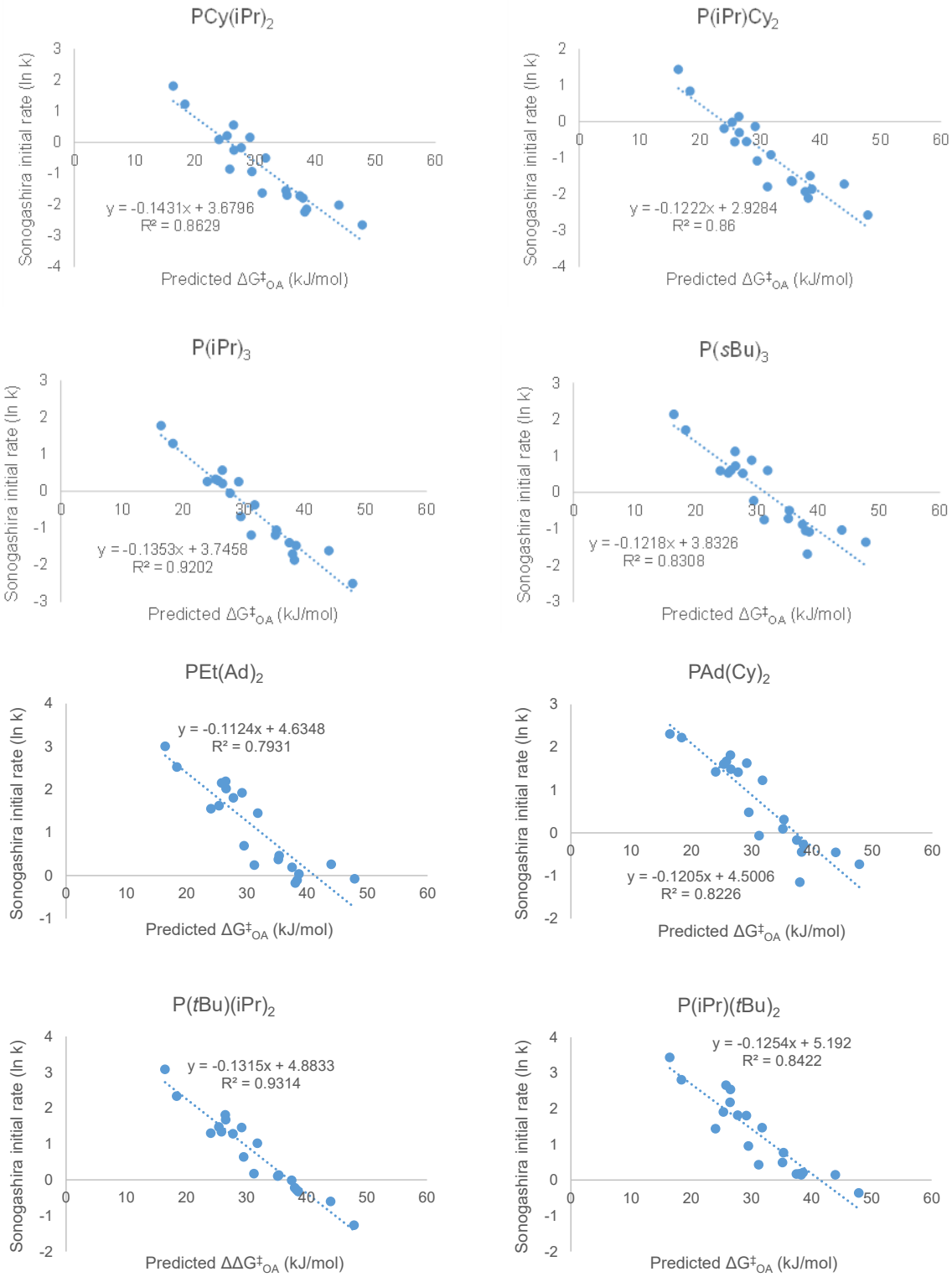


Fig. B74 continued. Linear correlations between ln k for Sonogashira reactions and predicted $\Delta G^{\ddagger}_{OA}$ for each substrate across 10 ligands used only with substrate set #1.

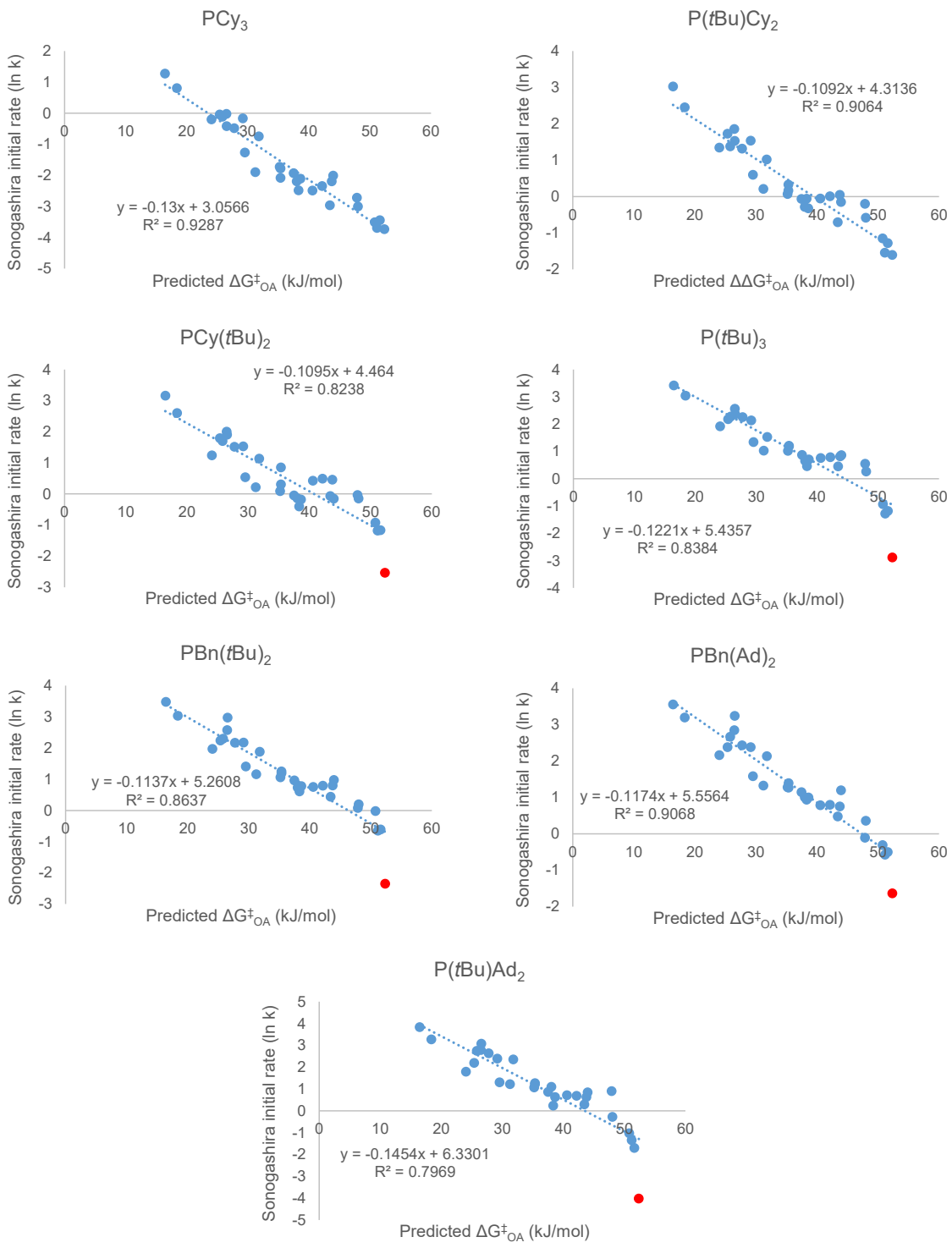


Fig. B75. Linear correlations between ln *k* for Sonogashira reactions and predicted $\Delta G^{\ddagger}_{OA}$ across 7 ligands used with substrate sets #1 & #2. Red points are outliers (1-bromo-2,4,6-triisopropylbenzene), which are included in the regression analysis.

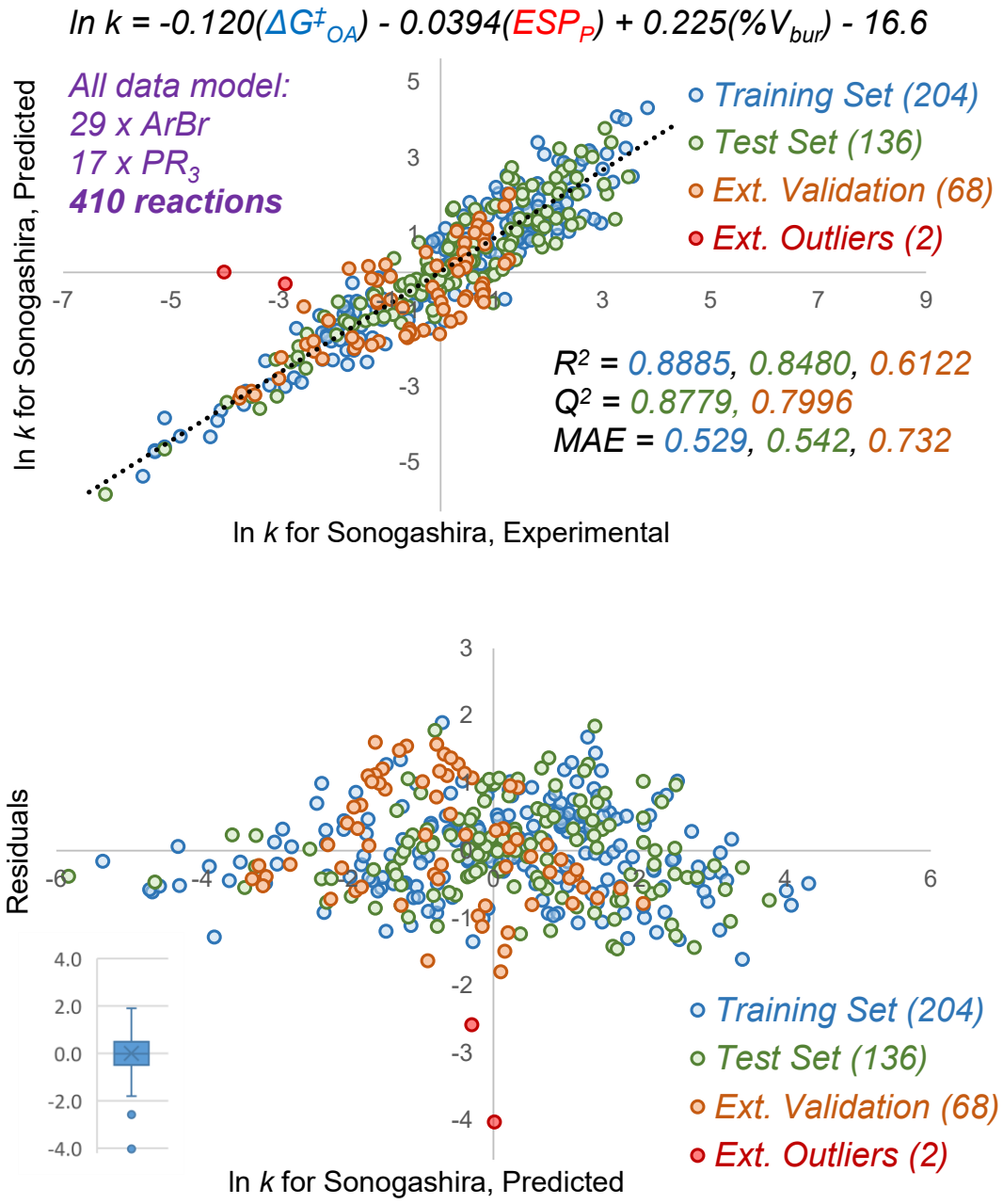


Fig. B76. Unified Sonogashira coupling rate predictive model constructed using the electronic and steric descriptors of free phosphines and predicted $\Delta G^{\ddagger}_{OA}$ for the entire set of 410 reactions, including plot of experimental versus predicted ln k values (top) and plot of predicted versus residuals with corresponding box plot inset (bottom), showing two outliers (red data points in main plots).

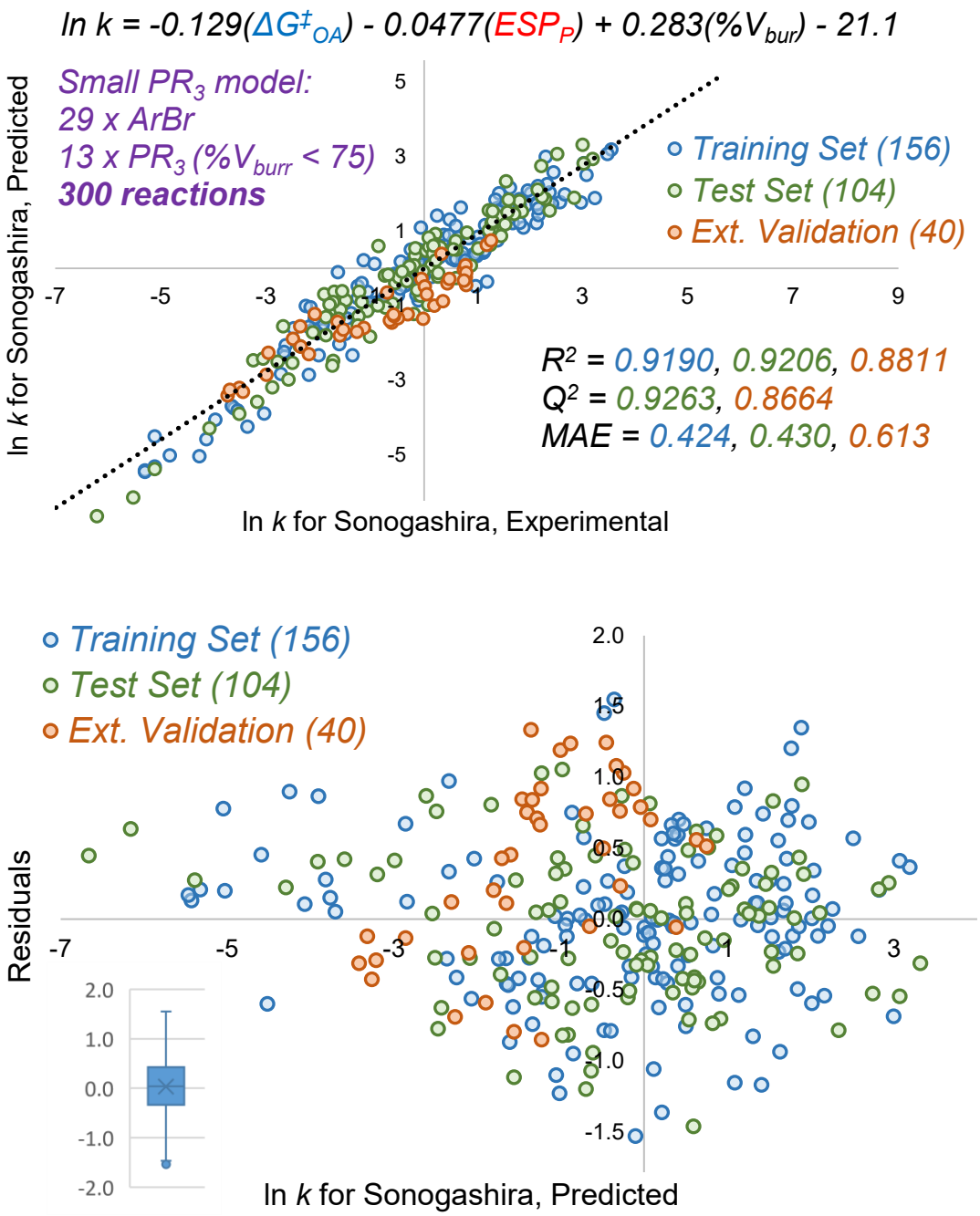


Fig. B77. Unified Sonogashira coupling rate predictive model constructed using the electronic and steric descriptors of free phosphines and predicted $\Delta G^{\ddagger}_{OA}$ for the set of phosphines with $\%V_{bur} < 75$ (300 reactions), including plot of experimental versus predicted ln k values (top) and plot of predicted versus residuals with corresponding box plot inset (bottom), showing no significant outliers.

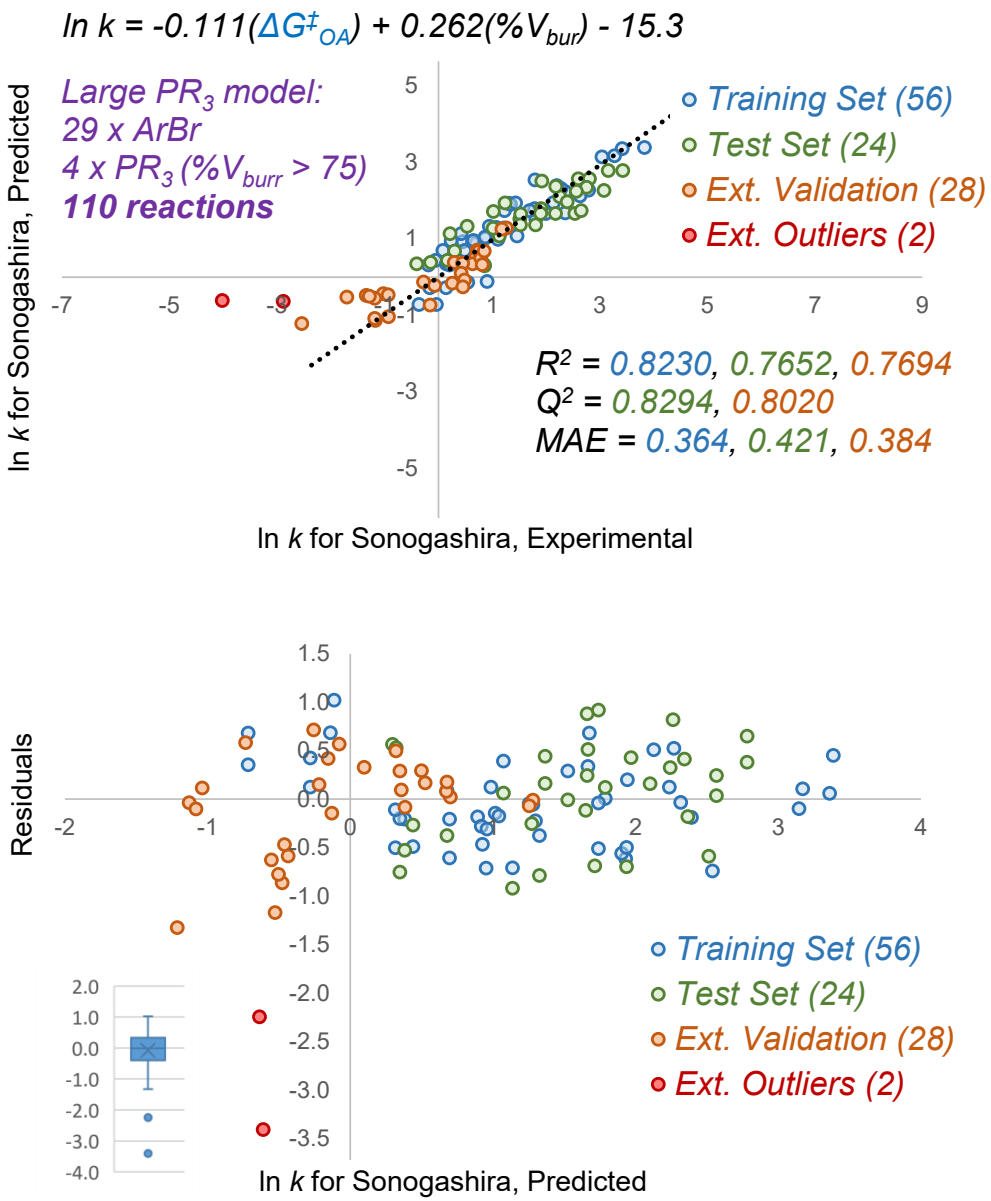


Fig. B78. Unified Sonogashira coupling rate predictive model constructed using the electronic and steric descriptors of free phosphines and predicted $\Delta G^{\ddagger}_{OA}$ for the set of phosphines with $\%V_{bur} > 75$ (110 reactions), including plot of experimental versus predicted ln k values (top) and plot of predicted versus residuals with corresponding box plot inset (bottom), showing two outliers (red data points in main plots).

To build the unified model shown in Fig. 4 in the main text, we evaluated several molecular descriptors related to the 17 phosphine ligands. Molecular ESP values and $\%V_{bur}$ were calculated for the corresponding PdL_2 and PdL complexes, as well as the "free" phosphines (Table B6).

To evaluate these different catalyst descriptors, we constructed similar multivariate linear regression models using a combination of catalyst ESP and $\%V_{bur}$ with the predicted

$\Delta G^{\ddagger}_{OA}$ for each substrate. The training/test sets were generated by a random 60/40 split of the substrate set #1 data, and the substrate #2 data was retained as an external test set. The models corresponding to catalyst descriptors for PdL_2 and PdL are shown in Figs. B79 and B80, respectively. Note that for PdL, the $\%V_{bur}$ term has a very small coefficient when included in the regression analysis; therefore, we opted to exclude it from the model shown in Fig. B77. Notably, both alternative catalyst descriptors underperform the free phosphine descriptors (Figs. 3.4C-3.4E, main text) with respect to linear correlation and predictive ability with this data set.

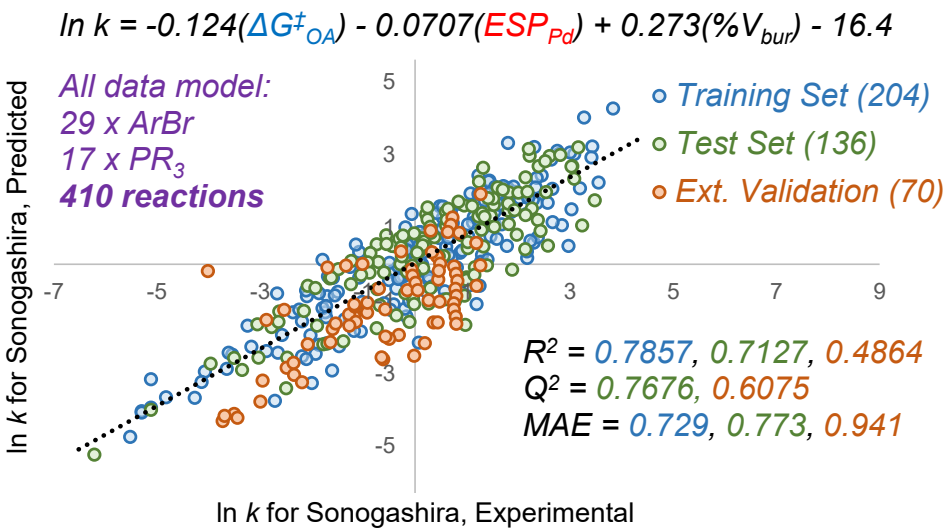


Fig. B79. Unified Sonogashira coupling rate predictive model constructed using the electronic and steric descriptors of PdL_2 and predicted $\Delta G^{\ddagger}_{OA}$ for the entire set of 410 reactions.

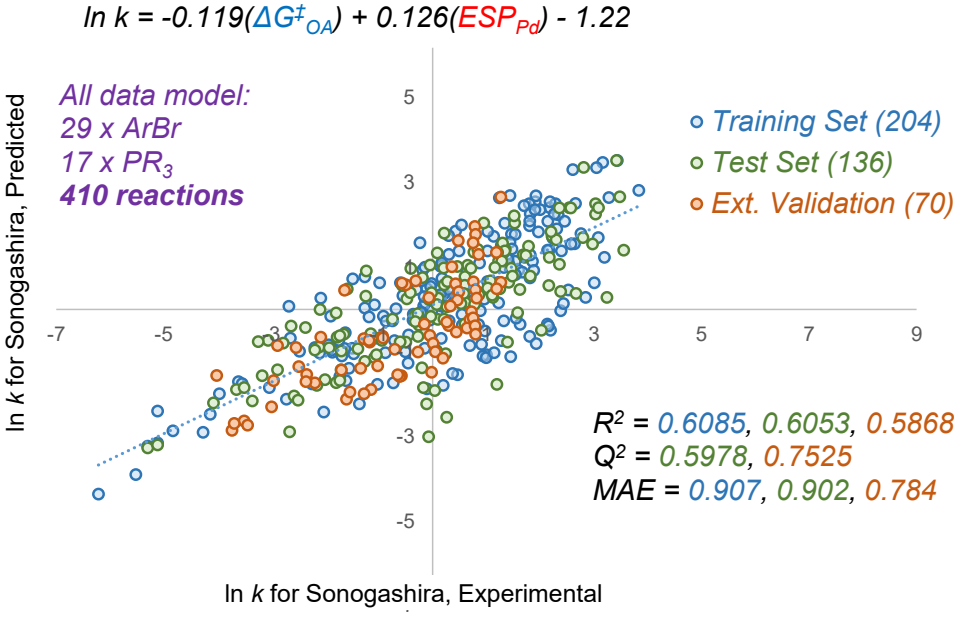


Fig. B80. Unified Sonogashira coupling rate predictive model constructed using the electronic and steric descriptors of PdL and predicted $\Delta G^{\ddagger}_{OA}$ for the entire set of 410 reactions.

Site Selectivity Predictions

Molecular descriptors for each of the substrates shown in Table B10 were obtained following the procedure used for the oxidative addition substrates. The predicted $\Delta G^{\ddagger}_{OA}$ for each C–X (X = Cl, Br) site was calculated using the equation from Fig. 3.2C.

Table B10 – Molecular descriptors and predicted $\Delta G^{\ddagger}_{OA}$ for multihalogenated heterocycles. Purple sphere indicates observed major site of cross-coupling. The original experimental data refers to References 21–30 (Lu-PhD-thesis_Ch.3 SI tables.xlsx).

Experimental and computational data for the additional substrates in section 3.4

For the substrates studied in section 3.4: 10 aryl chlorides, 7 (hetero)aryl iodide, 2-pyridyl triflate and 2-pyridyl tosylate (19 in total), the experimental data for competition reactions and measured relative rates, and the molecular descriptors refer to the electronically available Tables B11 to B13.

Table B11 – ^{31}P NMR chemical shifts and the peak area ratios of the 34 competition reactions to determine the $\Delta G^{\ddagger}_{OA}$ of the 19 substrates in THF (Lu-PhD-thesis_Ch.3 SI tables.xlsx).

Table B12 – Observed $\Delta G^{\ddagger}_{OA}$ for the 19 substrates determined from competition reactions in THF. All $\Delta G^{\ddagger}_{OA}$ values are given relative to 2-bromo-5-nitropyridine (the $\Delta G^{\ddagger}_{OA}$ of this substrate is set to 0.00 kJ mol^{-1}) (Lu-PhD-thesis_Ch.3 SI tables.xlsx).

Table B13 – Calculated molecular descriptors for the 19 substrates (Lu-PhD-thesis_Ch.3 SI tables.xlsx).
** pKa of HCl, HBr and HOTf is 0.2, -4.4, -11.3, respectively.[17]

Cross Validation for the four-descriptor predictive model from Fig. 3.10B, main text.

To evaluate the four-descriptor predictive model from Fig. 3.10B in main text, we performed cross-validation by doing five random 70/30 training/test data splits. Excellent linear correlation was achieved between the observed and predicted $\Delta G^{\ddagger}_{OA}$, as indicated by the range of R^2 from 0.95 to 0.96 for training set, Q^2 from 0.89 to 0.97, and MAE from 1.71 to 2.92 kJ mol^{-1} for test set. A plot for one such 70/30 split is shown in Fig. 3.10D in the main text, and the plots for other four divisions are shown in Figure B81. The good agreement between the observed and predicted $\Delta G^{\ddagger}_{OA}$ obtained from this random split cross-validation has indicated that this four-descriptor multivariate linear regression model is appropriately fitted, without overfitting issues.

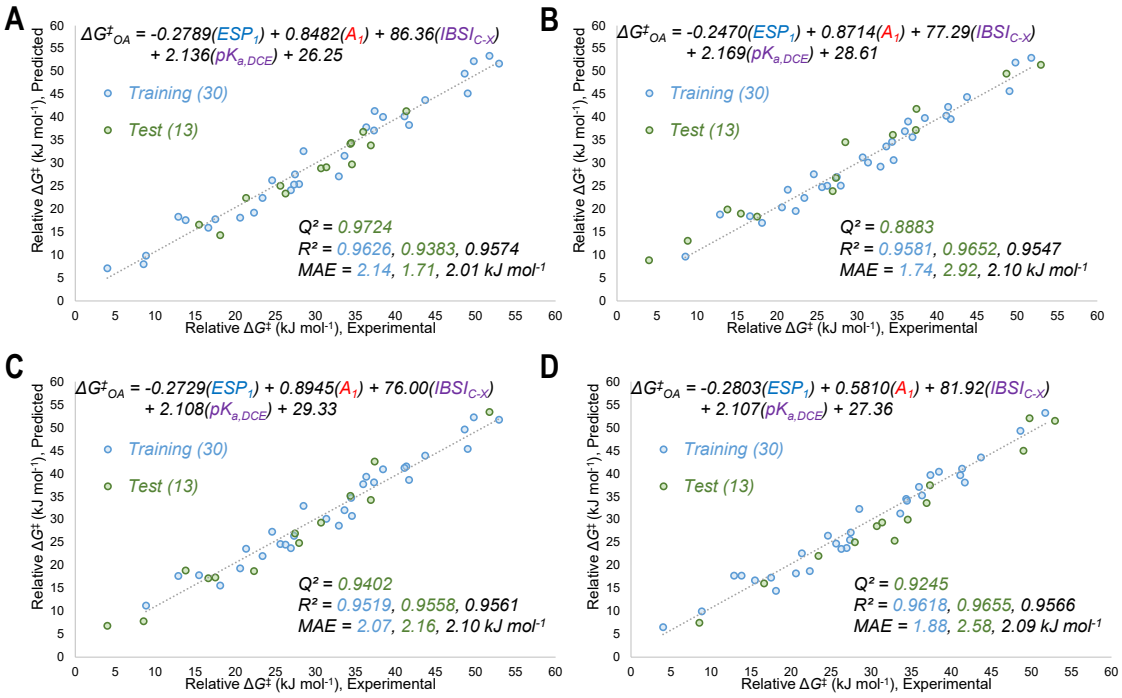


Fig. B81. (A) – (D) Multivariate linear regression model from one of the five 70/30 random split divisions (2/5) – (5/5).

Reference

(1) Sengmany, S.; Lebre, J.; Le Gall, E.; Léonel, E. Selective Mono-Amination of Dichlorodiazines. *Tetrahedron* **2015**, *71* (29), 4859–4867. https://doi.org/10.1016/j.tet.2015.05.056.

(2) Cui, L.-C.; Zhang, Z.-Q.; Lu, X.; Xiao, B.; Fu, Y. Pd-Catalyzed Cross-Coupling of 1,1-Diborylalkanes with Aryl Triflates. *RSC Adv.* **2016**, *6* (57), 51932–51935. https://doi.org/10.1039/C6RA09959A.

(3) Uchiyama, M.; Kobayashi, Y.; Furuyama, T.; Nakamura, S.; Kajihara, Y.; Miyoshi, T.; Sakamoto, T.; Kondo, Y.; Morokuma, K. Generation and Suppression of 3-/4-Functionalized Benzynes Using Zinc Ate Base (TMP−Zn−ate): New Approaches to Multisubstituted Benzenes. *J. Am. Chem. Soc.* **2008**, *130* (2), 472–480. https://doi.org/10.1021/ja071268u.

(4) Fantoni, T.; Bernardoni, S.; Mattellone, A.; Martelli, G.; Ferrazzano, L.; Cantelmi, P.; Corbisiero, D.; Tolomelli, A.; Cabri, W.; Vacondio, F.; Ferlenghi, F.; Mor, M.; Ricci, A. Palladium Catalyst Recycling for Heck-Cassar-Sonogashira Cross-Coupling Reactions in Green Solvent/Base Blend. *ChemSusChem* **2021**, *14* (12), 2591–2600. https://doi.org/10.1002/cssc.202100623.

(5) Dogga, B.; Kumar, C. S. A.; Joseph, J. T. Palladium-Catalyzed Reductive Carbonylation of (Hetero) Aryl Halides and Triflates Using Cobalt Carbonyl as CO Source. *Eur. J. Org. Chem.* **2021**, *2021* (2), 309–313. https://doi.org/10.1002/ejoc.202001328.

(6) Taeufer, T.; Pospech, J. Palladium-Catalyzed Synthesis of N,N-Dimethylanilines via Buchwald–Hartwig Amination of (Hetero)Aryl Triflates. *J. Org. Chem.* **2020**, *85* (11), 7097–7111. https://doi.org/10.1021/acs.joc.0c00491.

(7) Quesnelle, C. A.; Snieckus, V. Directed Ortho Metalation (DoM)-Linked Corriu–Kumada, Negishi, and Suzuki–Miyaura Cross-Coupling Protocols: A Comparative Study. *Synthesis* **2018**, *50* (22), 4395–4412. https://doi.org/10.1055/s-0037-1611053.

(8) Zhu, Z.; Gong, Y.; Tong, W.; Xue, W.; Gong, H. Ni-Catalyzed Cross-Electrophile Coupling of Aryl Triflates with Thiocarbonates via C–O/C–O Bond Cleavage. *Org. Lett.* **2021**, *23* (6), 2158–2163. https://doi.org/10.1021/acs.orglett.1c00313.

(9) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA Quantum Chemistry Program Package. *J. Chem. Phys.* **2020**, *152* (22), 224108. https://doi.org/10.1063/5.0004608.

(10) Pence, H. E.; Williams, A. ChemSpider: An Online Chemical Information Resource. *J. Chem. Educ.* **2010**, *87* (11), 1123–1124. https://doi.org/10.1021/ed100697w.

(11) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminformatics* **2012**, *4* (1), 17. https://doi.org/10.1186/1758-2946-4-17.

(12) Lu, T.; Chen, F. Quantitative Analysis of Molecular Surface Based on Improved Marching Tetrahedra Algorithm. *J. Mol. Graph. Model.* **2012**, *38*, 314–323. https://doi.org/10.1016/j.jmgm.2012.07.004.

(13) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33* (5), 580–592. https://doi.org/10.1002/jcc.22885.

(14) *J. A. Hirsch Topics in Stereochemistry , John Wiley & Sons, Ltd, 1967, Pp. 199–222*.

(15) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(16) Klein, J.; Khartabil, H.; Boisson, J.-C.; Contreras-García, J.; Piquemal, J.-P.; Hénon, E. New Way for Probing Bond Strength. *J. Phys. Chem. A* **2020**, *124* (9), 1850–1860. https://doi.org/10.1021/acs.jpca.9b09845.

(17) Paenurk, E.; Kaupmees, K.; Himmel, D.; Kütt, A.; Kaljurand, I.; Koppel, I. A.; Krossing, I.; Leito, I. A Unified View to Brønsted Acidity Scales: Do We Need Solvated Protons? *Chem. Sci.* **2017**, *8* (10), 6964–6973. https://doi.org/10.1039/C7SC01424D.

(18) Hansch, Corwin.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91* (2), 165–195. https://doi.org/10.1021/cr00002a004.
(19) Maes, B. U. W.; Verbeeck, S.; Verhelst, T.; Ekomié, A.; von Wolff, N.; Lefèvre, G.; Mitchell, E. A.; Jutand, A. Oxidative Addition of Haloheteroarenes to Palladium(0): Concerted versus SNAr-Type Mechanism. *Chem. – Eur. J.* **2015**, *21* (21), 7858–7865. https://doi.org/10.1002/chem.201406210.
(20) Anjali, B. A.; Suresh, C. H. Interpreting Oxidative Addition of Ph–X (X = CH3, F, Cl, and Br) to Monoligated Pd(0) Catalysts Using Molecular Electrostatic Potential. *ACS Omega* **2017**, *2* (8), 4196–4206. https://doi.org/10.1021/acsomega.7b00745.
(21) Almond-Thynne, J.; Blakemore, D. C.; Pryde, D. C.; Spivey, A. C. Site-Selective Suzuki–Miyaura Coupling of Heteroaryl Halides – Understanding the Trends for Pharmaceutically Important Classes. *Chem. Sci.,* **2017***,8, 40-62*. https://doi.org/10.1039/C6SC02118B.
(22) Yang, C.-G.; Liu, G.; Jiang, B. Preparing Functional Bis(Indole) Pyrazine by Stepwise Cross-Coupling Reactions: An Efficient Method to Construct the Skeleton of Dragmacidin D. *J. Org. Chem.* **2002**, *67* (26), 9392–9396. https://doi.org/10.1021/jo026450m.
(23) Yang, W.; Wang, Y.; Corte, J. R. Efficient Synthesis of 2-Aryl-6-Chloronicotinamides via PXPd2-Catalyzed Regioselective Suzuki Coupling. *Org. Lett.* **2003**, *5* (17), 3131–3134. https://doi.org/10.1021/ol035188g.
(24) Blaise, E.; Kümmerle, A. E.; Hammoud, H.; de Araújo-Júnior, J. X.; Bihel, F.; Bourguignon, J.-J.; Schmitt, M. Access to 4-Alkylaminopyridazine Derivatives via Nitrogen-Assisted Regioselective Pd-Catalyzed Reactions. *J. Org. Chem.* **2014**, *79* (21), 10311–10322. https://doi.org/10.1021/jo501930s.
(25) Ji, J.; Li, T.; Bunnelle, W. H. Selective Amination of Polyhalopyridines Catalyzed by a Palladium−Xantphos Complex. *Org. Lett.* **2003**, *5* (24), 4611–4614. https://doi.org/10.1021/ol0357696.
(26) Keylor, M. H.; Niemeyer, Z. L.; Sigman, M. S.; Tan, K. L. Inverting Conventional Chemoselectivity in Pd-Catalyzed Amine Arylations with Multiply Halogenated Pyridines. *J. Am. Chem. Soc.* **2017**, *139* (31), 10613–10616. https://doi.org/10.1021/jacs.7b05409.
(27) Sicre, C.; Alonso-Gómez, J.-L.; Cid, M. M. Regioselectivity in Alkenyl(Aryl)-Heteroaryl Suzuki Cross-Coupling Reactions of 2,4-Dibromopyridine. A Synthetic and Mechanistic Study. *Tetrahedron* **2006**, *62* (48), 11063–11072. https://doi.org/10.1016/j.tet.2006.09.040.
(28) Garg, N. K.; Sarpong, R.; Stoltz, B. M. The First Total Synthesis of Dragmacidin D. *J. Am. Chem. Soc.* **2002**, *124* (44), 13179–13184. https://doi.org/10.1021/ja027822b.
(29) Dai, X.; Chen, Y.; Garrell, S.; Liu, H.; Zhang, L.-K.; Palani, A.; Hughes, G.; Nargund, R. Ligand-Dependent Site-Selective Suzuki Cross-Coupling of 3,5-Dichloropyridazines. *J. Org. Chem.* **2013**, *78* (15), 7758–7763. https://doi.org/10.1021/jo401096u.
(30) Scott, N. W. J.; Ford, M. J.; Jeddi, N.; Eyles, A.; Simon, L.; Whitwood, A. C.; Tanner, T.; Willans, C. E.; Fairlamb, I. J. S. A Dichotomy in Cross-Coupling Site Selectivity in a Dihalogenated Heteroarene: Influence of Mononuclear Pd, Pd Clusters, and Pd Nanoparticles—the Case for Exploiting Pd Catalyst Speciation. *J. Am. Chem. Soc.* **2021**, *143* (25), 9682–9693. https://doi.org/10.1021/jacs.1c05294.

Appendix C: Supporting Information for Chapter 4

Other Supporting information for this chapter include the following:

Lu-PhD-thesis_Ch.4 SI tables. (separate file; *.xlsx, electronically available at *https://doi.org/10.20383/103.0795*)
Extended tables in Microsoft Excel format (.xlsx) that contain list of competition reactions, measured $\Delta G^{\ddagger}_{OA}$ and molecular descriptors.

Lu-PhD-thesis_Ch.4_MLR models. (separate file; *.xlsx, electronically available at *https://doi.org/10.20383/103.0795*)
Extended tables in Microsoft Excel format (.xlsx) that contain predicted $\Delta G^{\ddagger}_{OA}$ calculations, and statistical analysis for multivariate linear regression models.

General Considerations

Materials

All solvents, reagents, and organic substrates were used as purchased from commercial suppliers without further purification with the following exceptions. N-(2,6-dichloropyridin-3-yl)acetamide[1], 2,6-dichloro-3-methoxypyridne[2], 2-chloro-3/5/6-pyridyl-triflate[3], 2-bromo-5-pyridyl-triflate[3] and 4-bromo-phenyl-triflate[4] were prepared using published procedures. 2,6-dichloropyridin-3-yl acetate was prepared using the procedures in *Synthesis of dichloropyridine*. Bis(tricyclohexylphosphine)palladium(0) was purchased from Strem Chemicals and used as received. All reactions were performed inside an MBraun glovebox under an N_2 atmosphere.

Analysis and Spectroscopy

NMR analysis refer to section *Analysis and Spectroscopy* in Appendix B.

LCMS analysis was performed on a Waters LCMS system. The Waters Acquity class H UPLC system is equipped with an autosampler, quaternary pump system, column oven, a photodiode array detector and a QDa Mass Spectrometer. The chromatogram was recorded using an ACQUITY UPLC® BEH C18 column (2.1 × 50 mm, particle size 1.7 µm) with 0.1% formic acid (FA) in water as the mobile phase A and 0.1% FA in acetonitrile as the mobile phase B. The data was processed using Masslynx.

Flash column purification of all the chloro-pyridines in *Synthesis of chlorinated pyridines* section was performed using a Biotage Selekt system. The system is equipped with a built-in QR reader for Biotage Sfar columns, RFID reader for Selekt collection racks, UV detector, pump, fraction collector, and touch screen. The column information is as follows: Biotage Sfar Silica 60 µm, 25 g, part No. FSRS-0445-0025.

Experimental Details for Intermolecular Oxidative Addition Competition Reactions

The relative rates of 100 substrates in THF, 49 substrates in 1:1 THF/DMF and 50 substrates in toluene were determined using the competition experimentation approach; the experimental details refer to section *Experimental Details for Oxidative Addition Competition Studies* in Appendix B.

Table C1 contains the ^{31}P NMR chemical shifts and the peak area ratios of 136 competition reactions in THF; Table C2 contains the ^{31}P NMR chemical shifts and the peak area ratios of 60 competition reactions in 1:1 THF/DMF; Table C3 contains the ^{31}P NMR chemical shifts and the peak area ratios of 59 competition reactions in toluene. All 31P NMR chemical shifts are referenced to triphenylphosphine (-6.00 ppm) as an internal chemical shift standard.

Table C1 – ^{31}P NMR chemical shifts and the peak area ratios of 136 competition reactions in THF (Lu-PhD-thesis_Ch.4 SI tables.xlsx).

Table C2 – ^{31}P NMR chemical shifts and the peak area ratios of 60 competition reactions in 1:1 THF/DMF (Lu-PhD-thesis_Ch.4 SI tables.xlsx).

Table C3 – ^{31}P NMR chemical shifts and the peak area ratios of 59 competition reactions in toluene (Lu-PhD-thesis_Ch.4 SI tables.xlsx).

Experimental Details of reaction progress kinetic analysis for the touchstone reactions

We have published an oxidative addition reactivity dataset containing 79 measured relative rates ($\Delta G^{\ddagger}_{OA}$) of (hetero)aryl (pseudo)halides to $Pd(PCy_3)_2$ in THF[4]. We set $\Delta G^{\ddagger}_{OA}$=0 kJ mol^{-1} for the oxidative addition of 2-bromo-5-nitropyridine, the fastest reaction in our dataset, and all other oxidative addition rates are recorded as relative to it. The complete list of the competition reactions and the measured $\Delta G^{\ddagger}_{OA}$ refer to the ESI of this publication. Here we measured the absolute oxidative addition rates for the touchstone reactions in each solvent by reaction progress kinetic analysis; with the measured $\Delta G^{\ddagger}_{OA}$ for the touchstone reactions, we calibrated all the relative rates and obtained a quantitative reactivity scale of the absolute $\Delta G^{\ddagger}_{OA}$ values in each solvent.

The experimental procedures described below were used to measure the $\Delta G^{\ddagger}_{OA}$ values of the touchstone reactions in THF, 1:1 THF/DMF and toluene.

Two of the oxidative additions to $Pd(PCy_3)_2$ were chosen as the touchstone reactions and their absolute reaction rates were measured by kinetic analysis. The $\Delta G^{\ddagger}_{OA}$ (obtained from the *pseudo* first order rate constant) determined for one of those touchstone reactions (2-chloro-6-methylpyridine in THF, 2-chloro-5-methylpyridine in 1:1 THF/DMF and 2-chloro-6-methylpyridine in toluene) was used to calibrate all the relative rates determined by competition reactions, giving the absolute $\Delta G^{\ddagger}_{OA}$ values for the entire array of substrates. The other touchstone reaction (4-bromoanisole in THF, 4-bromotoluene in 1:1 THF/DMF and 2-chloro-4-methylpyridine in toluene) was used to validate the accuracy of the rate constants determined from this competition experimentation approach, by comparing the

reaction rates determined from kinetic analysis to those obtained from competition experiments.

All the following procedures were performed inside an MBraun glovebox under an N_2 atmosphere. Bis(tricyclohexylphosphine)palladium(0) ($Pd(PCy_3)_2$) was used as the palladium source. Stock solutions were prepared for each component in each solvent system: the concentration of the substrate stock solutions was 0.75 M, and the concentration of $Pd(PCy_3)_2$ stock solution was 0.0375 M. The oxidative additions were conducted at room temperature at initial concentrations of 0.15 M of the substrate and 0.0075 M of $Pd(PCy_3)_2$ in THF and toluene; the initial concentrations were reduced by half (0.075 M of the substrate and 0.00375 M of $Pd(PCy_3)_2$) in 1:1 THF/DMF due to low solubility of $Pd(PCy_3)_2$ in DMF.

200 μL of the $Pd(PCy_3)_2$ stock solution and 600 μL solvent were transferred into a J young NMR tube containing a capillary filled with internal standard PPh_3 in C_6D_6. The tube was seal properly with a septa cap and it was transferred outside glovebox. The reaction progress was monitored by ^{31}P NMR. The J young tube was loaded to an NMR spectrometer to take the ^{31}P NMR peak area ratio of unreacted $Pd(PCy_3)_2$ to PPh_3 at T_0 of the reaction. Then 200 μL of the substrate stock solution was injected into the tube using a syringe and the NMR tube was inverted twice to have the components mix well. Then ^{31}P NMR spectra were recorded to monitor the reaction progress until the oxidative addition was close to completion (the remaining concentration of $Pd(PCy_3)_2$ is less than 10% compared to its T_0 concentration). The rate of each oxidative addition was measured in duplicate.

A [$Pd(PCy_3)_2$] versus time plot was constructed, and an exponential fit applied. Using a *pseudo* first-order rate law assumption, the rate constant (k) was calculated using Eq (C1). This k value was substituted into the Eyring equation to obtain the Gibbs free energy of this reaction ($\Delta G^{\ddagger}_{OA}$) according to Eq (C2). The results of the reaction rate measurement experiments for the two touchstone reactions are summarized in Fig. C1 (THF), Fig. C2 (1:1 THF/DMF) and Fig. 4.2 in the main text (toluene).

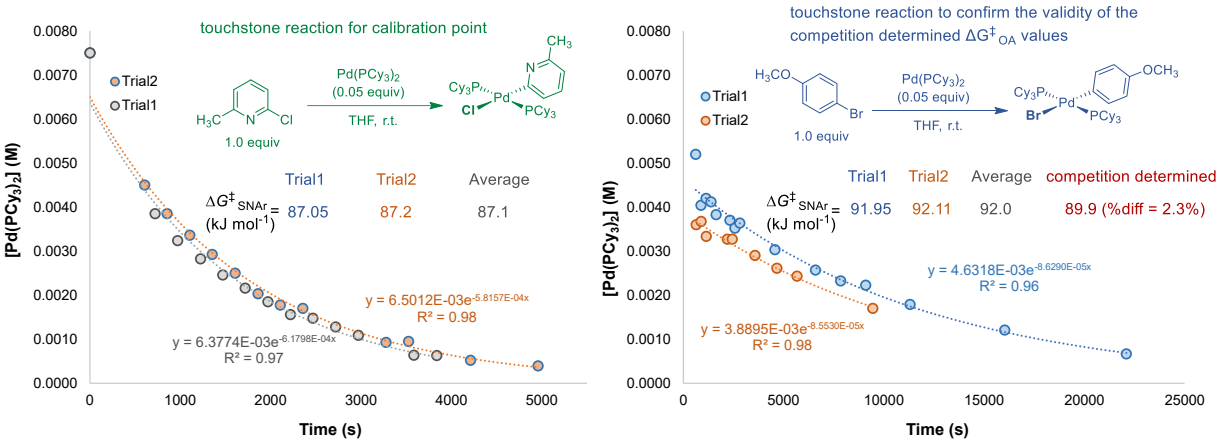


Figure C1. Touchstone reaction progress analysis under pseudo first order conditions in THF. The $\Delta G^{\ddagger}_{OA}$ of 2-chloro-6-methylpyridine **(left)** was used as the calibration point to obtain the absolute rates of the entire array of substrates, and the $\Delta G^{\ddagger}_{OA}$ of 4-bromoanisole **(right)** was used to confirm the validity of the competition determined $\Delta G^{\ddagger}_{OA}$ for this substrate.

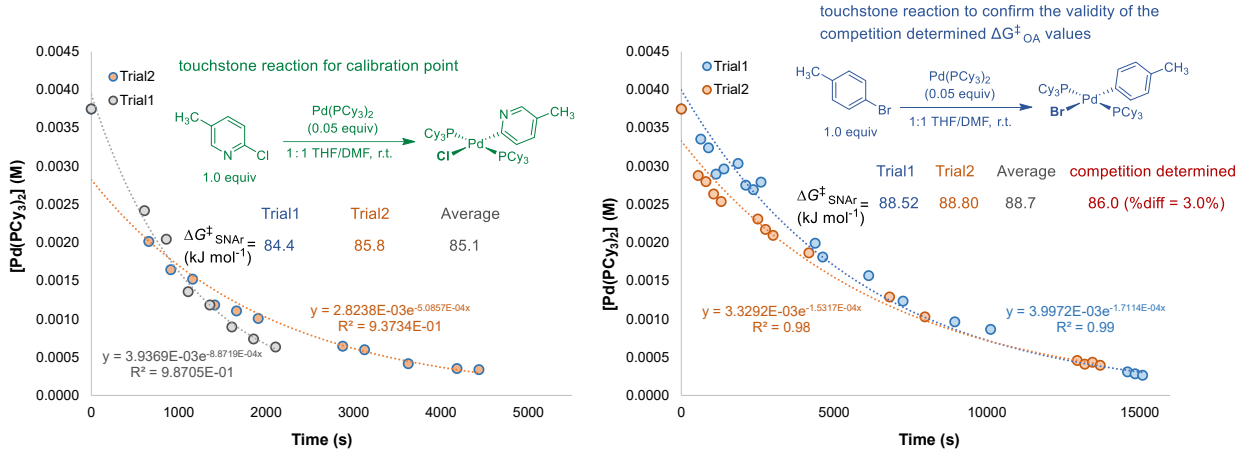


Figure C2. Touchstone reaction progress analysis under pseudo first order conditions in 1:1 THF/DMF. The $\Delta G^{\ddagger}_{OA}$ of 2-chloro-5-methylpyridine **(left)** was used as the calibration point to obtain the absolute rates of the entire array of substrates, and the $\Delta G^{\ddagger}_{OA}$ of 4-bromotoluene **(right)** was used to confirm the validity of the competition determined $\Delta G^{\ddagger}_{OA}$ for this substrate.

According to the *pseudo* first-order rate law assumption, the rate constant (k) can be determined by Eq (C1):

$$[Pd(PCy3)2]_t = [Pd(PCy3)2]_{t=0} e^{-k't}, where\ k' = k[substrate] \qquad Eq(C1)$$

Then the Gibbs free energy ($\Delta G^{\ddagger}_{OA}$) is determined by the Eyring equation Eq (C2) (transmission coefficient κ assumed to be 1):

$$k = \frac{k_B T}{h} \exp\left(-\frac{\Delta G^{\ddagger}}{RT}\right) \Rightarrow \Delta G^{\ddagger} = -RTln\frac{kh}{k_B T} \qquad Eq(C2)$$

$$where\ R\ is\ the\ ideal\ gas\ constant: 8.314\ J \times (K \times mol)^{-1};$$
$$T\ is\ the\ reaction\ temperature: 299.55\ K;$$
$$k_B\ is\ the\ Boltzmann\ constant: 1.38 \times 10^{-23}\ J \times k^{-1};$$
$$h\ is\ Planck\ constant: 6.63 \times 10^{-34}\ J \times s$$

Calculating absolute $\Delta G^{\ddagger}_{OA}$ for Each Substrate

All of the competition reactions were conducted under *pseudo* first-order conditions, with the two substrates being present in large excess ([substrate]:[Pd source] =20:1). The ratio of the reaction rates can be expressed as Eq (C3):

$$\frac{r_1}{r_2} = \frac{k_1[Pd][Substrate_1]}{k_2[Pd][Substrate_2]} = \frac{k_1[Substrate_1]}{k_2[Substrate_2]} \qquad Eq(C3)$$

The concentrations of the two substrates are assumed to stay constant throughout the reaction; then, Eq (C3) can be simplified to Eq (C4):

$$\frac{r_1}{r_2} = \frac{k_1}{k_2} = \frac{\frac{d[Product_1]}{dt}}{\frac{d[Product_2]}{dt}} = \frac{[Product_1]}{[Product_2]} \qquad Eq(C4)$$

In the $^{31}P\{^{1}H\}$ qNMR spectroscopy analysis, $\frac{[Product_1]}{[Product_2]}$ is equal to the peak area ratio of the two oxidative addition products from the ^{31}P qNMR spectrum. By substituting the Eyring equation (Eq (C5)):

$$k = \frac{k_B T}{h} \exp\left(-\frac{\Delta G^{\ddagger}}{RT}\right) \qquad Eq(C5)$$

into Eq (C4), the relative activation energy ($\Delta\Delta G^{\ddagger}_{OA}$) of the two oxidative addition reactions from the competition can be calculated by Eq (C6):

$$\frac{k_1}{k_2} = \exp\left(\frac{\Delta G^{\ddagger}{}_2 - \Delta G^{\ddagger}{}_1}{RT}\right) \Rightarrow \Delta\Delta G^{\ddagger}_{OA} = \ln\left(\frac{k_1}{k_2}\right) RT \Rightarrow$$
$$\Delta\Delta G^{\ddagger}_{OA} = \ln\left(\frac{Peak\ area_{Product1}}{Peak\ area_{Product2}}\right) RT \qquad Eq(C6)$$

$$where\ \Delta\Delta G^{\ddagger}_{OA} = \Delta G^{\ddagger}_2 - \Delta G^{\ddagger}_1$$

The $\Delta\Delta G^{\ddagger}_{OA}$ values determined from the competition experiments were related to the absolute $\Delta G^{\ddagger}_{OA}$ determined from kinetic analysis of the touchstone reaction in each solvent; therefore, all other $\Delta G^{\ddagger}_{OA}$ values are given relative to the touchstone reaction.

The measured absolute $\Delta G^{\ddagger}_{OA}$ of substrates in each solvent are given in the tables below.

Table C4 – Measured absolute $\Delta G^{\ddagger}_{OA}$ values of the 100 substrates in THF (Lu-PhD-thesis_Ch.4 SI tables.xlsx).

Table C5 – Measured absolute $\Delta G^{\ddagger}_{OA}$ values of the 49 substrates in 1:1 THF/DMF (Lu-PhD-thesis_Ch.4 SI tables.xlsx).

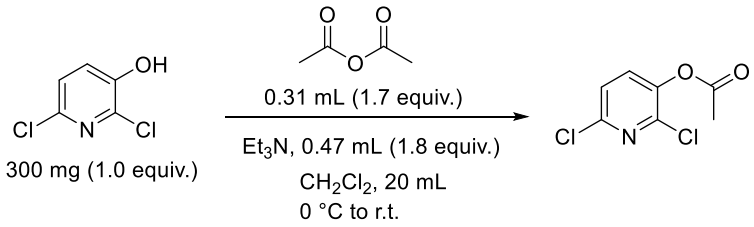
Table C6 – Measured absolute $\Delta G^{\ddagger}_{OA}$ values of the 50 substrates in toluene (Lu-PhD-thesis_Ch.4 SI tables.xlsx).

Synthesis of Dichloropyridine

2,6-dichloropyridin-3-yl acetate was prepared by the following procedures.

A 50 mL round bottom flask containing a stir bar immersed in an ice bath was charged with 300 mg 2,6-dichloropyridin-3-ol (1.0 equiv.), 0.47 mL Et_3N and 15 mL CH_2Cl_2. 0.31 mL (1.7 equiv.) acetic anhydride was dissolved in 5 mL CH_2Cl_2 then this solution was added dropwise to the round bottom flask; the reaction mixture was stirred from 0 °C to room temperature overnight.

On the next day, the reaction mixture, 30 mL ethyl acetate and 30 mL saturated $NaHCO_3$ solution was transferred into a separatory funnel for extraction. The organic phase was collected and washed with saturated $NaHCO_3$ (30 mL) and saturated NaCl solution (30 mL) then dried over Mg_2SO_4, filtered, and evaporated under vacuum. Then the crude products were isolated by automated flash chromatography (Biotage Selekt) on silica gel using hexanes/ethyl acetate. Pure 2,6-dichloropyridin-3-yl acetate was obtained (287 mg, white solid).



Scheme. C1. Reaction scheme for synthesis of 2,6-dichloropyridin-3-yl acetate.

Characterization of 2,6-dichloropyridin-3-yl acetate

287 mg of a white solid was obtained (76% yield).

^{1}H NMR: (500 MHz, $CDCl_3$): δ 2.40 (s, 3H, 3 x $COCH_3$), 7.32 (d, 1H, 1 x Py-H), 7.49 (d, 1H, 1 x Py-H).

$^{13}C\{^1H\}$ NMR: (126 MHz, $CDCl_3$): δ 20.63 (1 x CO**C**H_3), 82.53 (1 x O**C**$(CH_3)_3$), 123.83 (1 x Py), 134.41 (1 x Py), 142.98 (1 x Py), 143.50 (1 x Py), 146.65 (1 x Py), 167.79 (1 x **C**OCH_3).

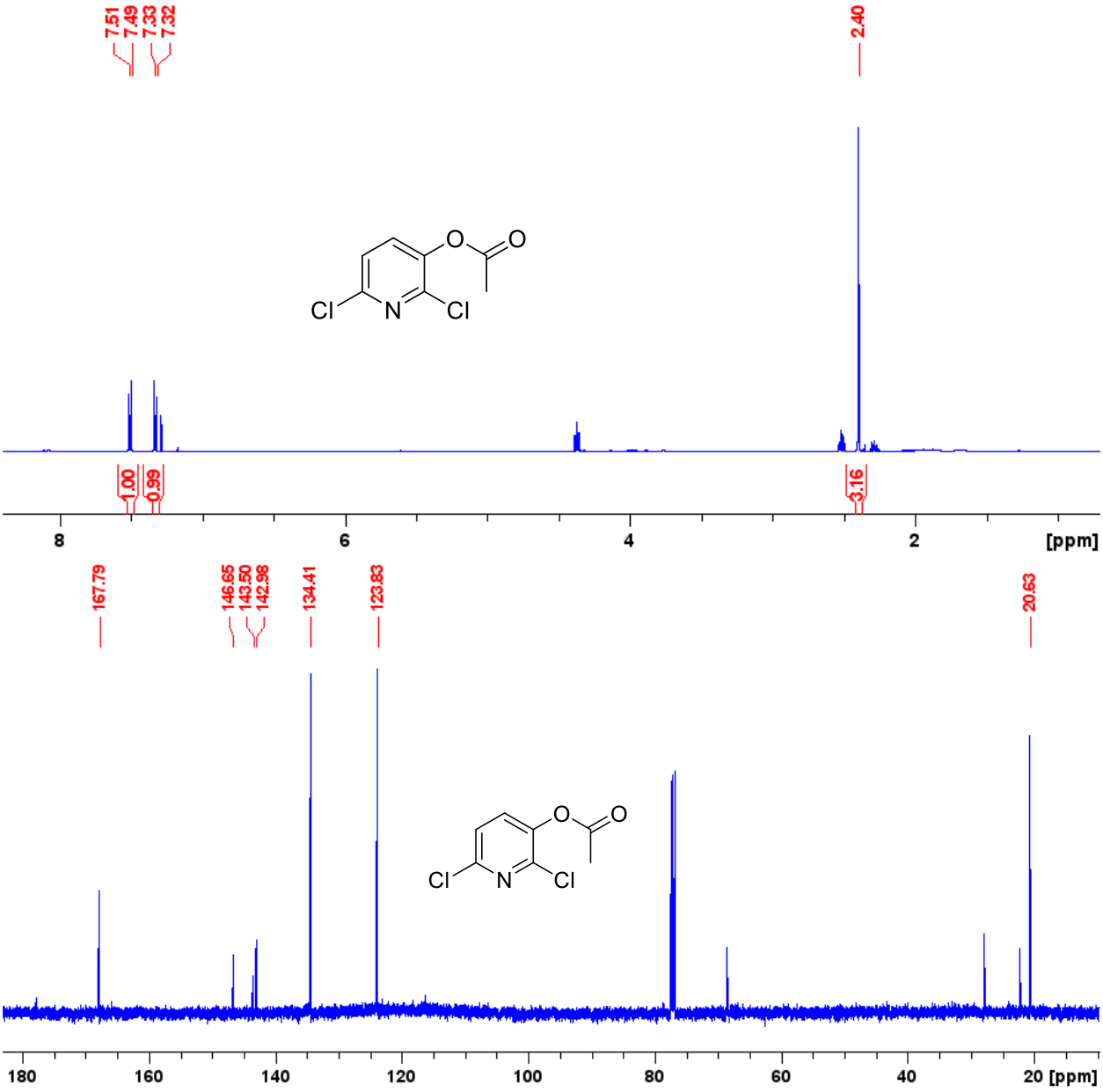


Fig. C3. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **2,6-dichloropyridin-3-yl acetate**.

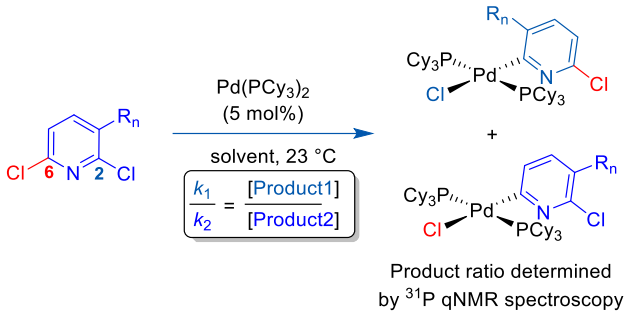
Experimental Details for Oxidative Addition Intramolecular Site-selectivity Determination

Site-selectivity of the 8 2,6-dichloro-3-R-pyridines (R = NH_2, NHCOMe, OH, OMe, OAc, NO_2, CN, CHO) oxidative addition to $Pd(PCy_3)_2$ in different solvents (toluene, THF, 1:1 THF/DMF, etc.) were determined as the following procedures.

Bis(tricyclohexylphosphine)palladium(0) ($Pd(PCy_3)_2$) was used as the palladium source. Stock solutions were prepared for each component: the concentration of the

substrate stock solutions was 0.375 M, and the concentration of $Pd(PCy_3)_2$ stock solution was 0.0375 M.

In a 4 mL vial containing a stirbar, an aliquot of the substrate stock solution (100 µL, 0.0375 mmol) was diluted with additional reaction solvent (700 µL), followed by addition of an aliquot of the $Pd(PCy_3)_2$ stock solution (200 µL, 0.00750 mmol). The resulting solution was mixed for 2-18 hours, then a 600 µL sample was transferred to an NMR tube containing a capillary filled with PPh_3 in C_6D_6. A ^{31}P qNMR spectrum was recorded and the ratio of the two products is determined as the peak area ratio.



Scheme. C2. Experiment design for intramolecular Oxidative Addition Competition Reactions for 2,6-dichloro-3-R-pyridines.

Further experiments were carried out to confirm the structures of the major oxidative addition complexes from all the intramolecular competition reactions except for 2,6-dichloro-pyridin-3-yl acetate, which yields a 1:1 ratio of C_2/C_6 products in all the 3 solvents. For 2,6-dichloro-3-R-pyridines (R=NH_2, NHCOMe, OMe, NO_2, CN and CHO), structural characterization of the oxidative addition complexes refers to section *Preparative Scale Synthesis and structural elucidation of the Oxidative Addition Complexes from the substrates in Fig. 5.6* in Appendix D. For 2,6-dichloro-pyridin-3-ol, the experimental details refer to the below section.

Site-selectivity of the 5 halo(hetero)aryl triflates in Table 4.1 oxidative addition to $Pd(PCy_3)_2$ in different solvents (toluene, THF, 1:1 THF/DMF) were determined as the following procedures.

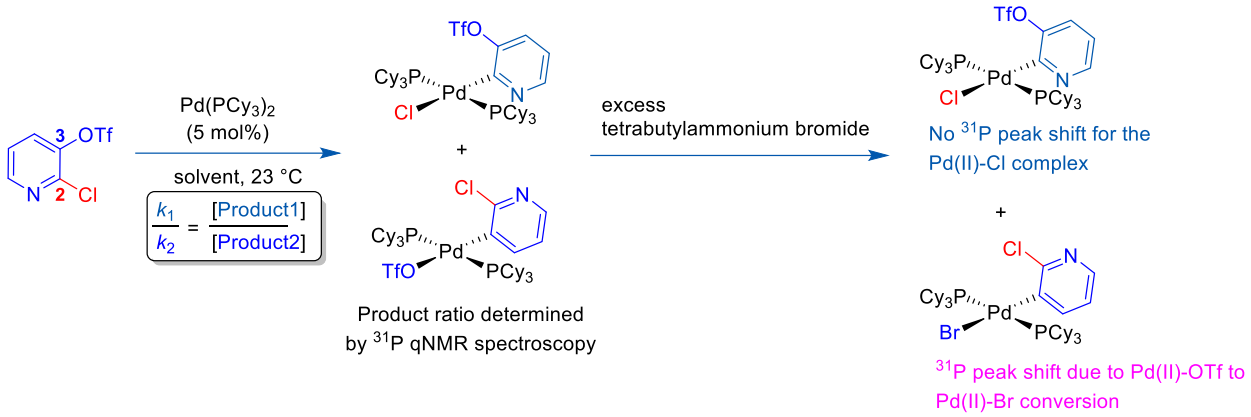
Bis(tricyclohexylphosphine)palladium(0) ($Pd(PCy_3)_2$) was used as the palladium source. Stock solutions were prepared for each component: the concentration of the substrate stock solutions was 0.375 M, and the concentration of $Pd(PCy_3)_2$ stock solution was 0.0375 M. Stock solution of tetrabutylammonium bromide was prepared and the concentration was 0.075 M.

In a 4 mL vial containing a stirbar, an aliquot of the substrate stock solution (100 µL, 0.0375 mmol) was diluted with additional reaction solvent (700 µL), followed by addition of an aliquot of the $Pd(PCy_3)_2$ stock solution (200 µL, 0.00750 mmol). The resulting solution was mixed for 2-18 hours, then a 600 µL sample was transferred to an NMR tube containing a capillary filled with PPh_3 in C_6D_6. A first ^{31}P qNMR spectrum was recorded and the ratio of the two products is determined as the peak area ratio.

Assignment of the ^{31}P NMR peaks to the corresponding oxidative addition complexes was performed as follows: excess amount of tetrabutylammonium bromide (3.0 equiv. to the amount of oxidative addition complexes) was added into the NMR tube containing

reaction solution and the tube was allowed to stand overnight for a fully Pd(II)-OTf to Pd(II)-Br conversion (scheme 3).

A second ^{31}P qNMR spectrum was recorded and compared to the first one: the peak of which the chemical shift remained belongs to the Pd(II) complex at C-Cl/Br insertion, and the peak shifted belongs to the Pd(II) complex at C-OTf insertion. The ^{31}P NMR spectra of the 5 halo(hetero)aryl triflates are shown in Figs. C4 – C8 (a representative set of ^{31}P NMR spectra in one of the solvents for each substrate).



Scheme. C3. Experiment design for intramolecular Oxidative Addition Competition Reactions for halo(hetero)aryl triflates; the reaction scheme for 2-chloropyridin-3-yl-triflate is shown as a representative example.

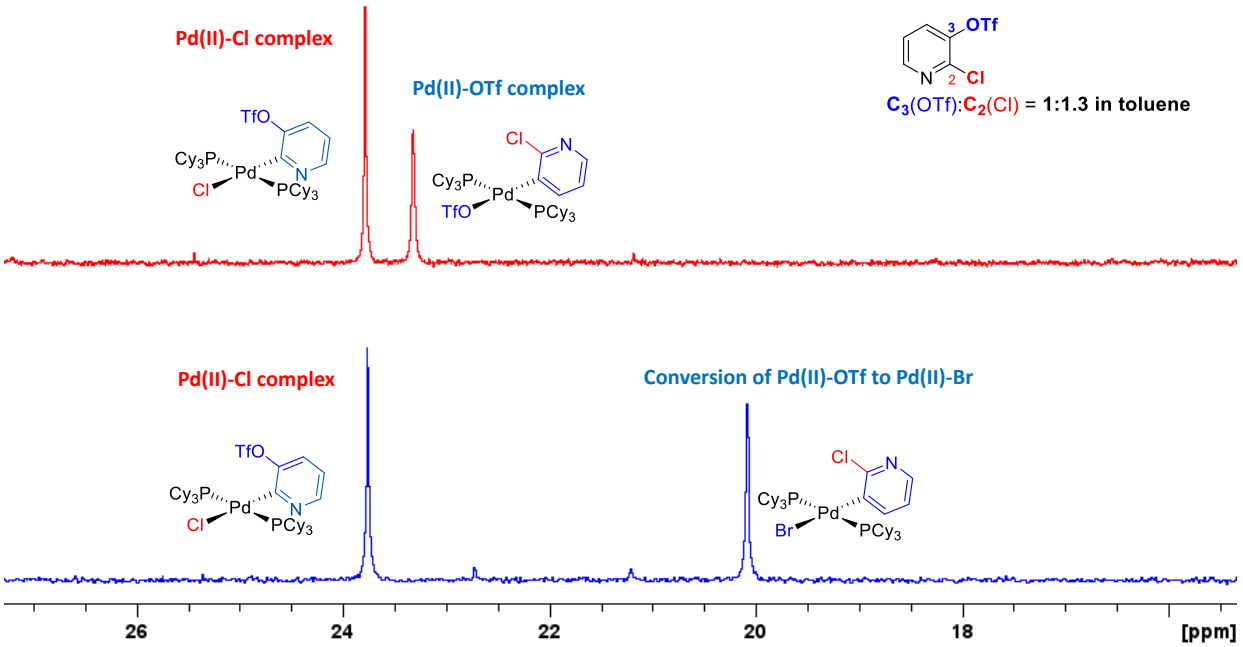


Fig. C4. Top: ^{31}P{^{1}H} NMR spectrum of Pd(II) complexes from **2-chloropyridin-3-yl-triflate** oxidative addition to $Pd(PCy_3)_2$ in **toluene**; Bottom: ^{31}P{^{1}H} qNMR spectrum of the Pd(II) complexes after excess amount of tetrabutylammonium bromide was added.

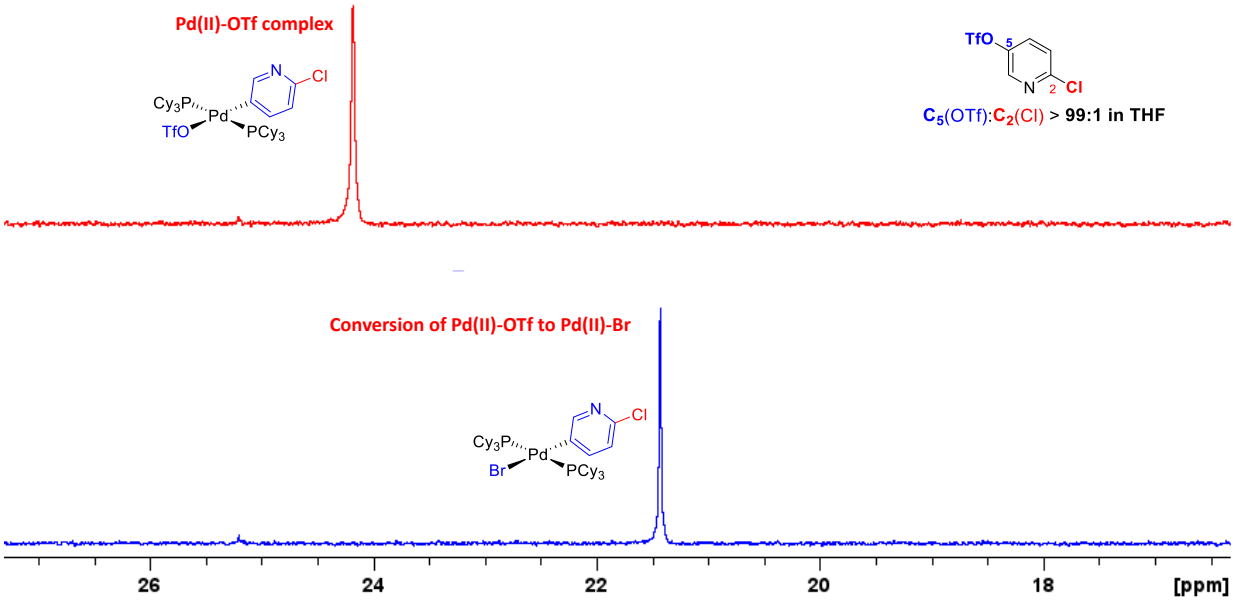


Fig. C5. Top: $^{31}P\{^{1}H\}$ NMR spectrum of Pd(II) complexes from **2-chloropyridin-5-yl-triflate** oxidative addition to $Pd(PCy_3)_2$ in **THF**; Bottom: $^{31}P\{^{1}H\}$ qNMR spectrum of the Pd(II) complexes after excess amount of tetrabutylammonium bromide was added.

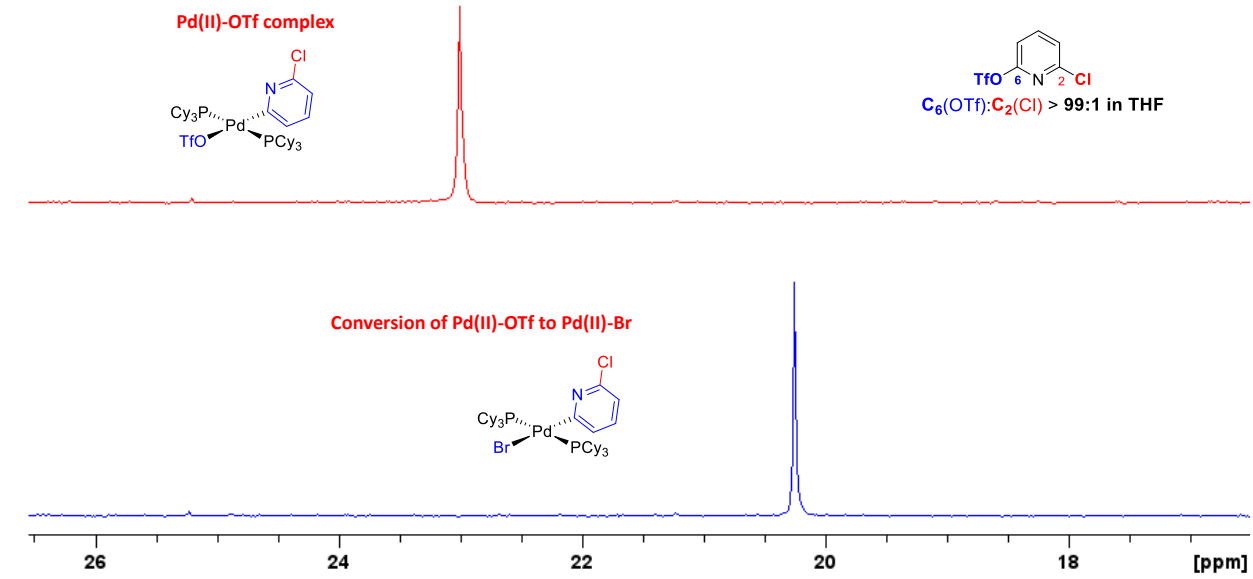


Fig. C6. Top: $^{31}P\{^{1}H\}$ NMR spectrum of Pd(II) complexes from **2-chloropyridin-6-yl-triflate** oxidative addition to $Pd(PCy_3)_2$ in **THF**; Bottom: $^{31}P\{^{1}H\}$ qNMR spectrum of the Pd(II) complexes after excess amount of tetrabutylammonium bromide was added.

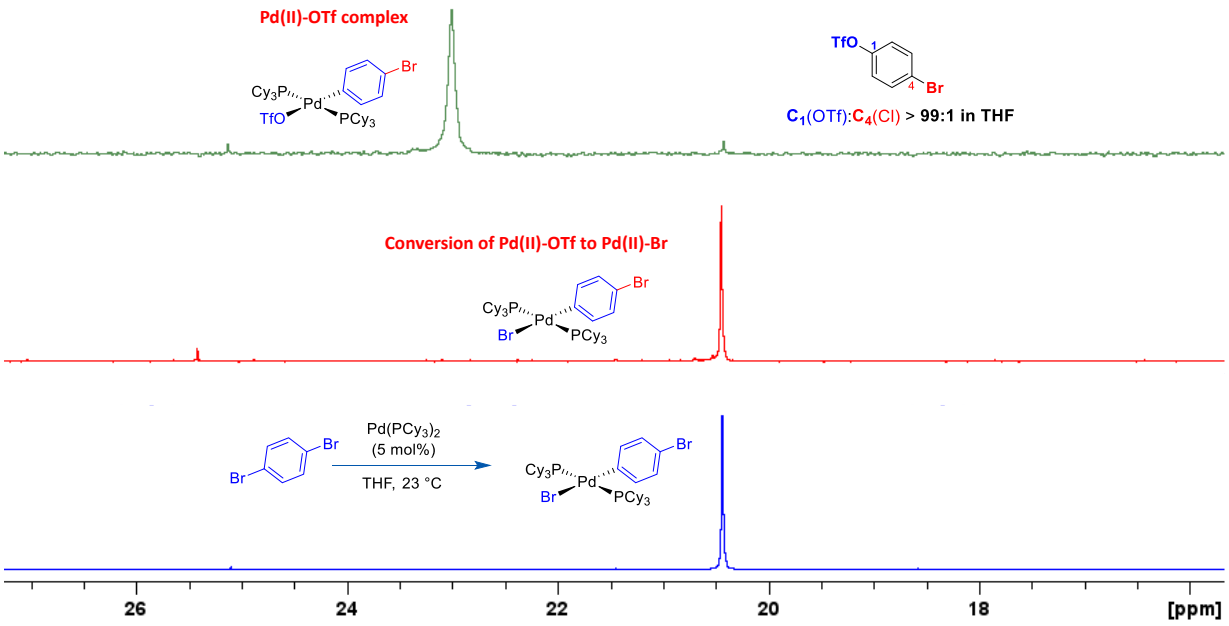


Fig. C7. Top: $^{31}P\{^{1}H\}$ NMR spectrum of Pd(II) complexes from **4-bromophenyl triflate** oxidative addition to $Pd(PCy_3)_2$ in **THF**; Middle: $^{31}P\{^{1}H\}$ qNMR spectrum of the Pd(II) complexes after excess amount of tetrabutylammonium bromide was added; Bottom: $^{31}P\{^{1}H\}$ qNMR spectrum of Pd(II) complexes from **1,4-dibromobenzene** oxidative addition to $Pd(PCy_3)_2$ in THF; the chemical shift of this Pd(II) complex is same as the one from the middle spectrum (20.4 ppm), which double confirms that 4-bromophenyl triflate oxidative addition to $Pd(PCy_3)_2$ gives the C-OTf insertion Pd(II) complex.

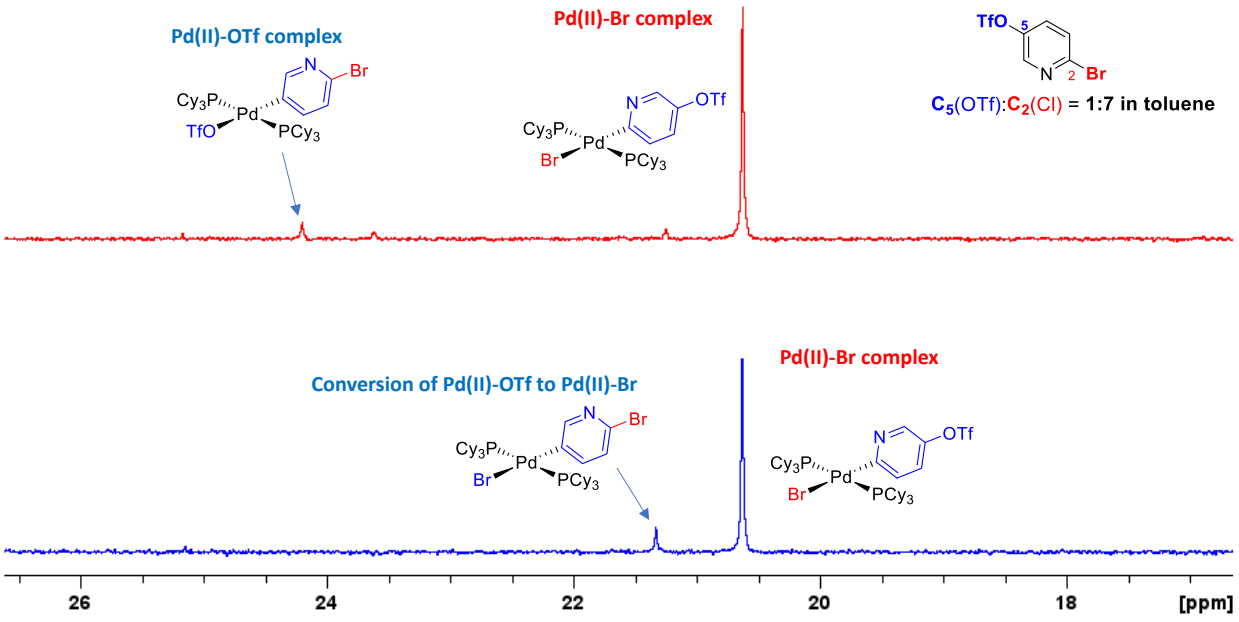


Fig. C8. Top: $^{31}P\{^{1}H\}$ NMR spectrum of Pd(II) complexes from **2-bromopyridin-5-yl-triflate** oxidative addition to $Pd(PCy_3)_2$ in **toluene**; Bottom: $^{31}P\{^{1}H\}$ qNMR spectrum of the Pd(II) complexes after excess amount of tetrabutylammonium bromide was added.

Structural Characterization of the Oxidative Addition Complexes from 2,6-dichloro-pyridin-3-ol

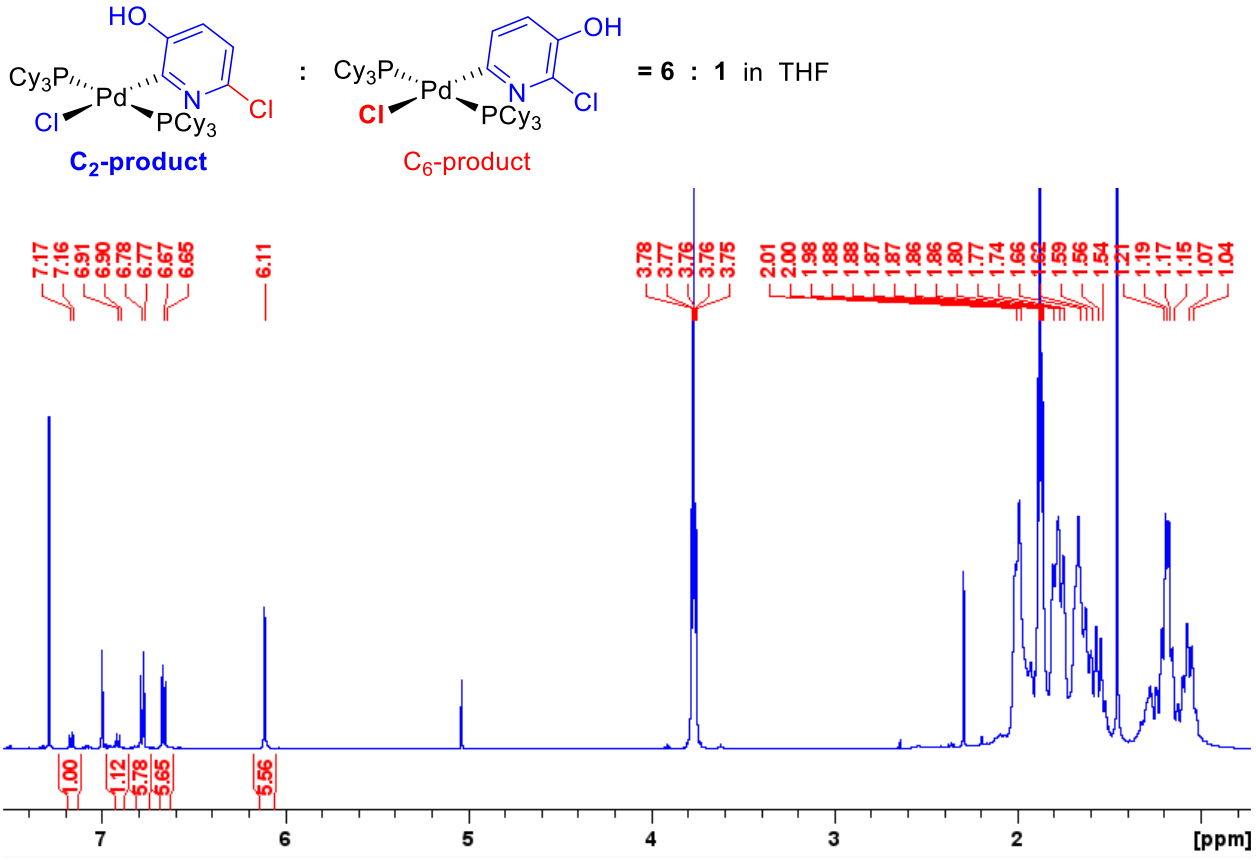
To identify the major and minor oxidative addition complexes and to determine the C_2/C_6 selectivity in different solvents, the following procedures were performed in each of the three solvents.

Stock solution of 2,6-dichloro-pyridin-3-ol was prepared in 0.375 M in THF.

In the glovebox, a 20 mL vial containing a stir bar was charged with 10 mg $Pd(PCy_3)_2$ (0.015 mmol, 1.0 equiv.), 40 µL of the substrate stock solution (0.015 mmol, 1.0 equiv.), and 4 mL of the solvent (toluene, THF, 1:1 THF/DMF). The reaction mixture was stirred at room temperature overnight.

On the next day, the solvent was evaporated under vacuum using a rotary evaporator and the remaining substance was dissolved in $CDCl_3$ for NMR characterization: ^{1}H, ^{31}P and ^{31}P-HMBC were recorded.

2,6-dichloro-pyridin-3-ol oxidative addition to $Pd(PCy_3)_2$ in **THF**



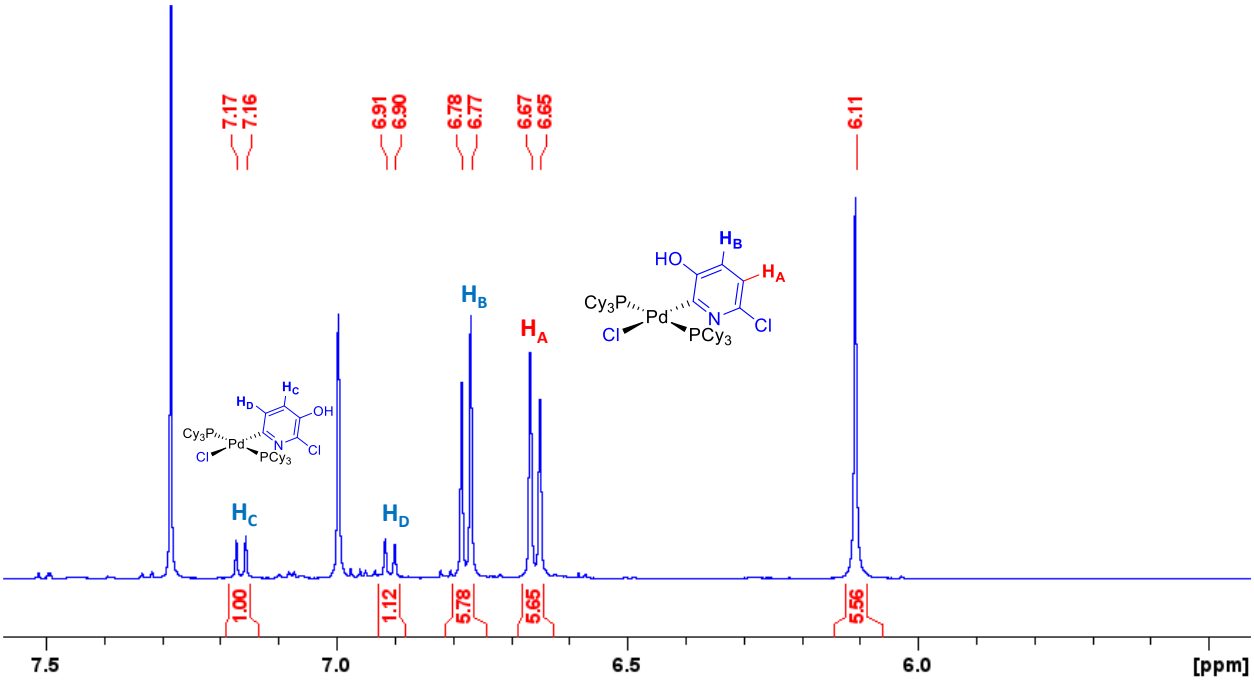
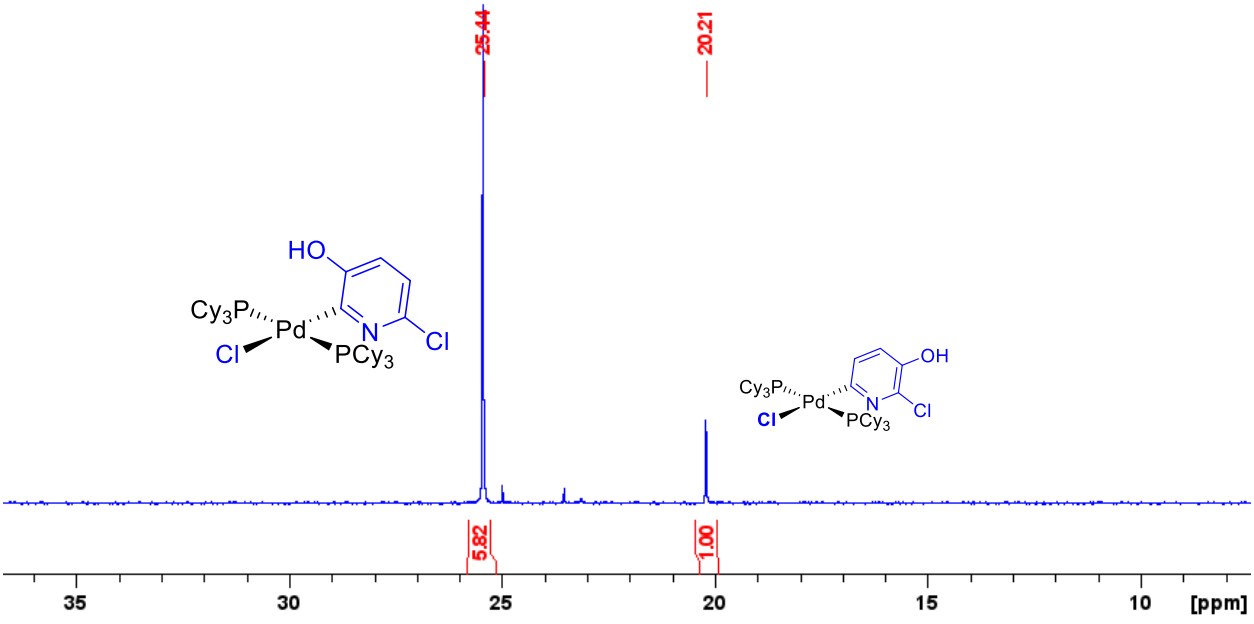


Fig. C9. Top: ^{1}H (500 MHz, $CDCl_3$) NMR spectrum of the oxidative addition complexes from reaction in THF; Bottom: Expansion of the pyridyl proton region from this ^{1}H spectrum.

The regioselective products are identified by the ^{1}H chemical shift. Because of the strong electron donation effect of the PdL_2, the proton para to the reactive carbon is expected to be the most upfield pyridyl proton (H_A from the C_2-product). An expanded pyridyl proton region is shown in Fig.S9: the most upfield peak (δ 6.65) belongs to the proton para to the reactive carbon, and it is the H_A from the C_2-product. The peak area ratio indicates that oxidative addition at C_2 gives the major product and at C_6 gives the minor product.



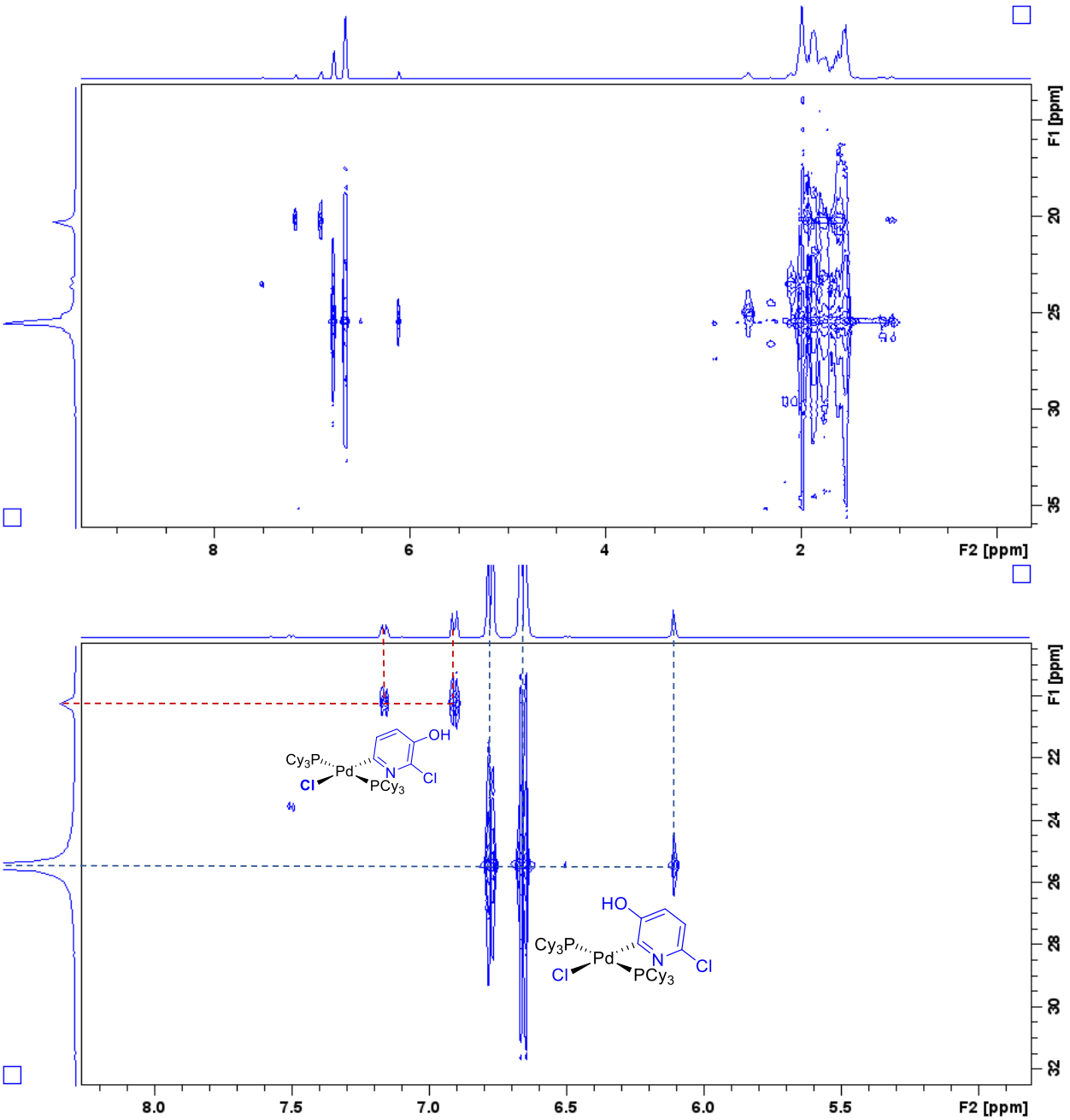
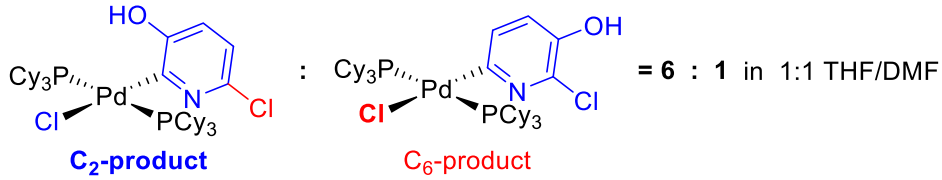


Fig. C10. Top: $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) of the oxidative addition complexes from reaction in THF; Middle: the accompanying long range ^{1}H-^{31}P HMBC NMR spectrum; Bottom: expanded pyridyl proton- phosphorus correlation region.

2,6-dichloro-pyridin-3-ol oxidative addition to $Pd(PCy_3)_2$ in **1:1 THF/DMF**



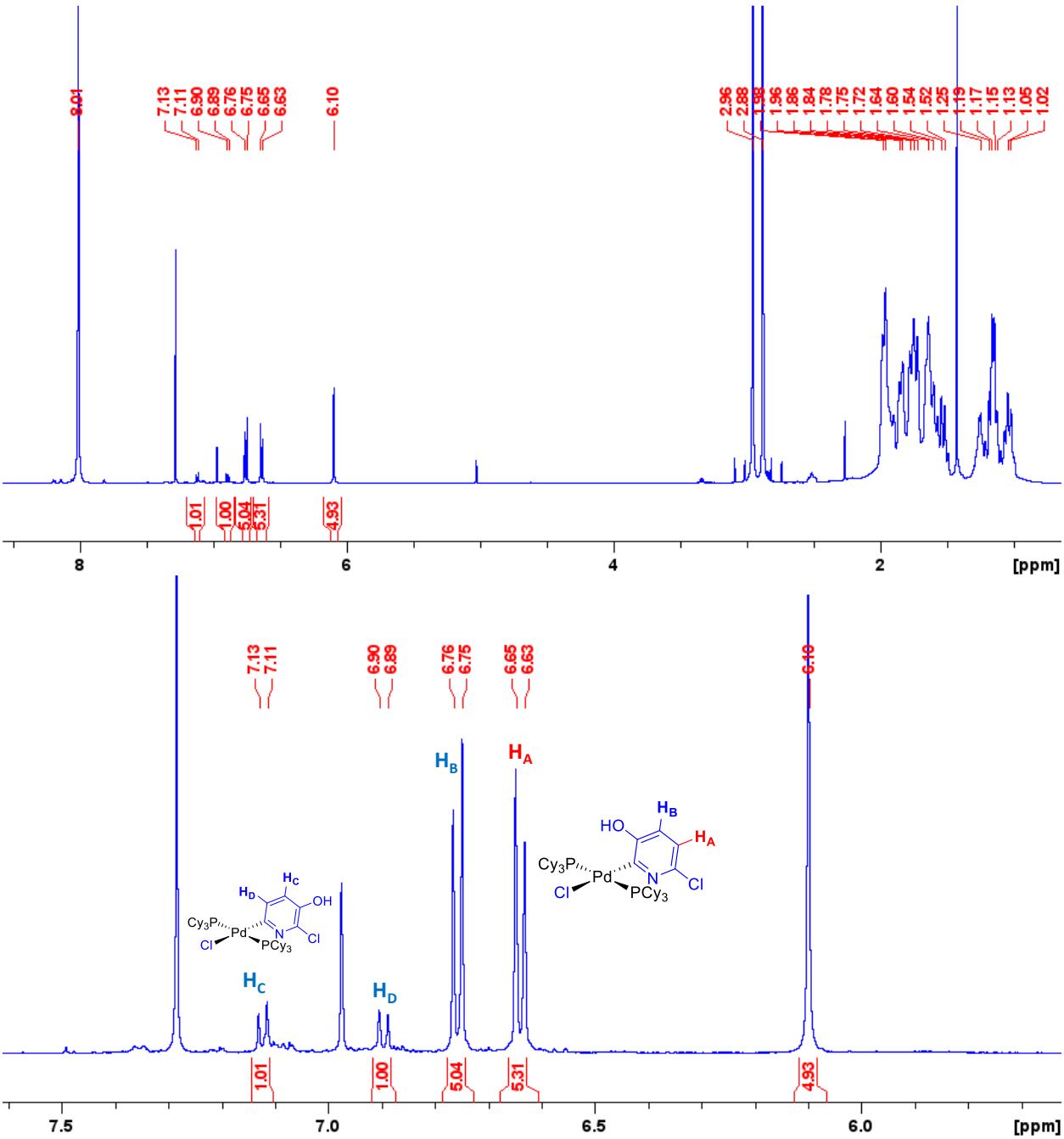
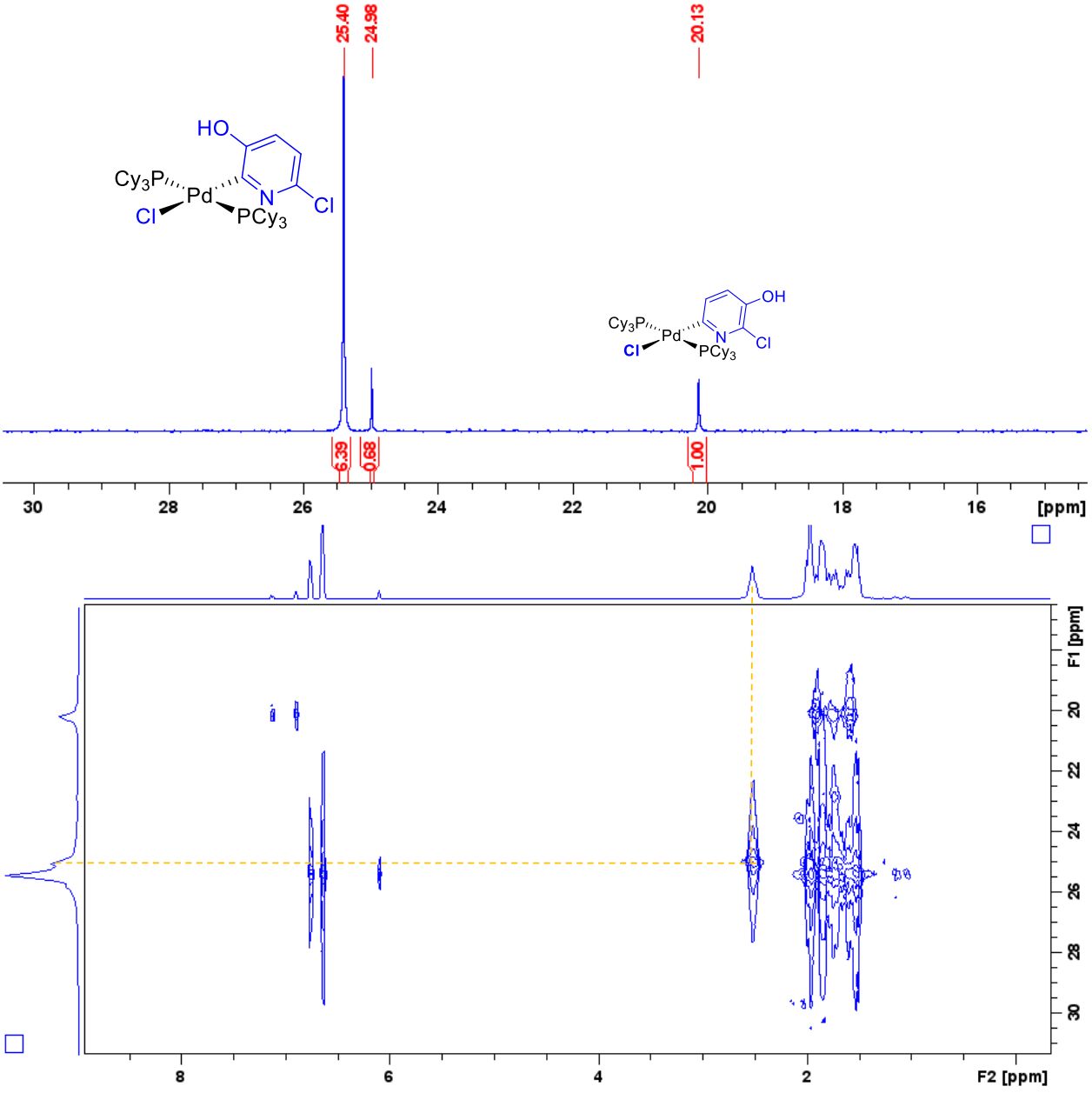


Fig. S11. Top: ^1H (500 MHz, $CDCl_3$) NMR spectrum of the oxidative addition complexes from reaction in 1:1 THF/DMF; Bottom: Expansion of the pyridyl proton region from this ^1H spectrum.


25.40
24.98
20.13
HO
Cy3P
Pd
N
Cl
Cl
PCy3
OH
6.39
0.68
1.00
30
28
26
24
22
20
18
16
[ppm]
F1 [ppm]
F2 [ppm]
8
6
4
2

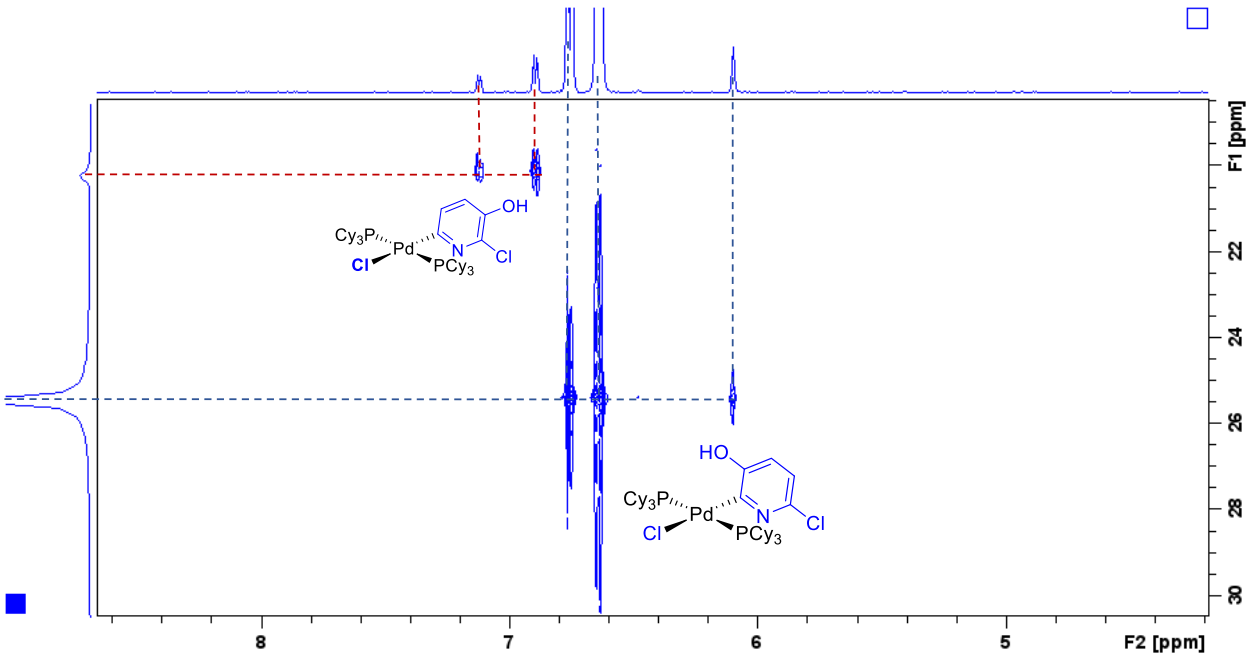
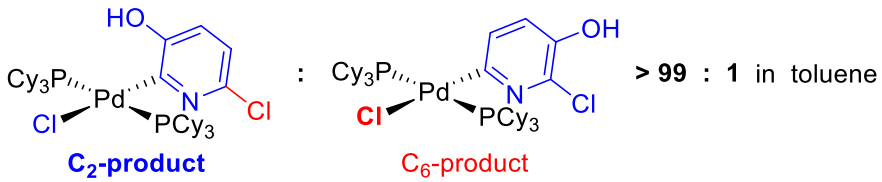


Fig. C12. Top: $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) of the oxidative addition complexes from reaction in 1:1 THF/DMF; Middle: the accompanying long range ^{1}H-^{31}P HMBC NMR spectrum; Bottom: expanded pyridyl proton- phosphorus correlation region. The small peak at 24.98 ppm does not come from any of the oxidative addition complex, instead it is possibly from $(PCy_3)_2Pd$-solvent coordination from the ^{1}H-^{31}P HMBC.

2,6-dichloro-pyridin-3-ol oxidative addition to $Pd(PCy_3)_2$ in **toluene**



Reaction in toluene only forms one oxidative addition complex as indicated by both the ^{1}H and ^{31}P spectra, and the chemical shift (^{1}H: δ 6.66, δ 6.76; ^{31}P: δ 25.4) gives that it is the C_2-product.

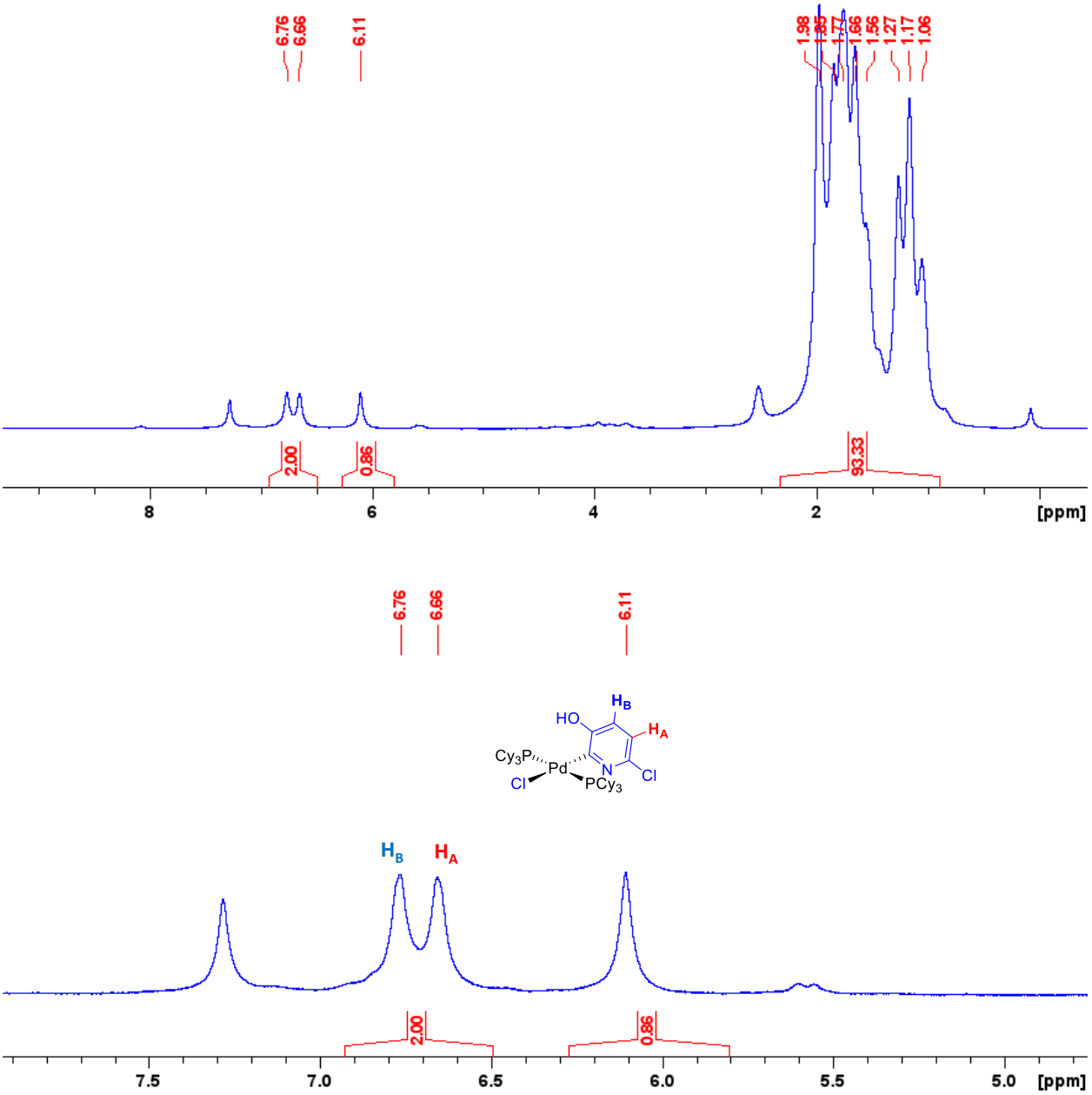
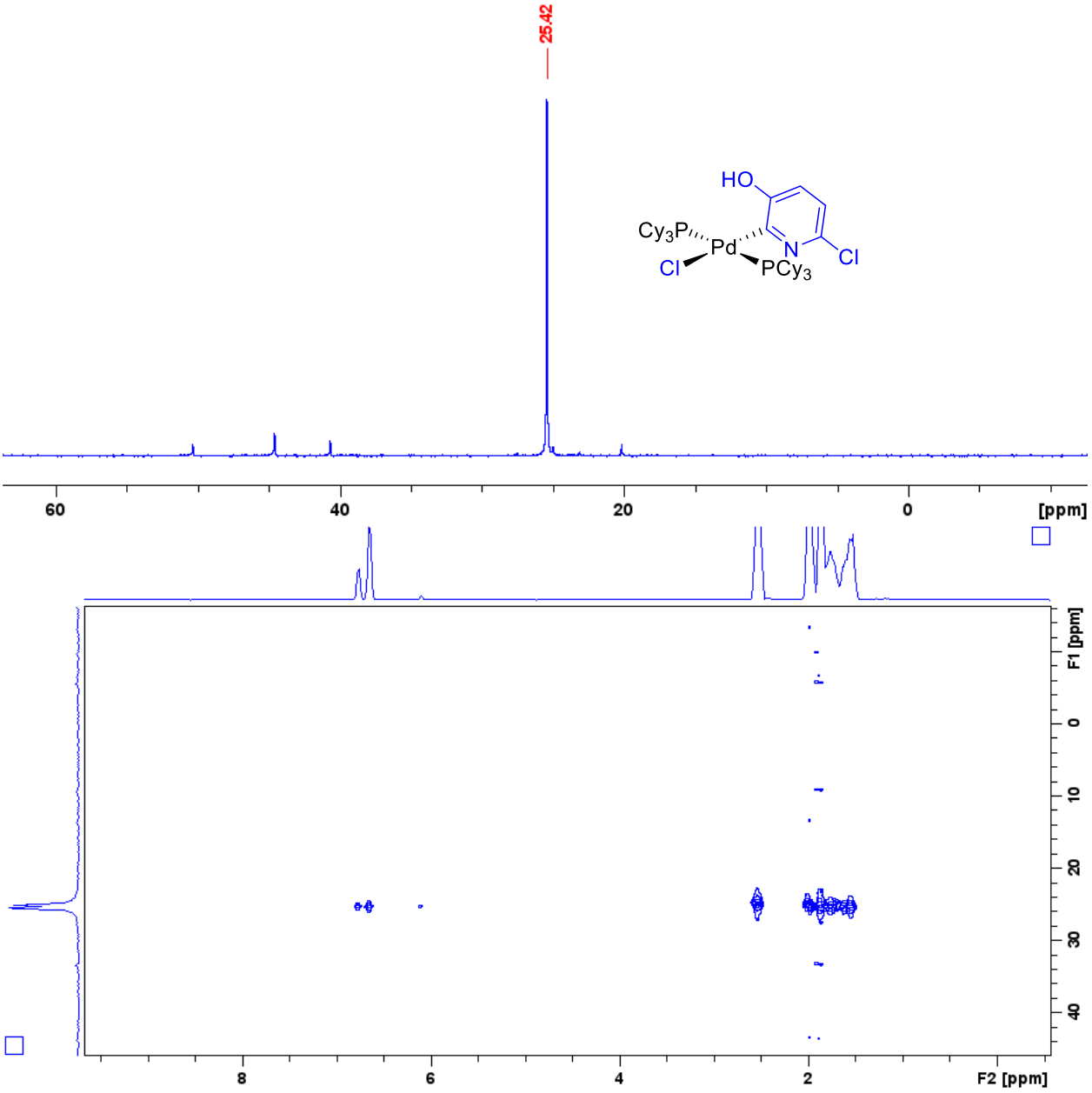


Fig. C13. Top: ^{1}H (500 MHz, $CDCl_3$) NMR spectrum of the oxidative addition complex from reaction in toluene; Bottom: Expansion of the pyridyl proton region from this ^{1}H spectrum.


25.42
HO
Cy3P
Pd
N
Cl
Cl
PCy3
60
40
20
0
[ppm]
F1 [ppm]
0
10
20
30
40
8
6
4
2
F2 [ppm]

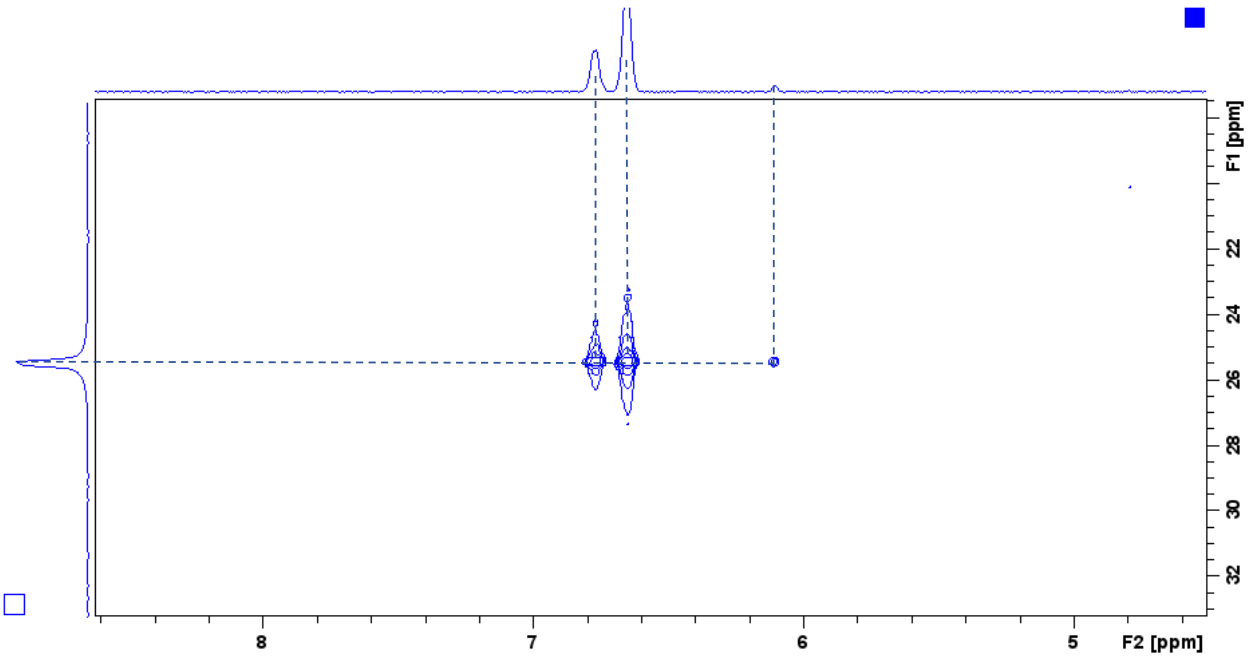


Fig. C14. Top: $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) of the oxidative addition complexes from reaction in toluene; Middle: the accompanying long range ^{1}H-^{31}P HMBC NMR spectrum; Bottom: expanded pyridyl proton- phosphorus correlation region. The small peak at 24.98 ppm does not come from any of the oxidative addition complex, instead it is possibly from $(PCy_3)_2Pd$-solvent coordination from the ^{1}H-^{31}P HMBC.

Computational Determination of Molecular Descriptors

Refer to section *Computational Determination of Molecular Descriptors* in Appendix D.

ESP values of the 3 2-chloropyridine derivatives (Figure 4.7) are given in Table C7; ESP values of the 3 2,6-dichloro-3-EDG-pyridines (Figure 4.8) are given in Table C8; molecular descriptors for the 3 2,6-dichloro-3-EWG-pyridines (Figure 4.9) are given in Table C9; and molecular descriptors for the 5 halo(hetero)aryl triflates (Table 4.1) are given in Table C10.

Table C7. Calculated ESP values at C_2 for the 3 2-chloropyridine derivatives in Figure 4.7.

Substrate	**CPCM(THF)**		**CPCM(DMF)**	
	***ESP*$_{C2}$ (kJ/mol) Substrate itself**	***ESP*$_{C2}$ (kJ/mol) Substrate - Explicit THF**	***ESP*$_{C2}$ (kJ/mol) Substrate itself**	***ESP*$_{C2}$ (kJ/mol) Substrate - Explicit DMF**
 2-chloro-3-aminopyridine	-4.867	-33.716	-9.502	-46.692

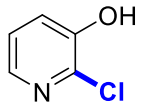
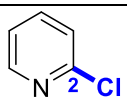
 2-chloropyridin-3-ol	7.065	-39.029	2.564	-62.644
 2-chloropyridine	21.247	7.035	18.143	-6.607

Table C8.
Calculated ESP values at C_2 and C_6 for the 3 2,6-dichloro-3-EDG-pyridines in Figure 4.8.

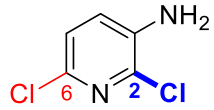
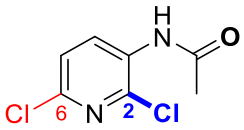
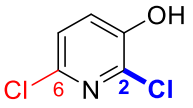
Substrate	**C–Cl site**	**CPCM(THF)**		**CPCM(DMF)**	
		ESP **(kJ/mol) Substrate itself**	***ESP*** **(kJ/mol) Substrate - Explicit THF**	***ESP*** **(kJ/mol) Substrate itself**	***ESP*** **(kJ/mol) Substrate - Explicit DMF**
 2,6-dichloropyridin-3-amine	C_2	27.549	3.615	27.175	-12.431
	C_6	2.910	-15.915	1.290	-23.164
 N-(2,6-dichloropyridin-3-yl)acetamide	C_2	91.978	74.475	92.578	39.070
	C_6	73.971	67.018	74.655	48.044
 2,6-dichloro-pyridin-3-ol	C_2	39.281	-7.667	36.449	-29.036
	C_6	38.712	-0.331	36.826	-16.168

Table C9.
Calculated molecular descriptors for the 3 2,6-dichloro-3-EWG-pyridines in Figure 4.9.
** pKa of HCl is 0.2.[5]

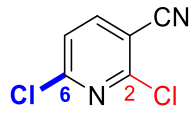
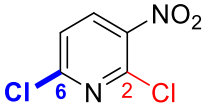
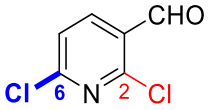
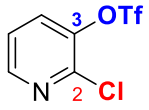
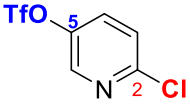
Substrate	C–Cl site	*ESP*$_1$ (kJ/mol)	*ESP*$_2$ (kJ/mol)	$A_1 + A_2$ (kJ/mol)	*IBSI*
 2,6-dichloro-3-cyanopyridine	C_2	116.039	-9.075	0.837	0.311
	C_6	112.994	-9.075	0.000	0.308
 2,6-dichloro-3-nitropyridine	C_2	119.347	-6.491	6.820	0.297
	C_6	117.298	-6.491	2.218	0.292
 2,6-dichloro-3-formylpyridine	C_2	92.434	-29.959	4.774	0.312
	C_6	96.377	-29.959	2.218	0.295

Table C10.
Calculated molecular descriptors for the 5 halo(hetero)aryl triflates in Table 4.1.
** pKa of HCl, HBr and HOTf is 0.2, -4.4, -11.3, respectively[5].

Substrate	C–X site	*ESP*$_1$ (kJ/mol)	*ESP*$_2$ (kJ/mol)	$A_1 + A_2$ (kJ/mol)	*IBSI*
 2-chloropyridin-3-yl triflate	C_2	81.696	-43.794	1.799	0.296
	C_3	98.782		2.092	0.448
 2-chloropyridin-5-yl triflate	C_2	73.514	-44.369	0	0.284
	C_5	89.504		0	0.430
	C_2	81.419	-25.393	1.799	0.295

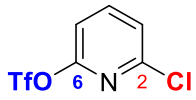
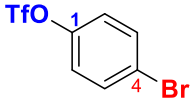
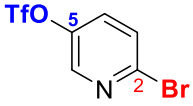
 2-chloropyridin-6-yl triflate	C_6	108.898		0	0.406
 4-bromophenyl triflate	C_4	52.210	44.249	0	0.234
	C_1	72.369		0	0.415
 2-bromopyridin-5-yl triflate	C_2	76.278	-37.086	0	0.225
	C_5	104.436		0	0.438

Table C7. Calculated molecular descriptors for the 5 halo(hetero)aryl triflates in Fig. 4.14 (main text) (Lu-PhD-thesis_Ch.4 SI tables.xlsx).
** pKa of HCl, HBr and HOTf is 0.2, -4.4, -11.3, respectively[5].

Transition State Calculations for Oxidative Addition by DFT

Refer to section *Transition State Calculations for Oxidative Addition by DFT* in Appendix D.

Transition states analysis was performed for oxidative addition of 2-pyridyl triflate and 2-chloropyridine in DMF, THF and toluene with CPCM solvation model. The energy diagrams of the 6 oxidative additions, as well as the transition state structures of 2-pyridyl triflates and 2-chloropyridine in DMF are shown in the main text 4.13; transition state structures of the two substrates in THF and toluene, including the bond length and bond angle of the essential atoms are shown in Figs. C15 – C18.

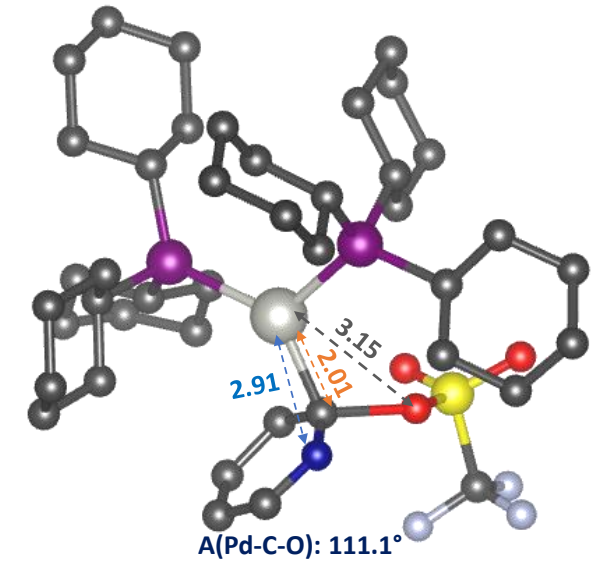


Fig. C15. Calculated transition state structure of oxidative addition of 2-pyridyl triflate to $Pd(PCy_3)_2$ in CPCM(THF). A concerted S_NAr-like transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

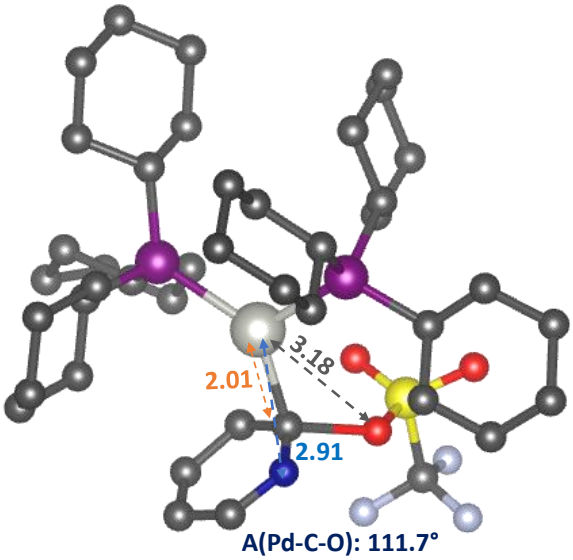


Fig. C16. Calculated transition state structure of oxidative addition of 2-pyridyl triflate to $Pd(PCy_3)_2$ in CPCM(toluene). A concerted S_NAr-like transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

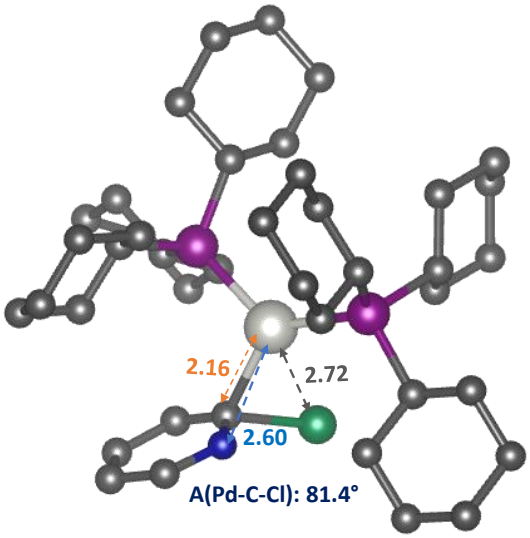


Fig. C17. Calculated transition state structure of oxidative addition of 2-chloropyridine to $Pd(PCy_3)_2$ in CPCM(THF). A stepwise nucleophilic displacement transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

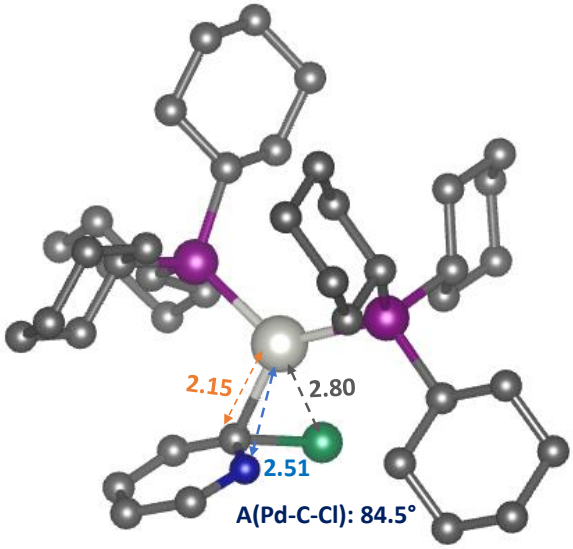


Fig. C18. Calculated transition state structure of oxidative addition of 2-chloropyridine to $Pd(PCy_3)_2$ in CPCM(toluene). A stepwise nucleophilic displacement transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

Cross Validation for the three reactivity models in different solvents (Fig. 4.3)

To evaluate the robustness of the three oxidative addition reactivity models in different solvents from Fig. 4.3B, main text, we performed cross-validation by doing five random 60/40 train-test splits for the THF model, and 70/30 train-test splits for the 1:1 DMF/THF model and toluene model due to the smaller number of data points. Excellent linear correlation was achieved between the observed and predicted $\Delta G^{\ddagger}_{OA}$ in all the three cases: the R^2 ranges from 0.92 to 0.95 for the training set, Q^2 from 0.92 to 0.95 and MAE

from 2.29 kJ mol^{-1} to 3.00 kJ mol^{-1} for test set for the THF model (Fig. C19); R^2 ranges from 0.93 to 0.95 for the training set, Q^2 from 0.89 to 0.91 and MAE from 2.79 kJ mol^{-1} to 3.50 kJ mol^{-1} for test set for the 1:1 DMF/THF model (Fig. C20); R^2 from 0.91 to 0.94 for the training set, Q^2 from 0.84 to 0.92 and MAE from 2.10 kJ mol^{-1} to 2.91 kJ mol^{-1} for test set for the toluene model (Fig. C21).

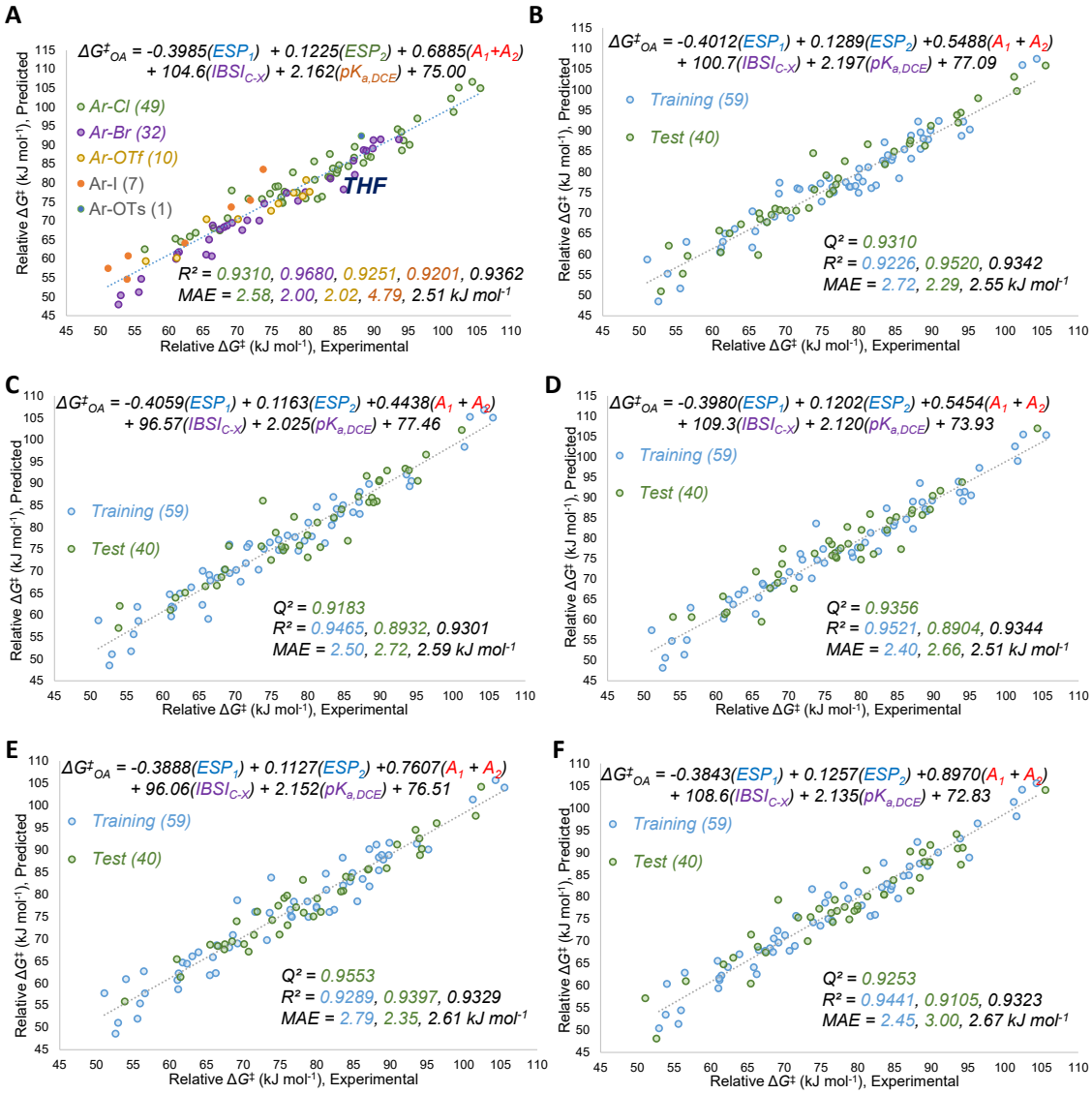


Fig. C19. Performance of the quantitative reactivity model for oxidative addition to Pd(0) in **THF**. (B) – (F) Multivariate linear regression model from one of the five 60/40 random split divisions (1/5) – (5/5).

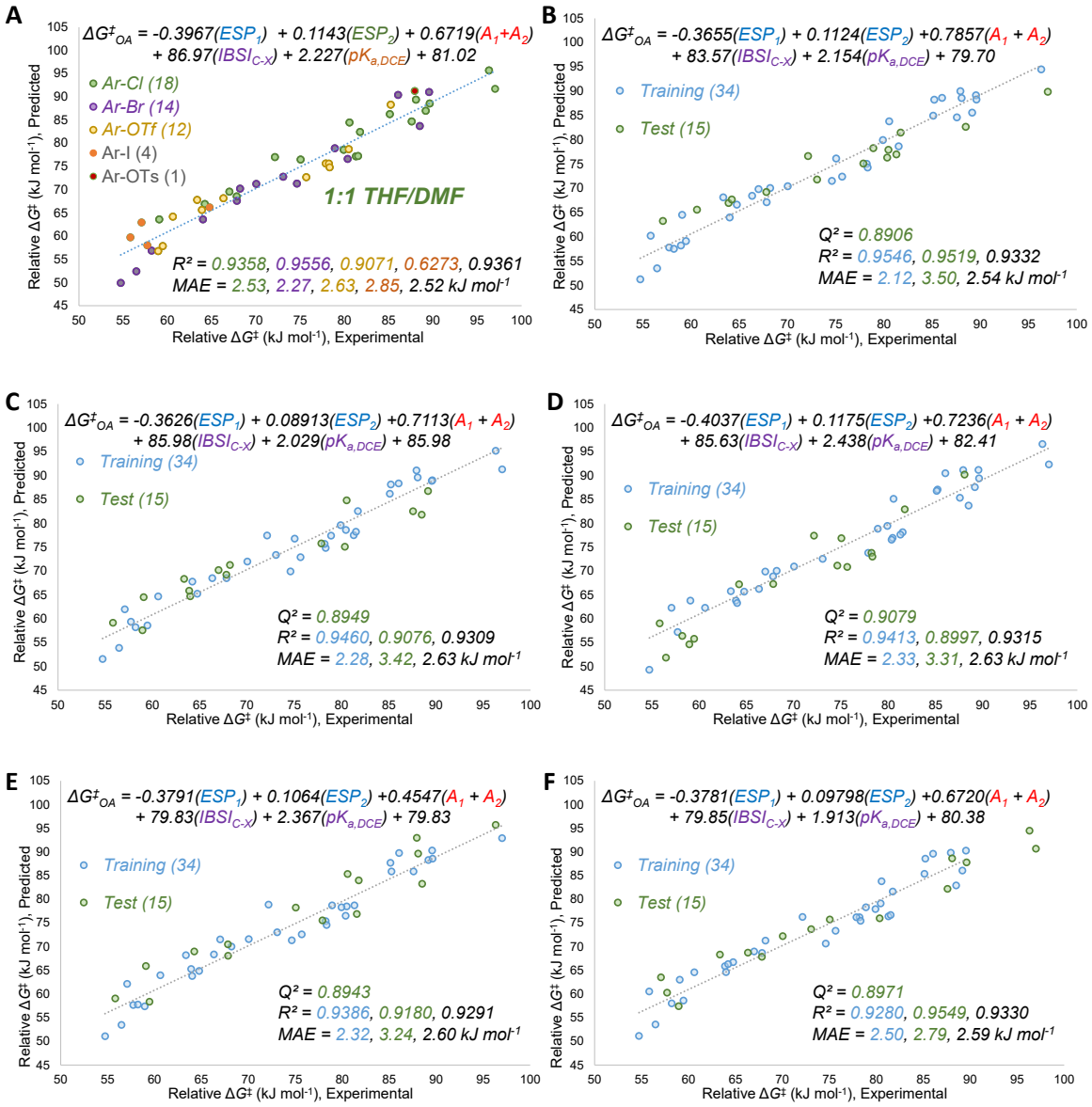


Fig. C20. Performance of the quantitative reactivity model for oxidative addition to Pd(0) in 1:1 **THF/DMF**. (B) – (F) Multivariate linear regression model from one of the five 70/30 random split divisions (1/5) – (5/5).

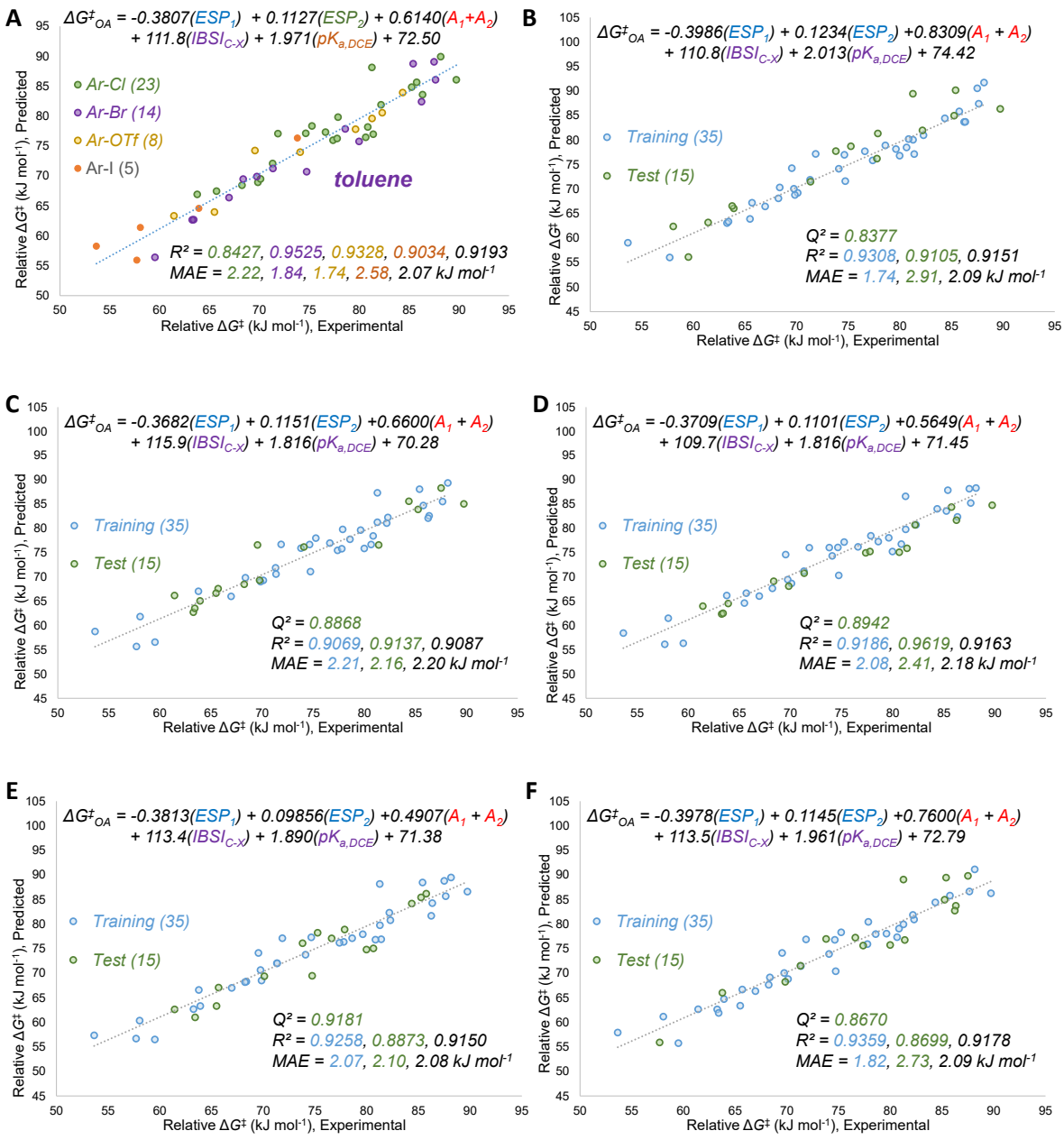


Fig. C21. Performance of the quantitative reactivity model for oxidative addition to Pd(0) in 1:1 **toluene**. (B) – (F) Multivariate linear regression model from one of the five 70/30 random split divisions (1/5) – (5/5).

Reference

(1) Rocaboy, R.; Dailler, D.; Baudoin, O. A Four-Step Synthesis of (±)-γ-Lycorane via Pd0-Catalyzed Double C(Sp2)–H/C(Sp3)–H Arylation. *Org. Lett.* **2018**, *20* (3), 772–775. https://doi.org/10.1021/acs.orglett.7b03909.

(2) Mccomas, C. C.; Liverton, N. J.; Habermann, J.; Koch, U.; Narjes, F.; Li, P.; Peng, X.; Soll, R.; Wu, H. Tetracyclic Heterocycle Compounds and Methods of Use Thereof for the Treatment of Viral Diseases. WO2013033900A1, March 14, 2013.

(3) Zhang, E.; Tang, J.; Li, S.; Wu, P.; Moses, J. E.; Sharpless, K. B. Chemoselective Synthesis of Polysubstituted Pyridines from Heteroaryl Fluorosulfates. *Chemistry – A European Journal* **2016**, *22* (16), 5692–5697. https://doi.org/10.1002/chem.201600167.

(4) Keaveney, S. T.; Kundu, G.; Schoenebeck, F. Modular Functionalization of Arenes in a Triply Selective Sequence: Rapid C(Sp2) and C(Sp3) Coupling of C−Br, C−OTf, and C−Cl Bonds Enabled by a Single Palladium(I) Dimer. *Angewandte Chemie International Edition* **2018**, *57* (38), 12573–12577. https://doi.org/10.1002/anie.201808386.

(5) Paenurk, E.; Kaupmees, K.; Himmel, D.; Kütt, A.; Kaljurand, I.; Koppel, I. A.; Krossing, I.; Leito, I. A Unified View to Brønsted Acidity Scales: Do We Need Solvated Protons? *Chem. Sci.* **2017**, *8* (10), 6964–6973. https://doi.org/10.1039/C7SC01424D.

Appendix D: Supporting Information for Chapter 5

Other Supporting information for this chapter include the following:

Lu-PhD-thesis_Ch.5 SI tables. (separate file; *.xlsx, electronically available at *https://doi.org/10.20383/103.0795*)
Extended tables in Microsoft Excel format (.xlsx) that contain list of competition reactions, measured $\Delta G^{\ddagger}_{OA}$ and molecular descriptors.

Lu-PhD-thesis_Ch.5_MLR models. (separate file; *.xlsx, electronically available at *https://doi.org/10.20383/103.0795*)
Extended tables in Microsoft Excel format (.xlsx) that contain predicted $\Delta G^{\ddagger}_{OA}$ calculations, and statistical analysis for multivariate linear regression models.

General Considerations

Materials

All solvents, reagents, and organic substrates were used as purchased from commercial suppliers without further purification with the following exceptions. N-(2-chloropyridin-3-yl)acetamide[1], N-(2,6-dichloropyridin-3-yl)acetamide[1] and 2,6-dichloro-3-methoxypyridne[2] were prepared using published procedures. N-(2-chloropyridin-5-yl)acetamide, 2-chloro-5-isopropoxypyridine, 2-chloro-3-(tert-butoxy)pyridine, 2-chloro-5-(tert-butoxy)pyridine, 2-chloro-3-isopropoxypyridine and 2-chloro-5-bromo-3-methoxypyridine were prepared using the procedures in the *Synthesis of chlorinated pyridines* section. Bis(tricyclohexylphosphine)palladium(0) was purchased from Strem Chemicals and used as received. All reactions were performed inside an MBraun glovebox under an N_2 atmosphere.

Analysis and Spectroscopy

NMR and HRMS analysis refer to section *Analysis and Spectroscopy* in Appendix B.

LCMS analysis and flash column purification refer to section *Analysis and Spectroscopy* in Appendix C.

Experimental Details for Intermolecular Oxidative Addition Competition Reactions

Refer to section *Experimental Details for Oxidative Addition Competition Studies* in Appendix B.

Table D1 contains the ^{31}P NMR chemical shifts and peak area ratios of the 9 competition reactions performed to determine the relative oxidative addition rates of the 6 2-chloropyridine derivatives in Figure 5.5, main text. The 2 additional substrates (2-chloropyridine and 2-chloro-5-methylpyridine) were used to relate the $\Delta\Delta G^{\ddagger}_{OA}$ of each competition reaction to our established reactivity scale to find the $\Delta G^{\ddagger}_{OA}$ of the 6 new

substrates. All ^{31}P NMR chemical shifts are referenced to triphenylphosphine (-6.00 ppm) as an internal chemical shift standard.

Table D1 – ^{31}P NMR chemical shifts and the peak area ratios of 9 competition reactions in THF (Lu-PhD-thesis_Ch.5 SI tables.xlsx).

Experimental Details of reaction progress kinetic analysis for the touchstone reactions

Refer to section *Experimental Details of reaction progress kinetic analysis for the touchstone reactions* in Appendix C.

Table D2 - Observed absolute $\Delta G^{\ddagger}_{OA}$ for the 9 substrates (2-chloro-3/4/5-EDG-pyridine, EDG = OMe, NH_2 and F)) determined previously (see Appendices B and C) and the 6 substrates (2-chloro-3/5-EDG-pyridine, EDG = NHCOMe, O*i*Pr and O*t*Bu) determined from the competition reactions listed in Table D1 (Lu-PhD-thesis_Ch.5 SI tables.xlsx).

Experimental Details for Oxidative Addition Intramolecular Site-selectivity Determination

Site-selectivity of the eight 2,6-dichloropyridine derivatives in Figure 5.6 and the four 2-chloro-5-bromopyridine derivatives in Figure 5.7 to $Pd(PCy_3)_2$ in THF were determined as the following procedures.

Bis(tricyclohexylphosphine)palladium(0) ($Pd(PCy_3)_2$) was used as the palladium source, and THF was used as the reaction solvent. Stock solutions were prepared for each component: the concentration of the substrate stock solutions was 0.375 M, and the concentration of $Pd(PCy_3)_2$ stock solution was 0.0375 M.

In a 4 mL vial containing a stirbar, an aliquot of the substrate stock solution (100 µL, 0.0375 mmol) was diluted with additional reaction solvent (THF, 700 µL), followed by addition of an aliquot of the $Pd(PCy_3)_2$ stock solution (200 µL, 0.00750 mmol). The resulting solution was mixed for 2-18 hours, then a 600 µL sample was transferred to an NMR tube containing a capillary filled with PPh_3 in C_6D_6. A ^{31}P qNMR spectrum was recorded and the ratio of the two products is determined as the peak area ratio.

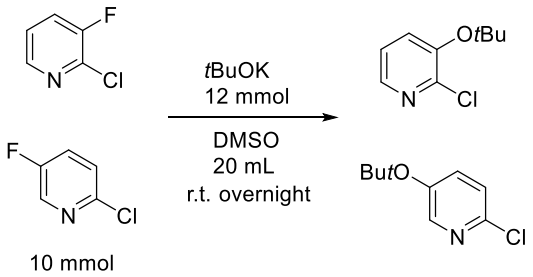
The observed site-selectivity of the 12 substrates is summarized in Table D4 along with the molecular descriptors. Structure elucidation of the regioselective products was carried out by NMR, HRMS and x-ray. The experimental details and characterization data refer to sections *Preparative Scale Synthesis and structural elucidation of the Oxidative Addition Complexes from the substrates in Fig. 5.6* and *Structural elucidation of the Oxidative Addition Complexes from the substrates in Fig. 5.7.*

Synthesis of chlorinated pyridines

2-chloro-3-(tert-butoxy)pyridine and 2-chloro-5-(tert-butoxy)pyridine were prepared by the following procedures.

In the glovebox, a 50 mL round bottom flask containing a stir bar was charged with 10 mmol of the substrate (1.0 equiv.) and 10 mL DMSO. 12 mmol *t*BuOK (1.2 equiv.) was dissolved into 10 mL DMSO and the solution was added dropwise into the round bottom flask. The reaction mixture was stirred at room temperature overnight.

On the next day, the reaction mixture was poured into 30 mL water, then extracted with ethyl acetate (3×20 mL). The combined ethyl acetate phase was washed using saturated NaCl solution (30 mL) then dried over Mg_2SO_4, filtered and evaporated under vacuum. Then the crude product was purified by automated flash chromatography (Biotage Selekt) on silica gel using hexanes/ethyl acetate to obtain the purified product.



Scheme. D1. Reaction scheme for synthesis of 2-chloro-3-(tert-butoxy)pyridine and 2-chloro-5-(tert-butoxy)pyridine.

Characterization of 2-chloro-3-(tert-butoxy)pyridine

504.6 mg of a pale yellow liquid was obtained (27% yield).

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.44 (s, 9H, 9 x $OC(CH_3)_3$), 7.14 (dd, 1H, 1 x Py-H), 7.40 (dd, 1H, 1 x Py-H), 8.08 (dd, 1H, 1 x Py-H).

$^{13}C\{^1H\}$ NMR: (126 MHz, $CDCl_3$): δ 28.85 (3 x $OC(\mathbf{C}H_3)_3$), 82.53 (1 x $O\mathbf{C}(CH_3)_3$), 122.61 (1 x Py), 131.52 (1 x Py), 143.20 (1 x Py), 146.98 (1 x Py), 148.73 (1 x Py).

HRMS (ESI): $[C_9H_{12}ClNO{\cdot}H]^+$ (hydrogen adduct): 186.06802 (calc'd), 186.06804 (found).

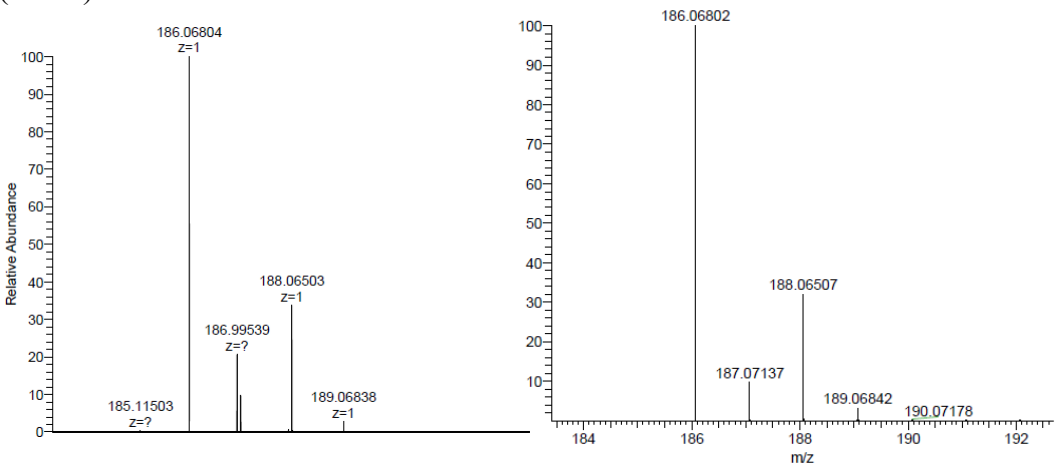


Fig. D1. Left: Experimental HRMS-ESI spectrum of [**Substrate**+H]$^+$. Right: Calculated HRMS isotope pattern for [**Substrate**+H]$^+$.

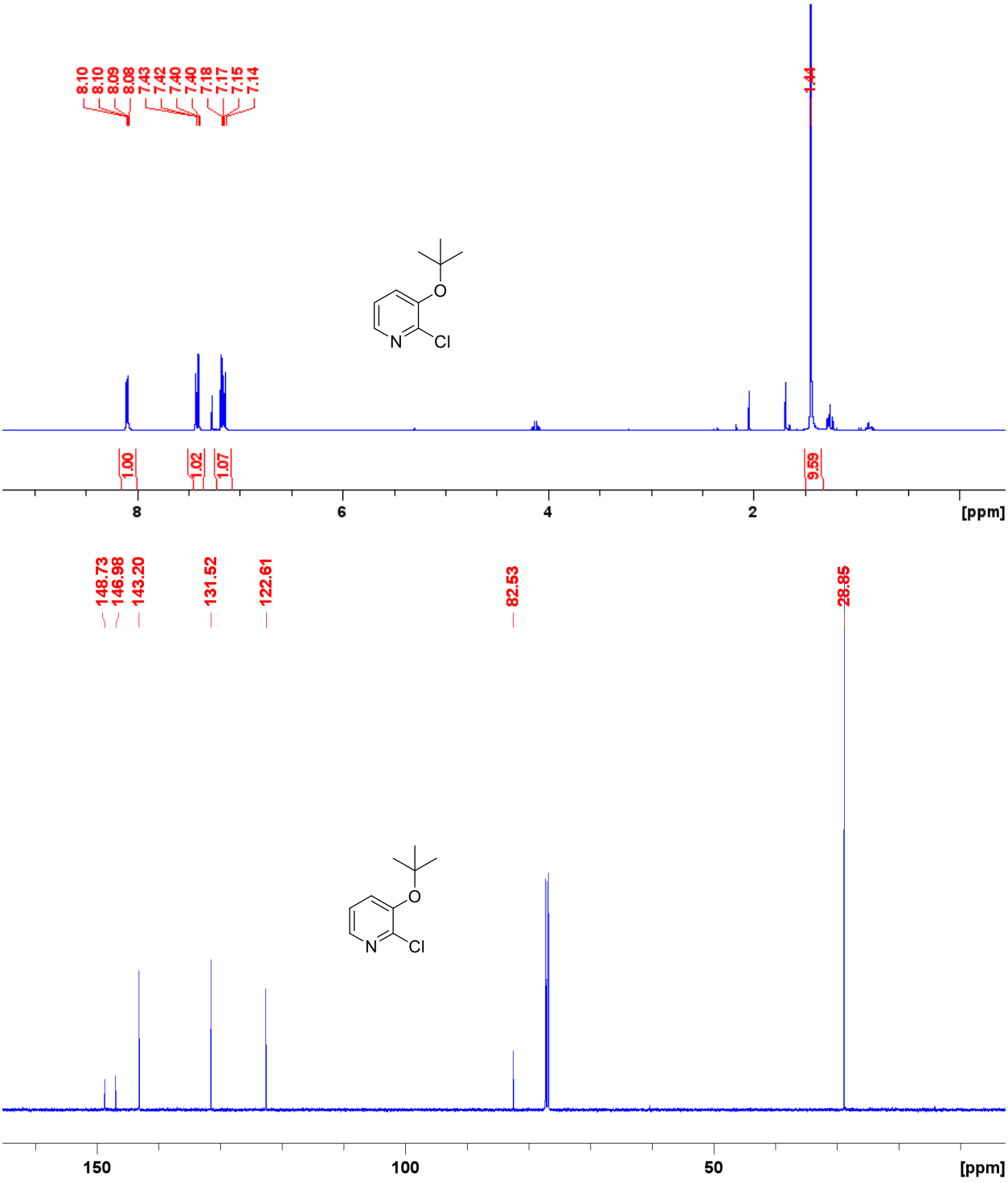


Fig. D2. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **2-chloro-3-(tert-butoxy)pyridine**.

Characterization of 2-chloro-5-(tert-butoxy)pyridine

473.9 mg of a pale-yellow liquid was obtained (26% yield).

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.28 (s, 9H, 9 x $OC(CH_3)_3$), 7.14-7.23 (m, 2H, 2 x Py-H), 8.01 (d, 1H, 1 x Py-H).

$^{13}C\{^{1}H\}$ NMR: (126 MHz, $CDCl_3$): δ 28.54 (3 x $OC(\mathbf{C}H_3)_3$), 80.17 (1 x $O\mathbf{C}(CH_3)_3$), 124.10 (1 x Py), 134.04 (1 x Py), 145.16 (1 x Py), 145.39 (1 x Py), 150.96 (1 x Py).

HRMS (ESI): $[C_9H_{12}ClNO{\cdot}H]^+$ (hydrogen adduct): 186.06802 (calc'd), 186.06807 (found).

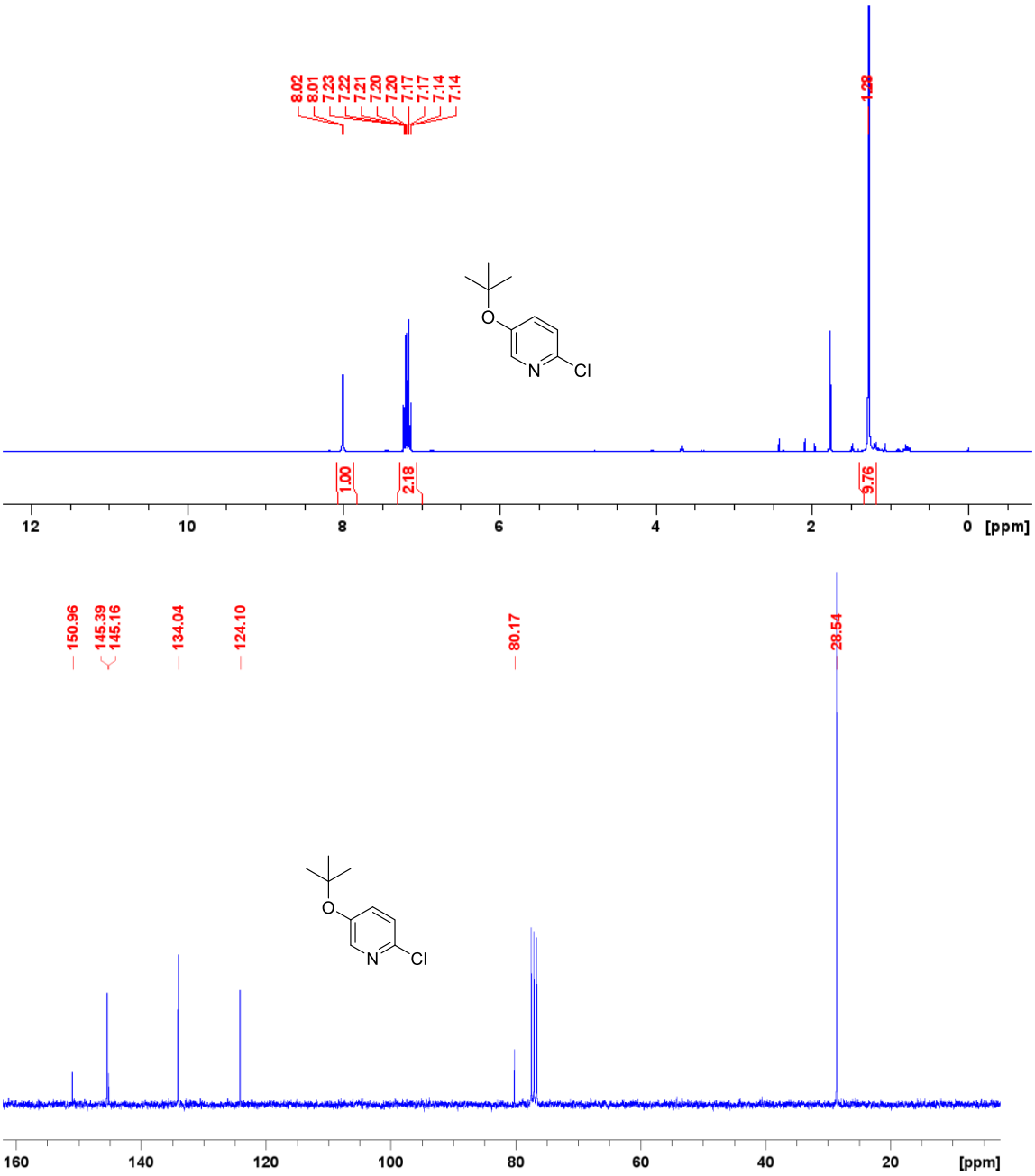


Fig. D3. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **2-chloro-5-(tert-butoxy)pyridine**.

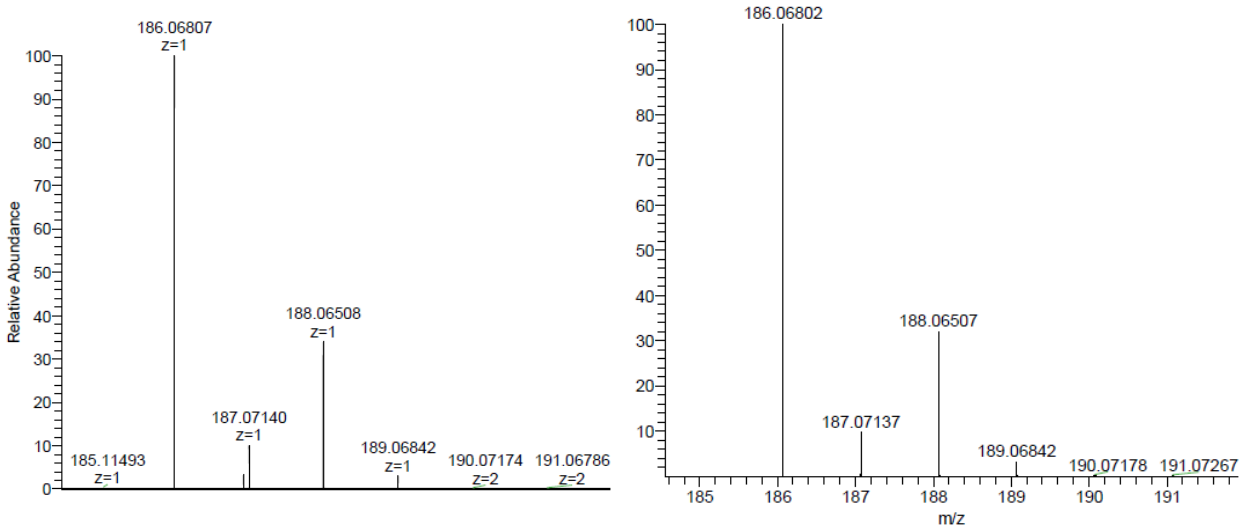
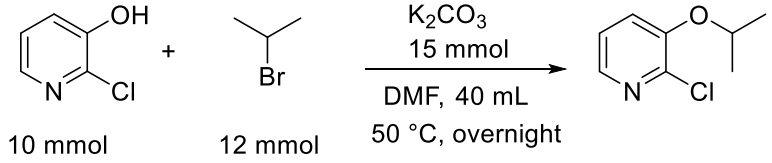


Fig. D4. Left: Experimental HRMS-ESI spectrum of [**Substrate**+H]$^+$; Right: Calculated HRMS isotope pattern for [**Substrate**+H]$^+$ of **2-chloro-5-(tert-butoxy)pyridine.**

2-chloro-3-isopropoxypyridine was prepared by the following procedures.

A 50 mL round bottom flask containing a stir bar was charged with 10 mmol of the substrate (1.0 equiv.), 12 mmol of 2-bromopropane, 15 mmol potassium carbonate and 40 mL DMF. The reaction mixture was stirred at 50 °C overnight.

On the next day, the reaction mixture was poured into 30 mL water, then extracted with ethyl acetate (3×20 mL). The combined ethyl acetate phase was washed using saturated NaCl solution (30 mL) then dried over Mg_2SO_4, filtered and evaporated under vacuum. Then the crude product was purified by automated flash chromatography (Biotage Selekt) on silica gel using hexanes/ethyl acetate to obtain the purified product.



Scheme. D2. Reaction scheme for synthesis of 2-chloro-3-isopropoxypyridine.

Characterization of 2-chloro-3-isopropoxypyridine

956 mg of a colorless liquid was obtained (56% yield).

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.40 (d, 6H, 6 x OCH(C**H**$_3$)$_2$), 4.54-4.61 (septet, 1H, 1 x OC**H**(CH$_3$)$_2$), 7.17-7.23 (m, 2H, 2 x Py-H), 7.98 (dd, 1H, 1 x Py-H).

^{13}C{^{1}H} NMR: (126 MHz, $CDCl_3$): δ 21.84 (2 x OCH(**C**H$_3$)$_2$), 72.37 (1 x O**C**H(CH$_3$)$_2$), 122.44 (1 x Py), 122.98 (1 x Py), 140.55 (1 x Py), 142.22 (1 x Py), 150.27 (1 x Py).

LCMS: 172 ([M+H]$^+$).

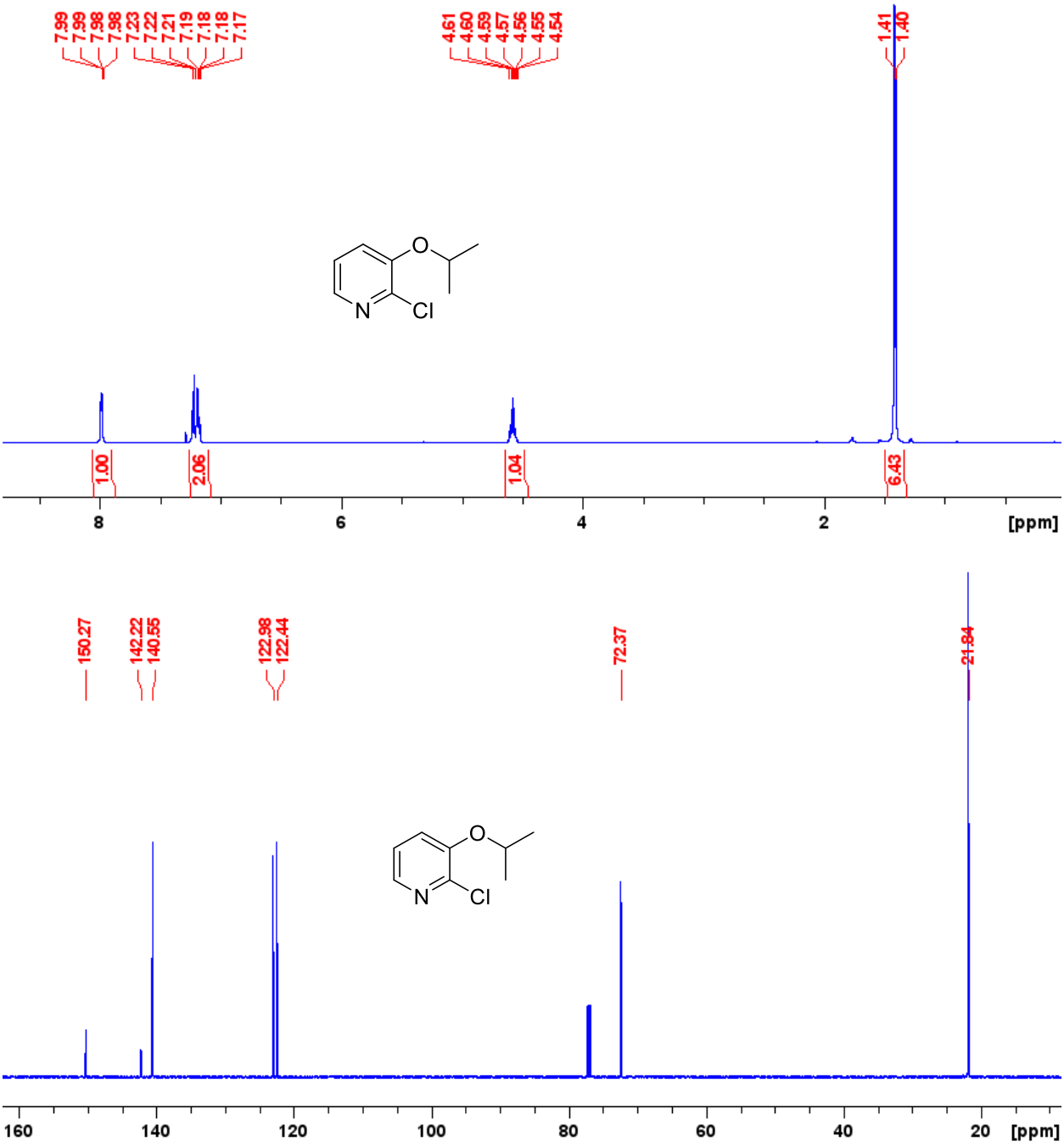


Fig. D5. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **2-chloro-3-isopropoxypyridine**.

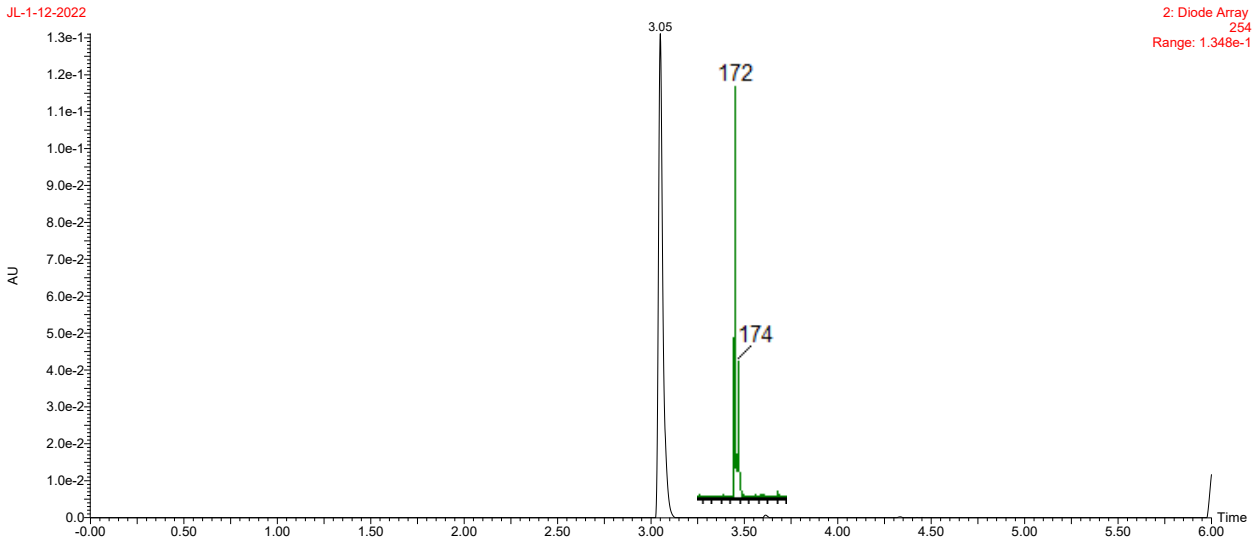
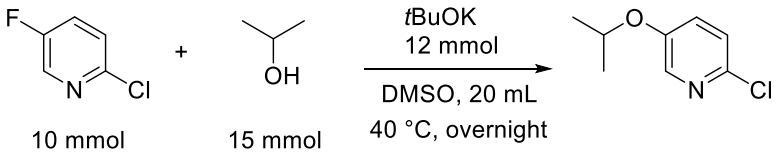


Fig. D6. LCMS chromatogram of **2-chloro-3-isopropoxypyridine**.

2-chloro-5-isopropoxypyridine was prepared by the following procedures.

In the glovebox, a 50 mL round bottom flask containing a stir bar was charged with 10 mmol of the substrate (1.0 equiv.), 15 mmol of isopropanol and 10 mL DMSO. 12 mmol *t*BuOK (1.2 equiv.) was dissolved into 10 mL DMSO and the solution was added dropwise into the round bottom flask. The reaction mixture was stirred at 40 °C overnight.

On the next day, the reaction mixture was poured into 30 mL water, then extracted with ethyl acetate (3×20 mL). The combined ethyl acetate phase was washed using saturated NaCl solution (30 mL) then dried over Mg_2SO_4, filtered and evaporated under vacuum. Then the crude product was purified by automated flash chromatography (Biotage Selekt) on silica gel using hexanes/ethyl acetate to obtain the purified product.



Scheme. D3. Reaction scheme for synthesis of 2-chloro-5-isopropoxypyridine.

Characterization of 2-chloro-5-isopropoxypyridine

903 mg of a colorless liquid was obtained (53% yield).

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.35 (m, 6H, 6 x OCH(C**H**$_3$)$_2$), 4.53 (m, 1H, 1 x OC**H**(CH$_3$)$_2$), 7.15-7.22 (m, 2H, 2 x Py-H), 8.03 (s, 1H, 1 x Py-H).

^{13}C{^{1}H} NMR: (126 MHz, $CDCl_3$): δ 21.78 (2 x OCH(**C**H$_3$)$_2$), 71.22 (1 x O**C**H(CH$_3$)$_2$), 124.39 (1 x Py), 126.06 (1 x Py), 138.10 (1 x Py), 142.15 (1 x Py), 153.33 (1 x Py).

HRMS (ESI): $[C_8H_{10}ClNO{\cdot}H]^+$ (hydrogen adduct): 172.05237 (calc'd), 172.05231 (found).

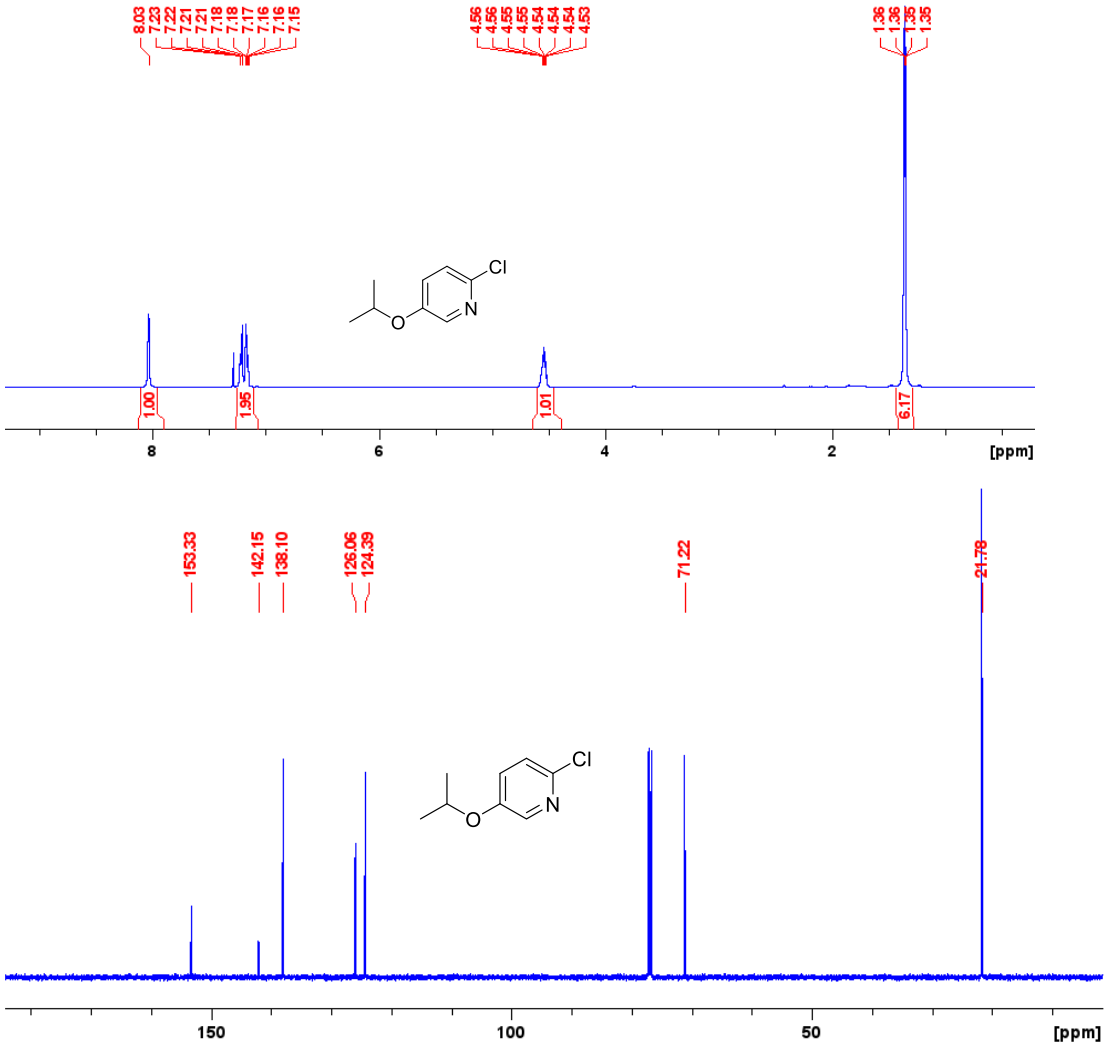


Fig. D7. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **2-chloro-5-isopropoxypyridine**.

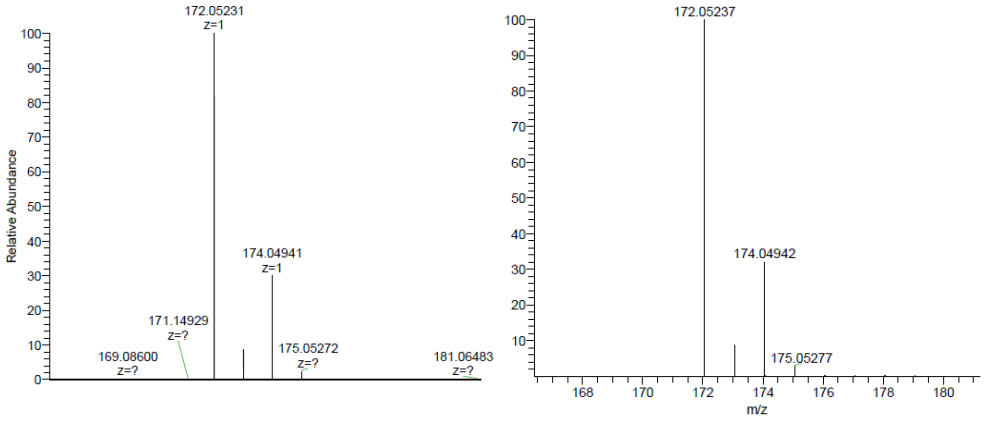
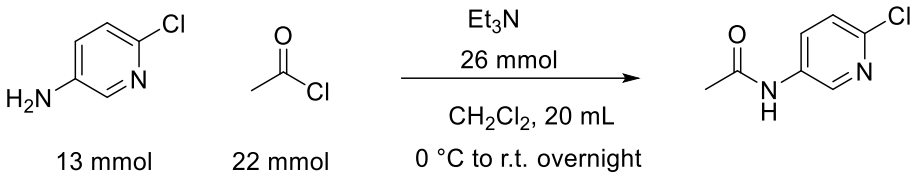


Fig. D8. Left: Experimental HRMS-ESI spectrum of [**Substrate**+H]$^+$. Right: Calculated HRMS isotope pattern for [**Substrate**+H]$^+$ of **2-chloro-5-isopropoxypyridine.**

N-(2-chloropyridin-5-yl)acetamide was prepared by the following procedures.

In open air, a 50 mL round bottom flask containing a stir bar was charged with 13 mmol of the substrate (1.0 equiv.), 26 mmol of triethylamine and 20 mL DMSO. The flask was immersed into an ice bath and 22 mmol of acetyl chloride (1.7 equiv.) was added dropwise while stirring. The reaction mixture was stirred overnight.

On the next day, the reaction mixture was poured into 30 mL ethyl acetate and washed with saturated $NaHCO_3$ solution (2×30 mL). The organic phase was dried over Mg_2SO_4, filtered and evaporated under vacuum. Then the crude product was purified by automated flash chromatography (Biotage Selekt) on silica gel using hexanes/ethyl acetate to obtain the purified product.



Scheme. D4. Reaction scheme for synthesis of N-(2-chloropyridin-5-yl)acetamide.

Characterization of N-(2-chloropyridin-5-yl)acetamide

169 mg of a bright yellow powder was obtained (8% yield).

^{1}H NMR: (500 MHz, d_6-DMSO): δ 2.06 (s, 3H, 3 x NHCOC**H**$_3$), 7.42 (d, 1H, 1 x Py-H), 8.03 (dd, 1H, 1 x Py-H), 8.55 (d, 1H, 1 x Py-H), 10.28 (s, 1H, 1 x N**H**COCH$_3$).

^{13}C{^{1}H} NMR: (126 MHz, d_6-DMSO): δ 23.75 (1 x NHCO**C**H$_3$), 124.09 (1 x Py), 129.47 (1 x Py), 135.44 (1 x Py), 140.08 (1 x Py), 143.24 (1 x Py), 168.91 (1 x NH**C**OCH$_3$).

HRMS (ESI): [$C_7H_7ClN_2O$•H]$^+$ (hydrogen adduct): 171.03197 (calc'd), 171.03191 (found).

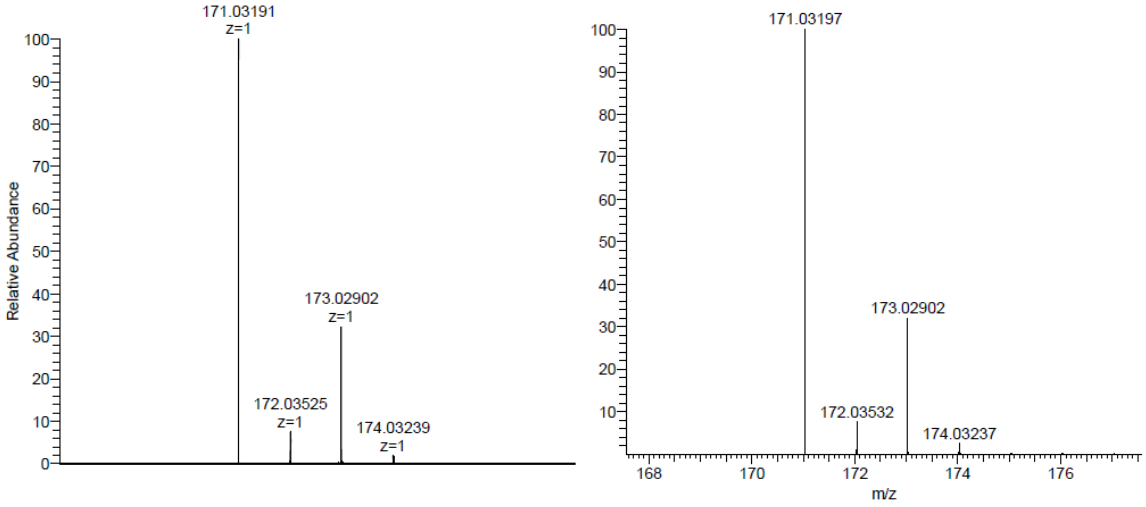


Fig. D9. Left: Experimental HRMS-ESI spectrum of [**Substrate**+H]$^+$. Right: Calculated HRMS isotope pattern for [**Substrate**+H]$^+$ of **N-(2-chloropyridin-5-yl)acetamide**.

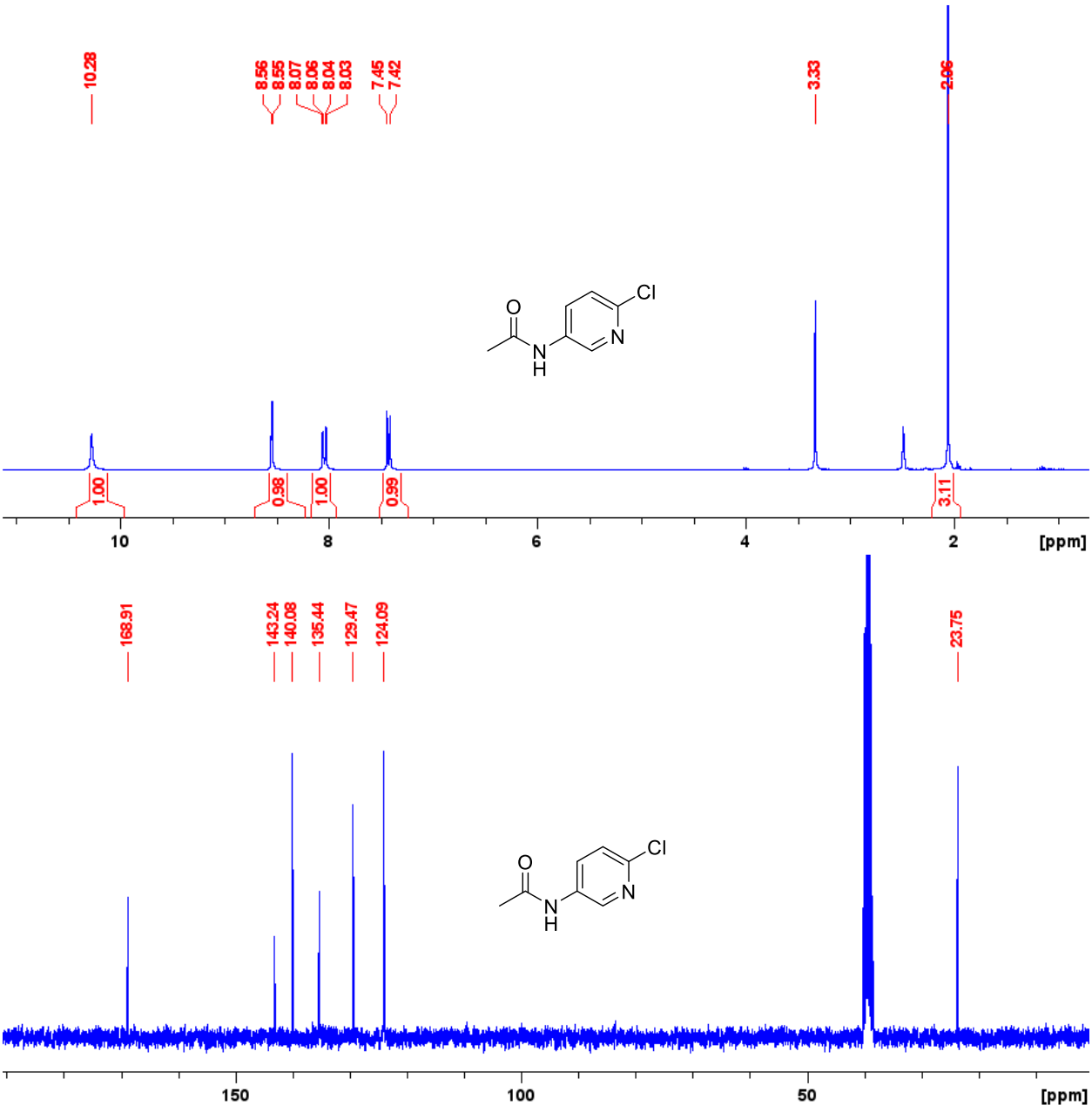
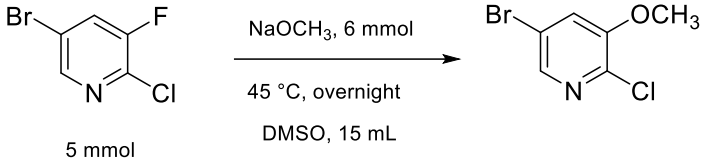


Fig. D10. ^{1}H (500 MHz, d_6-DMSO) and ^{13}C{^{1}H} (126 MHz, d_6-DMSO) NMR spectra of **N-(2-chloropyridin-5-yl)acetamide**. The singlet peak at 3.33 ppm in the ^{1}H spectrum comes from water in the substrate.

5-bromo-2-chloro-3-methoxypyridine was prepared by the following procedures.

In the glovebox, a 50 mL round bottom flask containing a stir bar was charged with 5 mmol of the substrate (1.0 equiv.) and 10 mL DMSO. 6 mmol sodium methoxide (1.2 equiv.) was dissolved into 5 mL DMSO and the solution was added dropwise into the round bottom flask. The reaction mixture was stirred at 45 °C overnight.

On the next day, the reaction mixture was poured into 30 mL water, then extracted with ethyl acetate (3×20 mL). The combined ethyl acetate phase was washed using saturated NaCl solution (30 mL) then dried over Mg_2SO_4, filtered and evaporated under vacuum. Then the crude product was purified by automated flash chromatography (Biotage Selekt) on silica gel using hexanes/ethyl acetate to obtain the purified product.



Scheme. D4. Reaction scheme for synthesis of 5-bromo-2-chloro-3-methoxypyridine.

Characterization of 5-bromo-2-chloro-3-methoxypyridine

415.6 mg of a white solid was obtained (34% yield).

^{1}H NMR: (500 MHz, d_6-benzene): δ 2.78 (s, 3H, 3 x OC**H**$_3$), 6.47 (d, 1H, 1 x Py-H), 7.85 (d, 1H, 1 x Py-H).

^{13}C{^{1}H} NMR: (126 MHz, d_6-benzene): δ 54.90 (1 x O**C**H$_3$), 118.76 (1 x Py), 121.22 (1 x Py), 139.66 (1 x Py), 140.61 (1 x Py), 151.71 (1 x Py).

LCMS: 222 ([M+H]$^+$).

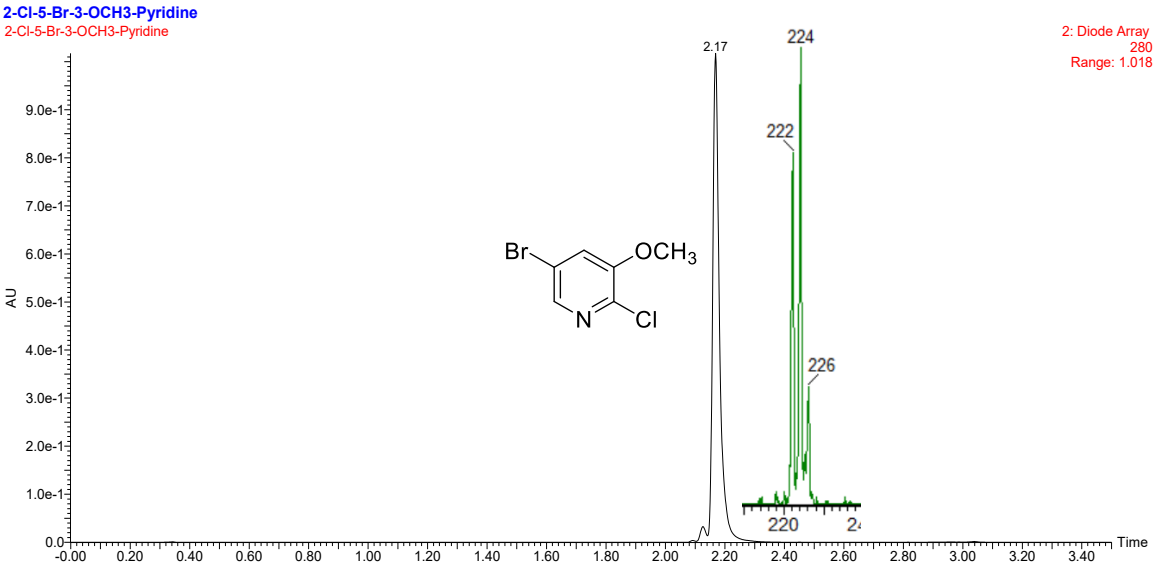


Fig. D11. LCMS chromatogram of **5-bromo-2-chloro-3-methoxypyridine**.

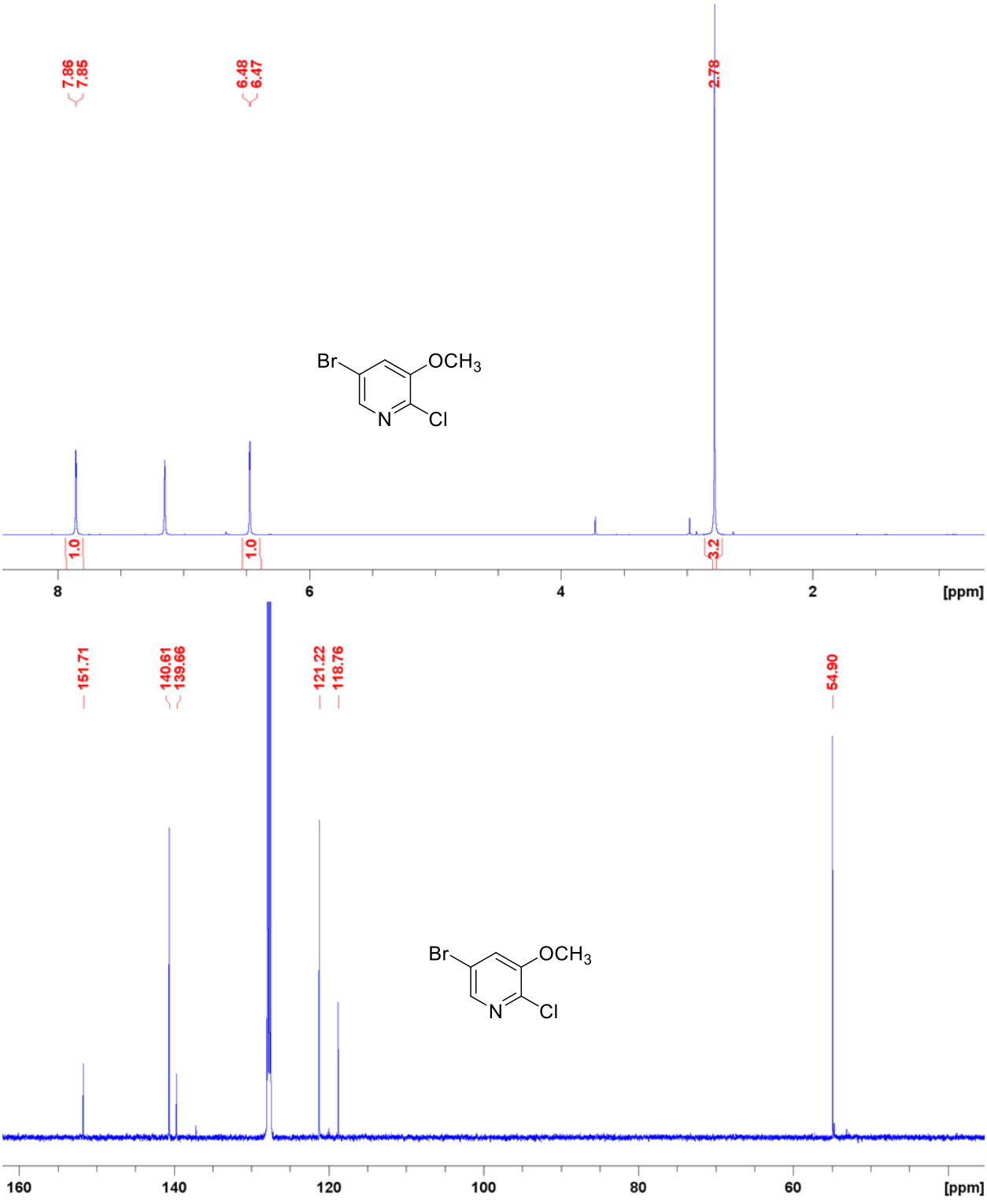


Fig. D12. ^{1}H (500 MHz, d_6-benzene) and ^{13}C{^{1}H} (126 MHz, d_6-benzene) NMR spectra of **5-bromo-2-chloro-3-methoxypyridine**.

Preparative Scale Synthesis and structural elucidation of the Oxidative Addition Complexes from the substrates in Fig. 5.6

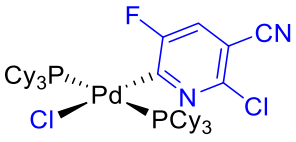
The eight Pd(II) oxidative addition complexes from the dichloro-pyridines in Fig. 5.6 (A1-A8) and one from N-(2-chloropyridin-3-yl)acetamide (S1) were isolated, purified, and characterized by NMR, HRMS using the following general procedure. X-ray structures were obtained for the four oxidative additional products from A1-A4.

All stock solutions were prepared in the following concentrations in THF: substrate (0.375 M), and $Pd(PCy_3)_2$ (0.0375 M).

In the glovebox, a 20 mL vial containing a stir bar was charged with 300 μL of the substrate stock solution (0.1125 mmol, 1.5 equiv.), 2 mL of the $Pd(PCy_3)_2$ stock solution (0.075 mmol, 1.0 equiv.), and 5 mL of THF. The reaction mixture was stirred at room temperature overnight. A $^{31}P\{^1H\}$ NMR spectrum was recorded to confirm the $Pd(PCy_3)_2$ had reacted to completion. The reaction mixture was then transferred outside glovebox for further purification.

The solvent was evaporated under vacuum using a rotary evaporator. The solid left in the vial was washed with either pentane or diethyl ether (2 x 7 mL) to remove unreacted substrate. The solid was dried under vacuum to give the oxidative addition product, and the isolated yield was recorded.

Characterization of Isolated Oxidative Addition Complexes



A1

The oxidative addition product **A1** was prepared by the general procedure using 2,6-dichloro-5-fluoronicotinonitrile. Pentane was used to remove the excess pyridine from its oxidative addition product. 40.5 mg of a pale-yellow solid was obtained (63% yield)

Crystals for X-ray diffraction were grown at 0 °C from THF/pentane (pentane as anti-solvent) by layering pentane on top of a concentrated solution of A1 in THF.

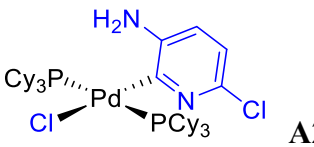
^1H NMR: (500 MHz, $CDCl_3$): δ 0.94-1.89 (m, 66H, 66 x Cy-H), 6.96 (d, 1H, 1 x Py-H).

$^{13}C\{^1H\}$ NMR: (126 MHz, $CDCl_3$): δ 25.61-34.15 (36 x Cy), 102.01 (1 x Py), 115.33 (1 x CN), 121.09 (1 x Py), 121.33 (1 x Py), 144.41 (1 x Py), 158.14 (1 x Py).

$^{19}F\{^1H\}$ NMR: (471 MHz, $CDCl_3$): δ -101.34.

$^{31}P\{^1H\}$ NMR: (203 MHz, $CDCl_3$): δ 23.63.

HRMS (ESI): $[C_{42}H_{67}ClFN_2P_2Pd]^+$ (target compound minus the chlorine): 821.34814 (calc'd), 821.34828 (found).



A2

The oxidative addition product **A2** was prepared by the general procedure using 2,6-dichloro-3-aminopyridine in **toluene** instead of THF, since we have observed much better C_2/C_6 selectivity in toluene (60:1) compared to that in THF (9:1). Diethyl ether was used

to remove the excess pyridine from its oxidative addition product. 39.8 mg of a light brown solid was obtained (64% yield)

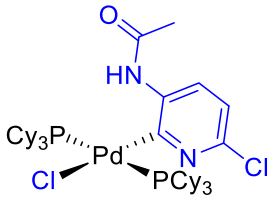
Crystals for X-ray diffraction were grown at 0 °C from chloroform/pentane (pentane as anti-solvent) by layering pentane on top of a concentrated solution of A2 in chloroform.

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.04-2.03 (m, 66H, 66 x Cy-H), 4.23 (s, 2H, 2 x NH_2), 6.37 (d, 1H, 1 x Py-H), 6.62 (d, 1H, 1 x Py-H).

^{13}C{^{1}H} NMR: (126 MHz, $CDCl_3$): δ 26.54-34.06 (36 x Cy), 116.43 (1 x Py), 118.33 (1 x Py), 138.40 (1 x Py), 144.37 (1 x Py), 165.65 (1 x Py).

^{31}P{^{1}H} NMR: (203 MHz, $CDCl_3$): δ 23.07.

HRMS (ESI): $[C_{41}H_{70}ClN_2P_2Pd]^+$ (target compound minus the chlorine): 793.37321 (calc'd), 793.37365 (found).



A3

The $Pd^0(PCy_3)_2$ used here was prepared by the following procedure: in glovebox, 0.075 mmol $Pd^2(OAc)_2(PCy_3)_2$ (1.0 equiv.) and 0.2625 mmol (3.5 equiv.) bis(pinacolato)diboron was dissolved in 6 mL toluene in a glass vial charged with a stir bar. The reaction mixture was stirred at 70 °C overnight. This approach of making $Pd(PCy_3)_2$ has been reported in the published work of xxx. The reaction mixture was brought into next step of reacting with the substrate for oxidative addition without further work-up of isolating the $Pd(PCy_3)_2$.

The oxidative addition product **A3** was prepared by adding 300 μL of the substrate stock solution (0.0825 mmol, 1.1 equiv.) into the above reaction mixture containing $Pd(PCy_3)_2$. The rest steps are same as stated in the general procedure. This reaction was conducted in **toluene** instead of THF, since we have observed much better C_2/C_6 selectivity of the oxidative addition of N-(2,6-dichloropyridin-3-yl)acetamide to $Pd(PCy_3)_2$ in toluene (60:1) compared to that in THF (8:1). Pentane was used to remove the excess pyridine from its oxidative addition product. 41.2 mg of a light brown solid was obtained (63% yield).

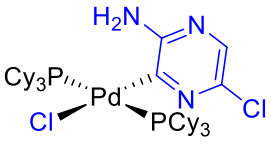
Crystals for X-ray diffraction were grown at 0 °C from THF/pentane (pentane as anti-solvent) by layering pentane on top of a concentrated solution of A3 in THF.

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.03-2.04 (m, 66H, 66 x Cy-H), 2.22 (s, 1H, 1 x NHCOC**H**$_3$), 6.84 (d, 1H, 1 x Py-H), 8.17 (d, 1H, 1 x Py-H), 8.91 (s, 1H, 1 x N**H**$COCH_3$). Protons from bis(pinacolato)diboron are detected at 1.28 ppm.

^{13}C{^{1}H} NMR: (126 MHz, $CDCl_3$): δ 24.89-34.06 (36 x Cy), 34.21 (1 x NHCO**C**H_3), 116.67 (1 x Py), 124.47 (1 x Py), 137.97 (1 x Py), 143.54 (1 x Py), 167.57 (1 x Py), 169.66 (1 x NH**C**OCH_3).

^{31}P{^{1}H} NMR: (203 MHz, $CDCl_3$): δ 24.08.

HRMS (ESI): $[C_{43}H_{72}ClN_2OP_2Pd]^+$ (target compound minus the chlorine): 835.38378 (calc'd), 835.38279 (found).



A4

The oxidative addition product **A4** was prepared by the general procedure using 3,5-dichloro-2-aminopyrazine. Diethyl ether was used to remove the excess pyridine from its oxidative addition product. 35.4 mg of a white solid was obtained (57% yield)

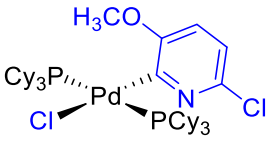
Crystals for X-ray diffraction were grown at 0 °C from THF/pentane (pentane as anti-solvent) by layering pentane on top of a concentrated solution of A4 in THF.

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.06-2.06 (m, 66H, 66 x Cy-H), 5.14 (s, 2H, 2 x NH_2), 7.45 (s, 1H, 1 x Py-H).

$^{13}C\{^{1}H\}$ NMR: (126 MHz, $CDCl_3$): δ 22.34-34.22 (36 x Cy), 132.09 (1 x Py), 136.89 (1 x Py), 155.30 (1 x Py), 165.41 (1 x Py).

$^{31}P\{^{1}H\}$ NMR: (203 MHz, $CDCl_3$): δ 23.23.

HRMS (ESI): $[C_{40}H_{69}ClN_3P2Pd]^+$ (target compound minus the chlorine): 794.36846 (calc'd), 794.36842 (found).



A5

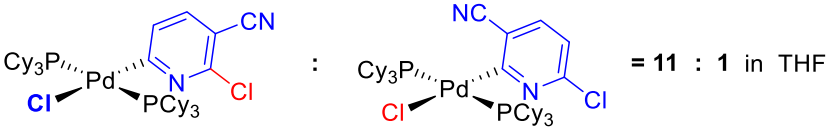
The oxidative addition product **A5** was prepared by the general procedure using 2,6-dichloro-3-methoxypyridine. THF was used to remove the excess pyridine from its oxidative addition product. 38.6 mg of a white solid was obtained (61% yield).

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.04-2.03 (m, 66H, 66 x Cy-H), 3.69 (s, 3H, 3 x OCH_3), 6.44 (d, 1H, 1 x Py-H), 6.76 (d, 1H, 1 x Py-H).

$^{13}C\{^{1}H\}$ NMR: (126 MHz, $CDCl_3$): δ 26.66-33.90 (36 x Cy), 54.17 (1 x OCH_3), 113.68 (1 x Py), 115.87 (1 x Py), 140.50 (1 x Py), 156.27 (1 x Py), 171.12 (1 x Py). [the A5 complex does not fully dissolve in $CDCl_3$, and the low sample concentration causes low peaks for the substituted carbons on the pyridyl ring. ^{13}C-^{1}H HSQC and ^{13}C-^{1}H HMBC were collected to confirm all the pyridyl carbon signals.]

$^{31}P\{^{1}H\}$ NMR: (203 MHz, $CDCl_3$): δ 21.18.

HRMS (ESI): $[C_{42}H_{71}ClNOP_2Pd]^+$ (target compound minus the chlorine): 808.37288 (calc'd), 808.37340 (found); $[C_{42}H_{71}Cl_2NOP_2Pd•H]+$ (hydrogen adduct): 844.34955 (calc'd), 844.34958 (found).



A6

The oxidative addition product **A6** was prepared by the general procedure using 2,6-dichloronicotinonitrile. Pentane was used to remove the excess pyridine from its oxidative addition product. 45.0 mg of a yellow solid was obtained (70% yield).

The regioselective products are identified by the ^{1}H chemical shift. Because of the strong electron donation effect of the PdL_2, the proton para to the reactive carbon should be the most upfield pyridyl protons. An expanded pyridyl protons region from the ^{1}H

spectrum of the isolated products from A6 is shown in Fig.D13: the most upfield peak (δ 6.87) belongs to the proton para to the reactive carbon, and it is the H_A from the C_2-product. The peak area ratio indicates that oxidative addition at C_2 forms the minor product and at C_6 forms the major product.

The chemical shift of the H_A, the proton para to the reactive carbon can be confirmed from the ^{1}H-NMR of one of the oxidative addition complexes we characterized in our prior work: the 2-chloro-3-cyanopyridine oxidative addition to $Pd(PCy_3)_2$. The complex and its ^{1}H spectrum are shown in Fig.D14. Besides the chemical shifts of the pyridyl protons, the coupling constants also provide structural information to confirm the proton assignment to each peak. The coupling constants of the dd peak at 6.82 ppm are 7.5 Hz and 4.8 Hz; and these large coupling constants indicate that it is the H_A coupled to two of its neighboring protons ($^3J_{AB}$ and $^3J_{AC}$). Compared to the coupling constants of the dd peak at 7.51 ppm, 4.8 Hz and 1.9 Hz, the small coupling constant of 1.9 Hz suggests that it is a four-bond coupling ($^4J_{CB}$) and this peak belongs to H_C.

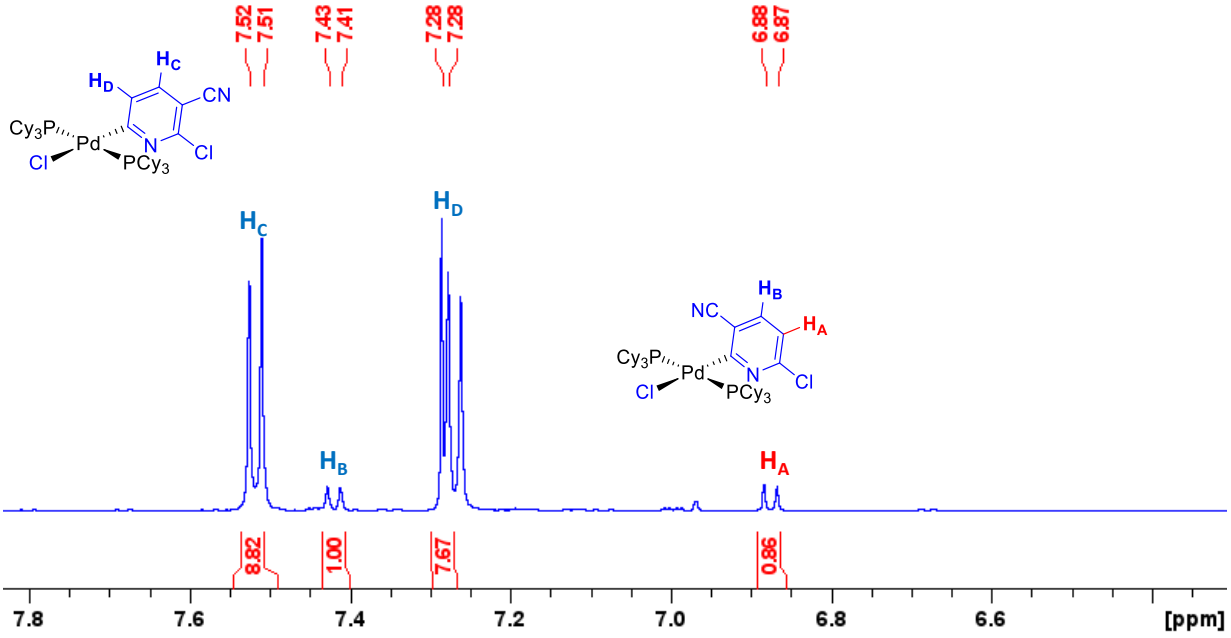


Fig. D13. An expanded pyridyl protons region from the ^{1}H (500 MHz, $CDCl_3$) NMR spectrum of **A6**.

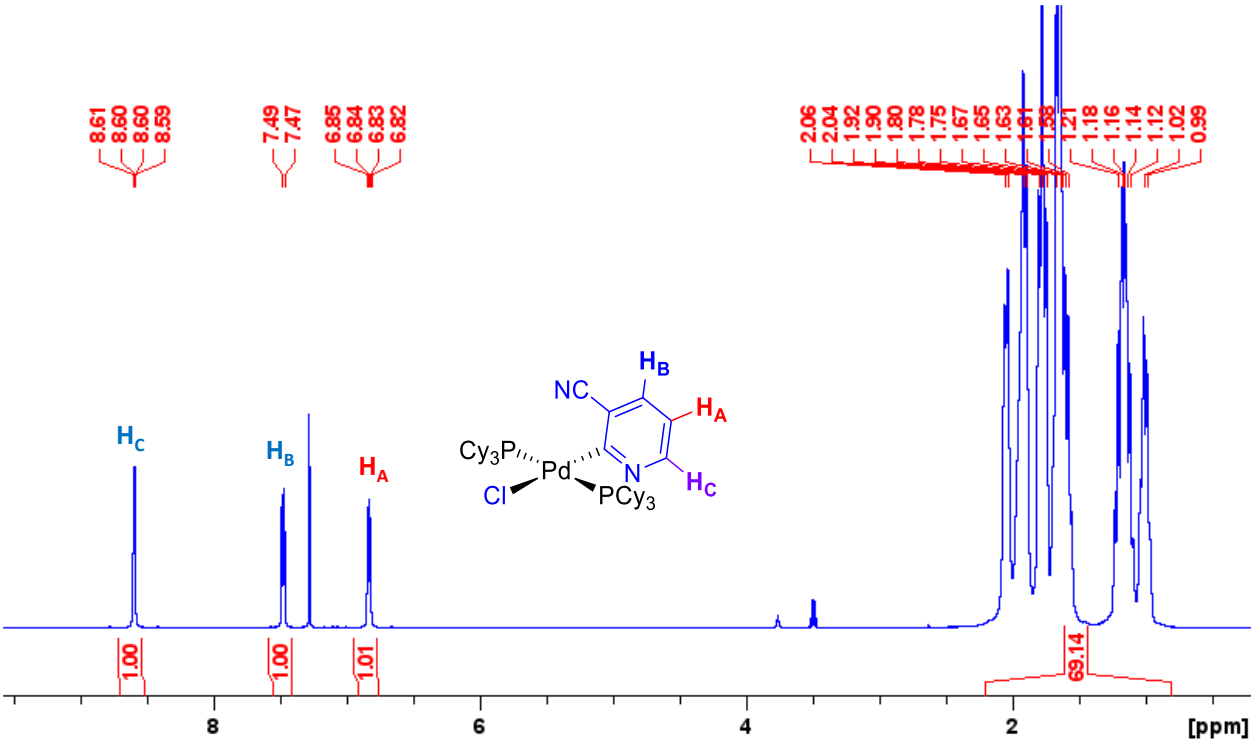


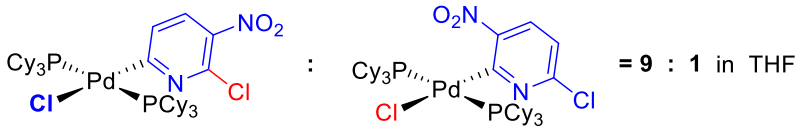
Fig. D14. ^{1}H (500 MHz, $CDCl_3$) NMR spectrum of the oxidative addition complex of 2-chloro-3-cyanopyridien to $Pd(PCy_3)_2$.

^{1}H NMR of the major product: (500 MHz, $CDCl_3$): δ 1.03-1,90 (m, 66H, 66 x Cy-H), pyridyl protons from the major C_6-product: 7.28 (d, 1H, 1 x Py-H, partially overlapped with the $CDCl_3$ peak), 7.51 (d, 1H, 1 x Py-H); pyridyl protons from the minor C_2-product: 6.87 (d, 1H, 1 x Py-H, partially overlapped with the $CDCl_3$ peak), 7.41 (d, 1H, 1 x Py-H); protons from THF are detected 3.76 ppm.

$^{13}C\{^{1}H\}$ NMR of the major product: (126 MHz, $CDCl_3$): δ 25.61-35.41 (36 x Cy), 101.66 (1 x Py), 116.31 (1 x CN), 132.89 (1 x Py), 134.63 (1 x Py), 148.40 (1 x Py), 196.04 (1 x Py). Peak at δ 67.97 is the carbon from THF.

$^{31}P\{^{1}H\}$ NMR: (203 MHz, $CDCl_3$): δ 21.17 (major C_6-product); δ 21.95 (minor C_2-product).

HRMS (ESI): $[C_{42}H_{68}ClN_2P_2Pd]^+$ (target compound minus the chlorine): 803.35756 (calc'd), 803.35765 (found).



C_6-product C_2-product

A7

The oxidative addition product **A7** was prepared by the general procedure using 2,6-dichloro-3-nitropyridine. Pentane was used to remove the excess pyridine from its oxidative addition product. 45.0 mg of a yellow solid was obtained (70% yield).

The regioselective products identification also uses the same ^{1}H chemical shift argument as to the A6 products. An expanded pyridyl protons region is shown in Fig. D15,

and the most upfield peak (δ 6.94) is the H_A from the minor C_2-product, as this proton is para to the reactive carbon.

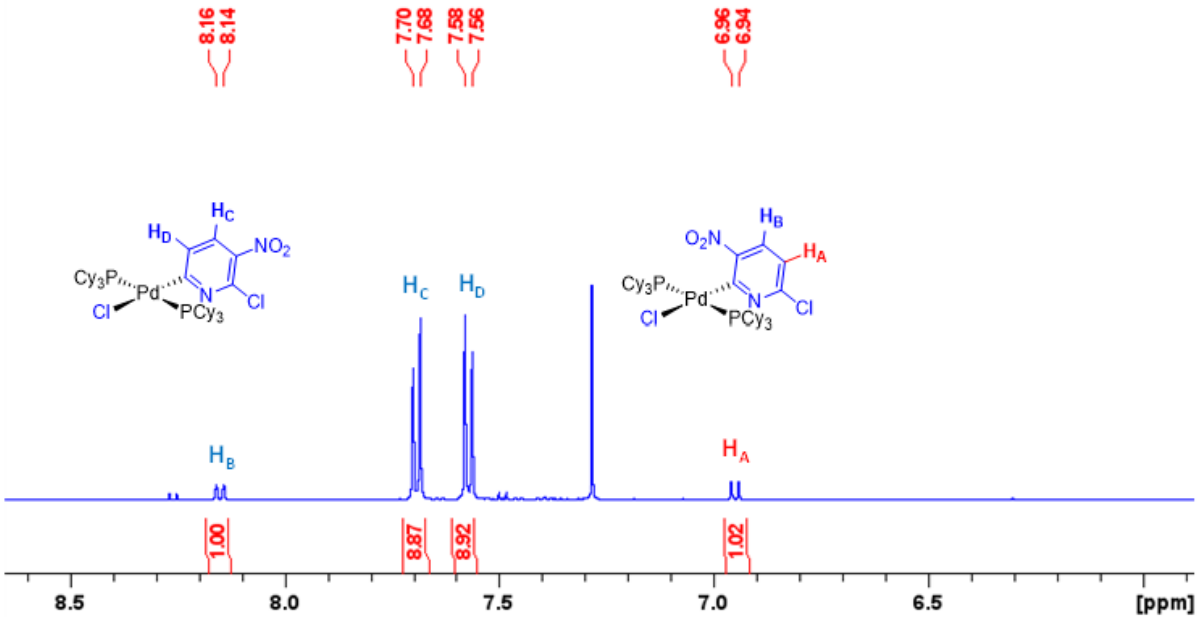
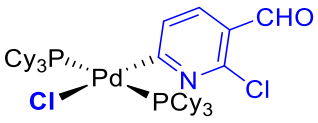


Fig. D15. An expanded pyridyl protons region from the ^1H (500 MHz, $CDCl_3$) NMR spectrum of **A7**.

^1H NMR: (500 MHz, $CDCl_3$): δ 1.04-2.03 (m, 66H, 66 x Cy-H), 4.23 (s, 2H, 2 x NH_2); pyridyl protons from the major C_6-product: 7.56 (d, 1H, 1 x Py-H), 7.68 (d, 1H, 1 x Py-H); pyridyl protons of the minor C_2-product: 8.14 (d, 1H, 1 x Py-H), 6.94 (d, 1H, 1 x Py-H).

$^{13}C\{^1H\}$ NMR: (126 MHz, $CDCl_3$): δ 22.35-34.13 (36 x Cy), 127.48 (1 x Py), 132.84 (1 x Py), 139.05 (1 x Py), 195.99 (1 x Py).

$^{31}P\{^1H\}$ NMR: (203 MHz, $CDCl_3$): δ 21.26 (major C_6-product); δ 19.74 (minor C_2-product).



A8

The oxidative addition product **A8** was prepared by the general procedure using 2,6-Dichloro-3-formylpyridine. Diethyl ether was used to remove the excess pyridine from its oxidative addition product. 50.4 mg of a bright yellow solid was obtained (80% yield).

The regioselective products identification also uses the same ^1H chemical shift argument as to the A6 and A7 regioselective major products. The ^1H NMR spectrum of the A8 product is shown in Fig.D16: the two pyridyl protons show up at 7.55 ppm are most likely from the oxidative addition product at the C_6 site, since the chemical shifts are close to those of the two pyridyl protons from A6 and A7 major products. Also there is no peak detected below 7.0 ppm which is the typical region for the pyridyl proton para to the $Pd^{(II)}$, so the oxidative addition product at C_2 site is unlikely to exist.

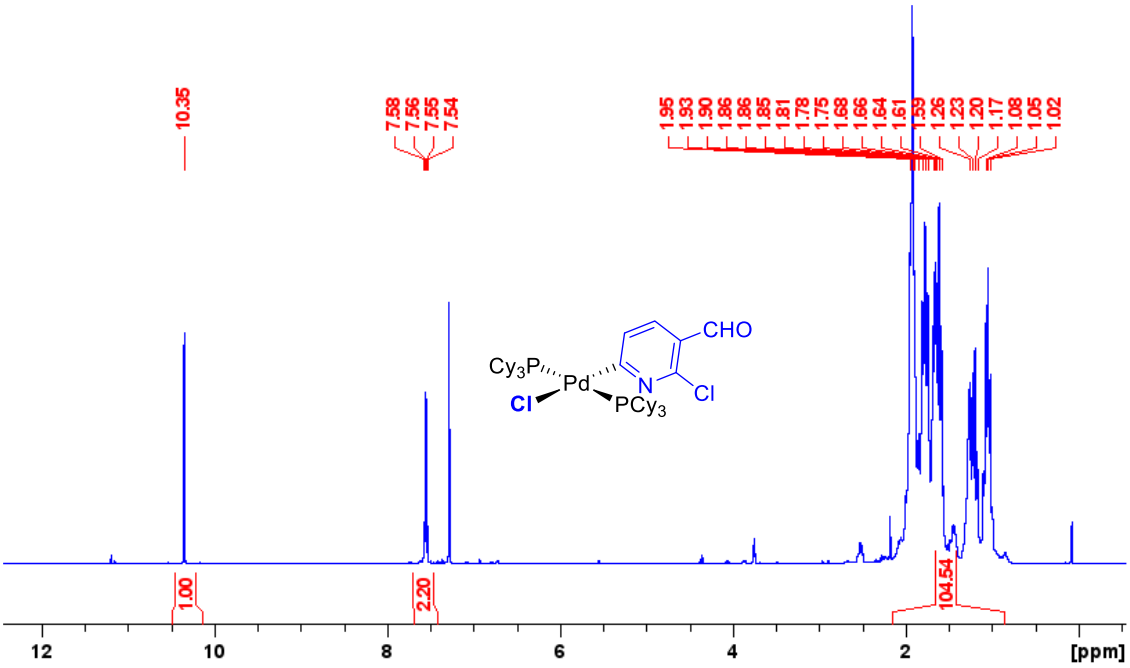


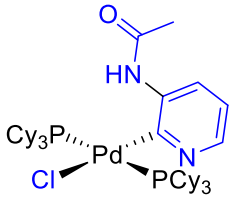
Fig. D16. ^{1}H (500 MHz, $CDCl_3$) NMR spectrum of **A8**.

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.02-1.95 (m, 66H, 66 x Cy-H), 7.54 (dd, 2H, 2 x Py-H), 10.35 (s, 1H, 1 x CHO).

$^{13}C\{^{1}H\}$ NMR: (126 MHz, $CDCl_3$): δ 26.61-35.61 (36 x Cy), 122.40 (1 x Py), 130.28 (1 x Py), 134.06 (t, 1 x Py), 149.96 (1 x Py), 190.89 (1 x Py), 198.61 (1 x CHO).

$^{31}P\{^{1}H\}$ NMR: (203 MHz, $CDCl_3$): δ 21.01.

HRMS (ESI): $[C_{42}H_{69}ClNOP_2Pd]^+$ (target compound minus the chlorine): 806.35723 (calc'd), 806.35764 (found); $[C_{42}H_{69}Cl_2NOP_2Pd•H]+$ (hydrogen adduct): 842.33390 (calc'd), 842.33405 (found).



S1

The oxidative addition product **S1** was prepared by the general procedure using N-(2-chloropyridin-3-yl)acetamide. Pentane was used to remove the excess pyridine from its oxidative addition product. 37.5 mg of a light-yellow solid was obtained (60% yield)

Crystals for X-ray diffraction were grown at 0 °C from THF/pentane (pentane as anti-solvent) by layering pentane on top of a concentrated solution of S1 in THF.

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.01-2.23 (m, 66H, 66 x Cy-H), 6.80 (dd, 1H, 1 x Py-H), 8.16 (d, 1H, 1 x Py-H), 8.23 (dd, 1H, 1 x Py-H), 9.01 (s, 1H, 1 x N**H**$COCH_3$).

$^{13}C\{^{1}H\}$ NMR: (126 MHz, $CDCl_3$): δ 24.97-34.15 (36 x Cy), 34.22 (1 x NHCO**C**H_3), 117.50 (1 x Py), 122.17 (1 x Py), 138.90 (1 x Py), 143.97 (1 x Py), 167.31 (1 x Py), 167.63 (1 x NH**C**OCH_3).

$^{31}P\{^{1}H\}$ NMR: (203 MHz, $CDCl_3$): δ 24.13.

HRMS (ESI): $[C_{43}H_{74}ClN_2OP_2Pd]^+$ (target compound minus the chlorine hydrogen adduct): 837.39943 (calc'd), 837.39885 (found).

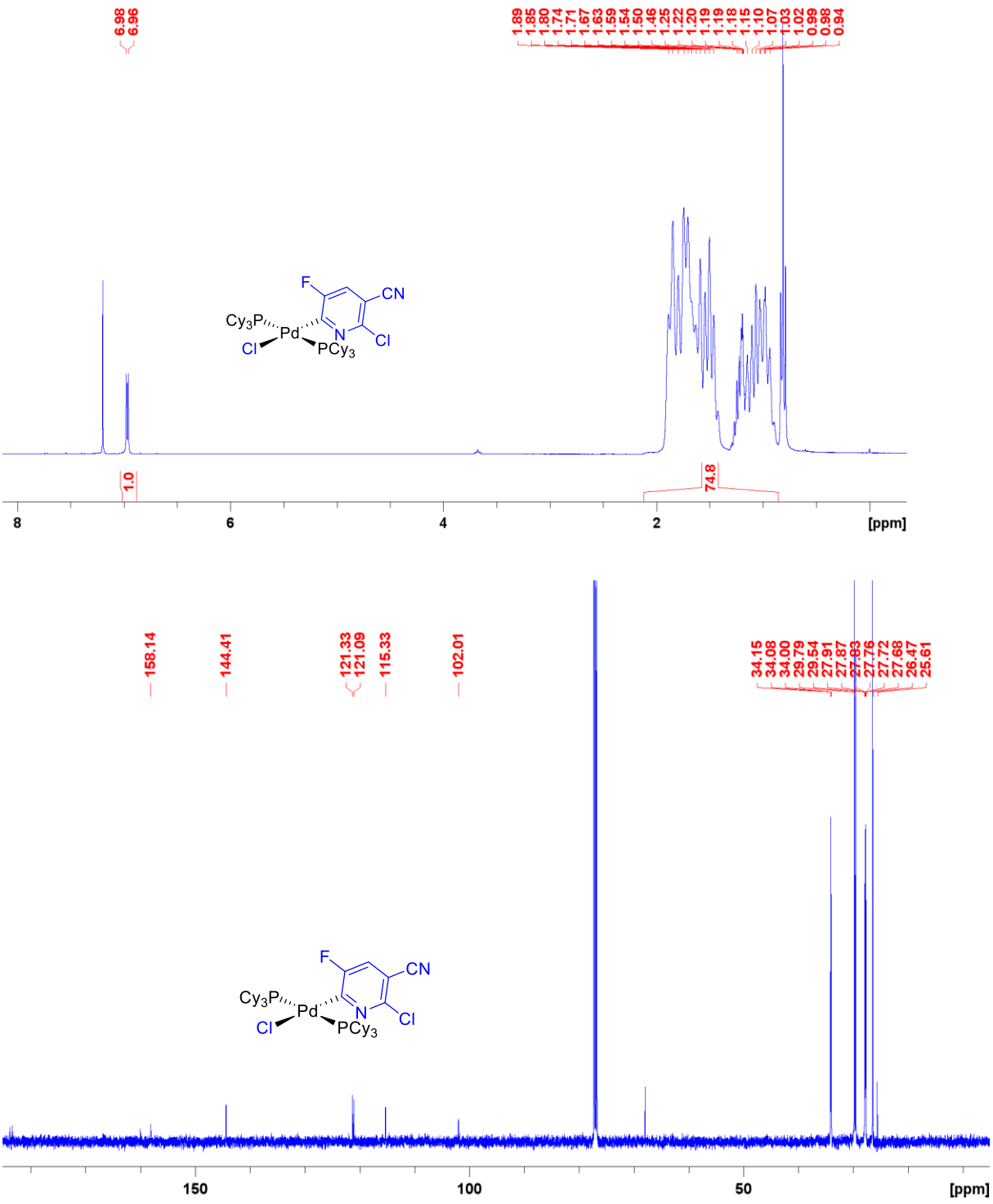


Fig. D17. ^1H (500 MHz, $CDCl_3$) and $^{13}C\{^1H\}$ (126 MHz, $CDCl_3$) NMR spectra of **A1**.

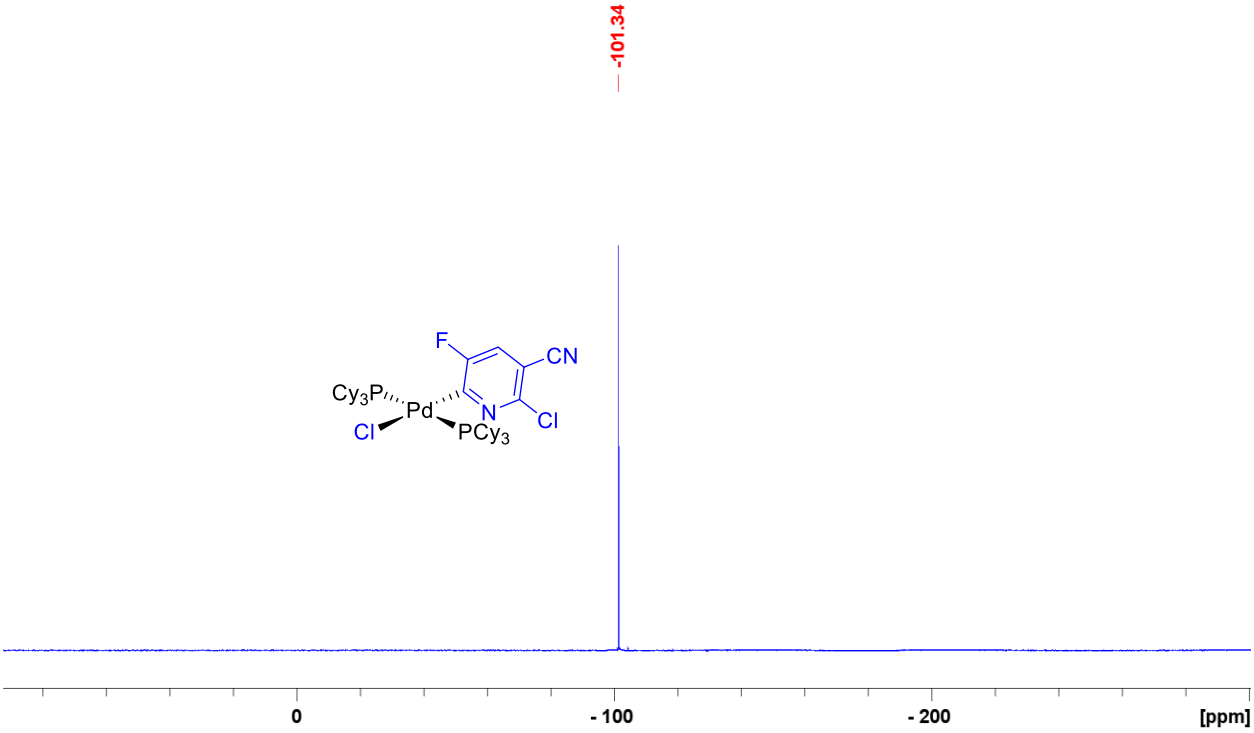


Fig. D18. $^{19}F\{^{1}H\}$ (471 MHz, $CDCl_3$) NMR spectrum of **A1**.

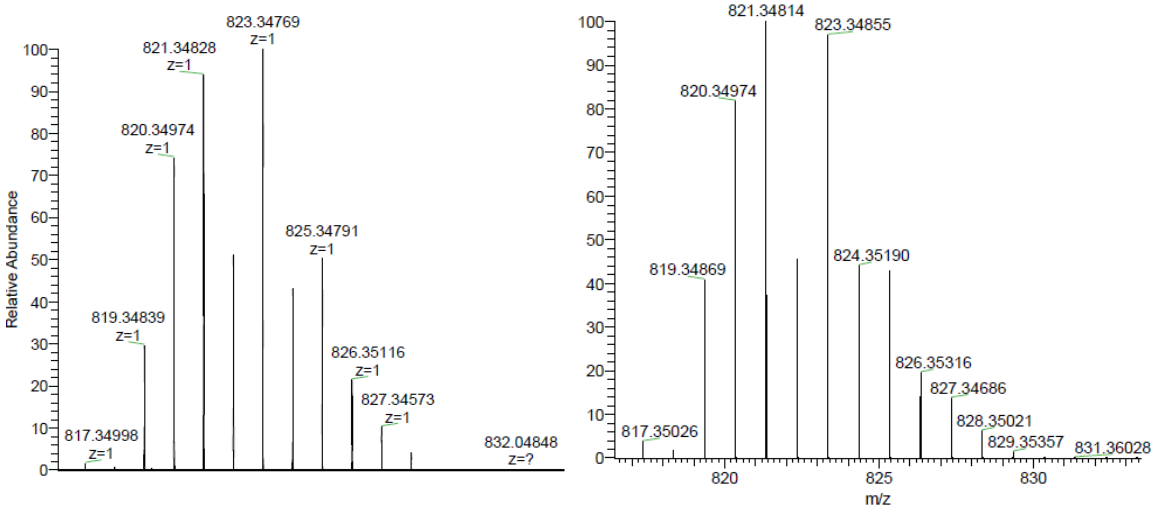


Fig. D19. Left: Experimental HRMS-ESI spectrum of [**A1**-Cl]$^{+}$. Right: Calculated HRMS isotope pattern for [**A1**-Cl]$^{+}$.

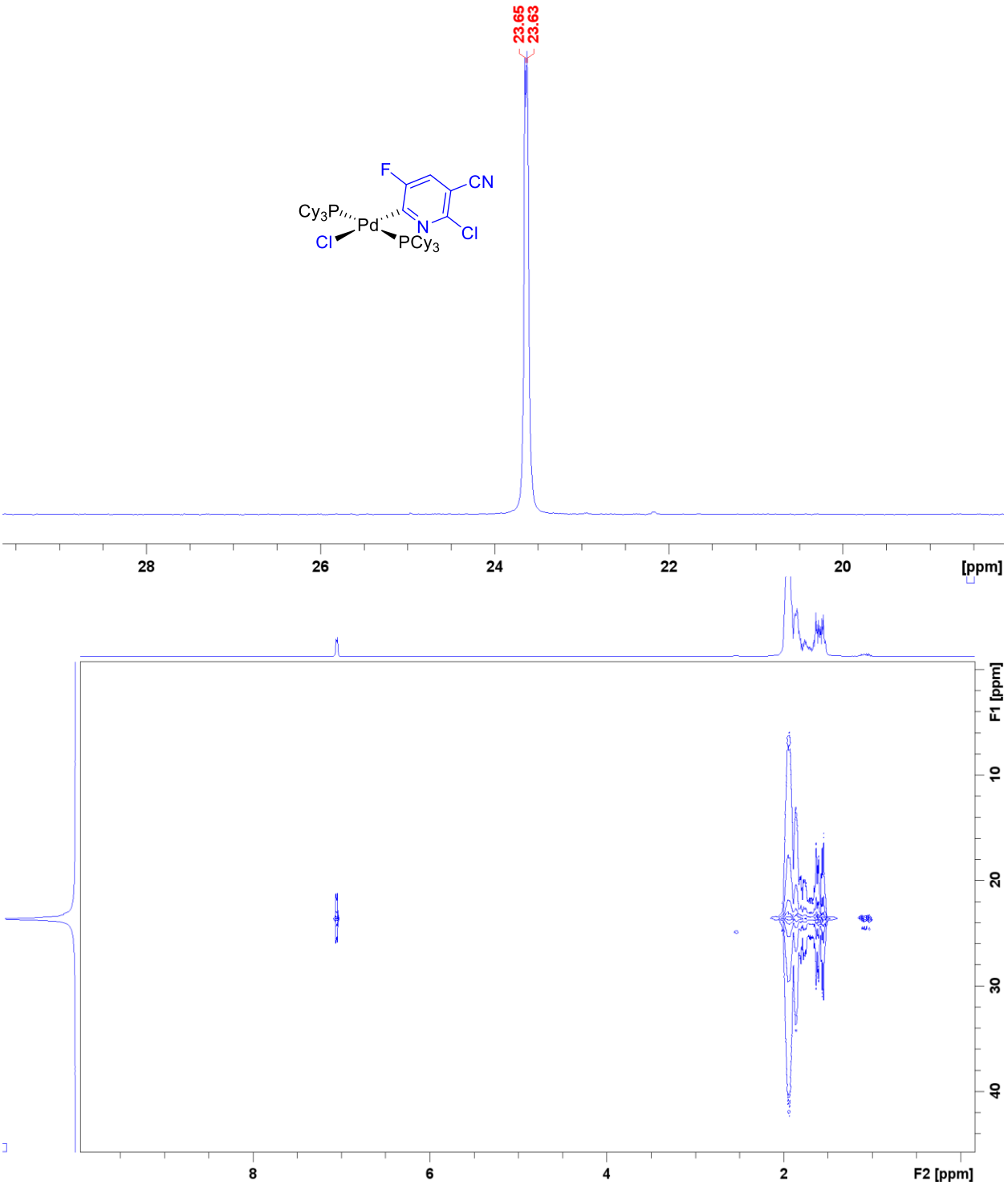


Fig. D20. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **A1**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

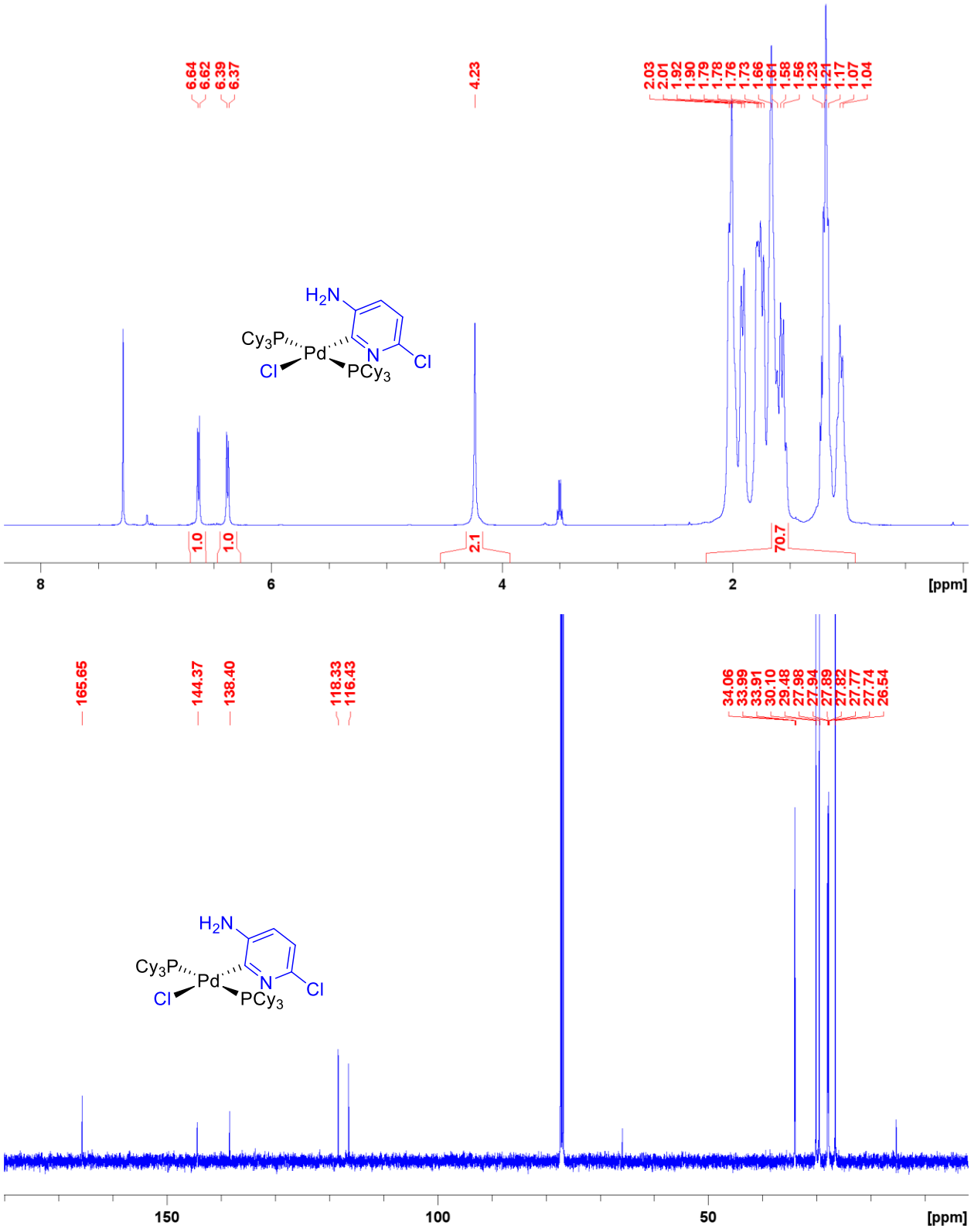


Fig. D21. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **A2**.

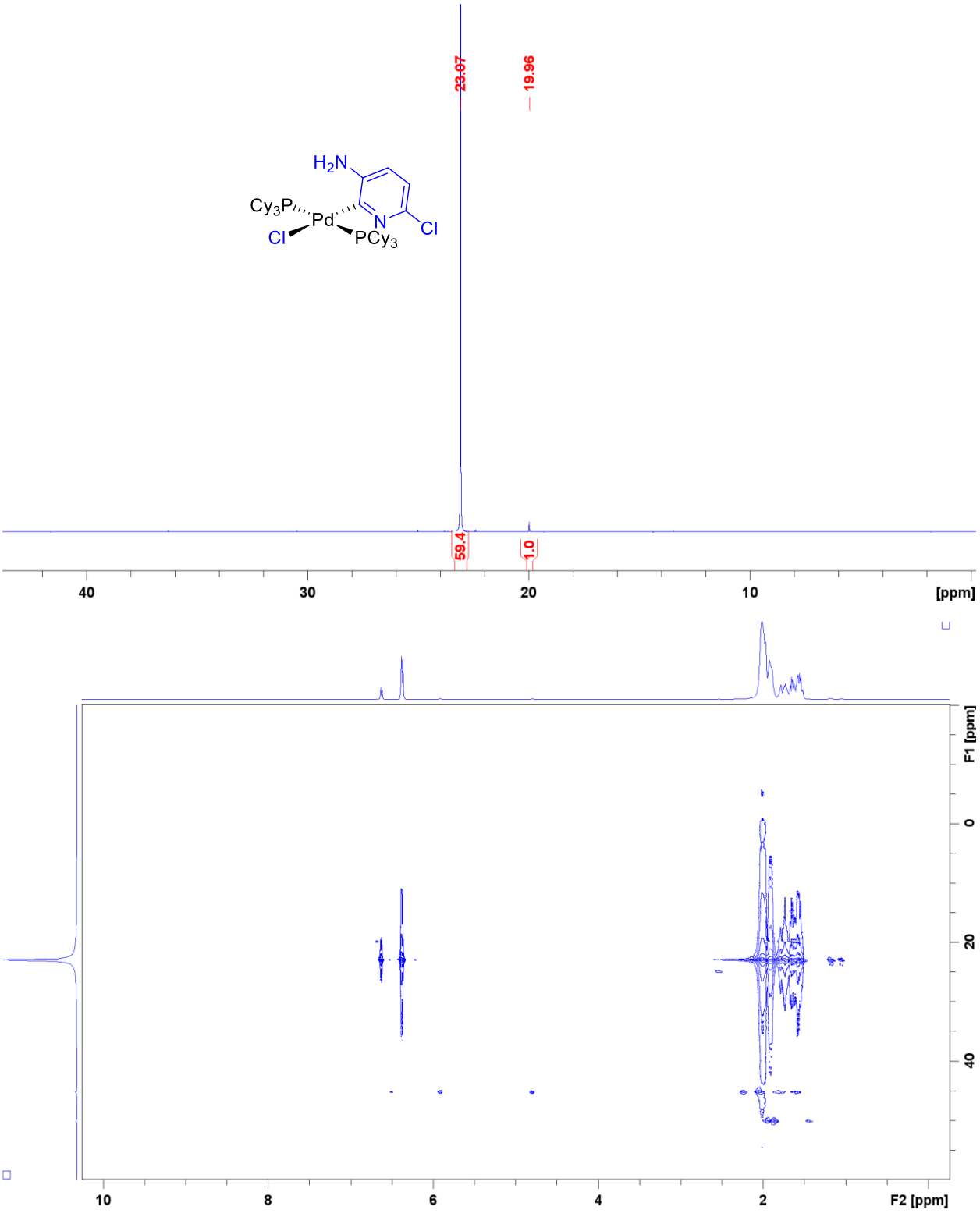


Fig. D22. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **A2**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

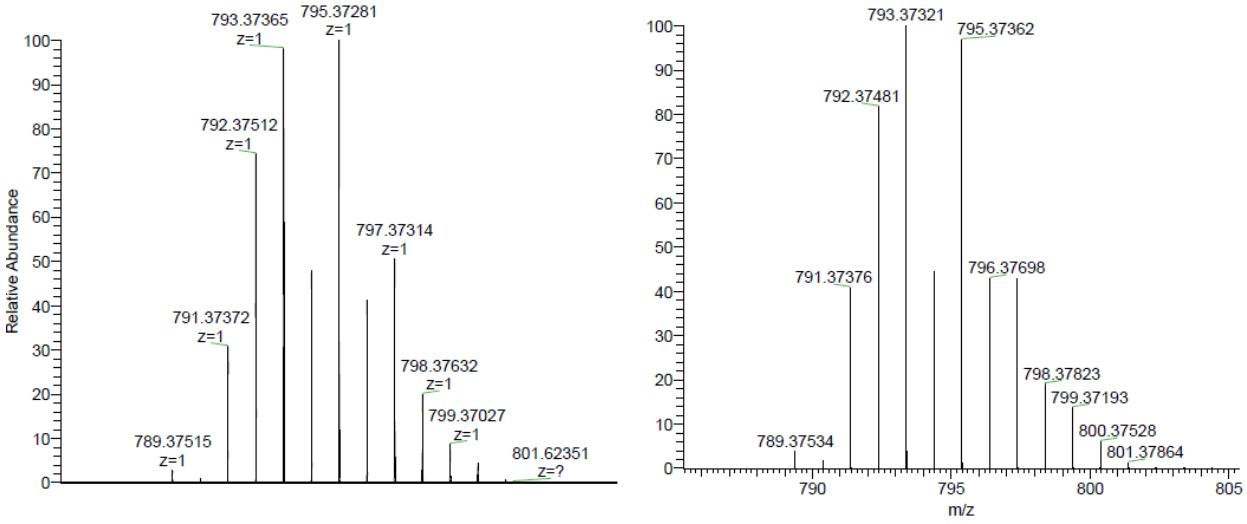


Fig. D23. Left: Experimental HRMS-ESI spectrum of [**A2**-Cl]$^+$. Right: Calculated HRMS isotope pattern for [**A2**-Cl]$^+$.

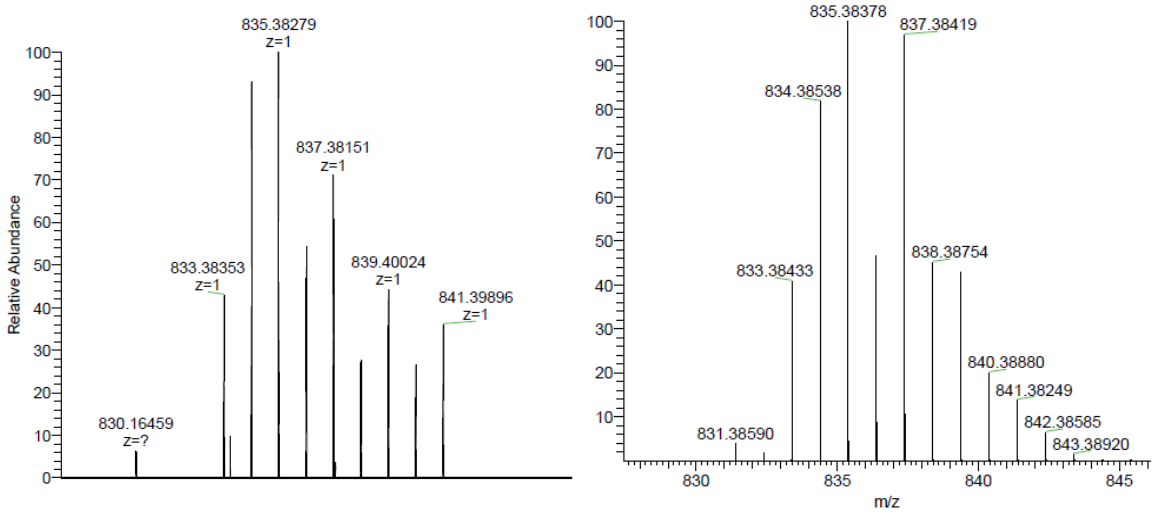


Fig. D24. Left: Experimental HRMS-ESI spectrum of [**A3**-Cl]$^+$. Right: Calculated HRMS isotope pattern for [**A3**-Cl]$^+$.

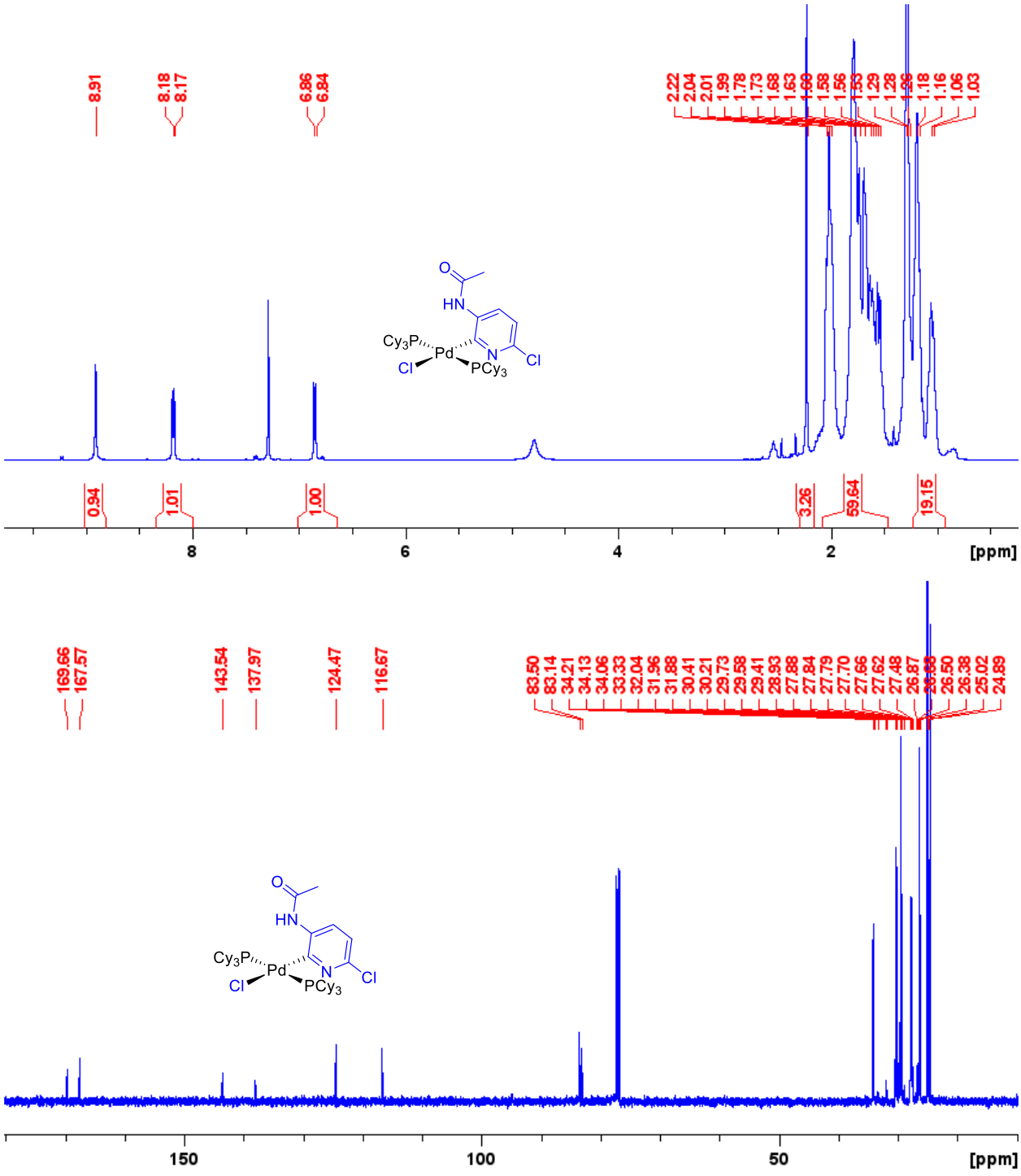


Fig. D25. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **A3**.

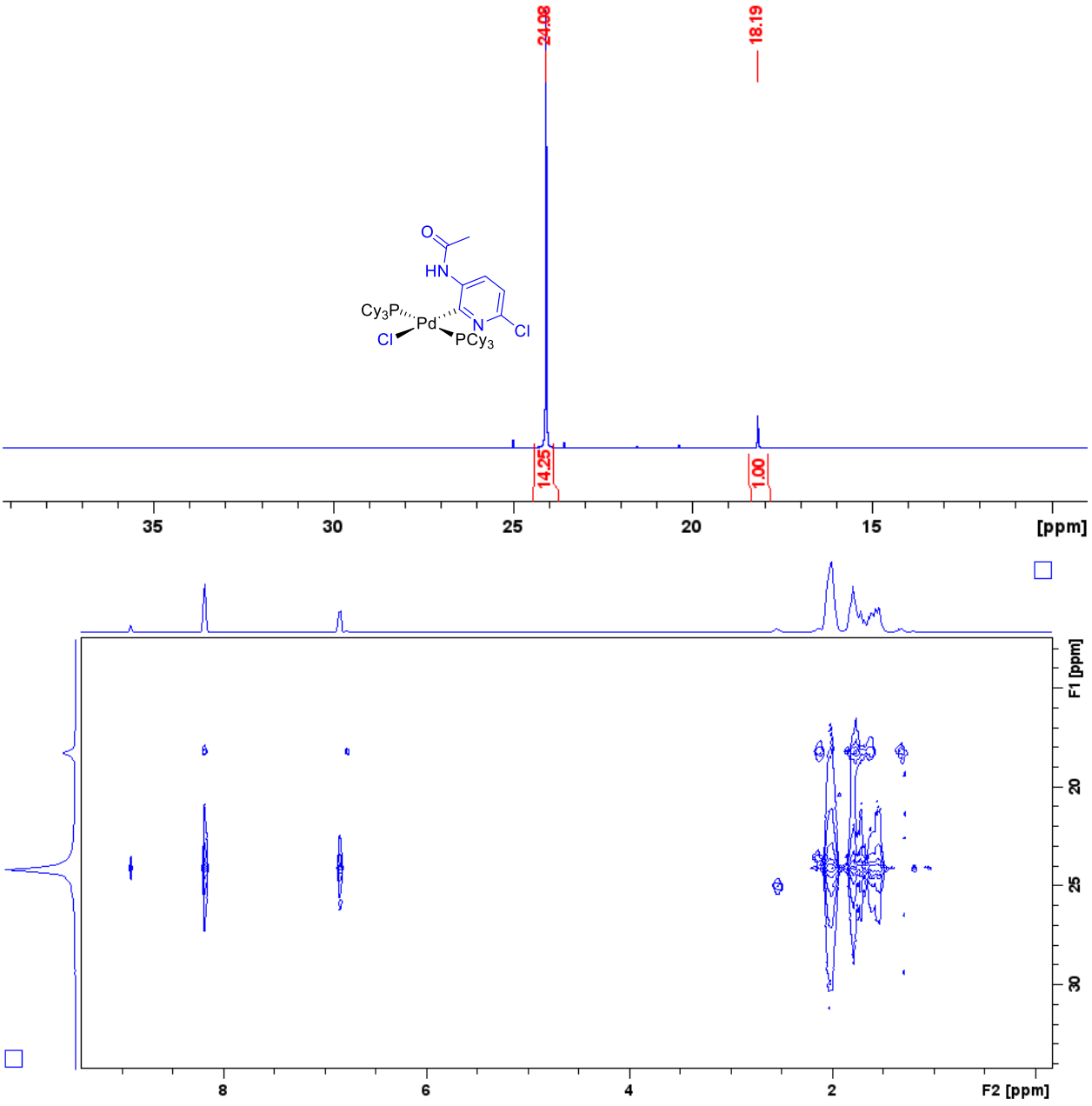


Fig. D26. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **A3**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

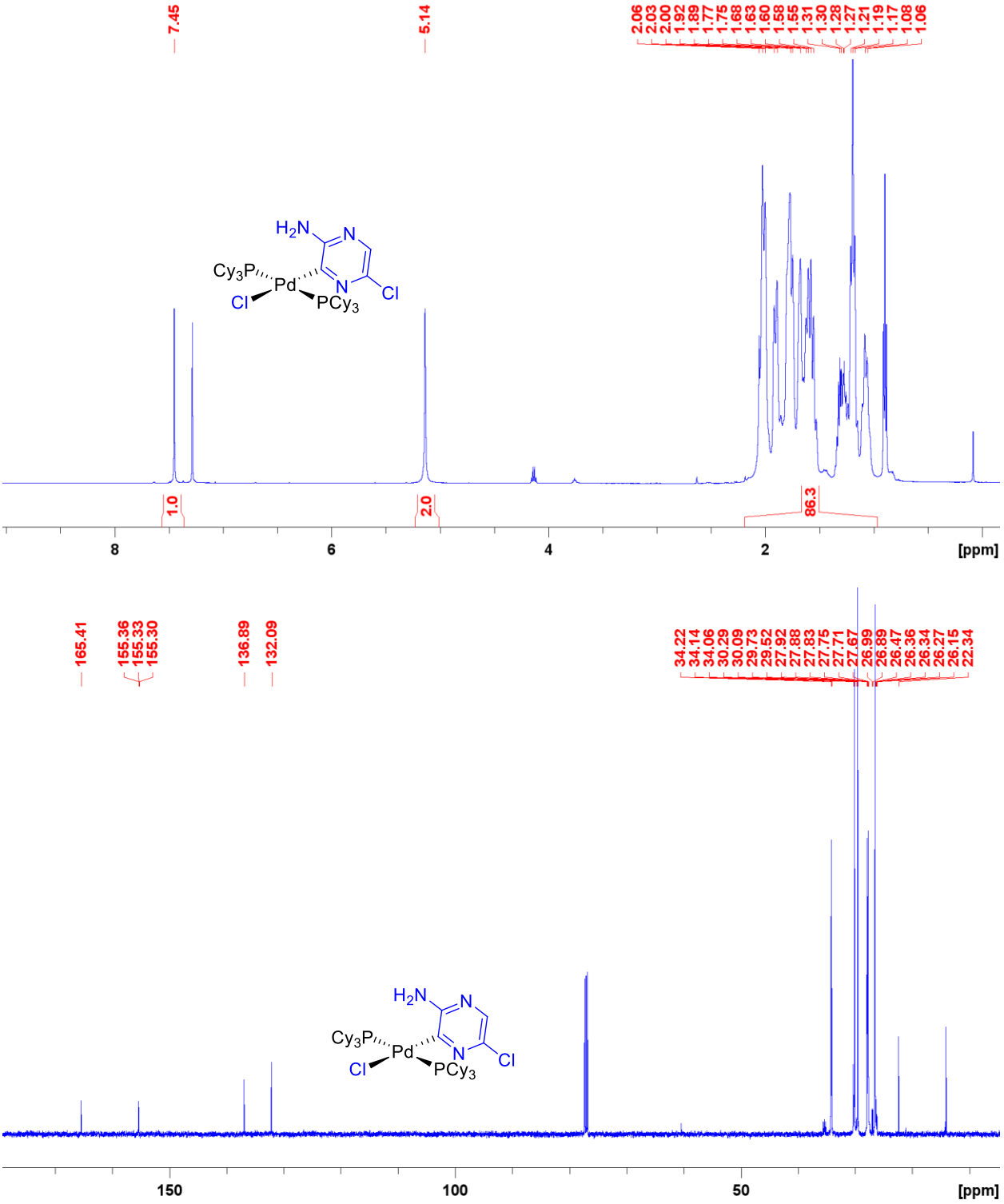


Fig. D27. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **A4**.

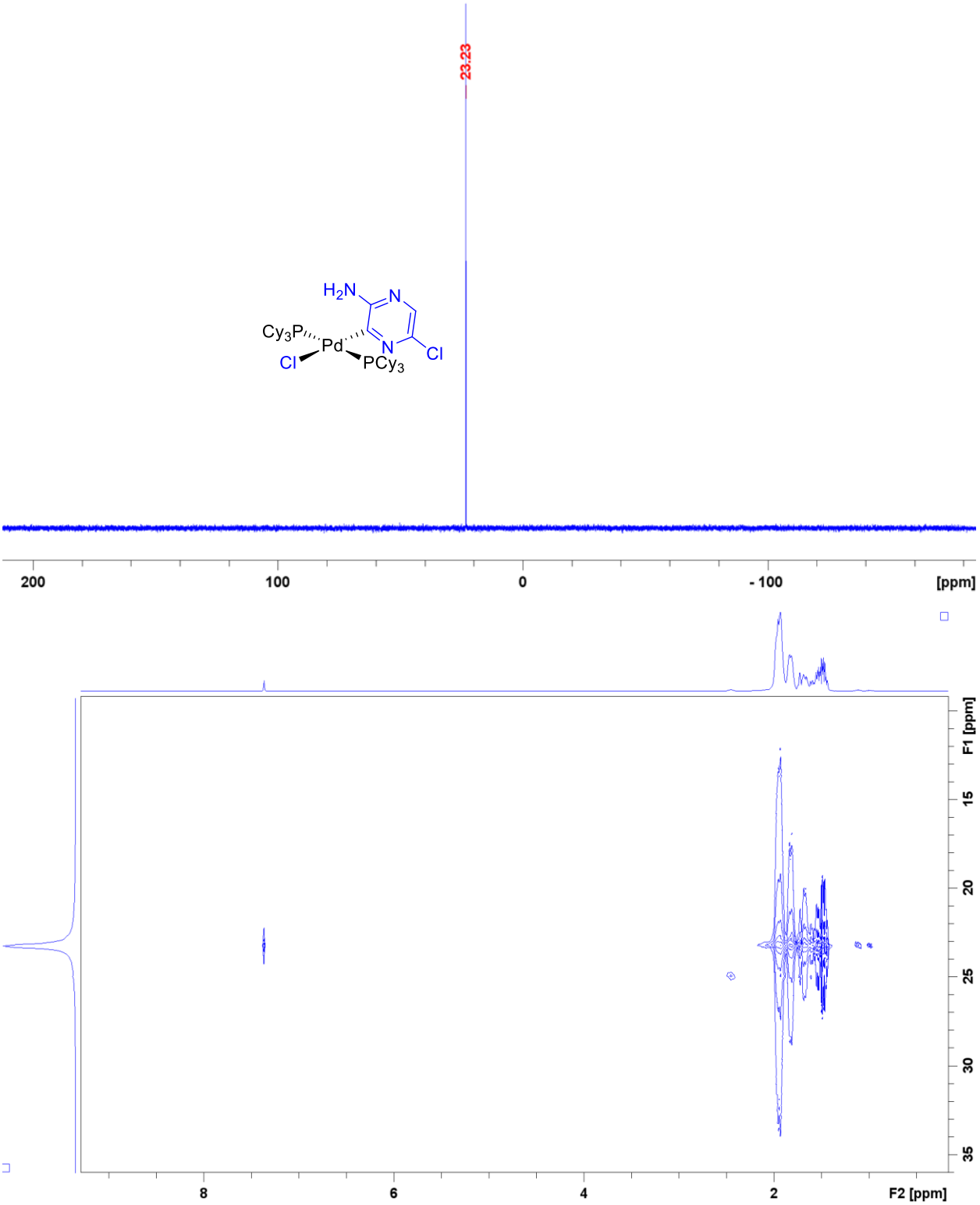


Fig. D28. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **A4**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

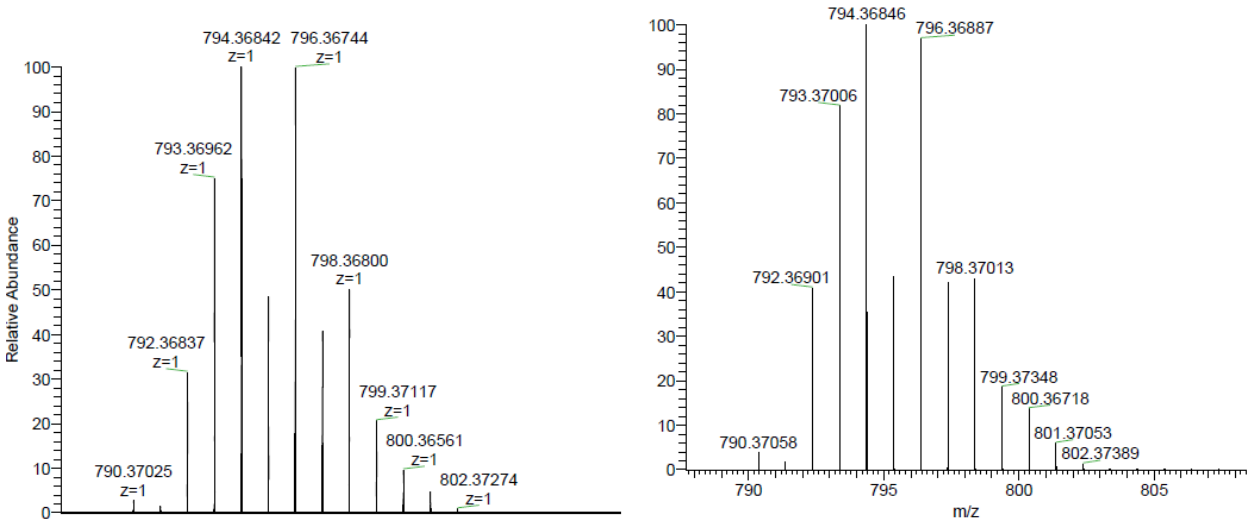


Fig. D29. Left: Experimental HRMS-ESI spectrum of [**A4**-Cl]$^+$. Right: Calculated HRMS isotope pattern for [**A4**-Cl]$^+$.

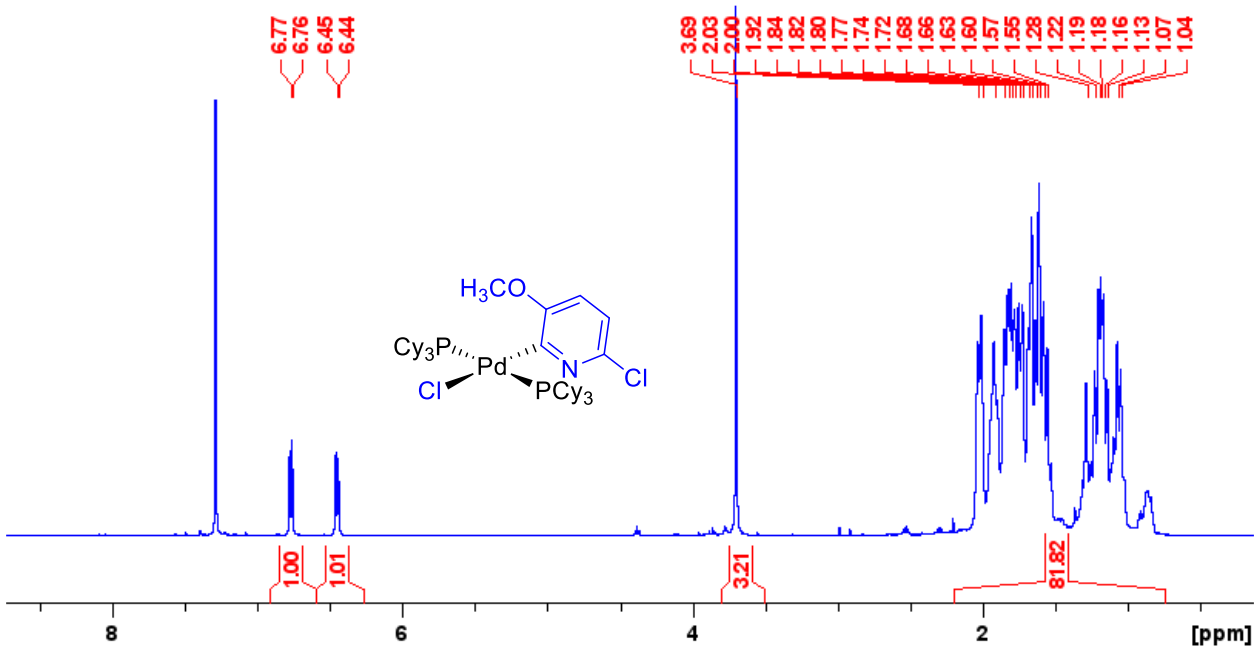


Fig. D30. ^{1}H (500 MHz, $CDCl_3$) NMR spectrum of **A5**.

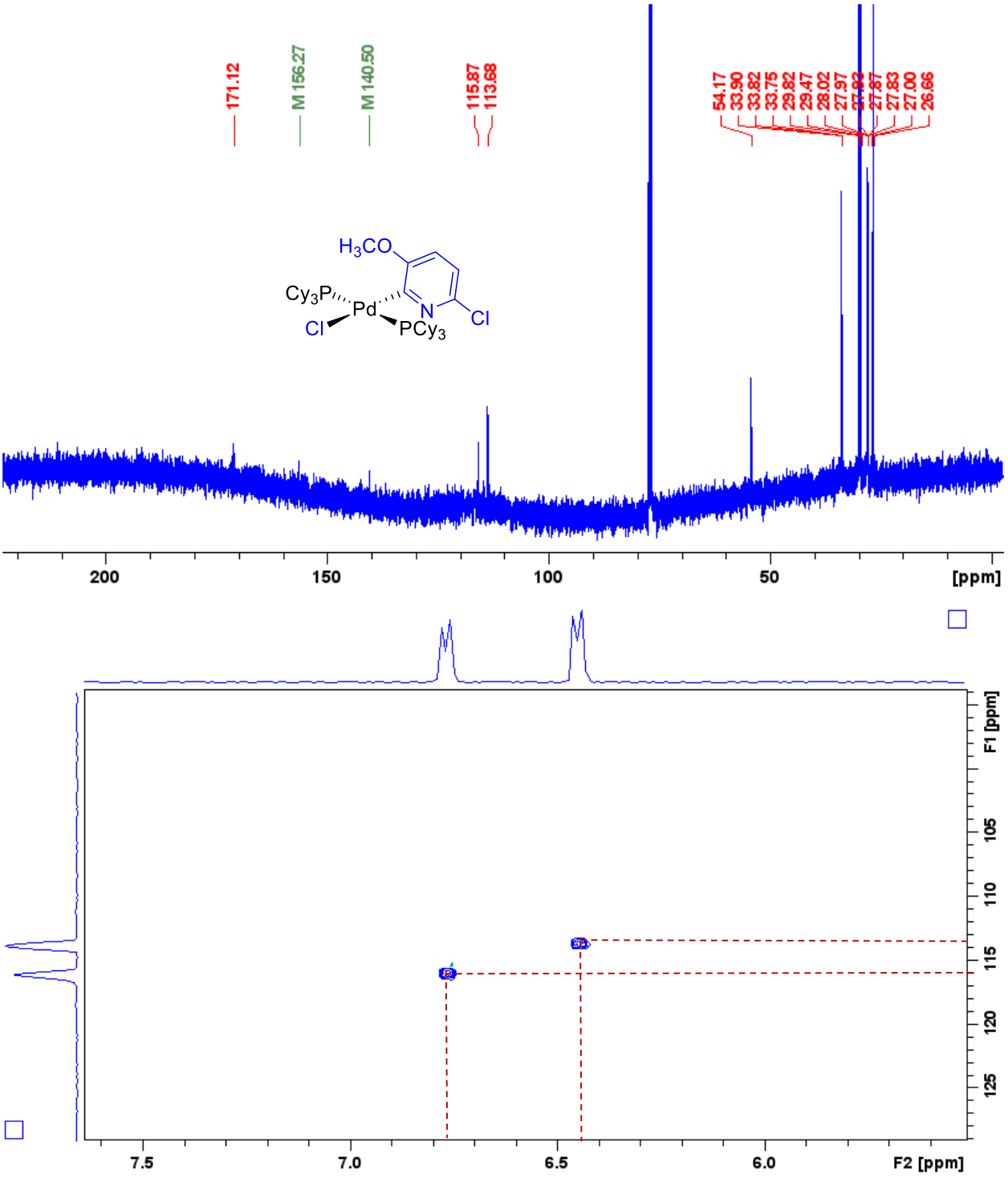


Fig. D31. $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectrum of **A5** with accompanying ^{1}H-^{13}C HSQC (middle) and ^{1}H-^{13}C HMBC (bottom) NMR spectra.

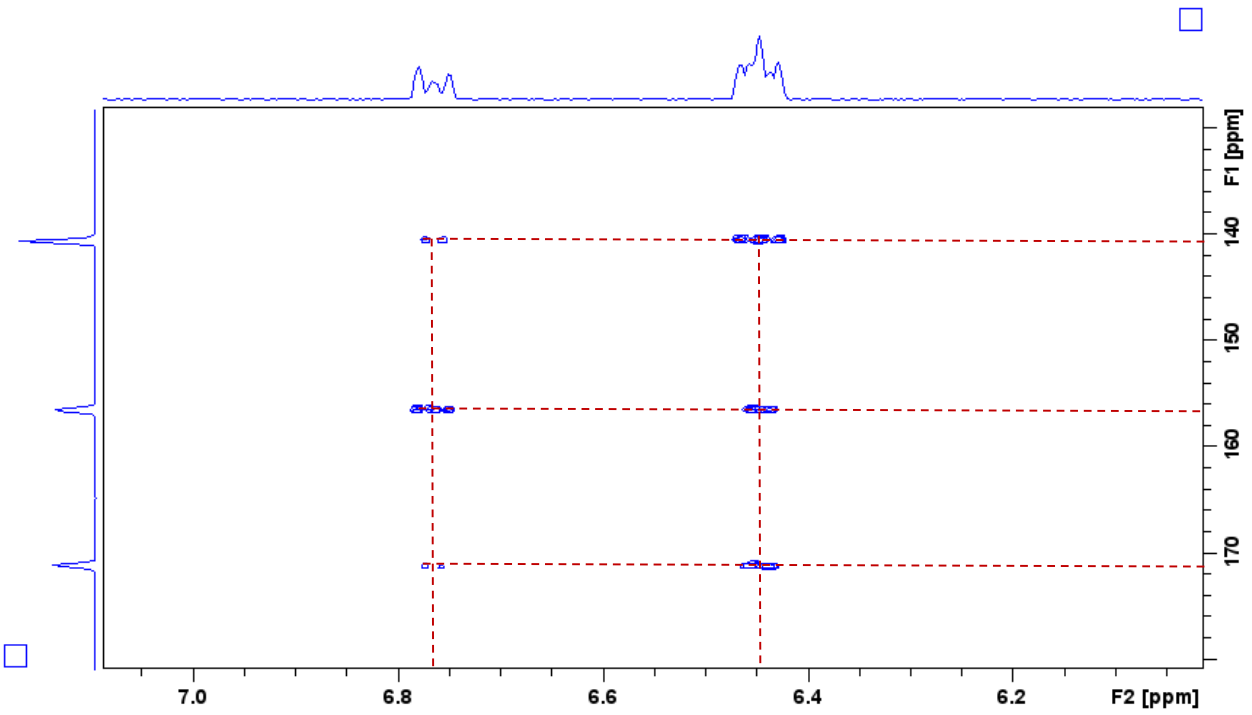


Fig. D31, continued. $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectrum of **A5** with accompanying ^{1}H-^{13}C HSQC (middle) and ^{1}H-^{13}C HMBC (bottom) NMR spectra.

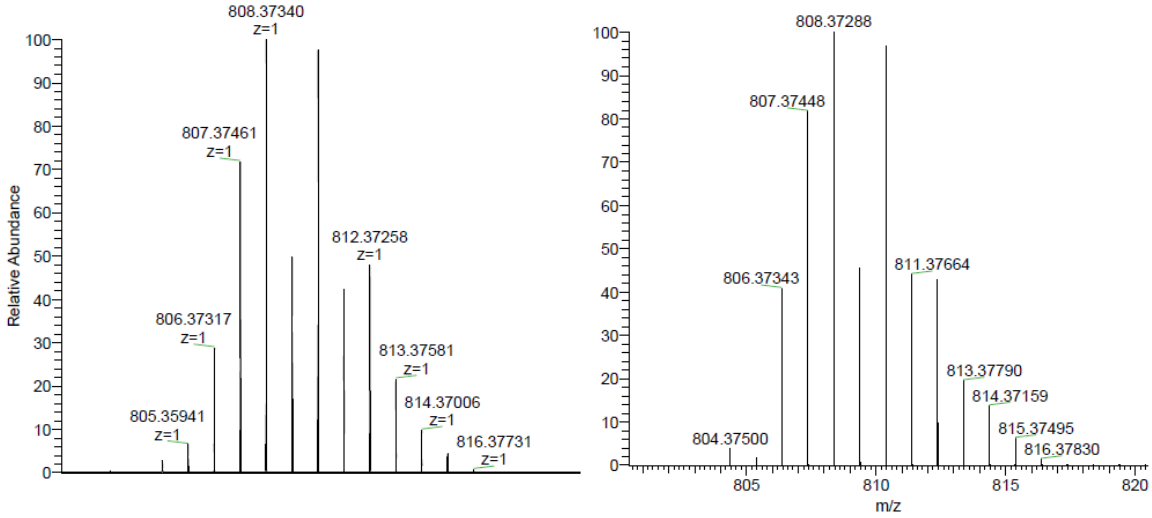


Fig. D32. Left: Experimental HRMS-ESI spectrum of [**A5**-Cl]$^{+}$. Right: Calculated HRMS isotope pattern for [**A5**-Cl]$^{+}$.

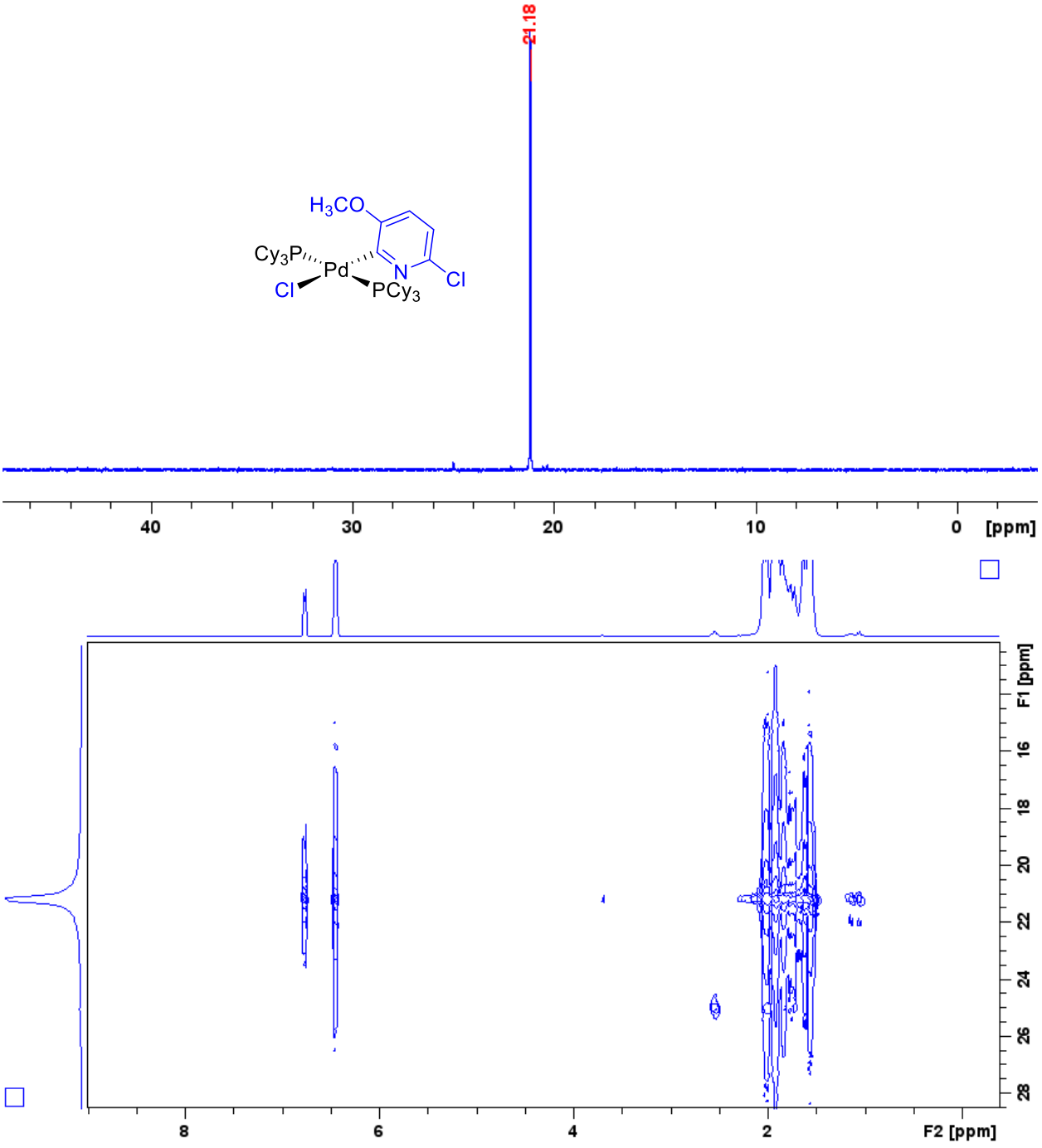


Fig. D33. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **A5** with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

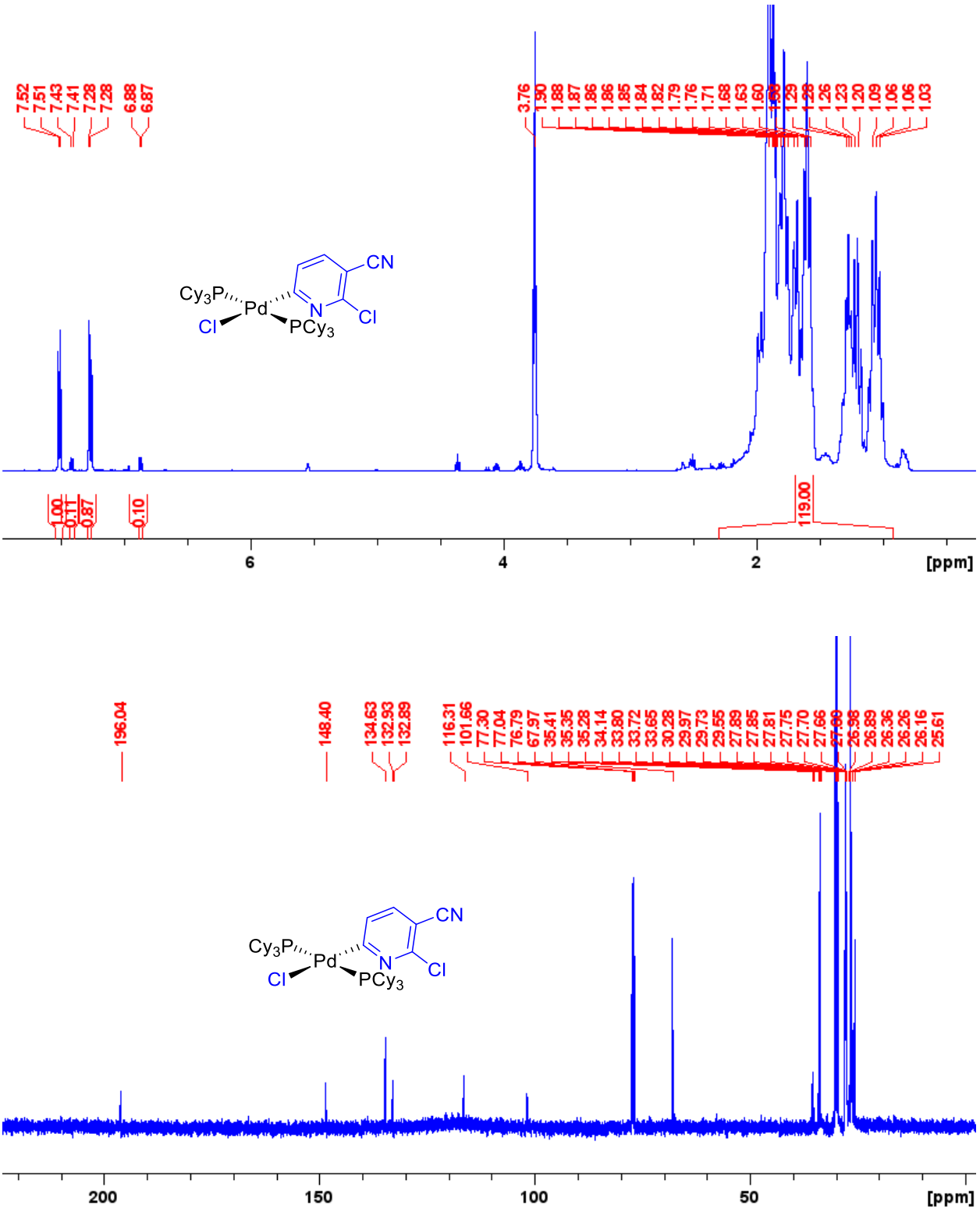


Fig. D34. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **A6**. The THF solvent peak is detected from both the ^{1}H (3.76 ppm) and ^{13}C (67.97 ppm) spectra.

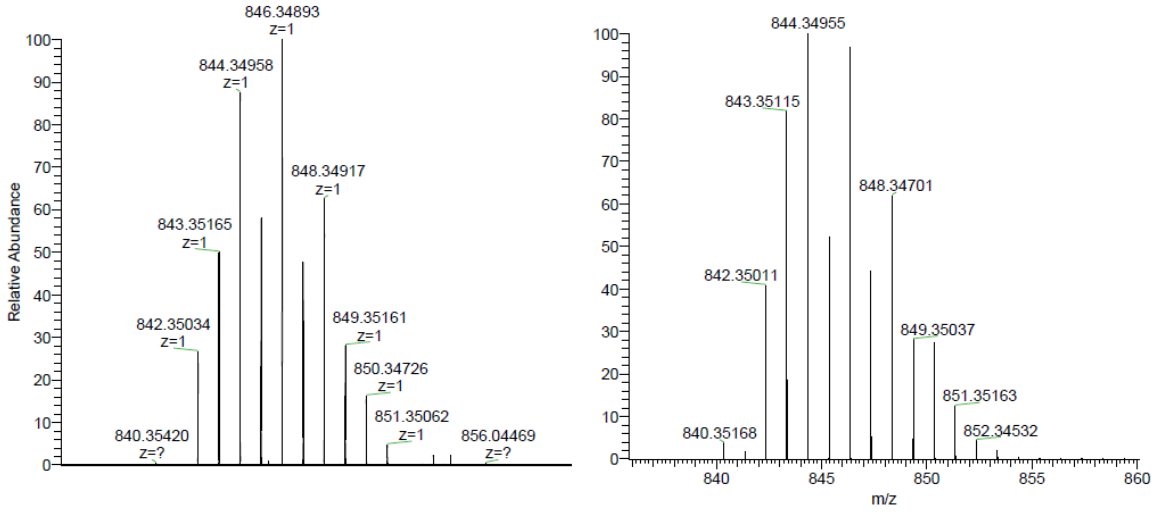


Fig. D35. Left: Experimental HRMS-ESI spectrum of [**A5**+H]$^+$. Right: Calculated HRMS isotope pattern for [**A5**+H]$^+$.

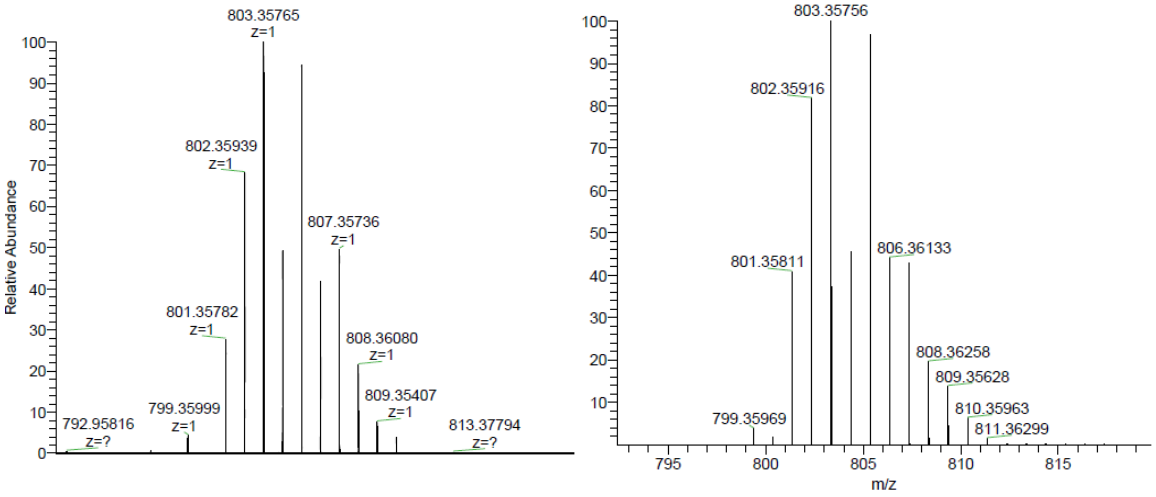


Fig. D36. Left: Experimental HRMS-ESI spectrum of [**A6**-Cl]$^+$. Right: Calculated HRMS isotope pattern for [**A6**-Cl]$^+$.

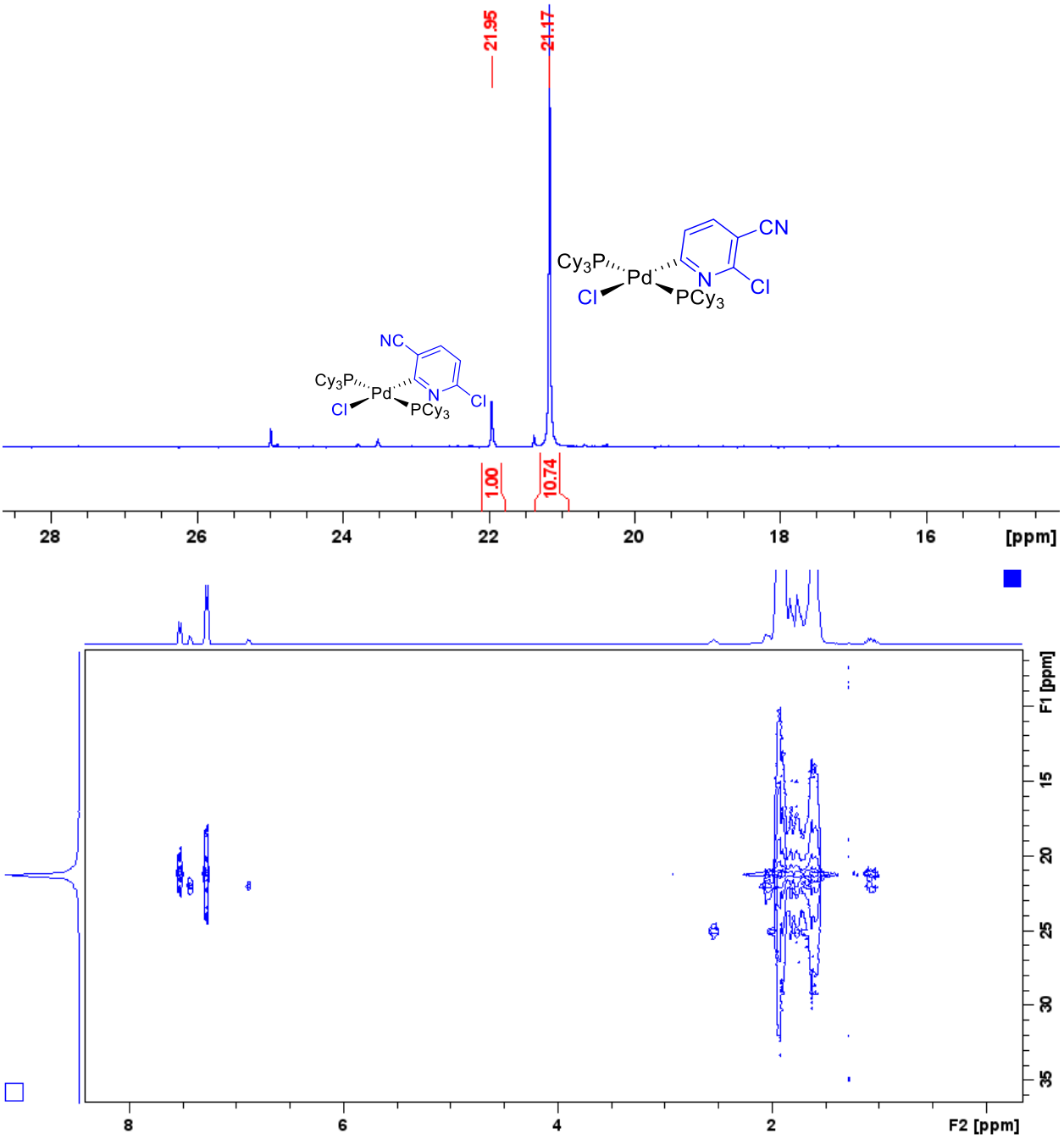


Fig. D37. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **A6**, the major product is detected at 21.17 ppm and the minor product is detected at 21.95 ppm (structures both shown next to their corresponding 31P peaks) with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

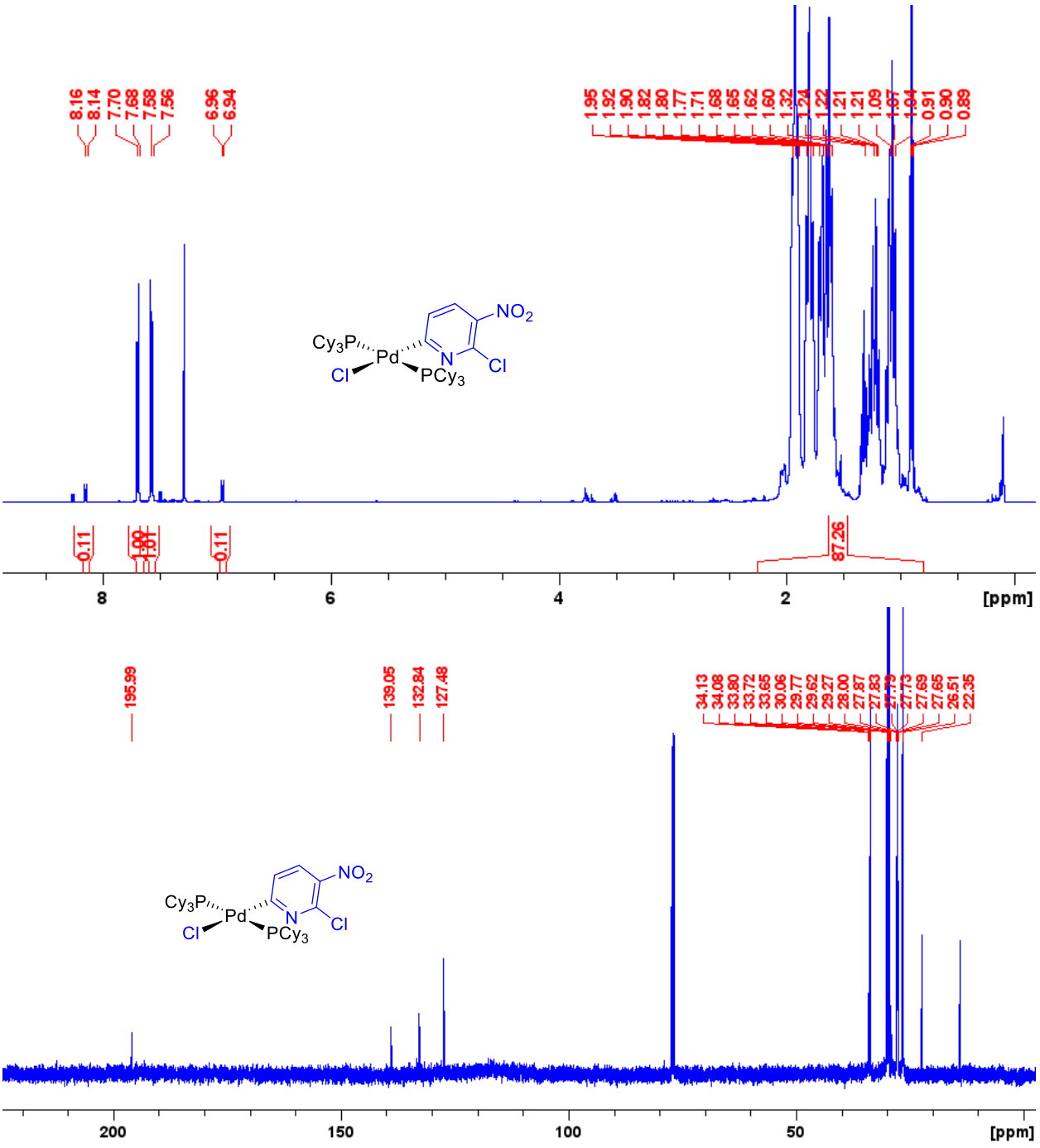


Fig. D38. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **A7**.

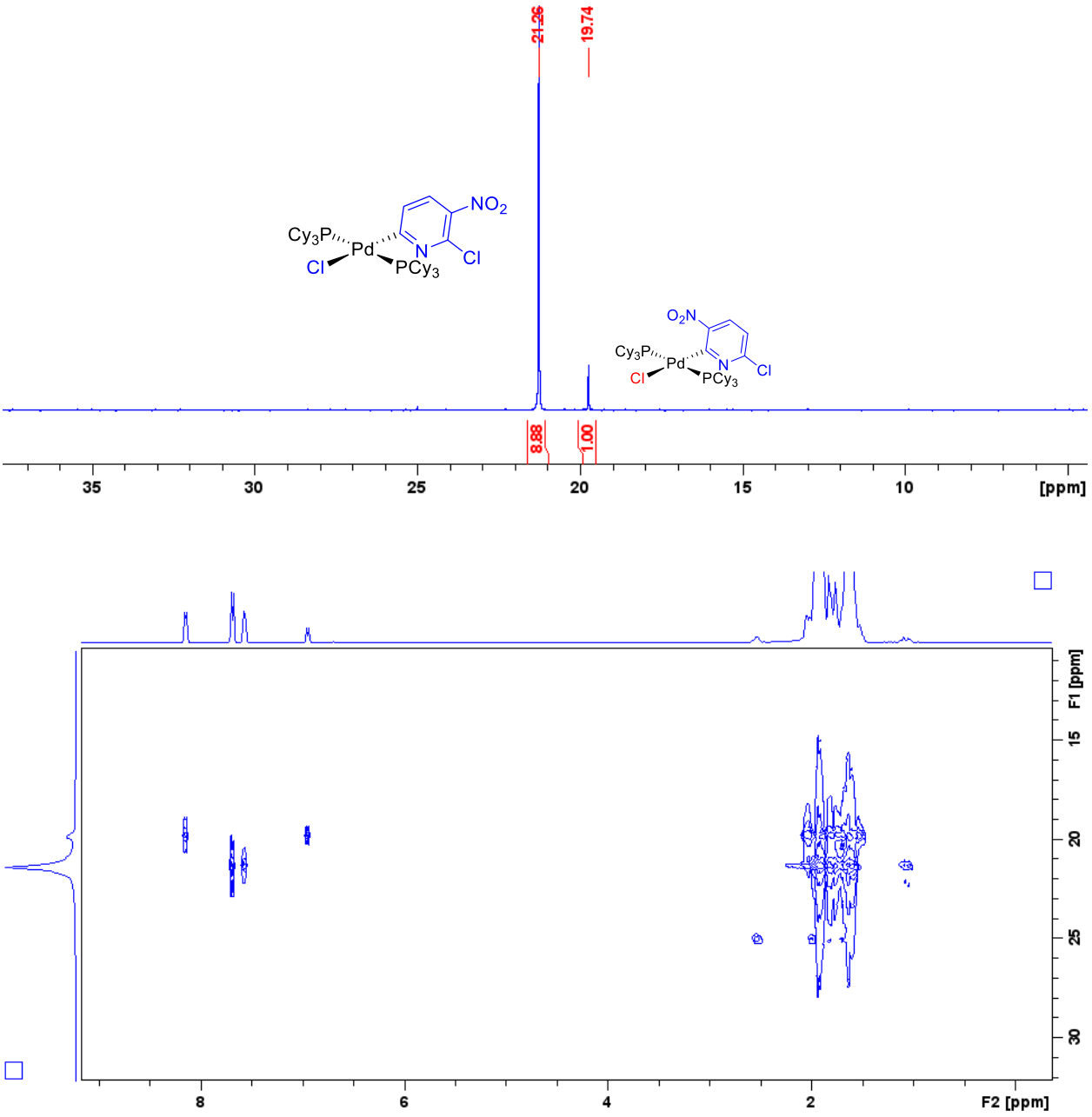


Fig. D39. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **A7**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

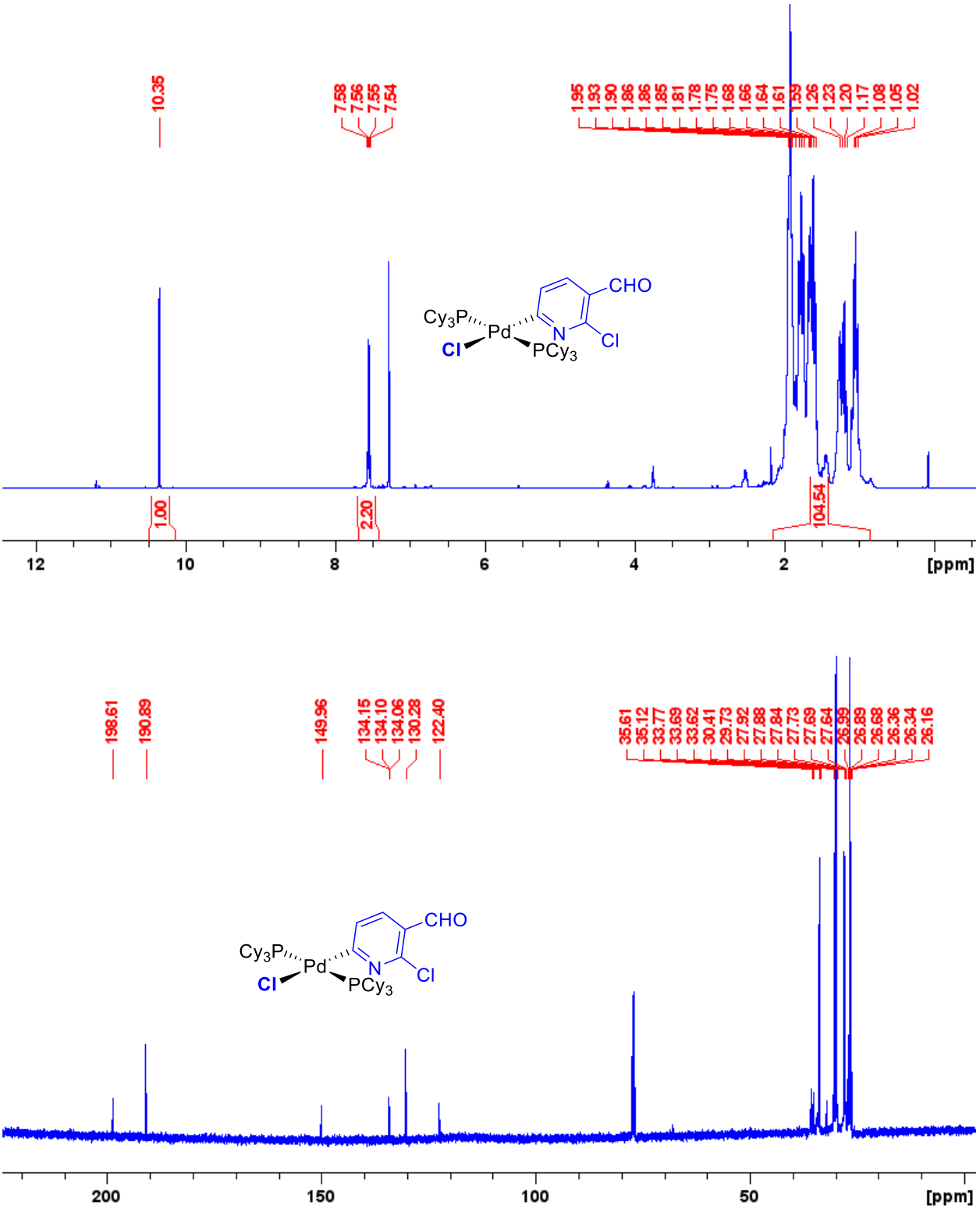


Fig. D40. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **A8**.

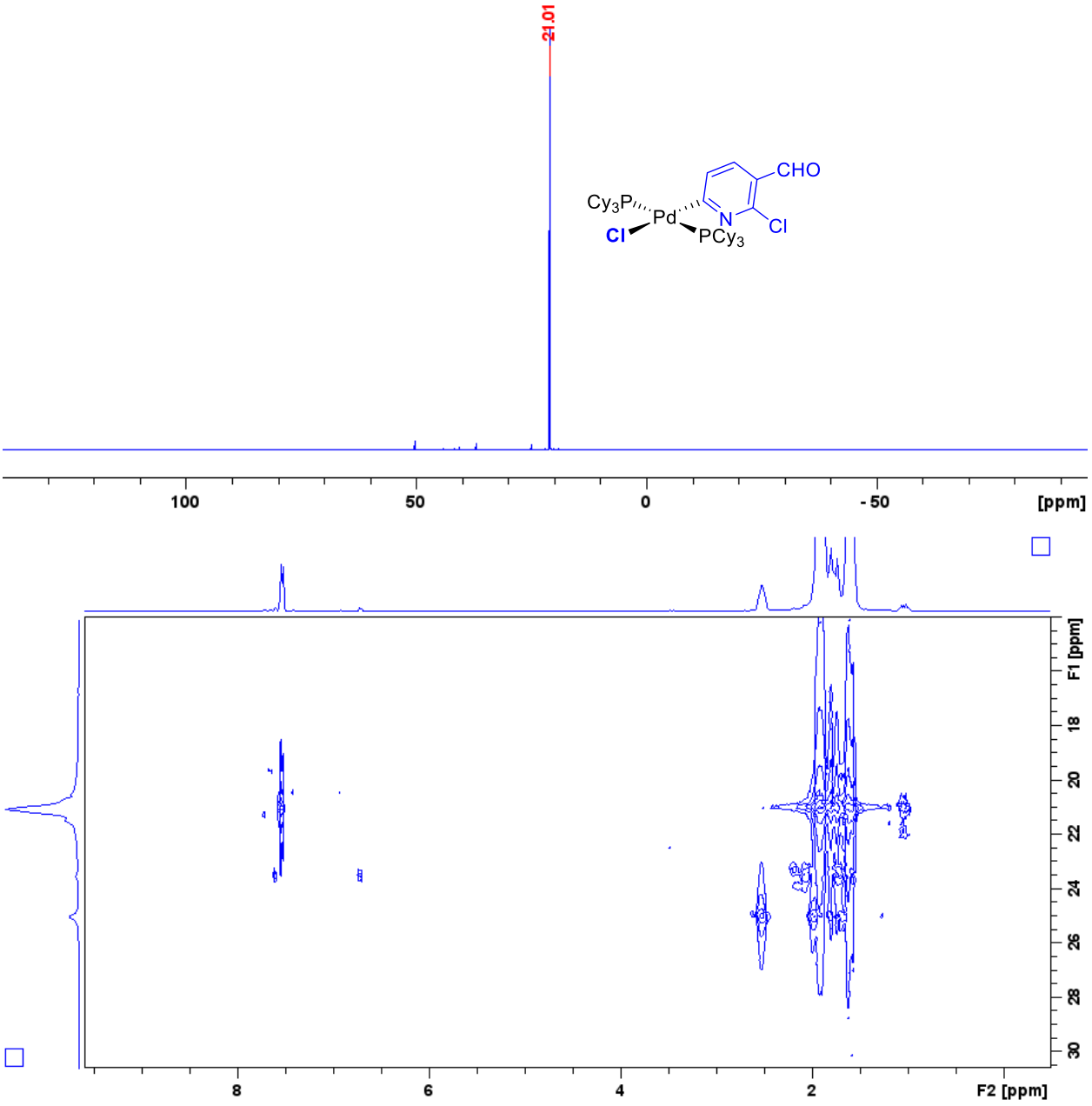


Fig. D41. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **A8** with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

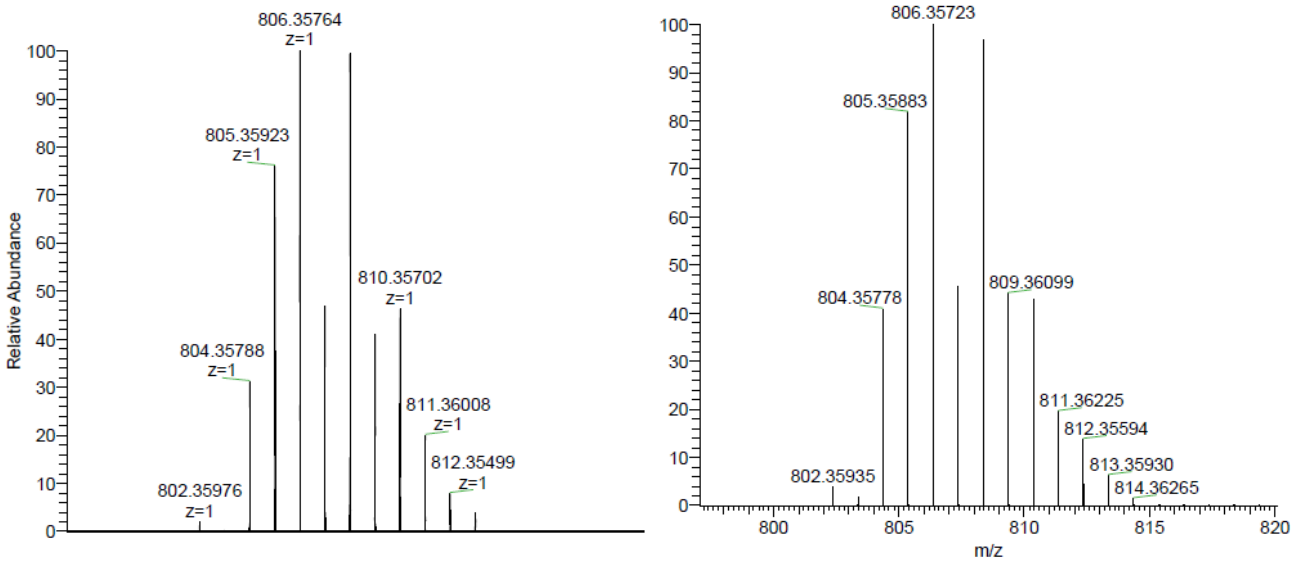


Fig. D42. Left: Experimental HRMS-ESI spectrum of [**A8**-Cl]$^+$. Right: Calculated HRMS isotope pattern for [**A8**-Cl]$^+$.

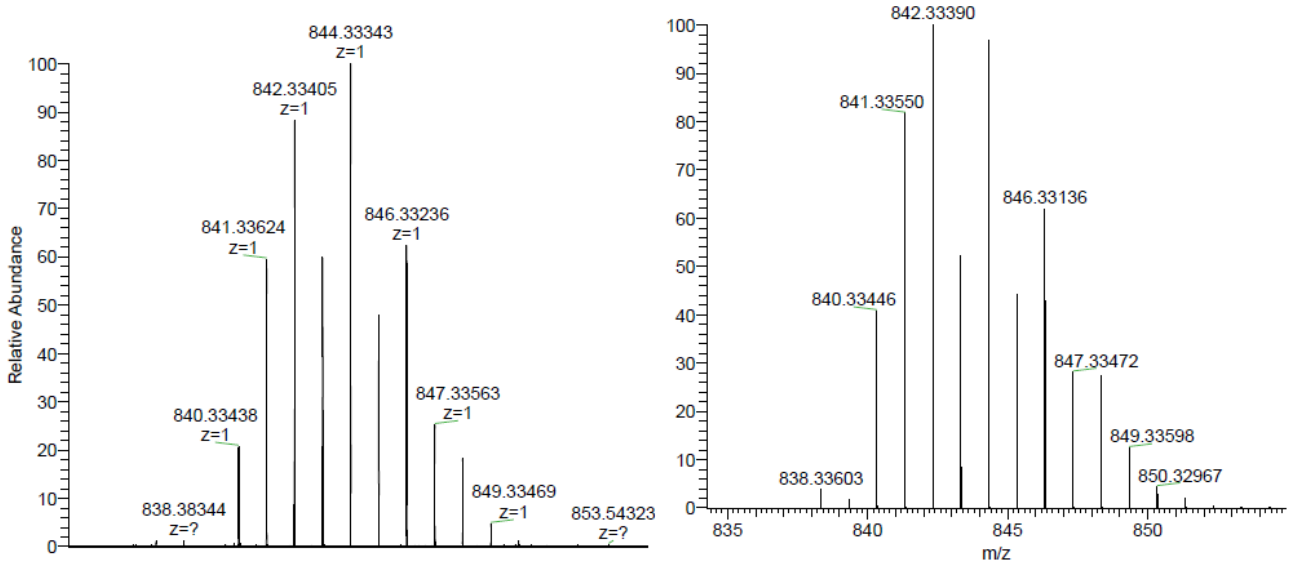


Fig. D43. Left: Experimental HRMS-ESI spectrum of [**A8**+H]$^+$. Right: Calculated HRMS isotope pattern for [**A8**+H]$^+$.

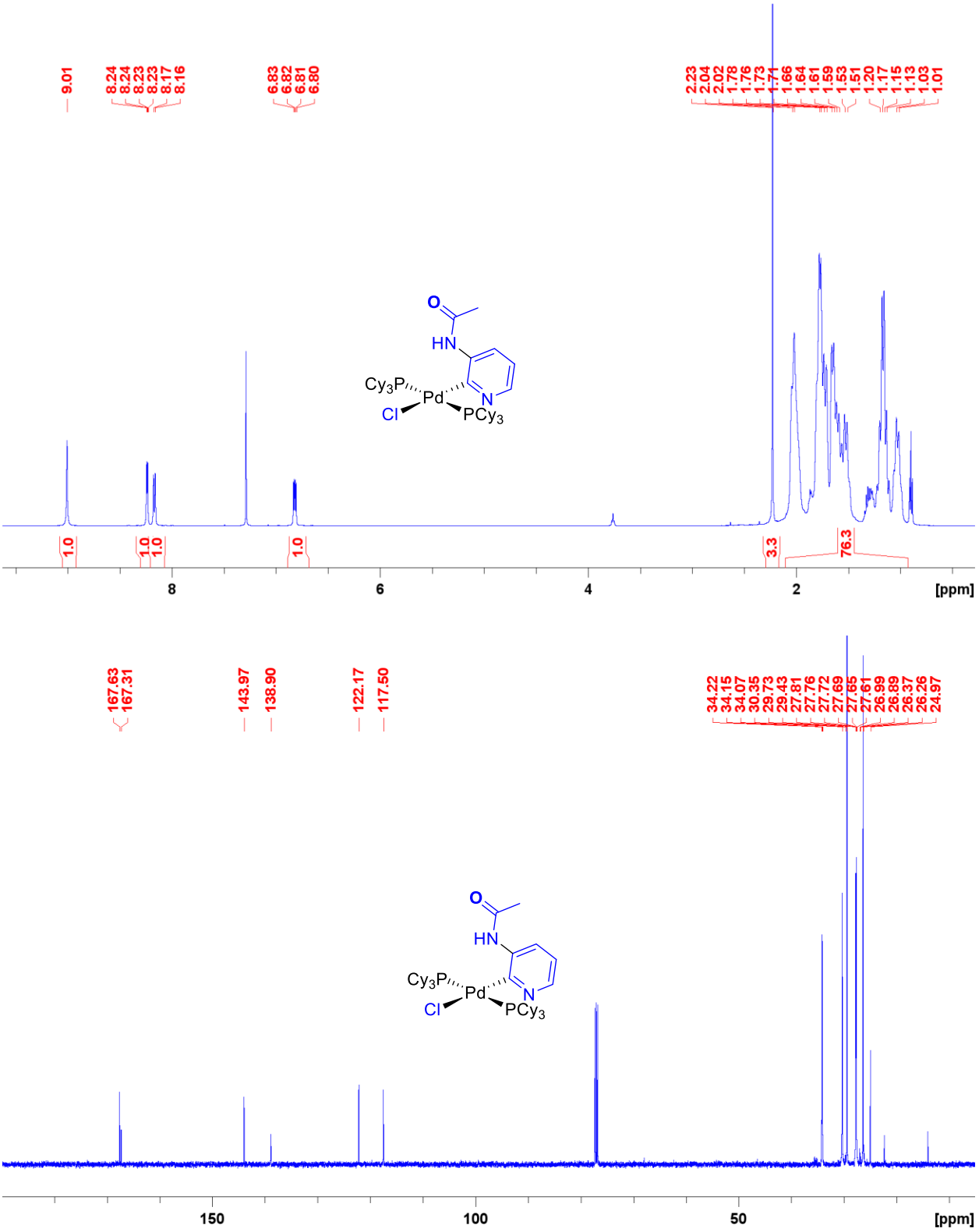


Fig. D44. ^{1}H (500 MHz, $CDCl_3$) and $^{13}C\{^{1}H\}$ (126 MHz, $CDCl_3$) NMR spectra of **S1**.

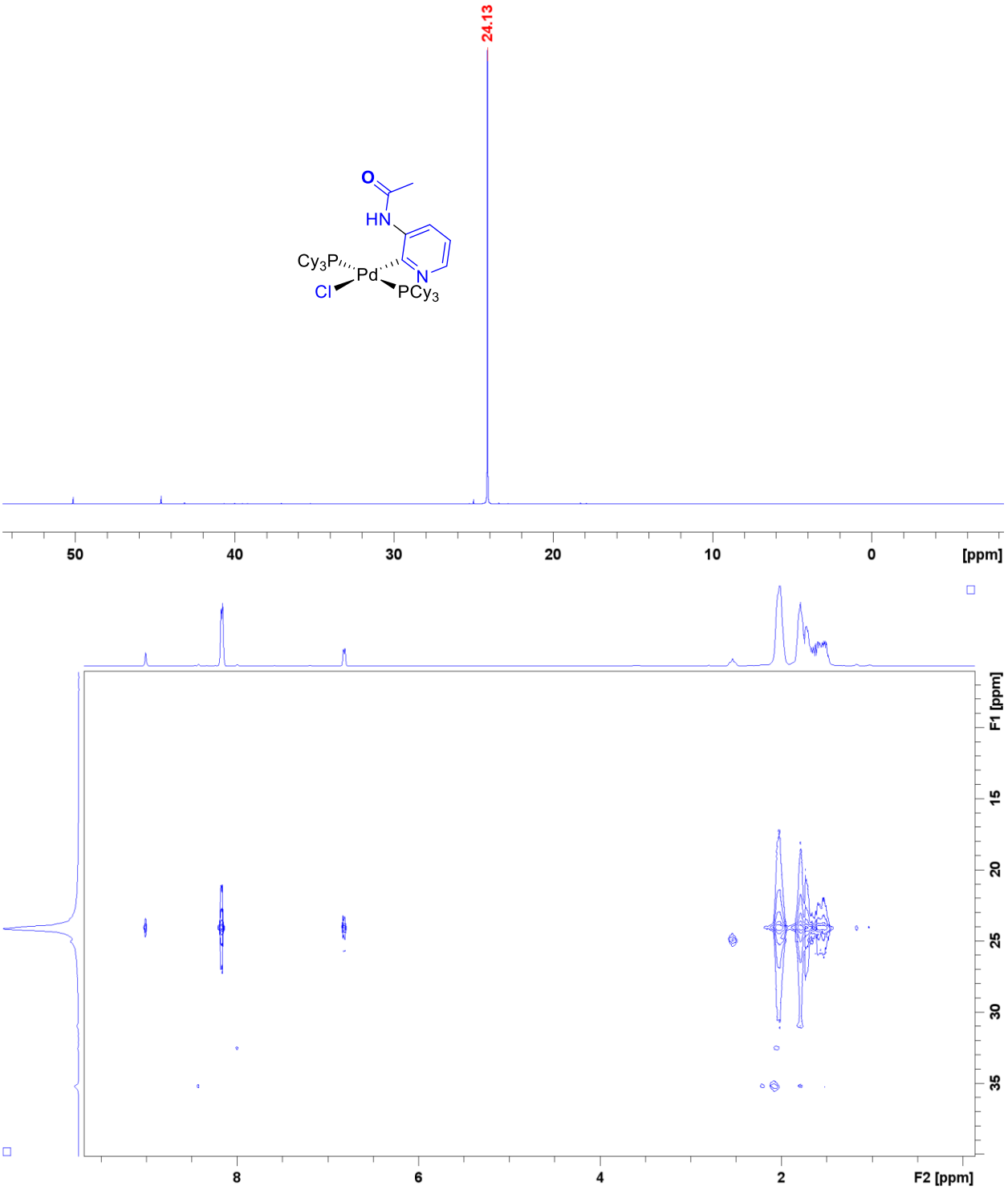


Fig. D45. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **S1**, with accompanying long range ^{1}H-^{31}P HMBC NMR spectrum.

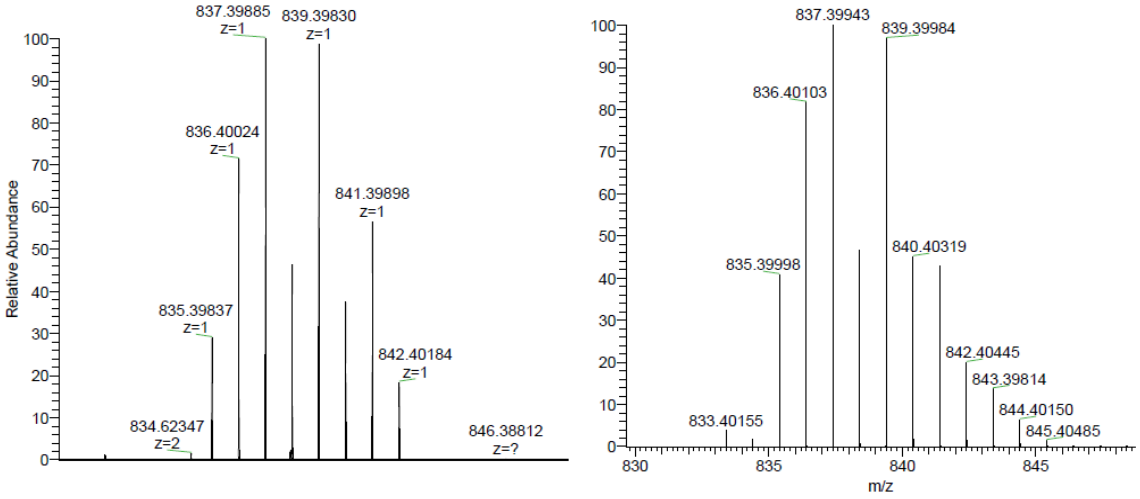
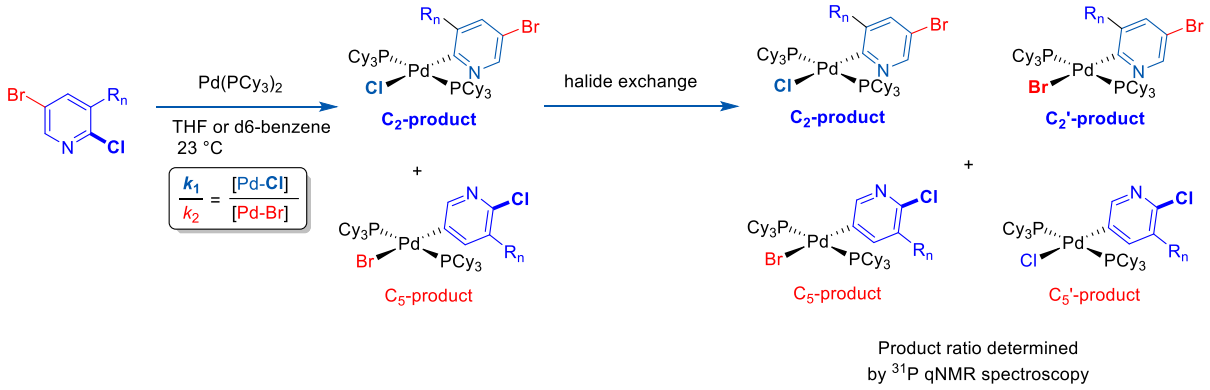


Fig. D46. Left: Experimental HRMS-ESI spectrum of [**S1**+H]$^+$. Right: Calculated HRMS isotope pattern for [**S1**+H]$^+$.

Structural elucidation of the Oxidative Addition Complexes from the substrates in Fig. 5.7

For the four 2-chloro-5-bromo-pyridine derivatives (B1-B4) in Fig. 5.7 carrying mixed halides, the two sites that are both reactive in oxidative addition to PdL_2. And due to halide exchange between the major and minor oxidative addition products, 4 products were detected in total as shown in Scheme D5.

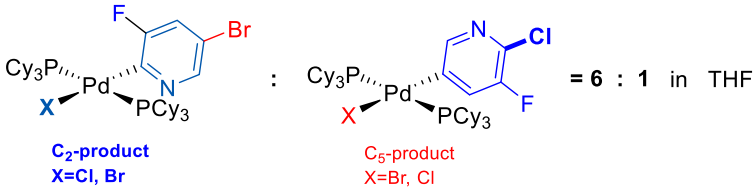


Scheme. D5. Formation of multiple oxidative addition products due to halide exchange.

The Pd(II) oxidative addition complexes from the substrates B3 and B4 were isolated, purified, and characterized by NMR using the general procedure in section *Preparative Scale Synthesis and structural elucidation of the Oxidative Addition Complexes from the substrates in Fig. 5.6* in Appendix D.

For substrates B1 and B2, the oxidative addition reactions were conducted in d6-benzene and the reaction mixtures were analyzed by NMR without further isolating the products.

Characterization of Isolated Oxidative Addition Complexes from substrate B3



B3

The oxidative addition complexes from **B3** were prepared by the general procedure using 2-chloro-5-bromo-3-fluoropyridine. Pentane was used to remove the excess pyridine from its oxidative addition product. 49.4 mg of a white solid was obtained (76% yield).

Two oxidative addition products were formed with a product ratio of 6:1 confirmed by the ^{1}H and ^{19}F NMR. The four peaks from the ^{31}P NMR indicate halide exchange between the major and minor oxidative addition products occurred.

The long range ^{1}H-^{31}P HMBC NMR spectrum shows up correlations between δ7.05 [^{1}H] to δ23.1 [^{31}P] and δ7.05 [^{1}H] to δ22.5 [^{31}P], which indicate the ^{31}P peaks at δ23.1 and δ22.5 are from the major oxidative addition product and its analogues from halide exchange. From the peak area ration of the ^{1}H-NMR, the peaks at δ7.05 and δ8.46 are both from the pyridyl protons of the major product, but the ^{1}H-^{31}P correlation for the ^{1}H peak at δ8.46 is missing.

We have isolated and characterized 5 oxidative addition complexes of halide pyridines to $Pd(PCy_3)_2$ in our previously published work, and the characterization data refers to the relating supporting information. From the long range ^{1}H-^{31}P HMBC NMR spectrums, we have noticed that the correlations between the pyridyl C_6-H to the phosphorus of the PCy_3 were missing in all the 5 oxidative addition complexes (Fig. D47A).

By relating to the ^{31}P-^{1}H correlation pattern of the oxidative addition complexes in Fig Sx.A, we drew a preliminary conclusion that C2-Cl is the major site of 2-chloro-5-bromo-3-fluoropyridine in oxidative addition to $Pd(PCy_3)_2$ as there is only one ^{31}P-^{1}H correlation observed. The C_5-Br is the minor site as the ^{31}P-^{1}H correlation is observed for both pyridyl protons (Fig. D47B).

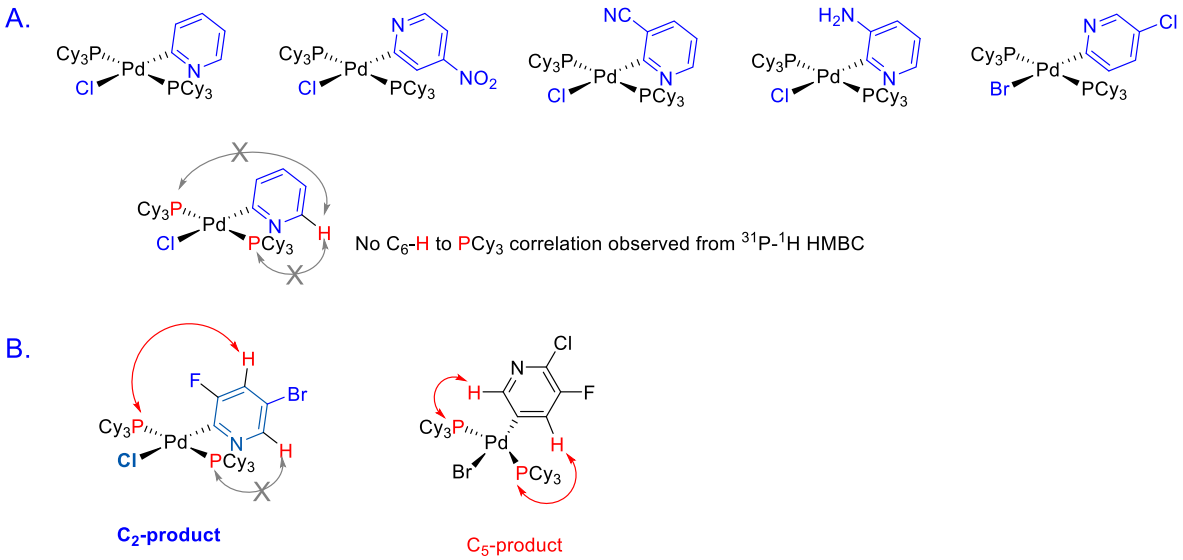
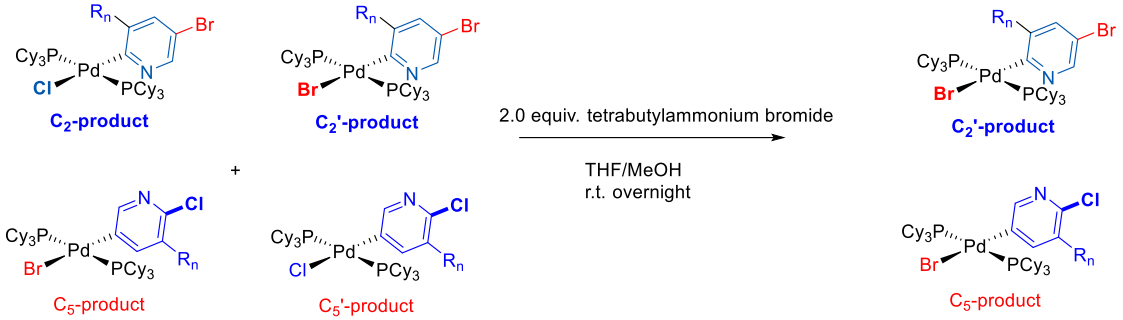


Fig. D47. Missing of the C_6-H to PCy_3 correlation from the ^{31}P-^{1}H HMBC of the five isolated oxidative addition products in Appendix B.

Then we collected further evidence for the major and minor products identification from halide exchange (Pd-Cl to Pd-Br) experiment by treating the oxidative addition products with bromide anion.

In the oxidative addition reaction mixture or the solution of the oxidative addition product, 2.0 equivalent of the tetrabutylammonium bromide was added (the stock solution was made in methanol for good solubility) and the reaction mixture was stirred overnight at room temperature.

As it is shown in scheme D6, the C_2-product will be converted into the C_2'-product, and as a result a 31P peak shift will be observed from C_2 to C_2'. The C_5'-product will be converted into the C_5-product with the accompanying 31P peak shift to indicate this conversion. Pd-Cl to Pd-Br takes place slowly, and after overnight of reaction, the Py-$Pd^{(2)}L_2$-Cl complexes will only be partially converted into its bromide analogue Py-$Pd^{(2)}L_2$-Br. Fig. D48 shows the overlap of the 31P spectra before and after the addition of tetrabutylammonium bromide in the 2-chloro-5-bromide-3-fluoropyridine (B3) oxidative addition product.



Scheme. D6. Identification of the regioselective oxidative addition products by halide exchange experiments.

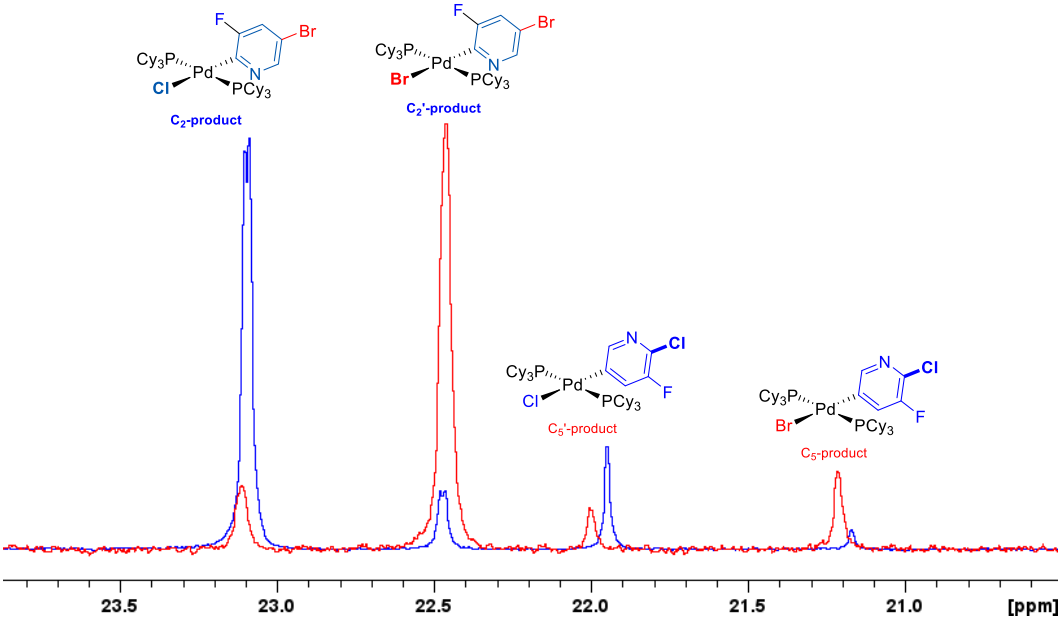


Fig. D48. Overlap of the ^{31}P spectra before (blue trace) and after (red trace) adding tetrabutylammonium bromide in the **B3** oxidative addition product and the $Pd^{(2)}$-complexes identified on top of the corresponding ^{31}P peaks.

^{1}H NMR: (500 MHz, $CDCl_3$): δ 1.05-1.97 (m, 66H, 66 x Cy-H); Pyridyl protons from the major C_2-procut: 7.05 (d, 1H, 1 x Py-H), 8.46 (s, 1H, 1 x Py-H); Pyridyl protons from the minor C_5-procut: 7.54 (d, 1H, 1 x Py-H), 8.21 (s, 1H, 1 x Py-H)

^{13}C{^{1}H} NMR: (126 MHz, $CDCl_3$): δ 22.35-34.61 (36 x Cy), 113.85 (1 x Py), 120.87 (1 x Py), 121.09 (1 x Py), 145.10 (1 x Py).

^{19}F{^{1}H} NMR: (471 MHz, $CDCl_3$): the major C_2-product: δ -97.29, δ -97.12; the minor C_5-product: δ -121.97, δ -122.28.

^{31}P{^{1}H} NMR: (203 MHz, $CDCl_3$): the major C_2-product: δ 23.09 (d), δ 22.46 (d); the minor C_5-product: δ 21.95, δ 21.17.

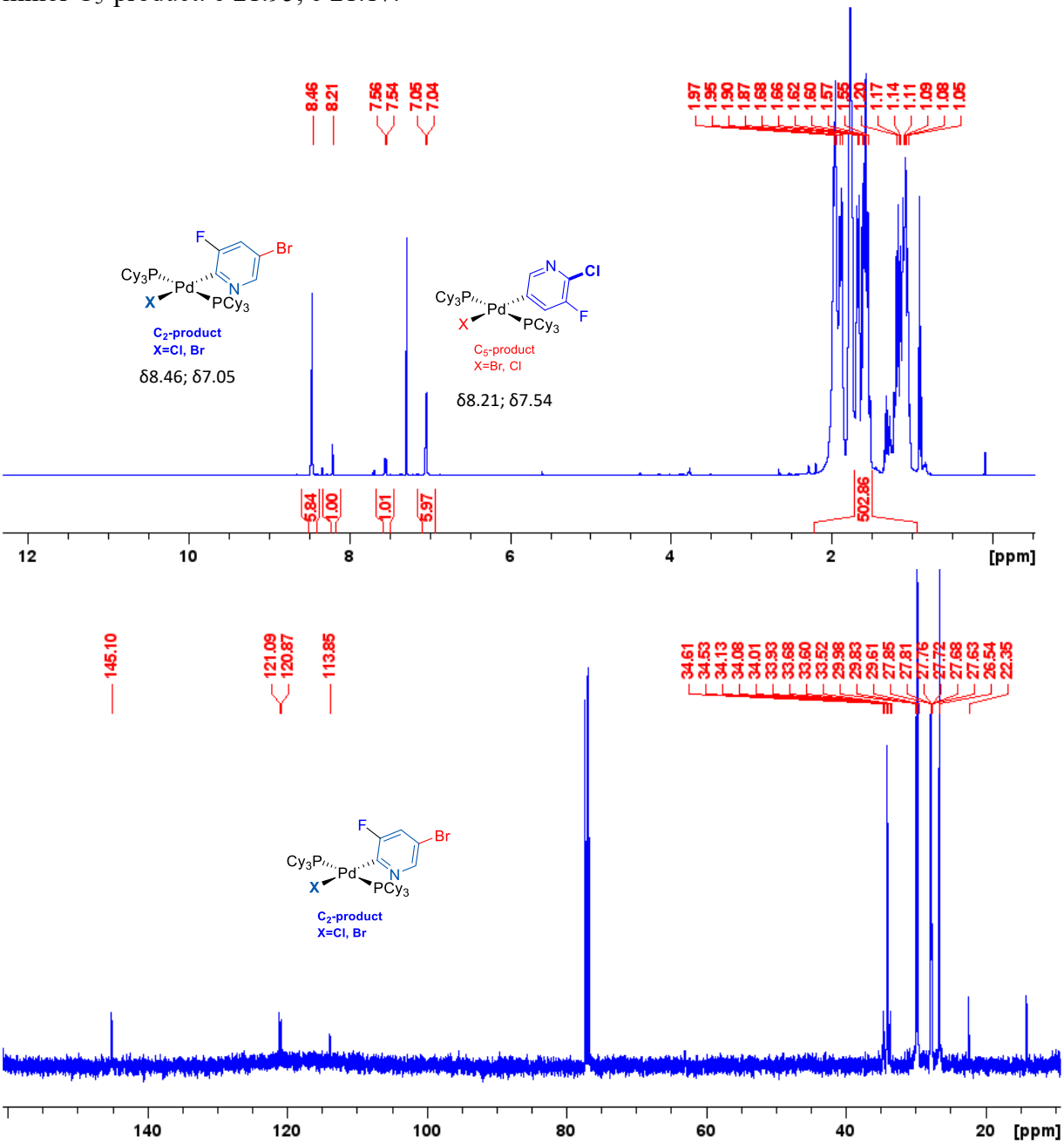


Fig. D49. ^{1}H (500 MHz, $CDCl_3$) and ^{13}C{^{1}H} (126 MHz, $CDCl_3$) NMR spectra of **B3**.

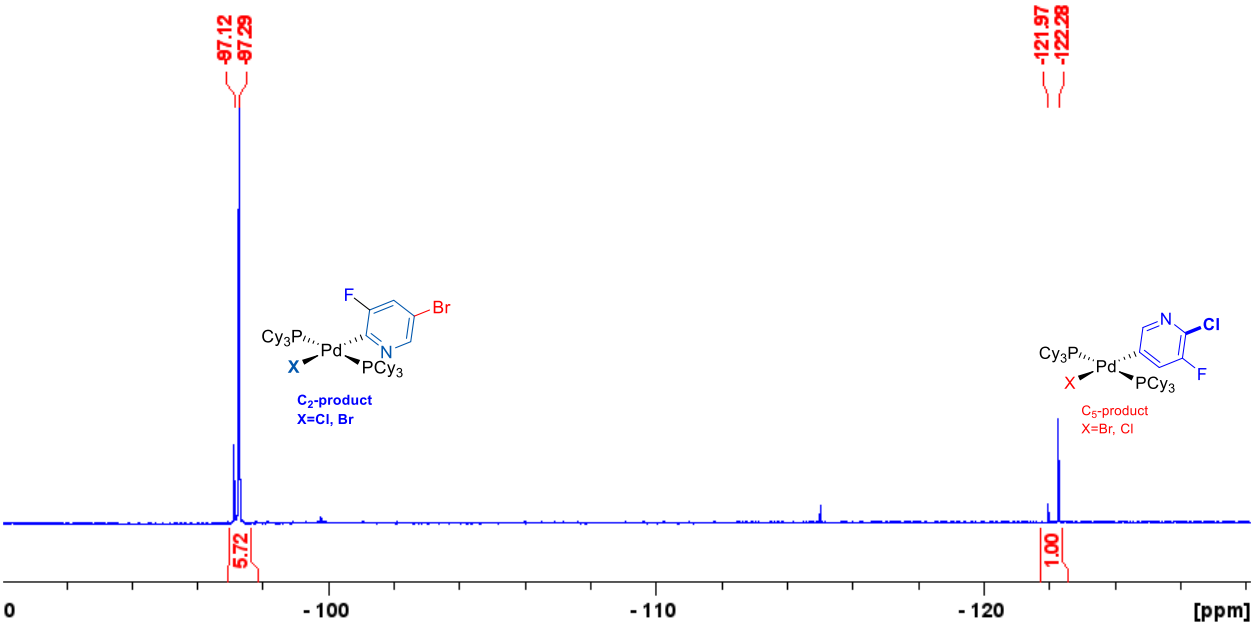


Fig. D50. $^{19}F\{^{1}H\}$ (471 MHz, $CDCl_3$) NMR spectrum of **B3**.

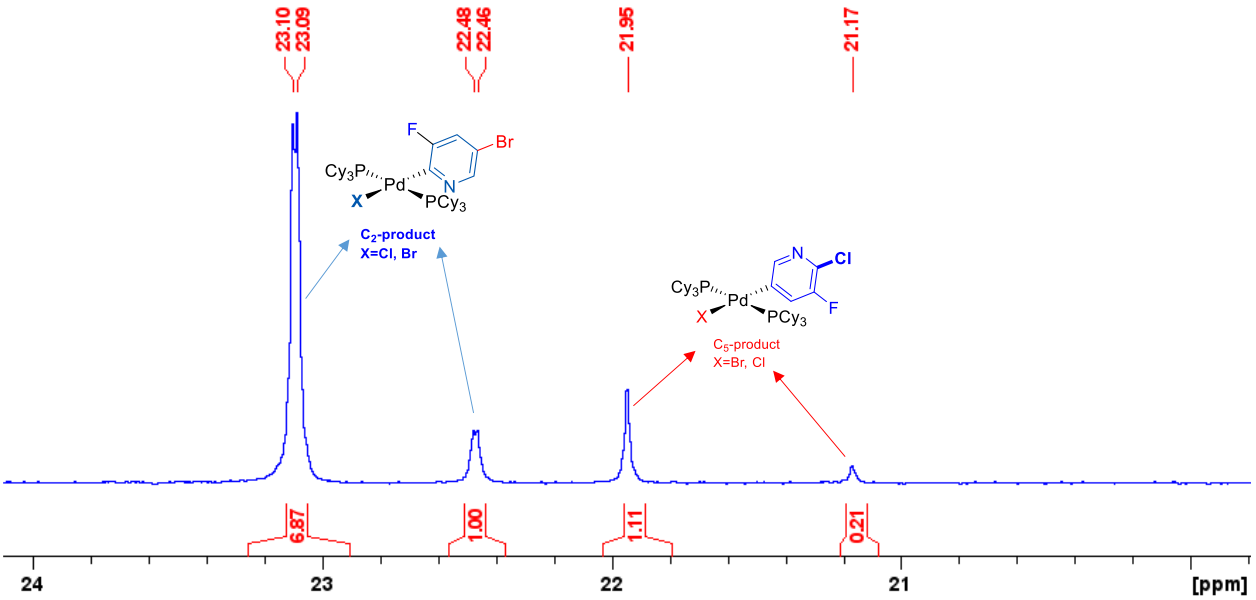


Fig. D51. $^{31}P\{^{1}H\}$ NMR (203 MHz, $CDCl_3$) spectrum of **B3**.

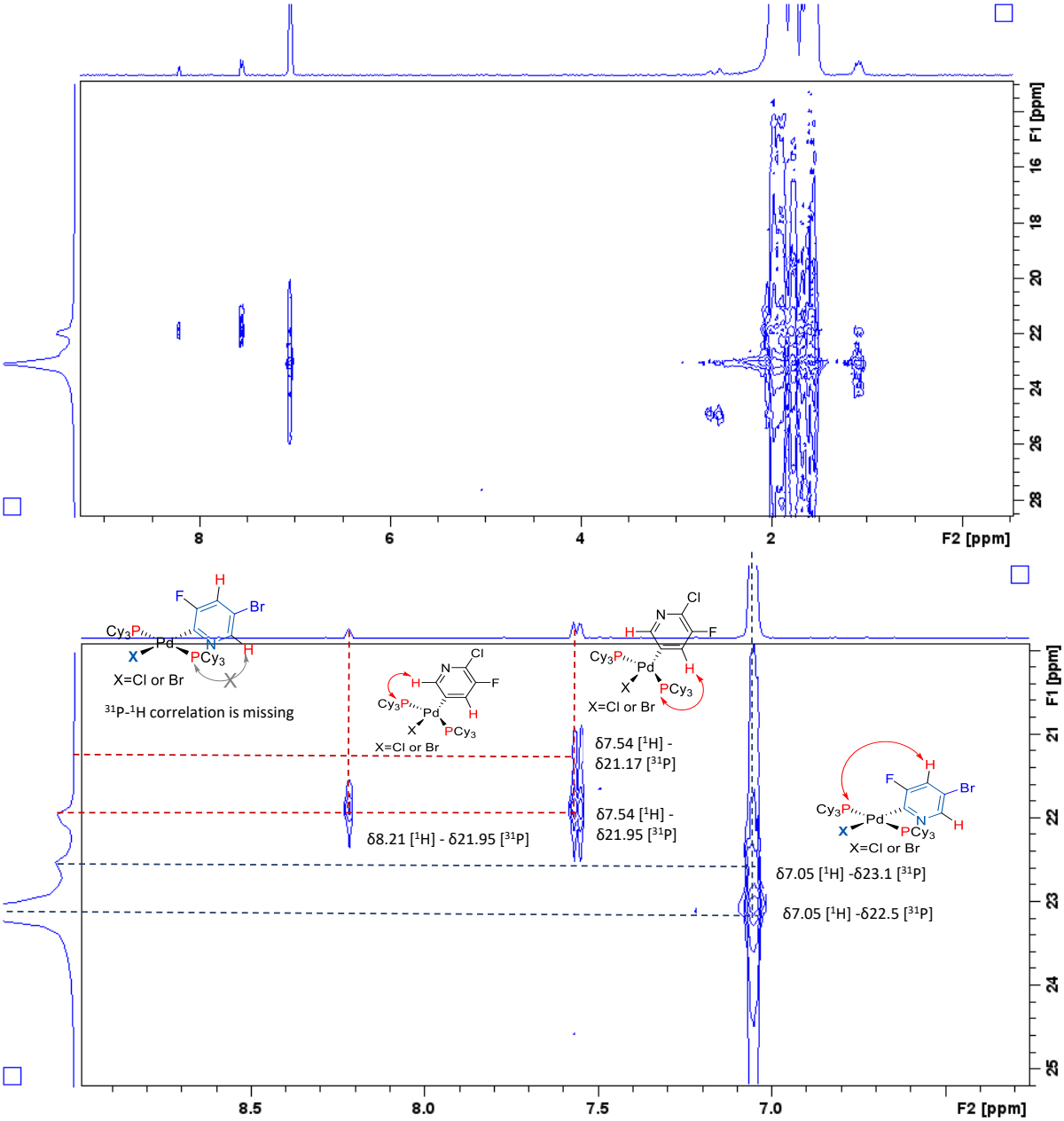
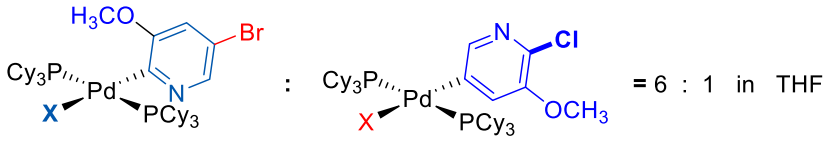


Fig. D51, continued. Top: the accompanying long range ^{1}H-^{31}P HMBC NMR spectrum; bottom: expansion of the pyridyl proton area and the ^{31}P-^{1}H correlation from the corresponding oxidative addition complexes.

Characterization of Isolated Oxidative Addition Complexes from substrate B4



B4

The oxidative addition complexes from **B4** were prepared by the general procedure using 2-chloro-5-bromo-3-methoxypyridine. Pentane was used to remove the excess pyridine from its oxidative addition product. 33.6 mg of a white solid was obtained (50% yield).

^{1}H NMR: (500 MHz, $CDCl_3$): δ 0.87-1.98 (m, 66H, 66 x Cy-H); protons from the C_2-procut: 3.68 (s 3H, 3 x OCH_3), 6.60 (s, 1H, 1 x Py-H), 8.21 (d, 1H, 1 x Py-H); protons from the C_5-procut: 3.82 (s 3H, 3 x OCH_3), 7.34 (d, 1H, 1 x Py-H), 7.95 (d, 1H, 1 x Py-H).

^{31}P{^{1}H} NMR: (203 MHz, $CDCl_3$): C_2-product: δ 21.65 and δ 20.99; C_5-product: δ 21.62 and δ 20.82.

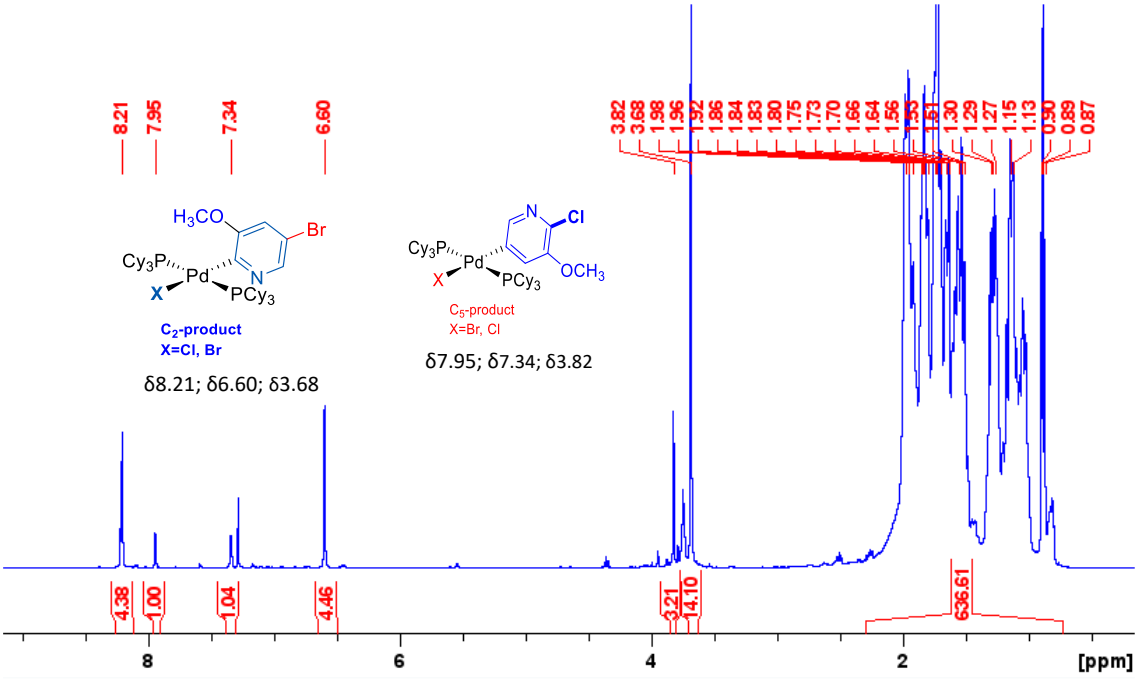


Fig. D52. ^{1}H (500 MHz, d_6-benzene) NMR spectrum of the reaction mixture of **B4** with $Pd(PCy_3)_2$.

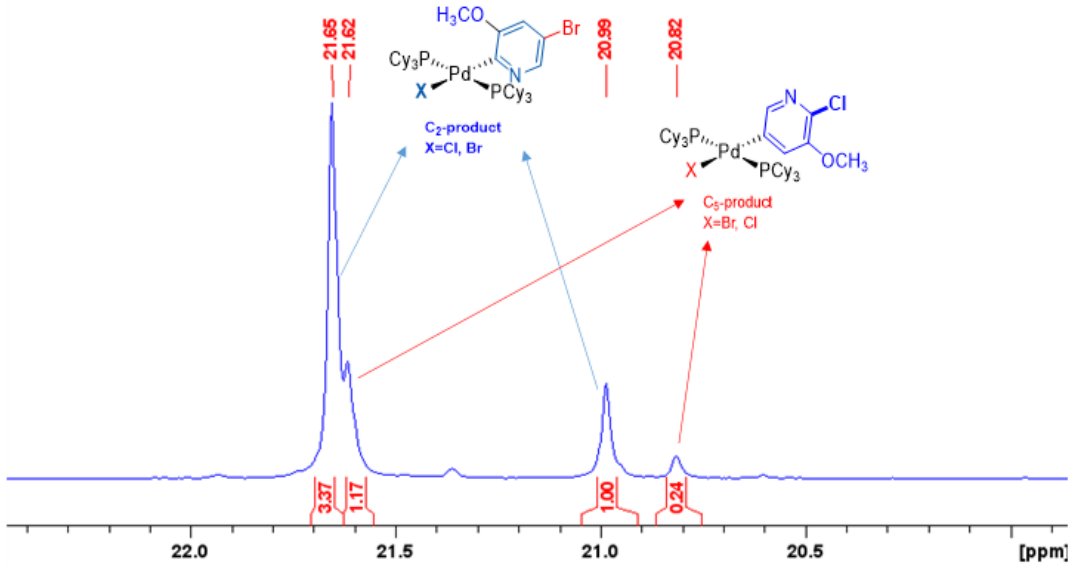


Fig. D53. ^{31}P{^{1}H} NMR (203 MHz, $CDCl_3$) spectrum of **B4**.

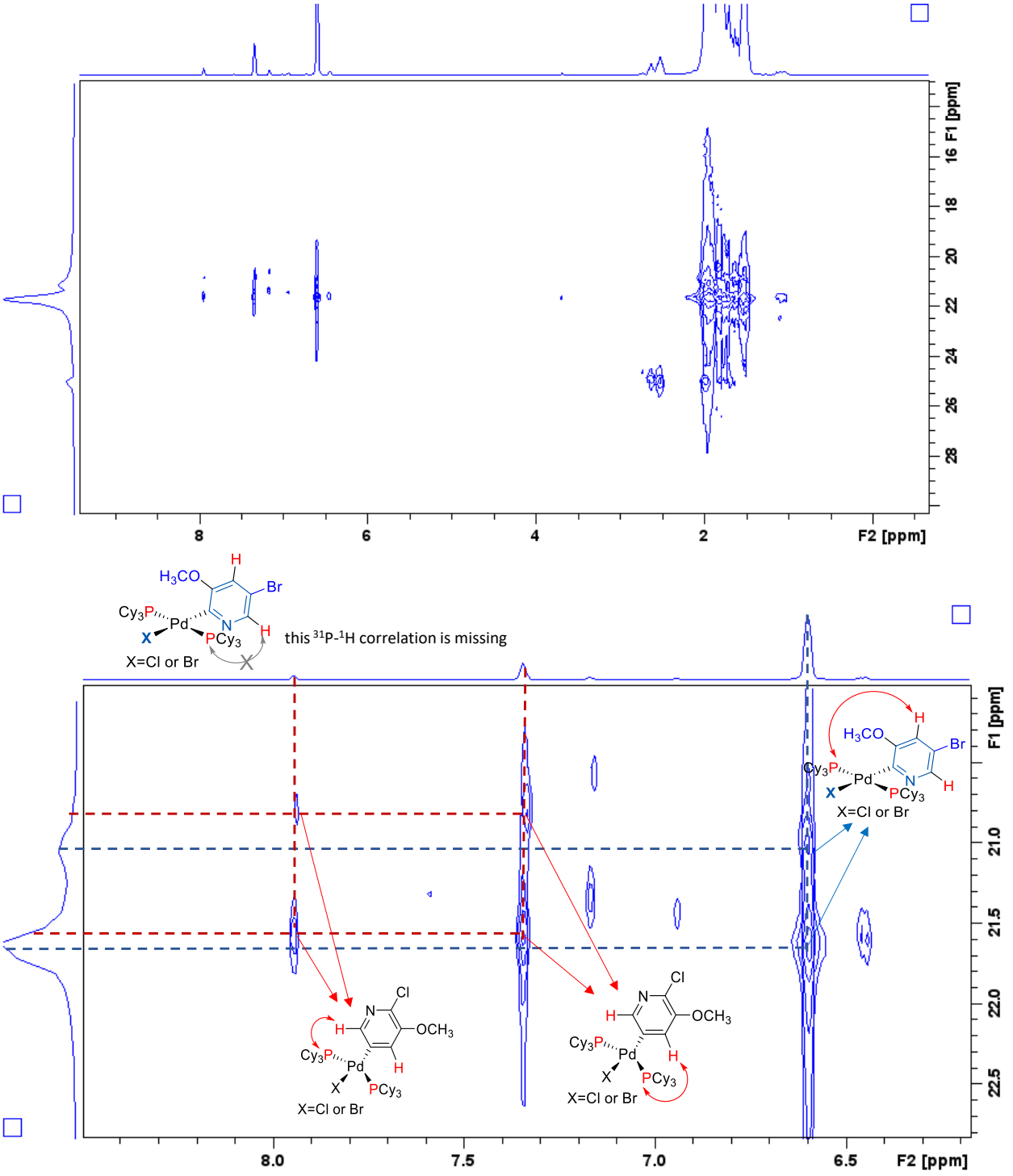


Fig. D53, continued. Top: the accompanying long range ^{1}H-^{31}P HMBC NMR spectrum; bottom: expansion of the pyridyl proton area and the ^{31}P-^{1}H correlation from the corresponding oxidative addition complexes.

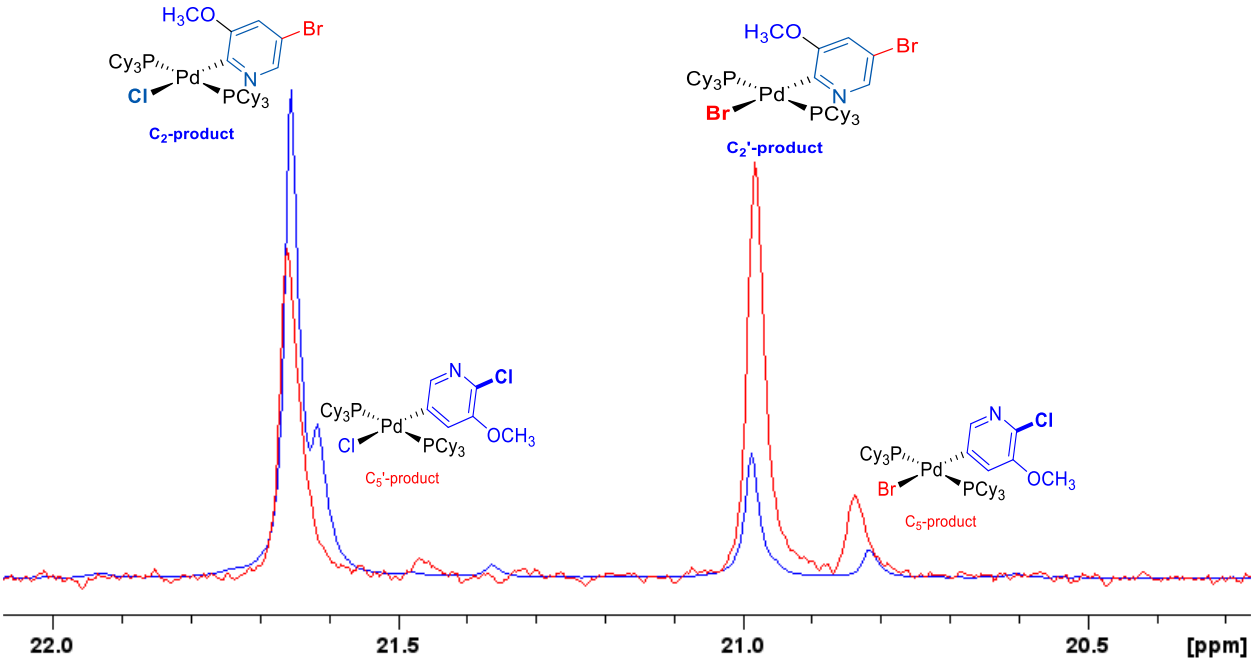
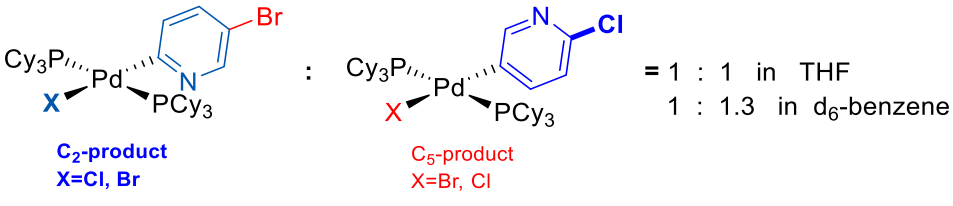


Fig. D54. Overlap of the ^{31}P spectra before (blue trace) and after (red trace) adding tetrabutylammonium bromide in the **B4** oxidative addition product and the $Pd^{(2)}$-complexes identified on top of the corresponding ^{31}P peaks.

Identification of the major and minor Oxidative Addition Products from substrates B1 and B2

In the glovebox, a 4 mL vial containing a stir bar was charged with 48 µL of the substrate stock solution in d6-benzene (0.18 mmol, 1.2 equiv.), 400 µL of the $Pd(PCy_3)_2$ stock solution in d6-benzene (0.15 mmol, 1.0 equiv.), and 200 µL of d6-benzene. The reaction mixture was stirred at room temperature overnight. On the next day, the reaction mixture was transferred into an NMR tube for ^{1}H, ^{31}P and ^{31}P-^{1}H HMBC analysis.



B1

^{1}H NMR: (500 MHz, d_6-benzene): δ 1.10-2.21 (m, 66H, 66 x Cy-H); protons from the C_2-procut: 7.01 (dd, 1H, 1 x Py-H), 7.26 (d, 1H, 1 x Py-H), 8.75 (d, 1H, 1 x Py-H); protons from the C_5-procut: 6.95 (dd, 1H, 1 x Py-H), 7.57 (d, 1H, 1 x Py-H), 8.71 (d, 1H, 1 x Py-H).

$^{31}P\{^{1}H\}$ NMR: (203 MHz, d_6-benzene): C_2-product: δ 21.30; C_5-product: δ 21.12.

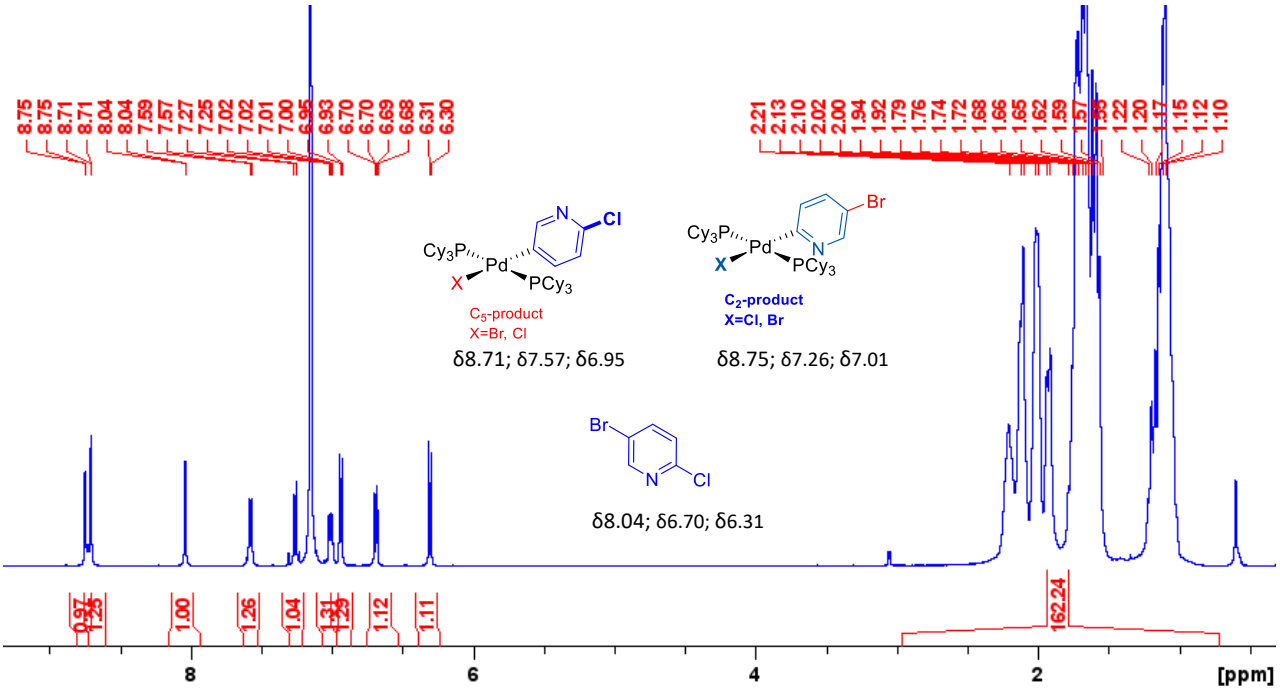


Fig. D55. ^{1}H (500 MHz, d_6-benzene) NMR spectrum of the reaction mixture of **B1** with $Pd(PCy_3)_2$.

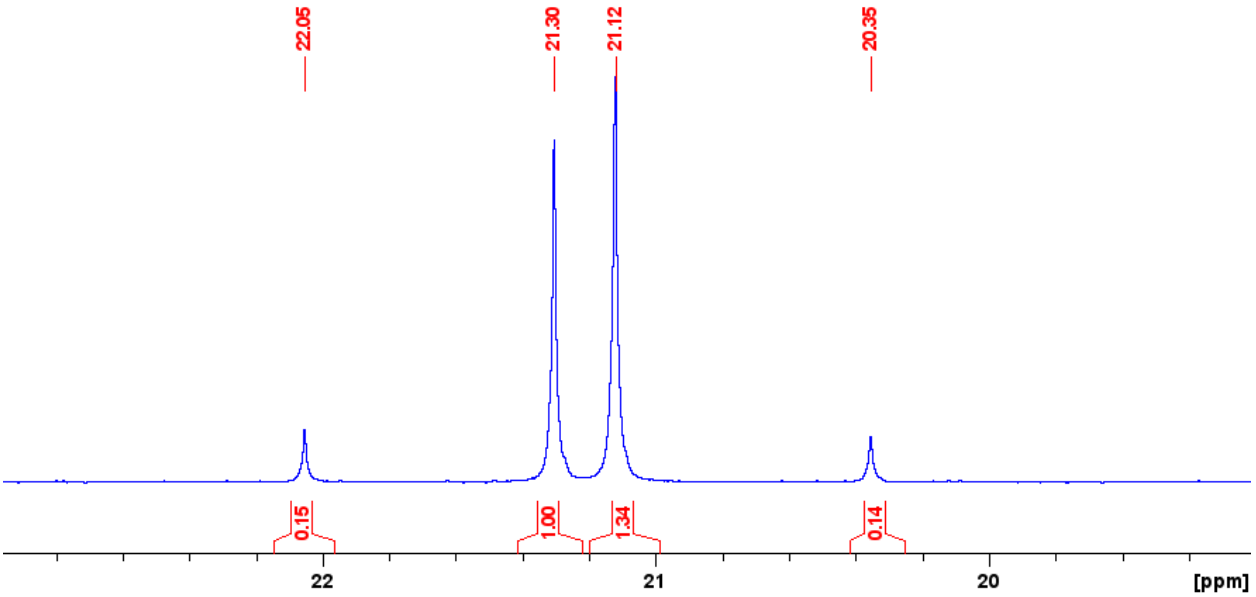


Fig. D56. $^{31}P\{^{1}H\}$ NMR (203 MHz, d_6-benzene) spectrum of the reaction mixture of **B2** with $Pd(PCy_3)_2$

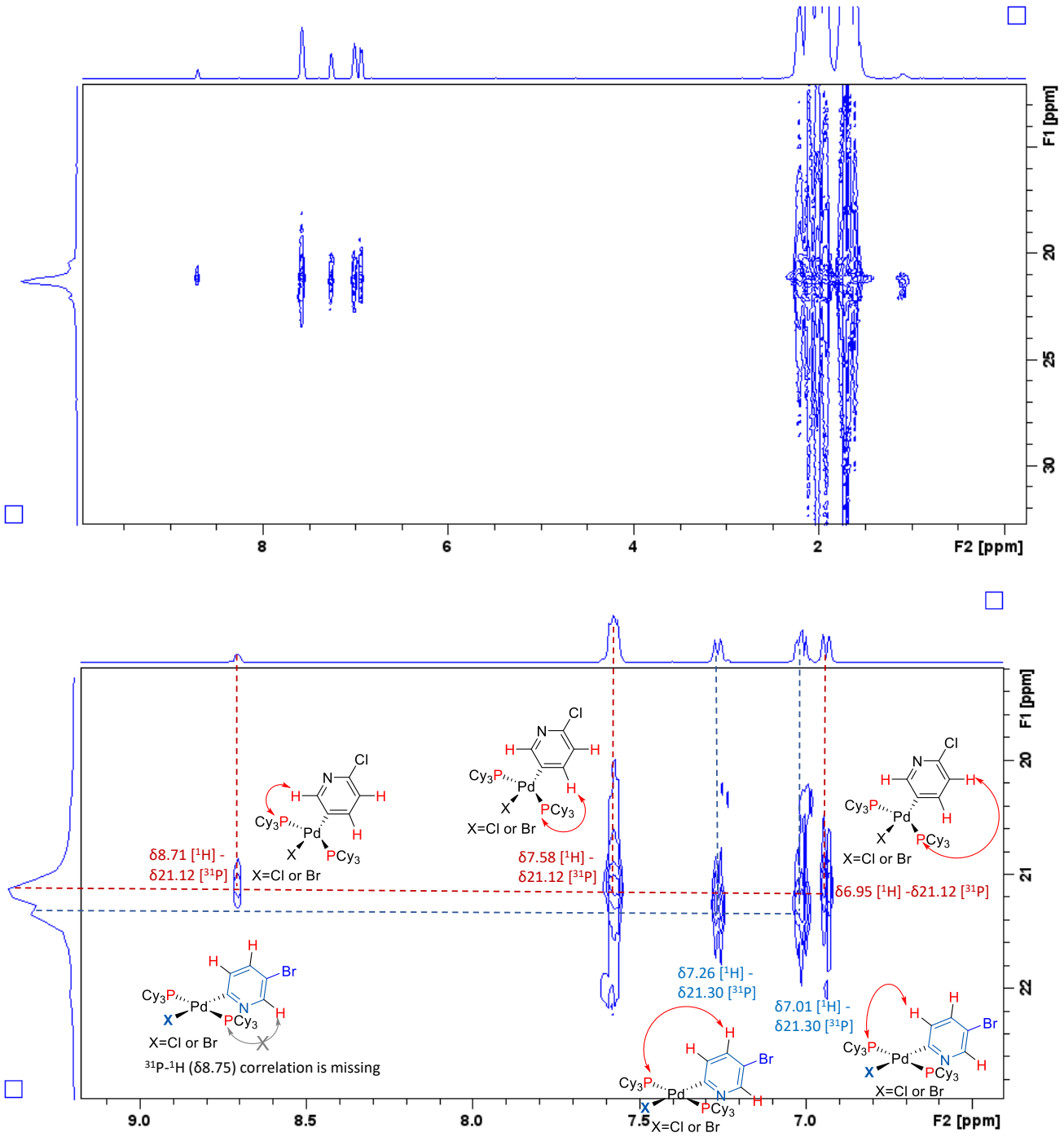
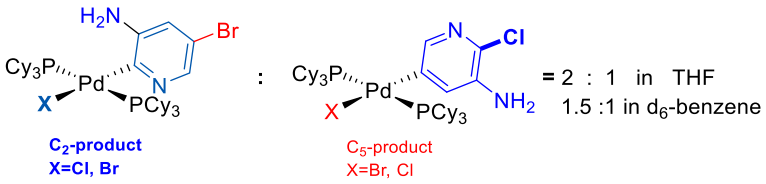


Fig. D56, continued. Top: the accompanying long range ^{1}H-^{31}P HMBC NMR spectrum; bottom: expansion of the pyridyl proton area and the ^{31}P-^{1}H correlation from the corresponding oxidative addition complexes.



B2

^{1}H NMR: (500 MHz, d_6-benzene): δ 1.08-2.20 (m, 66H, 66 x Cy-H); protons from the major C_2-procut: 4.09 (s, 2H, 2 x NH_2), 6.33 (s, 1H, 1 x Py-H), 8.35 (s, 1H, 1 x Py-H); protons from the minor C_5-procut: 3.13 (s, 2H, 2 x NH_2), 6.91 (d, 1H, 1 x Py-H), 8.11 (s, 1H, 1 x Py-H)

^{31}P{^{1}H} NMR: (203 MHz, d_6-benzene): the major C_2-product: δ 23.76, δ 22.80; the minor C_5-product: δ 21.81, δ 20.86.

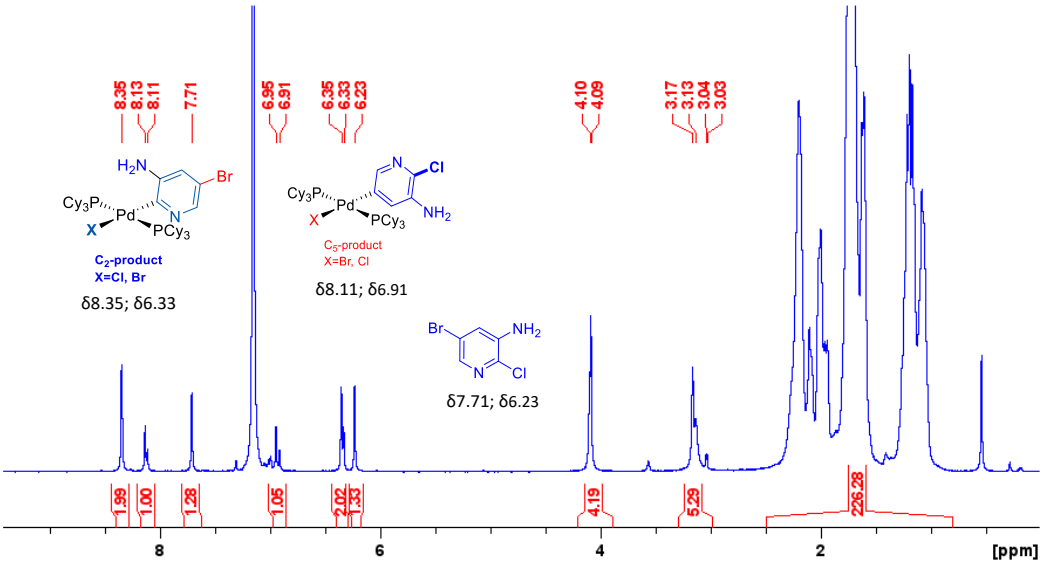


Fig. D57. ^{1}H (500 MHz, d_6-benzene) NMR spectrum of the reaction mixture of **B2** with $Pd(PCy_3)_2$.

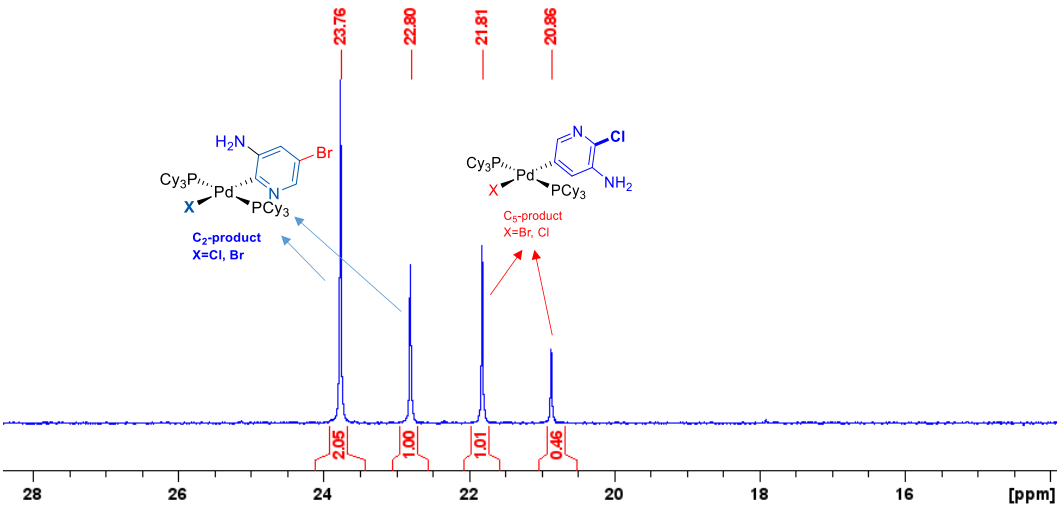


Fig. D58. ^{31}P{^{1}H} NMR (203 MHz, d_6-benzene) spectrum of the reaction mixture of **B2** with $Pd(PCy_3)_2$.

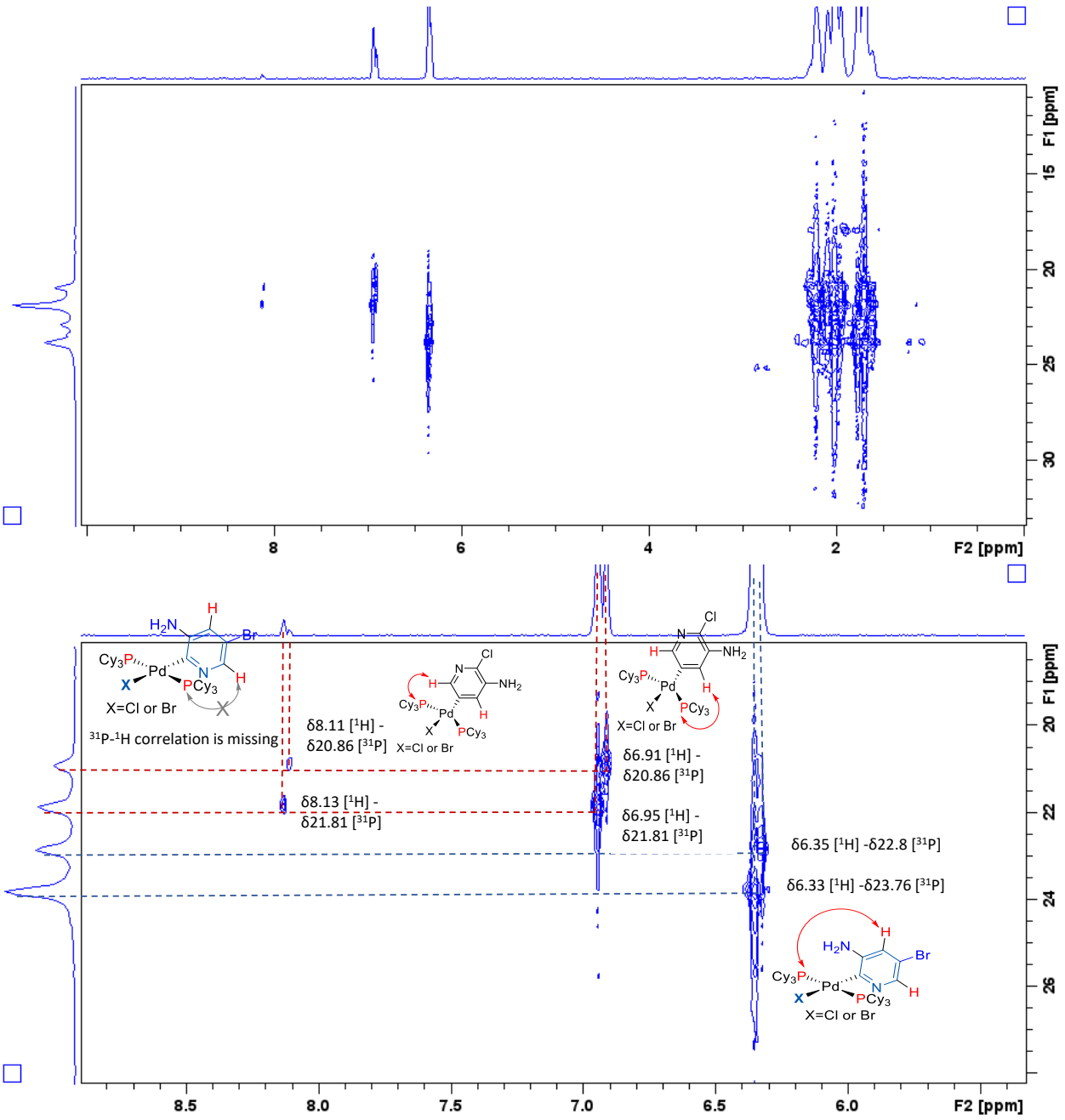
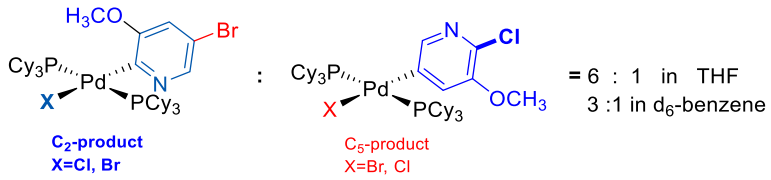


Fig. D58, continued. Top: the accompanying long range ^{1}H-^{31}P HMBC NMR spectrum; bottom: expansion of the pyridyl proton area and the ^{31}P-^{1}H correlation from the corresponding oxidative addition complexes.



B4

^{1}H NMR: (500 MHz, d_6-benzene): δ 1.05-2.21 (m, 66H, 66 x Cy-H); protons from the major C_2-procut: δ 3.22 (s, 3H, 3 x OCH_3), 6.54 (s, 1H, 1 x Py-H), 8.50 (d, 1H, 1 x Py-H); protons from the minor C_5-procut: δ 3.46 (s, 3H, 3 x OCH_3), δ 7.45 (s, 1H, 1 x Py-H) (it's 2 overlapping singlet peaks from the Pd$^{(2)}$-complex and its halide exchange analogue), δ 8.35 (d, 1H, 1 x Py-H).

^{31}P{^{1}H} NMR: (203 MHz, d_6-benzene): C_2-product: δ 22.06 and δ 21.13; C_5-product: δ 22.02 and δ 21.06.

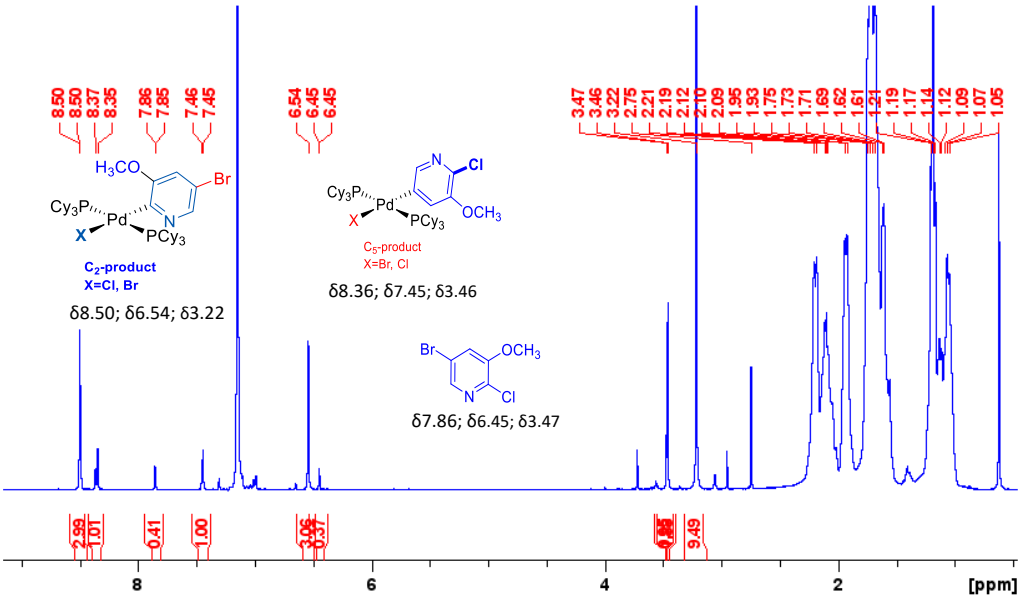


Fig. D59. ^{1}H (500 MHz, d_6-benzene) NMR spectrum of the reaction mixture of **B4** with $Pd(PCy_3)_2$.

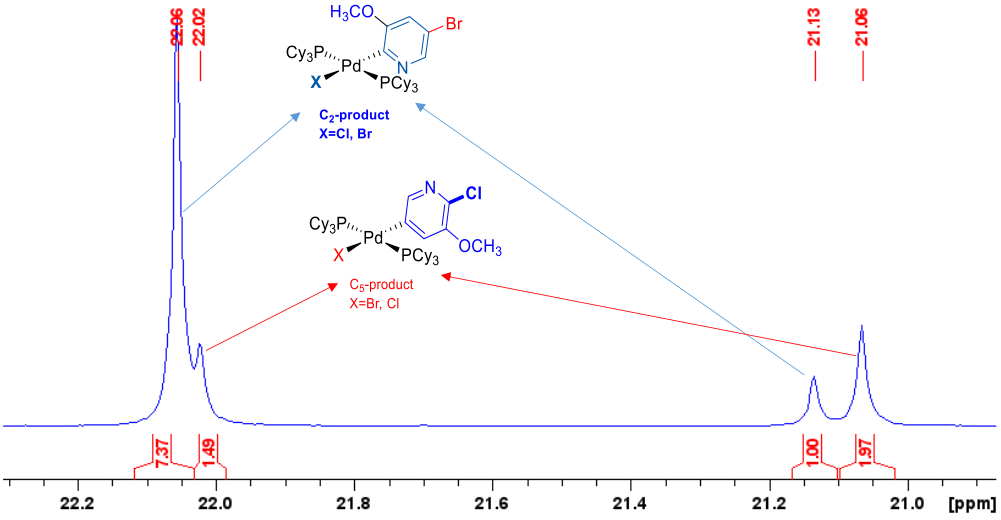


Fig. D60. ^{31}P{^{1}H} NMR (203 MHz, d_6-benzene) spectrum of the reaction mixture of **B2** with $Pd(PCy_3)_2$.

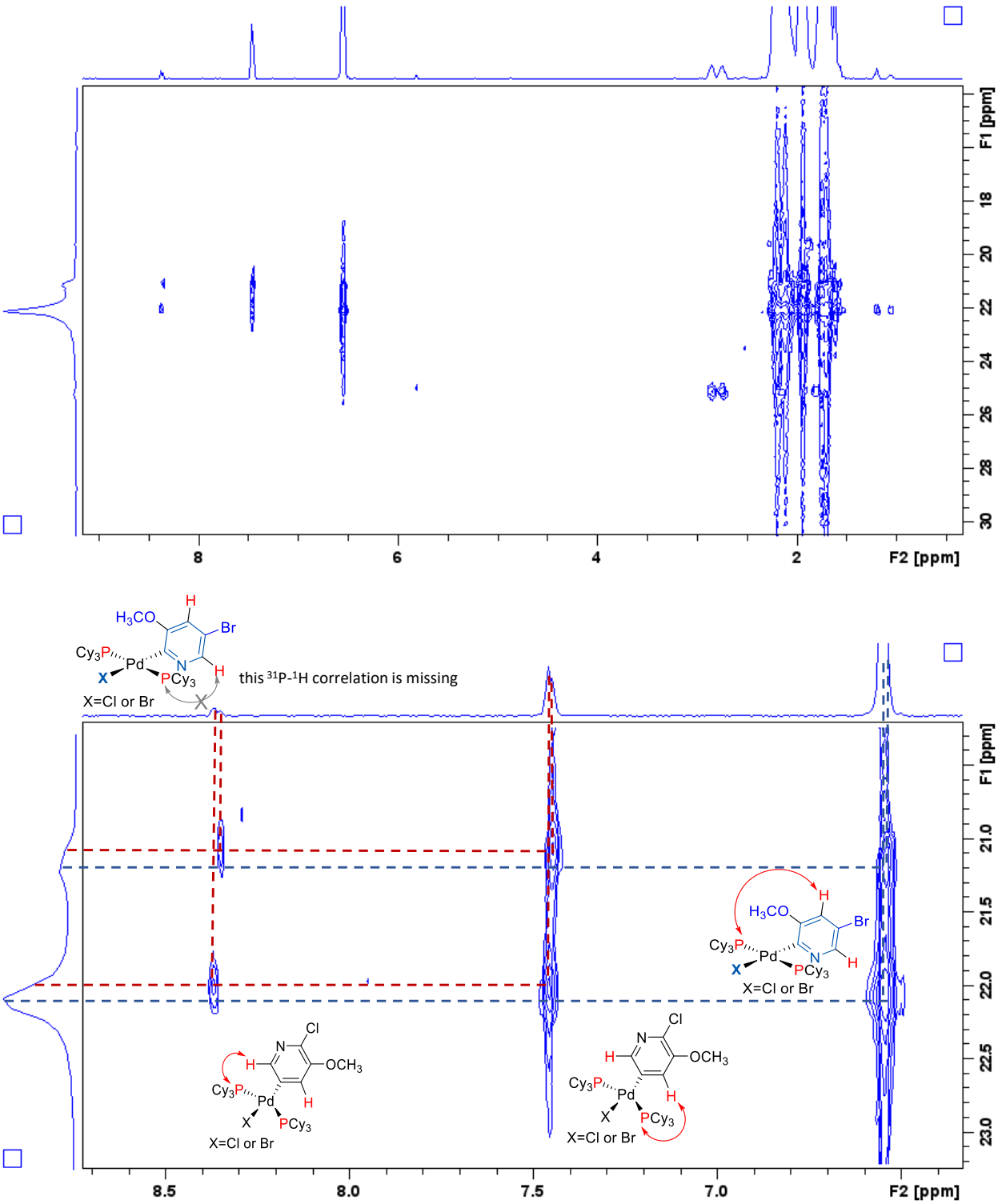


Fig. D60, continued. Top: the accompanying long range ^{1}H-^{31}P HMBC NMR spectrum; bottom: expansion of the pyridyl proton area and the corresponding ^{31}P-^{1}H correlation in the oxidative addition complexes.

Computational Determination of Molecular Descriptors including the %lumo coefficients

Our previously published multivariate linear regression model[3] takes 4 variables (molecular descriptors) to make accurate prediction of the oxidative addition outcome: average molecular electrostatic potential (ESP_1) at the reactive carbon, average molecular electrostatic potential (ESP_2) at the adjacent atom to the reactive center, sum of the A-values for substituents R_1 and R_2, and the intrinsic bond strength index (*IBSI*) of the C–X (X = Cl, Br) bond. The experimentally measured A-values are taken directly from published data, and the ESP and IBSI are from DFT calculations.

In this work we introduced two orbital-based descriptors: *%lumo(C)* and *%lumo(N)* to account for the frontier orbital symmetry effect for 2-halopyridine derivatives in oxidative addition. *%lumo(C)* is for the LUMO density at the reactive carbon, and it is a positive number; we use the LUMO coefficient of the nitrogen (*%lumo(N)*) with signs (+/-) to represent the orbital phase and the LUMO density at the nitrogen: a positive number indicate that it's a bonding interference between the C and N, and a negative number for antibonding interference.

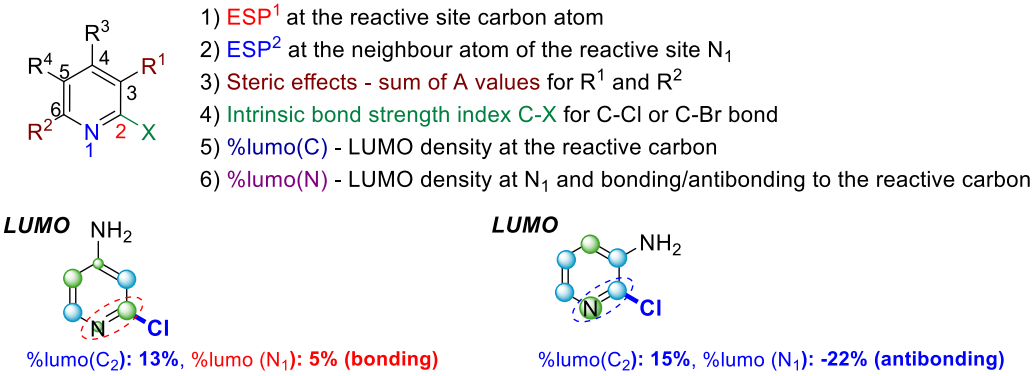


Fig. D61. The six molecular descriptors to construct the oxidative addition prediction model from Fig. 5.9B in main text; examples of the LUMO-based descriptors for bonding (2-chloro-4-aminopyridine) and antibonding (2-chloro-3-aminopyridine) symmetries through the C=N bond.

All geometry optimizations were performed using ORCA 4.0.1.2[4]. Initial substrate structures were either downloaded from the ChemSpider database[5] or generated using Avogadro version 4.1.[6] Geometry optimizations were performed in ORCA with a B3LYP/def-TVZPD approach. Wavefunction *.wfn* files were generated with the ORCA_2aim utility and imported in the Multiwfn program version 3.7[7,8] for calculation of electrostatic potential (*ESP*) and intrinsic bond strength index (*IBSI*). A-values[9] were obtained from published tables.

The molecular *ESP* was calculated at a 0.004 au isosurface of electron density. Surface properties for individual atoms were selected, including the atom-based surface area and the maximal, minimal and average *ESP* values at that surface Average *ESP* values at the reactive center (carbon) and its neighboring atom (nitrogen or carbon) were used as the electronic descriptors in construction of the oxidative addition predictive model. Electrostatic potential maps were plotted in VMD,[10] using electron density and *ESP* cube files generated by Multiwfn.

The *IBSI* was used as the bond strength descriptor between the reactive center (carbon) and the leaving group (Br, Cl, or OTf). *IBSI* is a recently proposed interpretation to quantify the bond strength of the covalent bonds.[11] The *IBSI* of the carbon-halogen bonds for each substrate was calculated using the Hirshfeld independent gradient model (IGMH) at high quality in Multiwfn.

The molecular descriptors for the fifteen 2-chloropyridine derivatives discussed in this chapter are given in Table D3; and for the eight 2,6-dichloropyridine derivatives, four 2-chloro-5-bromopyridine derivatives and the three external substrates in Fig. 5.11 are given in Table D4.

To obtain molecular orbital pictures, orbital cube files were generated in Multiwfn using .fch file as input, then plotted in VMD at a defined electron density isosurface. The LUMO coefficients were calculated by Multiwfn using .fch files as input and the functions under the category of *orbital composition analysis*.

Table D3 - Calculated molecular descriptors for the fifteen 2-chloropyridine derivatives discussed in Chapter 5 (Lu-PhD-thesis_Ch.5 SI tables.xlsx).

Table D4. Observed site-selectivity and calculated molecular descriptors for the eight 2,6-dichloropyridine derivatives, four 5-bromo-2-chloropyridine derivatives and the three external substrates in Fig. 5.11 (Lu-PhD-thesis_Ch.5 SI tables.xlsx).

Transition State Calculations for Oxidative Addition by DFT

All transition state calculations were performed in ORCA 4.0.1.2. Geometry optimization and frequency calculations for starting materials, intermediates, transition states and products were performed in THF with CPCM solvation model, at RI BP86 def2-SVP def2/J level with D3BJ dispersion for all atoms except for Pd, for which a def2-TZVP basis set was used. The optimized geometries of the starting materials, intermediates and products were confirmed by frequency calculations that no negative frequencies were found.

Transition state geometries were sought through relaxed scans along a simplified reaction coordinate (the breaking or forming bond length of the C-Cl bond) in CPCM(THF) at RI BP86 def2-SVP def2/J level with D3BJ dispersion for all atoms except for Pd, for which a def2-TZVP basis set was used. The geometry at the maximum energy point of the reaction coordinate scan was used as input geometry for continuing the transition state search via saddle-point optimization.

The optimization to a saddle point was performed in CPCM(THF) at RI BP86 def2-SVP def2/J level with D3BJ dispersion for all atoms except for Pd, for which a def2-TZVP basis set was used. The transition state structure was confirmed by vibrational analysis. The vibrational mode corresponding to the single imaginary frequency was visualized to ensure that it lied along the reaction coordinate bond stretch.

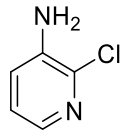
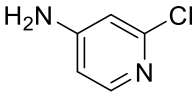
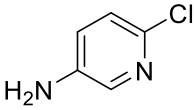
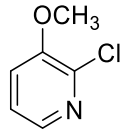
The free energies of the starting materials, intermediates and transition states were calculated as the sum of the electronic energies (single point energy at RI-B2PLYP D3 def2-TZVP def2-TZVP/C level) and the Gibbs free energy correction (obtained from frequency calculations stated above). Gibbs free energies of activation ($\Delta G^{\ddagger}_{SNAr}$) for the reactions were calculated as the difference in free energies between the transition state and the starting materials (unbound substrate and $Pd(PCy_3)_2$).

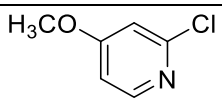
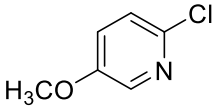
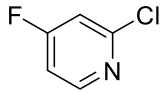
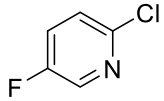
These labor-intensive transition state calculations were pursued for 9 2-chloropyridine derivatives and the results are summarized in Table D5. All transition state structures, including the bond length and bond angle of the essential atoms are shown in Figs. D62 – D70. All molecular images were generated in Vesta.[12] *HOMO* diagrams for the transition state were calculated for the 3 substrates of 2-chloro-x-aminopyridine (x=3,4,5) (Fig. 5.4, main text).

Charge analysis (electrostatic potentials and atomic charges) for the transition states was performed in Orca and Multiwfn. Since the effective core potentials are not supported by ORCA_2aim, which takes the gbw. file as input and generates the .wfn file. Single-point energy calculations with the all-electron relativistic ZORA-def2-TZVP basis set were carried out on the optimized transition state geometries to generate valid gbw. files for further calculations in Multiwfn.

Wavefunction *.wfn* files were generated with the ORCA_2aim utility and imported in the Multiwfn program version 3.7 for calculations of electrostatic potential (*ESP*) at 0.01 au isosurface of electron density and the Hirshfeld atomic charge.

Table D5. Energies of key species in the oxidative addition pathway (kJ/mol), relative to the unbound substrate and $Pd(PCy_3)_2$; and charge analysis for transition state geometries, where ΔESP is the difference in ESP at the Pd center from $Pd(PCy_3)_2$ to the transition state.

In CPCM(THF)		**Free energy (kJ/mol)**		**TS Charge analysis**		
Substrate	**TS type**	**Intermediate**	**Transition state**	**ΔESP at Pd (kJ/mol)**	***q*(Pd)[a]**	***q*(Py)[b]**
 2-chloro-3-aminopyridine	Stepwise Nucleophilic displacement	75.4	81.6	+89	+0.29e	-0.34e
 2-chloro-4-aminopyridine	Stepwise 3-centered	55.4	87.8	+83	+0.27e	-0.31e
 2-chloro-5-aminopyridine	Concerted 3-centered	Non-existent	94.9	+62	+0.24e	-0.20e
 2-chloro-3-methoxypyridine	Stepwise Nucleophilic displacement	69.9	80.4	+96	+0.32e	-0.42e

 2-chloro-4-methoxypyridine	Stepwise 3-centered	52.3	88.8	+76	+0.25e	-0.26e
 2-chloro-5-methoxypyridine	Concerted 3-centered	Non-existent	87.2	+79	+0.25e	-0.26e
 2-chloro-3-fluoropyridine	Stepwise Nucleophilic displacement	65.1	75.8	+114	+0.32e	-0.48e
 2-chloro-4-fluoropyridine	Stepwise 3-centered	33.74	84.86	+121	+0.30e	-0.44e
 2-chloro-5-fluoropyridine	Stepwise 3-centered	68.8	85.1	+96	+0.26e	-0.30e

[a] q(Pd) is the Hirshfeld atomic charge at the palladium of the transition state.
[b] q(Py) is the Hirshfeld charge carried by the pyridine (sum of the Hirshfeld atomic charges at all atoms of the 2-chloropyridine derivatives) of the transition state.

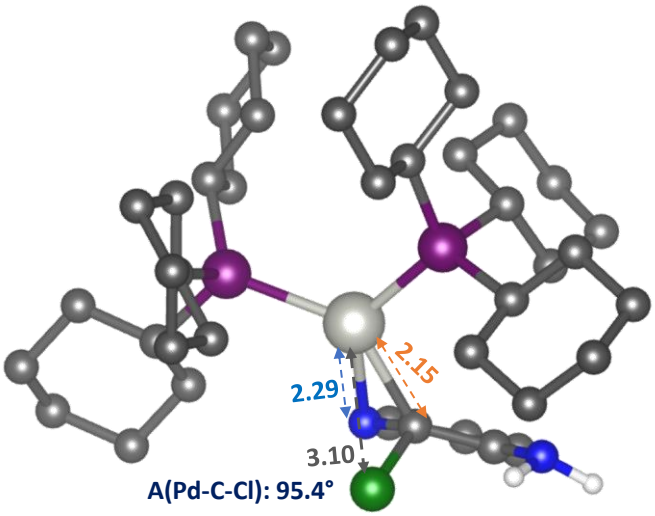


Fig. D62. Calculated transition state (TS) structure of oxidative addition of 2-chloro-3-aminopyridine to $Pd(PCy_3)_2$ in implicit THF. A stepwise nucleophilic displacement transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

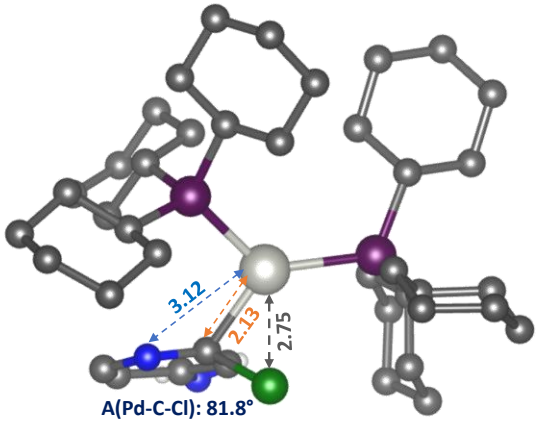


Fig. D63. Calculated TS structure of oxidative addition of 2-chloro-4-aminopyridine to $Pd(PCy_3)_2$ in implicit THF. A concerted 3-centered transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

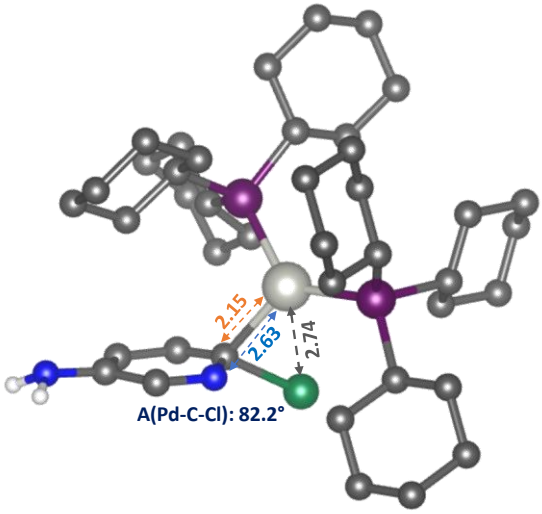


Fig. D64. Calculated TS structure of oxidative addition of 2-chloro-5-aminopyridine to $Pd(PCy_3)_2$ in implicit THF. A concerted 3-centered transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

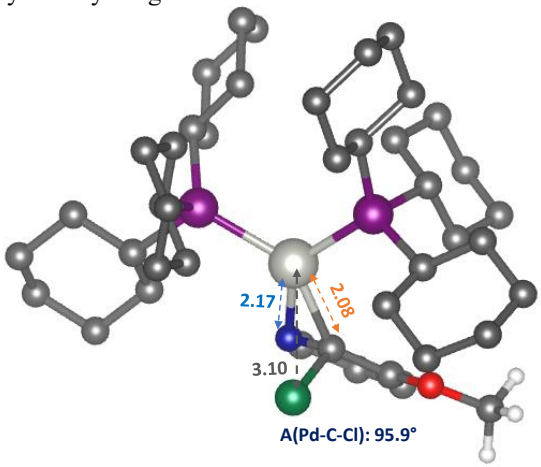


Fig. D65. Calculated TS structure of oxidative addition of 2-chloro-3-methoxypyridine to $Pd(PCy_3)_2$ in implicit THF. A stepwise nucleophilic displacement transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

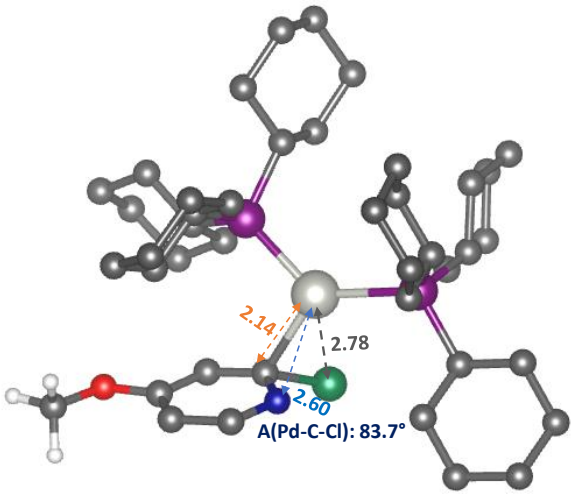


Fig. D66. Calculated TS structure of oxidative addition of 2-chloro-4-methoxypyridine to $Pd(PCy_3)_2$ in implicit THF. A concerted 3-centered transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

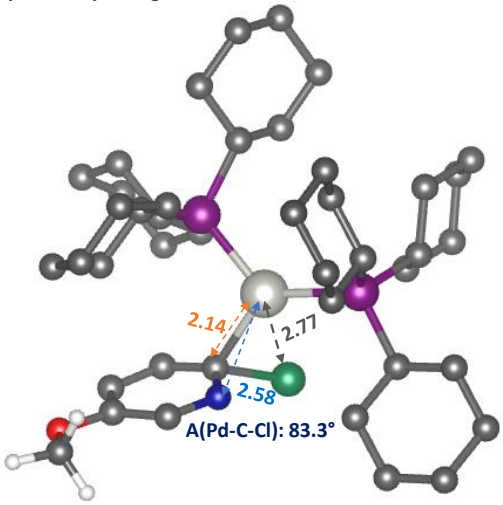


Fig. D67. Calculated TS structure of oxidative addition of 2-chloro-5-methoxypyridine to $Pd(PCy_3)_2$ in implicit THF. A concerted 3-centered transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

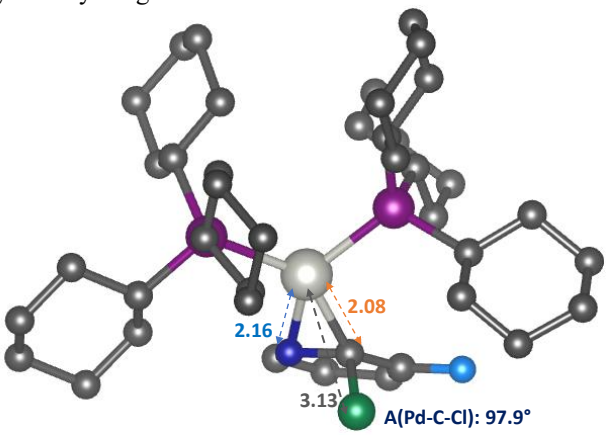


Fig. D68. Calculated TS structure of oxidative addition of 2-chloro-3-fluoropyridine to $Pd(PCy_3)_2$ in implicit THF. A stepwise nucleophilic displacement transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

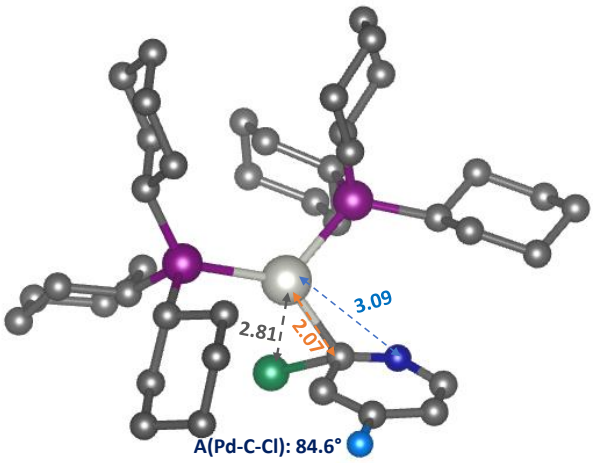


Fig. D69. Calculated TS structure of oxidative addition of 2-chloro-4-fluoropyridine to $Pd(PCy_3)_2$ in implicit THF. A stepwise 3-centered transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

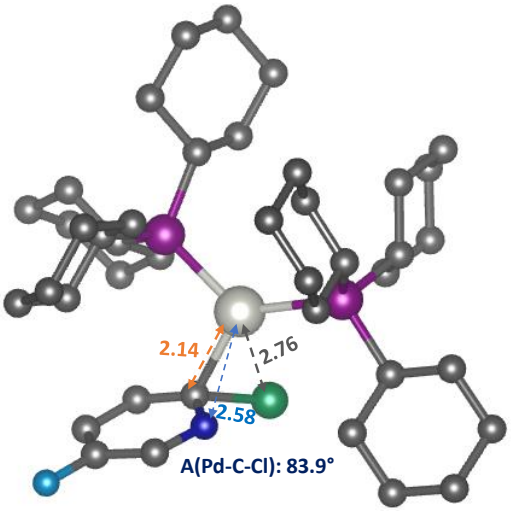


Fig. D70. Calculated TS structure of oxidative addition of 2-chloro-5-fluoropyridine to $Pd(PCy_3)_2$ in implicit THF. A stepwise 3-centered transition state has been found. Hydrogens on the pyridine and cyclohexyl rings are not shown.

Cross Validation for the six-descriptor predictive model from Fig. 5.9B (main text)

To evaluate the six-descriptor predictive model from Fig. 5.9B in main text, we performed cross-validation by doing five random 70/30 training/test data splits. Excellent linear correlation was achieved between the observed and predicted $\Delta G^{\ddagger}_{OA}$, as indicated by the range of R^2 from 0.83 to 0.95 for training set, Q^2 from 0.91 to 0.96, and MAE from 2.14 to 2.91 kJ mol^{-1} for test set. A plot for one such 70/30 split is shown in Fig. 5.9C in the main text, and the plots for other four divisions are shown in Figure D71. The good agreement between the observed and predicted $\Delta G^{\ddagger}_{OA}$ obtained from this random split cross-validation has indicated that our multivariate linear regression model is appropriately fitted, without overfitting issues.

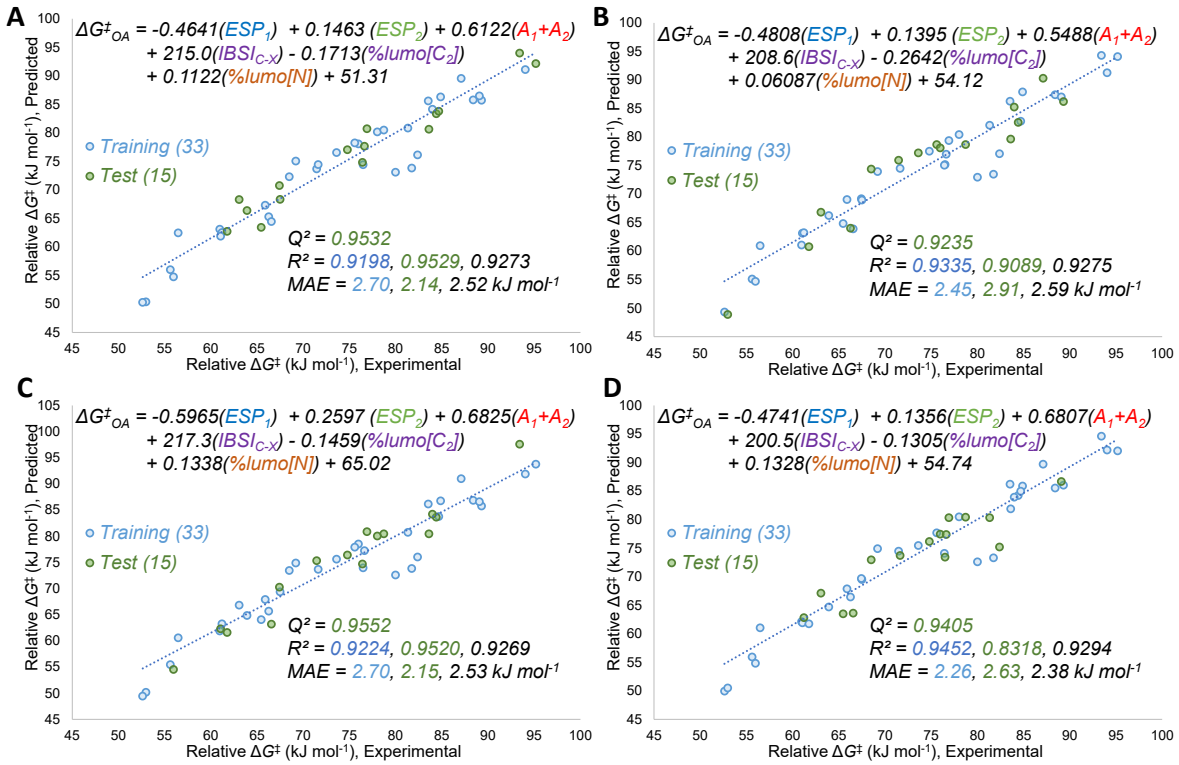


Fig. D71. (A) – (D) Multivariate linear regression model from one of the five 70/30 random split divisions (2/5) – (5/5).

References

(1) Rocaboy, R.; Dailler, D.; Baudoin, O. A Four-Step Synthesis of (±)-γ-Lycorane via Pd0-Catalyzed Double C(Sp2)–H/C(Sp3)–H Arylation. *Org. Lett.* **2018**, *20* (3), 772–775. https://doi.org/10.1021/acs.orglett.7b03909.

(2) Mccomas, C. C.; Liverton, N. J.; Habermann, J.; Koch, U.; Narjes, F.; Li, P.; Peng, X.; Soll, R.; Wu, H, Tetracyclic Heterocycle Compounds and Methods of Use Thereof for the Treatment of Viral Diseases. WO2013033900A1, 2013.

(3) Lu, J.; Donnecke, S.; Paci, I.; Leitch, D. C. A Reactivity Model for Oxidative Addition to Palladium Enables Quantitative Predictions for Catalytic Cross-Coupling Reactions. *Chem. Sci.* **2022**, *13* (12), 3477–3488. https://doi.org/10.1039/D2SC00174H.

(4) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA Quantum Chemistry Program Package. *J. Chem. Phys.* **2020**, *152* (22), 224108. https://doi.org/10.1063/5.0004608.

(5) Pence, H. E.; Williams, A. ChemSpider: An Online Chemical Information Resource. *J. Chem. Educ.* **2010**, *87* (11), 1123–1124. https://doi.org/10.1021/ed100697w.

(6) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminformatics* **2012**, *4* (1), 17. https://doi.org/10.1186/1758-2946-4-17.

(7) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33* (5), 580–592. https://doi.org/10.1002/jcc.22885.

(8) Lu, T.; Chen, F. Quantitative Analysis of Molecular Surface Based on Improved Marching Tetrahedra Algorithm. *J. Mol. Graph. Model.* **2012**, *38*, 314–323. https://doi.org/10.1016/j.jmgm.2012.07.004.

(9) *J. A. Hirsch Topics in Stereochemistry , John Wiley & Sons, Ltd, 1967, Pp. 199–222*.

(10) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(11) Klein, J.; Khartabil, H.; Boisson, J.-C.; Contreras-García, J.; Piquemal, J.-P.; Hénon, E. New Way for Probing Bond Strength. *J. Phys. Chem. A* **2020**, *124* (9), 1850–1860. https://doi.org/10.1021/acs.jpca.9b09845.

(12) Momma, K.; Izumi, F. VESTA: A Three-Dimensional Visualization System for Electronic and Structural Analysis. *J. Appl. Crystallogr.* **2008**, *41* (3), 653–658. https://doi.org/10.1107/S0021889808012016.