Enhanced Ligand Discovery through Generative AI and Latent-Space Exploration: Application to the Mizoroki-Heck Reaction

Authors: Wenxin Lu[1]†, Haote Li[1]†, Jan Paul Menzel[1], Abbigayle E. Cuomo[1], Andrea M. Nikolic[1], H. Ray Kelly[1,2], Yu Shee[1], Sanil Sreekumar[2], Frederic Buono[2], Jinhua J. Song[2], Robert H. Crabtree[1], Victor S. Batista[1]*, and Timothy R. Newhouse[1]*

Affiliations:

[1]Department of Chemistry, Yale University, New Haven, Connecticut 06511, United States.

[2]Chemical Development, Boehringer Ingelheim Pharmaceuticals Inc, Ridgefield, Connecticut 06877, United States.

*Corresponding authors. Email: timothy.newhouse@yale.edu, victor.batista@yale.edu

†These authors contribute equally.

Abstract: The identification of catalysts that promote chemical reactions is a critical challenge in the production of pharmaceuticals. One of the main bottlenecks in this process is the synthesis of vast libraries of precatalysts, although assessing catalyst effectiveness can be rapidly conducted through high-throughput experimentation. The rational design and development of high-performing precatalysts can circumvent this challenge and lead to important advances. In this study, we apply the transformer-based Kernel-Elastic Autoencoder (KAE) equipped with a conditioned latent space, enabling the targeted generation of ligands with desired steric and electronic properties. Our KAE model has facilitated the identification of a monodentate alkynylphosphine, dubbed MachinePhos A, as an effective precatalyst for forming carbon-carbon bond. Its utility was demonstrated experimentally in the Mizoroki-Heck reaction, using a variety of nitrogen-rich arenes pertinent to pharmaceutical applications.

One-Sentence Summary: Unanticipated ligand discovered with the generative model guided exploration

Main Text:

The discovery of new catalyst systems for specific organic reactions has led to important advances in organic synthesis (*1, 2*). In reactions catalyzed by transition-metals, precatalysts typically require modulation through the addition of suitable ligands (*3–5*). The most common methods for identifying new ligands are either experimental screening, via high-throughput experimentation (*6, 7*), or rational design, guided by calculated transition state structures (*8–10*). These approaches have resulted in considerable advances in organic synthesis. However, artificial intelligence (*11–14*) now appears ready to assist identifying promising ligand candidates. Here we demonstrate that generative modeling, by sampling from a latent-space, allows us to identify such ligands starting from desired properties rather than well-known structural motifs. The advantage of this approach is the emergence of unconventional structures that might not have been considered or investigated otherwise.

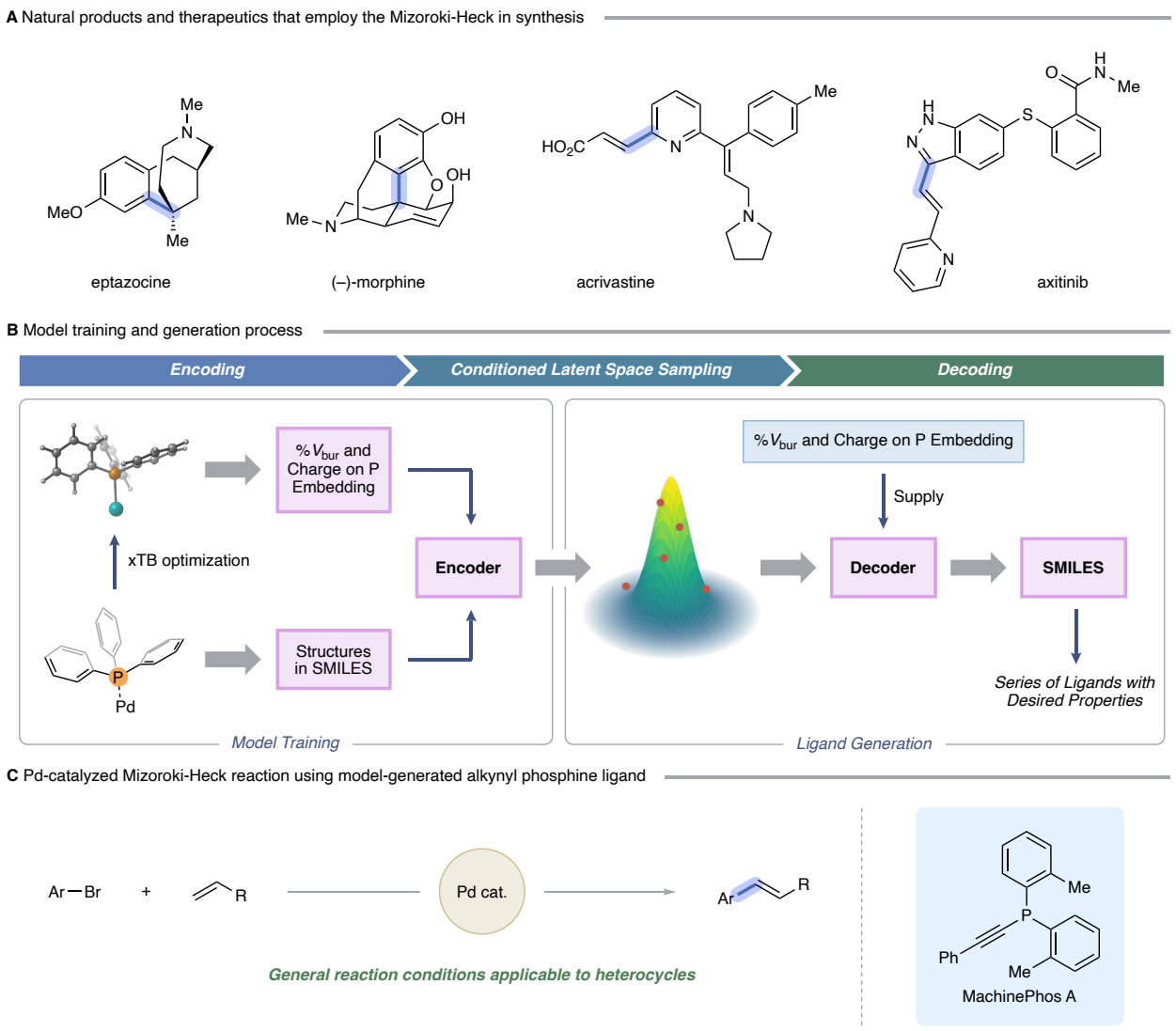

Fig. 1. Pd-Catalyzed Mizoroki-Heck Cross-Coupling Reaction and the Model-Enabled Generation of New Ligands

Our generative model, featuring a structured latent space, has enabled us to discover remarkably diverse ligand structures. This approach has particularly highlighted a series of monodentate alkynylphosphines (MAPs), a class of ligands previously overlooked, now revealed to have desirable catalytic properties. We synthesized and tested the prototype MAP in the Mizoroki-Heck reaction, and found it consistently achieved high yields on various pharmaceutically relevant heterocyclic substrates.

The Mizoroki-Heck reaction has seen widespread application to the synthesis of fine chemicals and materials (*15*). In small molecule synthesis, the intramolecular variant has permitted the synthesis both of sterically congested materials as notably demonstrated by Shibasaki (*16*) and Overman (*17*), and the formation of heterocyclic styrenes as shown in Figure 1A (*18*). Mechanistic study, ligand development, and exploration of the scope of this transformation have all served to advance its use in multistep synthesis (*19, 20*). The emergence of advanced artificial intelligence (AI) technologies is set to further enhance the versatility and efficiency of this widely used reaction. To discover new phosphines that could be broadly applicable in the Mizoroki-Heck reaction, we employed a machine learning (ML) approach, as outlined in Figure 1B.

ML-based generative models represent a unique, extremely useful area of artificial intelligence (*21*). The Generative Pre-trained Transformers (GPT) method, utilizing the Large Language Model (LLM) framework, is renowned for its multi-disciplinary achievements (*22, 23*). In the broad category of chemistry research, molecules can be effectively represented as strings of text (*24*). One common format is the Simplified Molecular-Input Line-Entry System (SMILES). Variational Autoencoders (VAEs), which can take SMILES strings as inputs, have gained growing attention (*25*). Leveraging LLM capabilities allows for the retrieval of SMILES-represented molecules with essential chemical properties from well-trained generative models. These models are applied across chemistry and can propose candidate structures with customized properties (*26*). The Kernel-Elastic AutoEncoder (KAE) applies a modified Maximum Mean Discrepancy (MMD) loss combined with a weighted reconstruction loss, differing from the KL-divergence used in VAEs (*27*). The advantage of KAE over VAE lies in its higher success rate in proposing new (valid) molecules, and its precision in identifying significant molecules within its latent space (reconstruction), facilitating clear unambiguous interpretation of each latent vector.

In the context of ligand design, various methods have been developed for the enumerating potential ligands motifs. A notable advancement is a data-driven approach for ligand screening utilizing the Kraken database with a vast collection of organophosphorus compounds (*28*). Most state-of-the-art approaches for *de novo* catalyst design use molecular fragments to generate and evaluate catalyst candidates (*29*). In contrast, in this work, over 300,000 ligands from the Kraken database were used to train the KAE with multiple targeted objectives (conditions) of interest, so that the model could learn to propose novel structures with desired properties. This methodology eliminates the need for manual search for candidate structures, a step which is traditionally followed by the calculation of a relevant property of interest (*30*). Consequently, human evaluation is guided by model-generated suggestions, free from potential biases that might be associated with naive chemical intuition (*31, 32*).

In this work, candidate ligands based on desired properties of interest based on experimentally limiting factors. Specifically, the Kraken database of phosphines encoded as SMILES was taken and post-labeled with calculated buried volumes to train our model. Considering that certain steric (%V_{bur}, *33, 34*) and electronic (Mulliken charge at phosphorus) properties could serve as predictive indicators of ligand performance, we produced a collection of bulky phosphine ligands labeled with these descriptors. The additional application of a conditioned

latent-space allows for the generation of sets of molecules predicted to have those desired properties (Figure 1B). Surprisingly, the monodentate alkynylphosphine motif emerged as an unconventional ligand in any context to promote catalytic cross-coupling reactions (*35*). The model contradicted our prior assumption that a sterically large ligand should not have a small substituent, such as the alkynyl group, that may itself coordinate to Pd. The revelation has led us to explore its potential in catalytic cross-coupling. This includes its ability to coordinate in an η^3 fashion (*36*), the possibility of alkyne pre-complexation followed by coordination to phosphorus, and likely reactivity under catalytically relevant conditions to attain a productive speciation of palladium. These possibilities were intriguing and merited further experimental investigation.

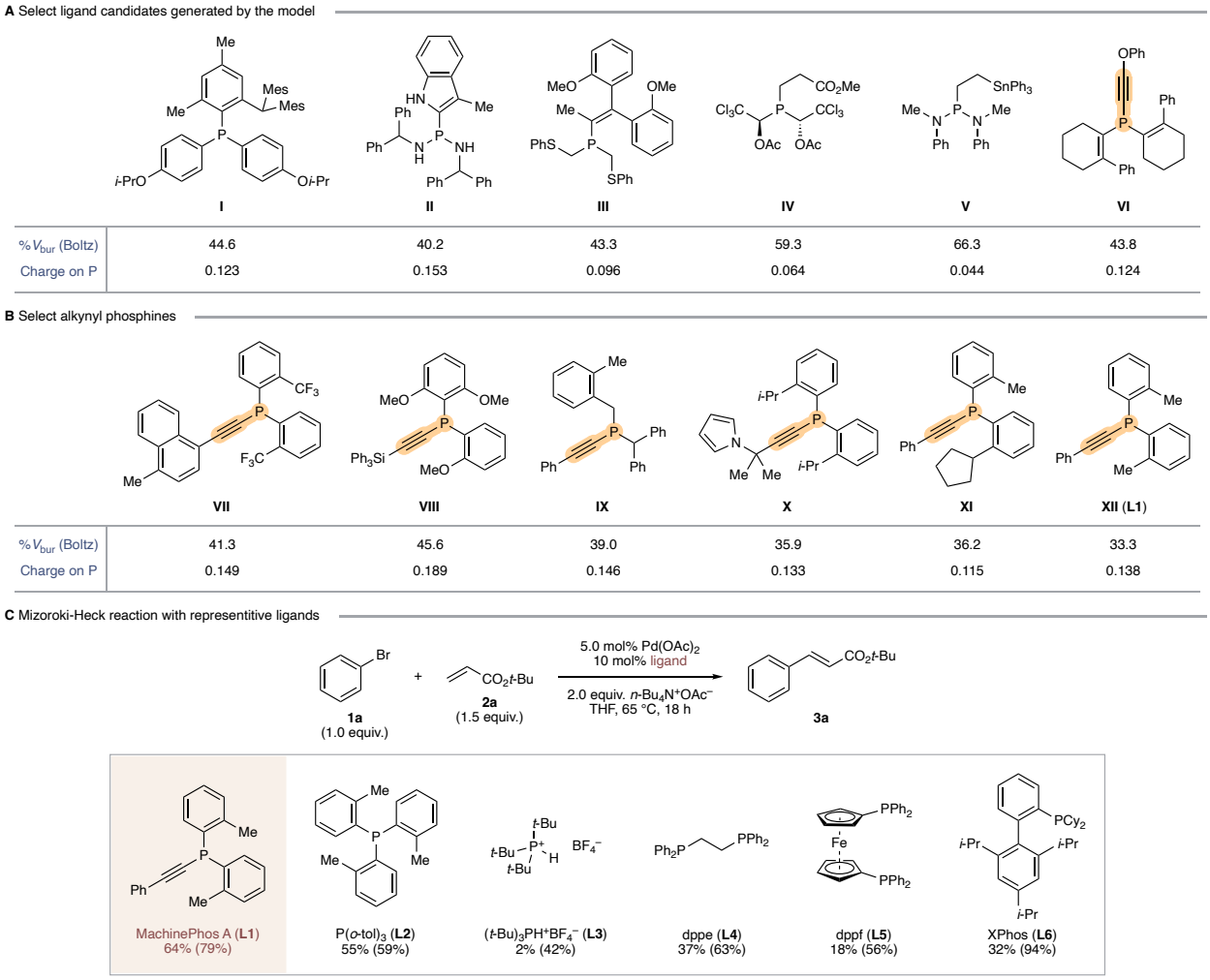

Fig. 2. Ligand Selection and Evaluation. A. Generated ligands with diverse architectures considered for evaluation. B. A variety of MAPs were considered for ligand synthesis. C. Evaluation of a MAP alongside several other commonly employed phosphines in the Mizoroki-Heck reaction.

Standard electronic structure methods, such as Density Functional Theory (DFT), are computationally demanding for analyzing large datasets of molecules on the order of 10^5 as necessary for training LLMs. Classical force field approaches, while suitable for large datasets, fail to account for electronic structure information and dispersion interactions, such as π-stacking,

that can significantly influence reaction outcomes. Therefore, the highly efficient semi-empirical GFN2-xTB method (*37*), which preserves essential electronic structure information, was used to derive molecular properties. RDkit (*38, 39*) was utilized to prepare geometries of the ligand-substrate complexes (see SI for details). These geometries were optimized using GFN2-xTB, and the buried volume was calculated with the Morfeus library (*40*). A high-throughput script was created to optimize the geometry of over 300,000 phosphine ligand candidates, incorporating an automatic recognition and expulsion of failed optimizations, such that structural, energetic, electronic and bonding data were efficiently processed for training and testing of the model.

With this model in hand, we found that the unusual ligand candidate structures shown in Figure 2A were suggested. These phosphines were analyzed based on their computed buried volume and partial atomic charges on the phosphorous atoms. The generated candidates exhibited mean $\%V_{bur}$ and Mulliken charge that correlate to the desired conditions (see SI Figure S1 and S2). In addition, these ligands were assessed based on their presumed synthetic accessibility, their expected air-stability, and their structural novelty. Common ligand types were excluded, particularly those resembling well-known classic ligands such as triaryl phosphines, e.g. P(*o*-tol)$_3$, or Buchwald-type biphenyl-containing phosphines (e.g. XPhos). Setting these aside, we focused on exploring the entirely novel ligand structures.

Among the ligand structures that met our basic design criteria, the MAP (**VI**) stood out, prompting us to refine the output of our model to highlight additional MAPs. We chose to synthesize a representative MAP (MachinePhos A, **L1**) that was readily accessed from commercially available materials and was adjustable for future design variations, such as aryl-substituted alkynes for electronic effect tuning and modifiable ortho-substituents on the aryl groups linked to phosphorus, aligning with Buchwald's methodologies. Unsurprisingly, we found that this structurally simple phosphine had a literature precedent (*41*); however, its potential in catalysis had been overlooked. A recent study has explored pincer tridentate ligands with N,P,N-coordination, incorporating an alkynyl group on phosphorous (*42*). We examined our alkynylphosphine (**L1**) in the Mizoroki-Heck reaction, and to our knowledge, this is the first example of a catalytic reaction being promoted by a MAP. With its bifunctional nature, this phosphine provides unique coordination possibly through the alkyne, potentially endowing it with novel and interesting properties.

We selected bromobenzene (**1a**) and *tert*-butyl acrylate (**2a**) as model substrates to evaluate the alkynyl phosphine ligand (**L1**). Since the Mizoroki-Heck reaction for this relatively simple model reaction had previously been reported in high yield, it served as a proof-of-concept of our approach. We assessed a selected set of reaction conditions by altering the solvent, base, temperature and reaction time, details of which are available in the Supplementary Material. As shown in Figure 2C, with *t*-butyl ammonium acetate in THF at 65 °C, the reaction achieved high conversion (79%) in 18 hours, with a yield of 64%. For comparison, various monodentate ligands and well-known bisphosphines were also tested under the same reaction conditions. These ligands afforded the product with yields ranging from 2%–55%. Moreover, in the absence of a phosphine ligand under optimized conditions, the product was only obtained in trace amounts (3%). Optimization with these alternative ligands would likely enhance the results.

We quantified the efficiency of the alkynyl phosphine ligand in coupling challenging substrates commonly encountered in the Mizoroki-Heck reaction, particularly those containing heterocycles with basic nitrogen atoms. These substrate types have often been reported previously under various reaction conditions, typically requiring individual optimization. We tested several nitrogen-containing heterocyclic aryl bromides as coupling partners to gauge the versatility of our

optimized conditions. For instance, anilines **3b** and **3c** were produced with yields of 66% and 62%, respectively. Evidently, the more electron-deficient and potentially coordinating aminopyrimidine also affords the product successfully, suggesting that the presence of basic groups does not adversely affect the yield. Similarly, dimethyl amino pyrimidine (**3d**) was yielded in similar yield (68%). Utilizing 5-bromopyrimidine resulted in the desired product (**3e**) with a moderate yield (51%). With 2-bromopyridine as the substrate, low conversion was observed, possibly due to the detrimental coordination of adjacent nitrogen atoms to the catalyst. In contrast, 2-bromo-6-methylpyridine facilitated the reaction smoothly, achieving a 62% yield for product **3f**. The reaction involving bromopyridazine with an *ortho*-methyl group yielded a lower yield (**3g**, 22%), likely due to the coordination of the additional nitrogen atom in the heterocycle. Conversely, reactions with substrates like 3- and 4-bromopyridine, as well as 6-bromoquinoline, were successful, producing the desired products (**3h**–**j**) in moderate to high yields. However, 2- and 4-bromoimidazole did not produce the desired product, whereas 5-bromoimidazole was tolerated in the reaction, albeit yielding only 14% for product **3k**.

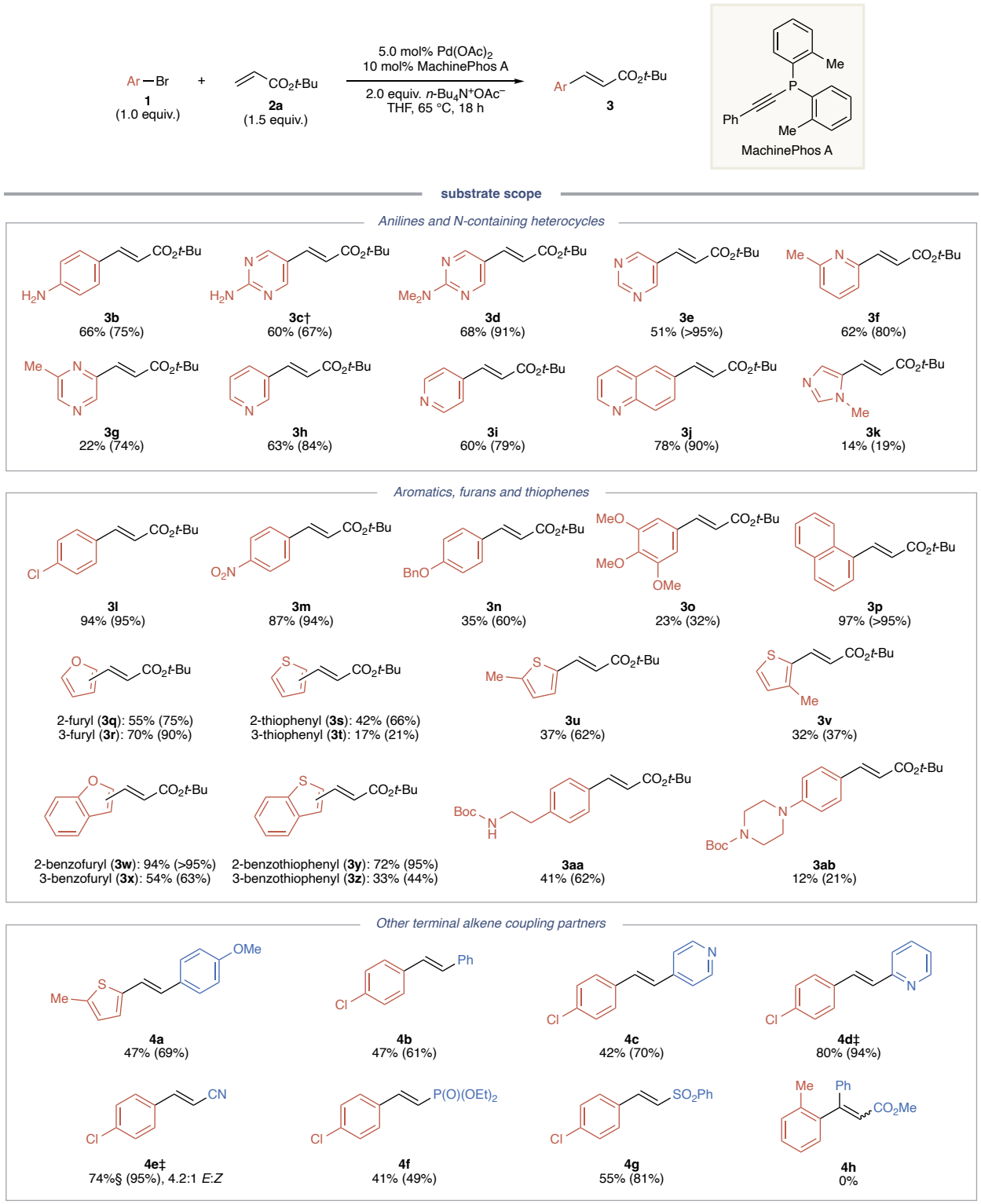

Fig. 3. Evaluation of the substrate scope for the optimized conditions. *Reactions were performed in 0.2 mmol scale. Isolated yields given. NMR conversion was shown in parentheses determined by ^{1}H NMR of the unpurified mixture using CH_2Br_2 as the internal standard. †The reaction was performed under 90 °C. ‡1-Chloro-4-iodobenzene was used. §Isolated yield of the *E*-isomer.

The scope of substrates was subsequently broadened to include other aryl bromides featuring phenyl or electron-rich heterocycles. The 4-chlorophenyl bromide underwent the reaction seamlessly, generating product **3l** with a high yield of 94%, demonstrating notable chemoselectivity for bromine over chlorine. A substrate with a more electron-withdrawing nitro group also produces the desired product **3m** in 87% yield. When evaluating bromophenyl substrates containing electron-rich substituents, such as benzyloxy and multiple methoxy groups (**3n**, **3o**), the outcomes indicated considerably lower conversion and yield. This suggests that the electron-rich groups might elevate the energy of the C–Br σ* bond, therefore hindering the oxidative addition step. Sterically hindered substrates, including 2,6-dimethyl-, 2,6-dimethoxy-, as well as the smaller 2,6-dichloro- and 2,6-difluorophenyl bromide, failed to generate products in significant yields (see SI). Conversely, 1-bromonaphthalene which is less hindered sterically, facilitated a smooth reaction, leading to product **3p** with a remarkable yield of 97%.

Comparing electron-rich heterocyclic substrates with electron-deficient nitrogen-containing ones, the 2- and 3-bromofuran and corresponding thiophene substrates without the fused arene ring could produce modest yield (**3q**–**t**). Notably, 3-bromothiophene produced **3t** in significantly lower yield compared to 2-bromothiophene, likely due to the electron-rich 3-substituent interfering with the oxidative addition and migratory insertion processes. The introduction of an additional methyl substituent to the thiophene (**3u**, **3v**) did not markedly affect the yield either positively or negatively. When using generally more electron-deficient benzofuran or benzothiophene substrates, higher yields were achieved (**3w**–**z**) in comparison to their monocyclic counterparts **3q**–**t**. In a different evaluation using 4-bromophenylethylamine, substrates with Boc-protected primary amines yielded product **3aa** in a moderate yield of 41%, whereas a more electron-rich piperazine substrate resulted in a lower yield of 12% for product **3ab**.

In the coupling reactions involving other alkene coupling partners, the combination of two electron-donating partners, 2-bromo-5-methylthiophene and *para*-MeO styrene resulted in product **4a** in 47% yield. Electron-neutral and electron-deficient styrenes were also effective in the reaction, producing **4b** and **4c** in 47% and 42% yield, respectively. The use of 2-vinylpyridine as a coupling partner afforded the product **4d** in low yield, indicating that the nitrogen atom in pyridine might engage in unproductively coordination with the catalyst. Employing a more reactive substrate, 1-chloro-4-iodobenzene, led to a significantly improved yield of 80% for product **4d**. The reaction between 1-chloro-4-iodobenzene and acrylonitrile yielded product **4e** in 74% yield, predominantly in the *E*-isomer form. The lower stereoselectivity (4.2:1 *E*/*Z*) is likely due to the small size of the cyano group, whereas other examples depicted in Figure 3, exclusively produced *E*-isomers. Vinyl phosphonate and vinyl sulfone served as effective coupling partners, delivering products **4f** and **4g** in moderate yields. However, the challenging substrate methyl cinnamate did not yield the desired product, likely due to the difficulties associated with migratory insertion.

This class of MAP ligands is expected to complement other frequently used ligands due to its straightforward synthesis and remarkable performance with many nitrogen-rich substrates, which are traditionally challenging in cross-coupling reactions. Many aspects of the mechanism are still being excavated under this complex system (*43*); however, a plausible pathway was explored using density functional theory (DFT). Through this approach, the potential binding modes of our MAP ligand type were examined with DFT, indicating that this ligand could participate in the reaction in mono- or bidentate modes.

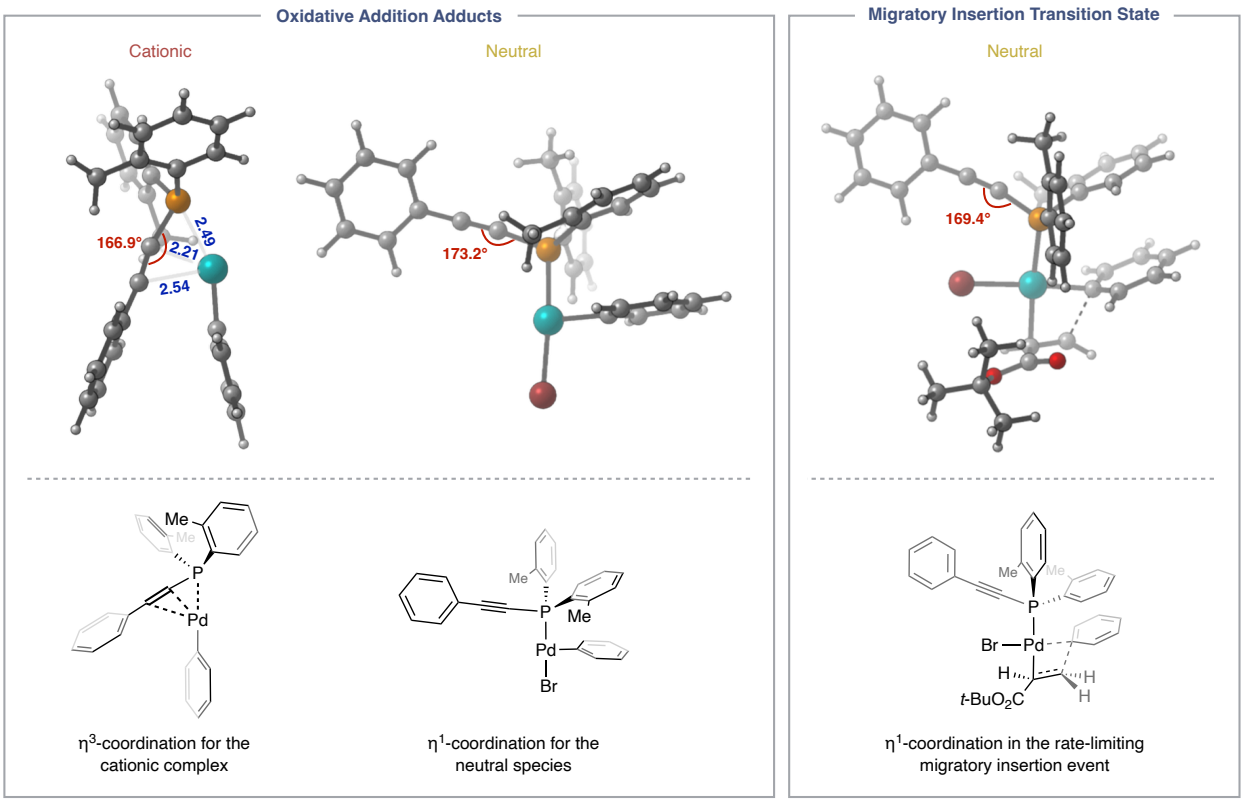

Fig. 4. Changing coordination mode of the alkynylphosphine motif as examined by Density Functional Theory in the cationic and neutral oxidative addition adducts with comparison to the migratory insertion transition state of the neutral pathway.

In the alternative cationic catalyst system, a distinct η^3-binding mode was observed, characterized by the involvement of the phosphorus center and both carbon atoms of the alkyne bonding to the metal. The calculated structure of the cationic complex [Pd(Ar)**L1**], reveals a Pd–P distance of 2.49 Å, with the alkyne bonding evidenced by distances of 2.21 Å to the proximal carbon and 2.54 Å to the distal carbon. This η^3-binding mode is reminiscent of that found in analogous all-carbon propargyl palladium complexes (*44*). The coordination of a bromide ion, leading back to the neutral pathway, results in a transition from the η^3- to the η^1-binding mode as shown in Figure 4 (*45–48*). Structures initially coordinated to the alkyne convert to the η^1-P-bound form upon optimization (*49*). Similarly, the coordination of the alkene coupling partner also results in a transition from the η^3- to η^1-binding of the MAP ligand (see SI). The concept of ligand hemilability, significant in catalysis (*50*), includes the adaptability of flexible bidentate ligands and the equilibration of propargyl palladium complexes. The shifting binding in this system may be apropos to other catalytic transformations.

At the design stage, these intricate mechanistic variabilities were not a concern, which highlights the advantage of employing a model trained on simple descriptors. In a system of such complexity, traditional physical organic approaches might be impractical, yet the machine learning method discussed here clearly remains capable for valuable insights.

References and Notes

1. R. H. Crabtree, *The Organometallic Chemistry of the Transition Metals* (Wiley, 7th ed. 2019).
2. J. F. Hartwig, J. P. Collman, *Organotransition metal chemistry: from bonding to catalysis* (University Science Books, 2010).
3. E. Negishi, Magical Power of Transition Metals: Past, Present, and Future (Nobel Lecture). *Angew. Chem. Int. Ed.* **50**, 6738–6764 (2011).
4. R. Martin, S. L. Buchwald, Palladium-Catalyzed Suzuki−Miyaura Cross-Coupling Reactions Employing Dialkylbiaryl Phosphine Ligands. *Acc. Chem. Res.* **41**, 1461–1473 (2008).
5. J. F. Hartwig, Evolution of a Fourth Generation Catalyst for the Amination and Thioetherification of Aryl Halides. *Acc. Chem. Res.* **41**, 1534–1544 (2008).
6. Y. Shen, J. E. Borowski, M. A. Hardy, R. Sarpong, A. G. Doyle, T. Cernak, Automation and computer-assisted planning for chemical synthesis. *Nat. Rev. Methods Primers* **1**, 23 (2021).
7. E. C. Hansen, D. J. Pedro, A. C. Wotal, N. J. Gower, J. D. Nelson, S. Caron, D. J. Weix, New ligands for nickel catalysis from diverse pharmaceutical heterocycle libraries. *Nat. Chem.* **8**, 1126–1130 (2016).
8. J. M. Ovian, P. Vojáčková, E. N. Jacobsen, Enantioselective transition-metal catalysis via an anion-binding approach. *Nature* **616**, 84–89 (2023).
9. A. J. Metrano, S. J. Miller, Peptide-Based Catalysts Reach the Outer Sphere through Remote Desymmetrization and Atroposelectivity. *Acc. Chem. Res.* **52**, 199–215 (2019).
10. G. Kiss, N. Çelebi-Ölçüm, R. Moretti, D. Baker, K. N. Houk, Computational Enzyme Design. *Angew. Chem. Int. Ed.* **52**, 5700–5725 (2013).
11. D. M. Anstine, O. Isayev, Generative Models as an Emerging Paradigm in the Chemical Sciences. *J. Am. Chem. Soc.* **145**, 8736–8750 (2023).
12. W. L. Williams, L. Zeng, T. Gensch, M. S. Sigman, A. G. Doyle, E. V. Anslyn, The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **7**, 1622–1637 (2021).
13. A. F. Zahrt, S. V. Athavale, S. E. Denmark, Quantitative Structure–Selectivity Relationships in Enantioselective Catalysis: Past, Present, and Future. *Chem. Rev.* **120**, 1620–1689 (2020).
14. A. E. Cuomo, S. Ibarraran, S. Sreekumar, H. Li, J. Eun, J. P. Menzel, P. Zhang, F. Buono, J. J. Song, R. H. Crabtree, V. S. Batista, T. R. Newhouse, Feed-Forward Neural Network for Predicting Enantioselectivity of the Asymmetric Negishi Reaction. *ACS Cent. Sci.* **9**, 1768–1774 (2023).
15. C. Torborg, M. Beller, Recent Applications of Palladium-Catalyzed Coupling Reactions in the Pharmaceutical, Agrochemical, and Fine Chemical Industries. *Adv. Synth. Catal.* **351**, 3027–3043 (2009).
16. T. Takemoto, M. Sodeoka, H. Sasai, M. Shibasaki, Catalytic asymmetric synthesis of benzylic quaternary carbon centers. An efficient synthesis of (-)-eptazocine. *J. Am. Chem. Soc.* **115**, 8477–8478 (1993).

17. C. Y. Hong, N. Kado, L. E. Overman, Asymmetric synthesis of either enantiomer of opium alkaloids and morphinans. Total synthesis of (-)- and (+)-dihydrocodeinone and (-)- and (+)-morphine. *J. Am. Chem. Soc.* **115**, 11028–11029 (1993).

18. J. Magano, J. R. Dunetz, Large-Scale Applications of Transition Metal-Catalyzed Couplings for the Synthesis of Pharmaceuticals. *Chem. Rev.* **111**, 2177–2250 (2011).

19. C. C. C. Johansson Seechurn, M. O. Kitching, T. J. Colacot, V. Snieckus, Palladium-Catalyzed Cross-Coupling: A Historical Contextual Perspective to the 2010 Nobel Prize. *Angew. Chem. Int. Ed.* **51**, 5062–5085 (2012).

20. K. C. Nicolaou, P. G. Bulger, D. Sarlah, Palladium-Catalyzed Cross-Coupling Reactions in Total Synthesis. *Angew. Chem. Int. Ed.* **44**, 4442–4489 (2005).

21. H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, A. Deac, A. Anandkumar, K. Bergen, C. P. Gomes, S. Ho, P. Kohli, J. Lasenby, J. Leskovec, T.-Y. Liu, A. Manrai, D. Marks, B. Ramsundar, L. Song, J. Sun, J. Tang, P. Veličković, M. Welling, L. Zhang, C. W. Coley, Y. Bengio, M. Zitnik, Scientific discovery in the age of artificial intelligence. *Nature* **620**, 47–60 (2023).

22. T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, "Language Models are Few-Shot Learners" in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin, Eds. (Curran Associates, Inc., 2020) vol. 33, pp. 1877–1901.

23. OpenAI, GPT-4 Technical Report. arXiv arXiv:2303.08774 [Preprint] (2023). http://arxiv.org/abs/2303.08774.

24. E. J. Corey, W. T. Wipke, Computer-Assisted Design of Complex Organic Syntheses: Pathways for molecular synthesis can be devised with a computer and equipment for graphical communication. *Science* **166**, 178–192 (1969).

25. C. W. Coley, W. H. Green, K. F. Jensen, Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **51**, 1281–1289 (2018).

26. R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).

27. H. Li, Y. Shee, B. Allen, F. Maschietto, V. Batista, Kernel-Elastic Autoencoder for Molecular Design. arXiv arXiv:2310.08685 [Preprint] (2023). http://arxiv.org/abs/2310.08685.

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

29. M. Foscato, V. R. Jensen, Automated in Silico Design of Homogeneous Catalysts. *ACS Catal.* **10**, 2354–2377 (2020).

30. G. Dos Passos Gomes, R. Pollice, A. Aspuru-Guzik, Navigating through the Maze of Homogeneous Catalyst Design with Machine Learning. *Trends in Chemistry* **3**, 96–110 (2021).

31. J. A. Hueffel, T. Sperger, I. Funes-Ardoiz, J. S. Ward, K. Rissanen, F. Schoenebeck, Accelerated dinuclear palladium catalyst identification through unsupervised machine learning. *Science* **374**, 1134–1140 (2021).

32. T. M. Karl, S. Bouayad-Gervais, J. A. Hueffel, T. Sperger, S. Wellig, S. J. Kaldas, U. Dabranskaya, J. S. Ward, K. Rissanen, G. J. Tizzard, F. Schoenebeck, Machine Learning-Guided Development of Trialkylphosphine Ni (I) Dimers and Applications in Site-Selective Catalysis. *J. Am. Chem. Soc.* **145**, 15414–15424 (2023).

33. A. C. Hillier, W. J. Sommer, B. S. Yong, J. L. Petersen, L. Cavallo, S. P. Nolan, A Combined Experimental and Theoretical Study Examining the Binding of *N* -Heterocyclic Carbenes (NHC) to the Cp*RuCl (Cp* = η^5-C_5Me_5) Moiety: Insight into Stereoelectronic Differences between Unsaturated and Saturated NHC Ligands. *Organometallics* **22**, 4322–4326 (2003).

34. S. H. Newman-Stonebraker, S. R. Smith, J. E. Borowski, E. Peters, T. Gensch, H. C. Johnson, M. S. Sigman, A. G. Doyle, Univariate classification of phosphine ligation state and reactivity in cross-coupling catalysis. *Science* **374**, 301–308 (2021).

35. H. Ito, H. Ohmiya, M. Sawamura, Construction of Methylenecycloheptane Frameworks through 7 -*Exo-Dig* Cyclization of Acetylenic Silyl Enol Ethers Catalyzed by Triethynylphosphine−Gold Complex. *Org. Lett.* **12**, 4380–4383 (2010).

36. J. R. Berenguer, E. Lalinde, M. T. Moreno, P. Montaño, Platinum(0) Complexes with Alkynylphosphane Ligands. *Eur. J. Inorg. Chem.* **2012**, 3645–3654 (2012).

37. M. Bursch, H. Neugebauer, S. Grimme, Structure Optimisation of Large Transition-Metal Complexes with Extended Tight-Binding Methods. *Angew. Chem. Int. Ed.* **58**, 11078–11087 (2019).

38. C. Bannwarth, S. Ehlert, S. Grimme, GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **15**, 1652–1671 (2019).

39. A. P. Bento, A. Hersey, E. Félix, G. Landrum, A. Gaulton, F. Atkinson, L. J. Bellis, M. De Veij, A. R. Leach, An open source chemical structure curation pipeline using RDKit. *J. Cheminform.* **12**, 51 (2020).

40. K. Jorner, kjelljorner/morfeus: v0.7.0, version v0.7.0, Zenodo (2022).

41. C. Ledesma, A. Alvarez-Larena, J. Suades, Rhenium carbonyl compounds with (diphenyl)phosphinoalkynes and a sterically hindered phosphinoalkyne. *J. Organomet. Chem.* **693**, 2775–2783 (2008).

42. W.-B. Xu, M. Sun, M. Shu, C. Li, Rhodium-Catalyzed Regio- and Enantioselective Allylic Amination of Racemic 1,2-Disubstituted Allylic Phosphates. *J. Am. Chem. Soc.* **143**, 8255–8260 (2021).

43. C. Amatore, A. Jutand, Anionic Pd(0) and Pd(II) Intermediates in Palladium-Catalyzed Heck and Cross-Coupling Reactions. *Acc. Chem. Res.* **33**, 314–321 (2000).

44. M. W. Baize, P. W. Blosser, V. Plantevin, D. G. Schimpff, J. C. Gallucci, A. Wojcicki, η^3-Propargyl/Allenyl Complexes of Platinum and Palladium of the Type $[(PPh_3)_2M(\eta^3\text{-}CH_2CCR)]^+$. *Organometallics* **15**, 164–173 (1996).

45. D. W. Lichtenberg, A. Wojcicki, Reactions of coordinated propargyl and allene ligands in cyclopentadienyliron dicarbonyl complexes. *J. Organomet. Chem.* **94**, 311–326 (1975).

46. T.-M. Huang, R.-H. Hsu, C.-S. Yang, J.-T. Chen, G.-H. Lee, Y. Wang, Organometallic Transformations from an η^3-Propargyl or η^2-Allenyl Ligand to η^3-Hydroxyallyl, η^3-Heterotrimethylenemethane, and η^6-Diallyl Ether Species. *Organometallics* **13**, 3657–3663 (1994).

47. W. Jiang, X. Yang, L. Lin, C. Yan, Y. Zhao, M. Wang, Z. Shi, Merging Visible Light Photocatalysis and P(III)-Directed C−H Activation by a Single Catalyst: Modular Assembly of P-Alkyne Hybrid Ligands. *Angew. Chem. Int. Ed.* **62**, e202309709 (2023).

48. Z. Li, Y. Fu, Q.-X. Guo, L. Liu, Theoretical Study on Monoligated Pd-Catalyzed Cross-Coupling Reactions of Aryl Chlorides and Bromides. *Organometallics* **27**, 4043–4049 (2008).

49. M. Ahlquist, G. Fabrizi, S. Cacchi, P.-O. Norrby, Palladium(0) alkyne complexes as active species: a DFT investigation. *Chem. Commun.* 4196 (2005).

50. K. Dong, R. Sang, Z. Wei, J. Liu, R. Dühren, A. Spannenberg, H. Jiao, H. Neumann, R. Jackstell, R. Franke, M. Beller, Cooperative catalytic methoxycarbonylation of alkenes: uncovering the role of palladium complexes with hemilabile ligands. *Chem. Sci.* **9**, 2510–2516 (2018).

Acknowledgments: This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAP0024372. We thank the Yale Center for Research Computing for guidance and use of the research computing infrastructure for Density Functional Theory calculations.

Funding: Boehringer Ingelheim is gratefully acknowledged for funding.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: code, and other electronic materials used in the analysis are deposited in a GitHub repository (Under Creation). Dataset created in this manuscript will be deposited as part of the supplementary material.

Supplementary Materials

Supplementary Text

Figs. S1 to S5

Tables S1 to S4

NMR Spectra

References (*51–87*)