

***In silico* rationalisation of selectivity and reactivity in Pd-catalysed C-H activation reactions**

Liwei Cao,^{a,b} Mikhail Kabeshov,^{c*} Steven V. Ley^c and Alexei A. Lapkin^{a,b,†}

^a*Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge
CB3 0AS, UK*

^b*Cambridge Centre for Advanced Research and Education in Singapore, CARES Ltd. 1 CREATE
Way, CREATE Tower #05-05, 138602 Singapore*

^c*Department of Chemistry, University of Cambridge, Lensfield Rd, Cambridge CB2 1EW, UK*

Abstract

A computational approach has been developed to automatically generate and analyse the structures of the intermediates of palladium catalysed carbon-hydrogen (C-H) activation reactions as well as to predict the final products. Implemented as a high-performance computing cluster tool, it has been shown to correctly choose the mechanism and rationalise regioselectivity of chosen examples from open literature reports. The developed methodology is capable of predicting reactivity of various substrates by differentiation between two major mechanisms - proton abstraction and the electrophilic aromatic substitution. An attempt has been made to predict new C-H activation reactions. This methodology can also be used for the automated reaction planning, as well as a starting point for microkinetic modelling.

Introduction

Periodically, our knowledge of chemistry is enriched with new transformations that provide significant breakthroughs by enabling new synthetic strategies. Such examples in recent years include olefins metathesis [1] as well as C-C and C-N coupling reactions [2], among the most obvious examples. While these reactions undoubtedly had very significant impacts on the development of much cleaner and efficient chemical synthesis strategies, the early days of all new transformations are invariably challenging, with very slow and protracted paths from the

* Present address: Benevolent AI, Minerva Building, Babraham Research Campus, Cambridge CB22 3AT

† Corresponding author; email: aal35@cam.ac.uk

initial discoveries to the demonstrations of broad substrate applicability and robustness, that are expected of industrial catalytic processes. Today, there exist a number of fairly recently (re)discovered transformations, that are of potential high industrial significance, and where one can observe the same problem of a lack of robustness. Thus, any approach that may speed-up the transition from a discovery of a new transformation to it becoming a robust synthetic strategy, is highly desired.

Recent years have seen the emergence of new methods of research in chemistry and process development, which include high-throughput experiments [3], autonomous self-optimising reactors [4], as well as predictions of reaction outcomes and of reaction conditions based on machine learning (ML) and artificial intelligence (AI) tools [5, 6]. Especially the methods of ML/AI for prediction of reaction outcomes are attracting a lot of attention. Prediction accuracies in the order of 70-80% for the reaction outcomes [6], and around 60-70% for reaction conditions [7], were recently demonstrated. While machine learning methods are showing great promise and continue to be improved upon, it is also clear that a ML model is unlikely to ever be able to compete in accuracy and interpretability with fully predictive mechanistic models, were it not for the prohibitively high cost of developing the mechanistic models based on accurate quantum chemical methods, such as the density functional theory (DFT) methods. Automation of DFT, as well as using results of DFT to develop less expensive predictive models, are the two approaches that may offer the alternatives to the fully data-driven statistical methods.

Here we demonstrate an approach that was developed to automate the DFT-level calculations of energies of the auto-generated reaction intermediates. These results were further used to generalize mechanistic knowledge of a class of reactions, and the developed models were used for *in silico* prediction of reactions outcomes. This approach was tested on the important for green chemistry class of C-H activation reactions. Whilst this study does not completely solve the problem of developing a robust chemical reaction, it offers an approach that is

complementary to efforts of developing machine learning models for predicting reaction outcomes.

C-H activation reactions allow conversion of relatively inexpensive and abundant hydrocarbons into the more sophisticated value-added molecules [8]. With the notion of step-economical and environmental friendly synthesis, direct functionalization of C-H bond is considered as a powerful strategy for the synthesis and derivatization of organic molecules [9]. Homogeneous catalysis employing transition metal complexes has been widely accepted as one of the most efficient ways to perform C-H activation-based synthesis with high selectivity under relatively mild conditions [10]. In particular, reactions involving palladium-catalyzed activation of sp^2 or sp^3 C-H bonds of arenes or alkanes have been extensively investigated due to their wide scope and functional group tolerance [11].

A number of different mechanisms are proposed in the literature, explaining the experimental observations on C-H activation reactions, depending on the nature of a ligand (L_n) and transition metal (M) in the catalytically active species (L_nM). These mechanisms include four elementary steps: oxidative addition, σ -bond metathesis, electrophilic substitution and 1,2-addition, respectively [12]. Even though the mechanisms are inherently different, three most important aspects should be primarily taken into account when classifying and rationalising C-H activation reactions:

- 1) the proximity of C-H bond to the transition metal;
- 2) the energy of C-H bond cleavage within the transition metal coordination sphere;
- 3) the energy of a new M-C bond formed and the thermodynamic stability of organometallic product.

With new developments in computational chemistry, mechanistic studies using density functional theory (DFT) provide valuable insights into the reactivity of organometallic complexes in C-H activation reaction. Moreover, along with the huge increase in computing power, this method becomes practically feasible to build model systems that provide

parameters of the actual experimental systems with acceptable accuracy [13]. However, there is a big challenge remaining which is to apply the computational analysis to a large number of mechanistically different transformations, both described and novel, in order to start generating accurate *in silico* reaction predictions. Here, we report an algorithm with high-performance computing (HPC) implementation, which has been developed to automatically generate and analyze the structures of the intermediates, and which allows prediction of the final products. The application of the developed methodology is in predicting reactivity for various substrates within a class of reactions. Using analysis of the computational data, a threshold to distinguish between two possible reaction mechanisms was established.

Computational methods

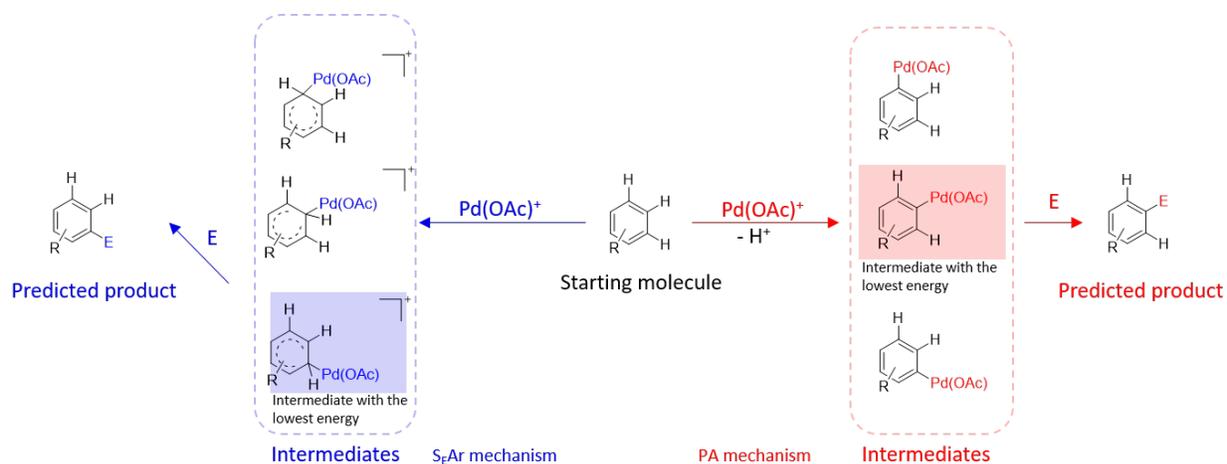
The NWChem, an open source software package, was used for the DFT calculations. It is easily scalable and designed to solve large scientific computational problems efficiently employing modern supercomputer clusters [14]. The structures were generated by the Python module developed in house and explained in detail elsewhere [15]. Electronic energies were evaluated using Becke's three-parameter hybrid B3LYP functional, while the molecular orbitals are expanded in triple-zeta all electron 6-31 set with added polarization and diffuse functions [6-31g(d,p)] [16]. B3LYP functional was proven to give accurate description of geometries, frequencies, relative stabilities of different conformers and the energy profile calculation [17]. Implementation of the tools is available at GitHub: https://github.com/sustainable-processes/Pd-catalysed_C-H_activation_reaction_prediction.

Results and Discussion

Computational approach to rationalise reactivity in Pd-catalyzed C-H bond activation reactions

Chemical reactivity is simultaneously influenced by many factors including catalyst, reactants, reaction conditions, and so on [18]. In order to achieve accurate and efficient reaction prediction, a mechanism-based method was chosen to direct quantum chemistry calculations and predictions, see Figure 1.

For the Pd(II)-catalysed C-H activation reactions, there are two main commonly accepted mechanisms: a) electrophilic aromatic substitution (S_EAr) mechanism and b) proton-abstraction (PA) mechanism. The key step for the electrophilic aromatic substitution is an electrophilic attack by Pd(II) onto the aromatic substrate that also defines the regioselectivity of the overall process [19]. The key feature of the proton abstraction (PA) mechanism [20] is that the formation of the metal-carbon bond (M-C) occurs simultaneously with the cleavage of the carbon-hydrogen (C-H) bond, while the hydrogen is being transferred to a basic center, Scheme 1.



Scheme 1. Schematic representation of the two mechanisms of Pd-catalyzed C-H activation reaction considered in this study.

Assuming the reaction proceeds through the formation of a relatively unstable intermediate (Figure 1) [21], Hammond postulate can be applied to the electrophilic substitution reactions. Hammond postulate states that a transition state will be structurally and energetically similar to the species (reactant, intermediate or product) nearest to it on the reaction path. In this case, the intermediates are likely to be close to, and resemble, transition states. Due to that, their relative energy of formation can be translated to relative reaction kinetic barriers and thus be used, as the first approximation, to predict distributions of the final products, as well as the relative reactivity of the substrates [22]. For the PA mechanism, it has not been shown that the Hammond postulate can also be employed. Nevertheless, it is still reasonable to propose that

the Hammond postulate can similarly be applied as a first approximation to produce *in silico* predictions.

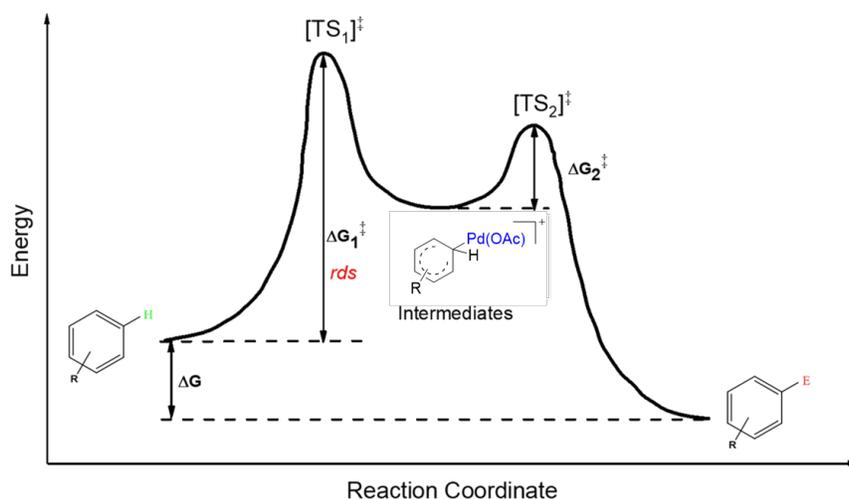


Figure 1. An approximate energy map for the electrophilic aromatic substitution mechanism.

Employing the Python module [15] and OpenBabel executables, the 3D structures of the most stable conformers were generated from 2D structure of a substrate. Subsequently, structures of all possible palladium intermediates representing both mechanisms (PA and $S_{E}Ar$) were built for each conformer. A quick geometry optimization (maximum number of iteration steps was set to 5) was then applied to refine the intermediates and discard the ones with high energy (energy cut off of $10 \text{ kcal} \cdot \text{mol}^{-1}$). Full geometry optimisation followed by the frequency and thermochemistry analysis was then performed for the selected intermediates to obtain electronic energies. Multiple error handlers were implemented in order to automatically reprocess computational analysis for the intermediates when initial geometry optimisation failed.

Literature validation

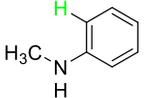
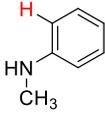
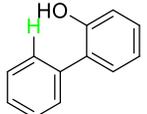
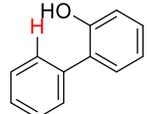
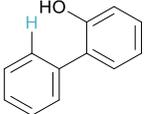
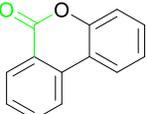
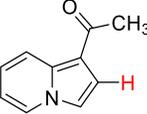
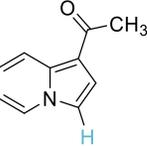
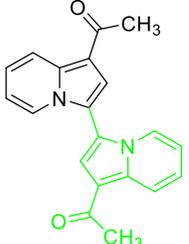
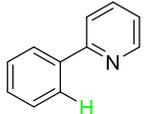
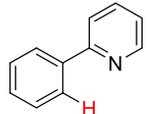
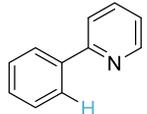
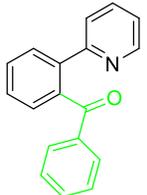
In order to test the developed algorithm, a representative literature data selection of Pd-catalyzed C-H activation reactions, consisting of reactant, reagents, and product structures as well as reaction conditions, was taken and analysed. Thus, twelve substrates shown in Table 1 were submitted to the algorithm, assuming that both mechanisms are possible. Using the relative energies of the intermediates obtained, the theoretically expected regioselectivity of

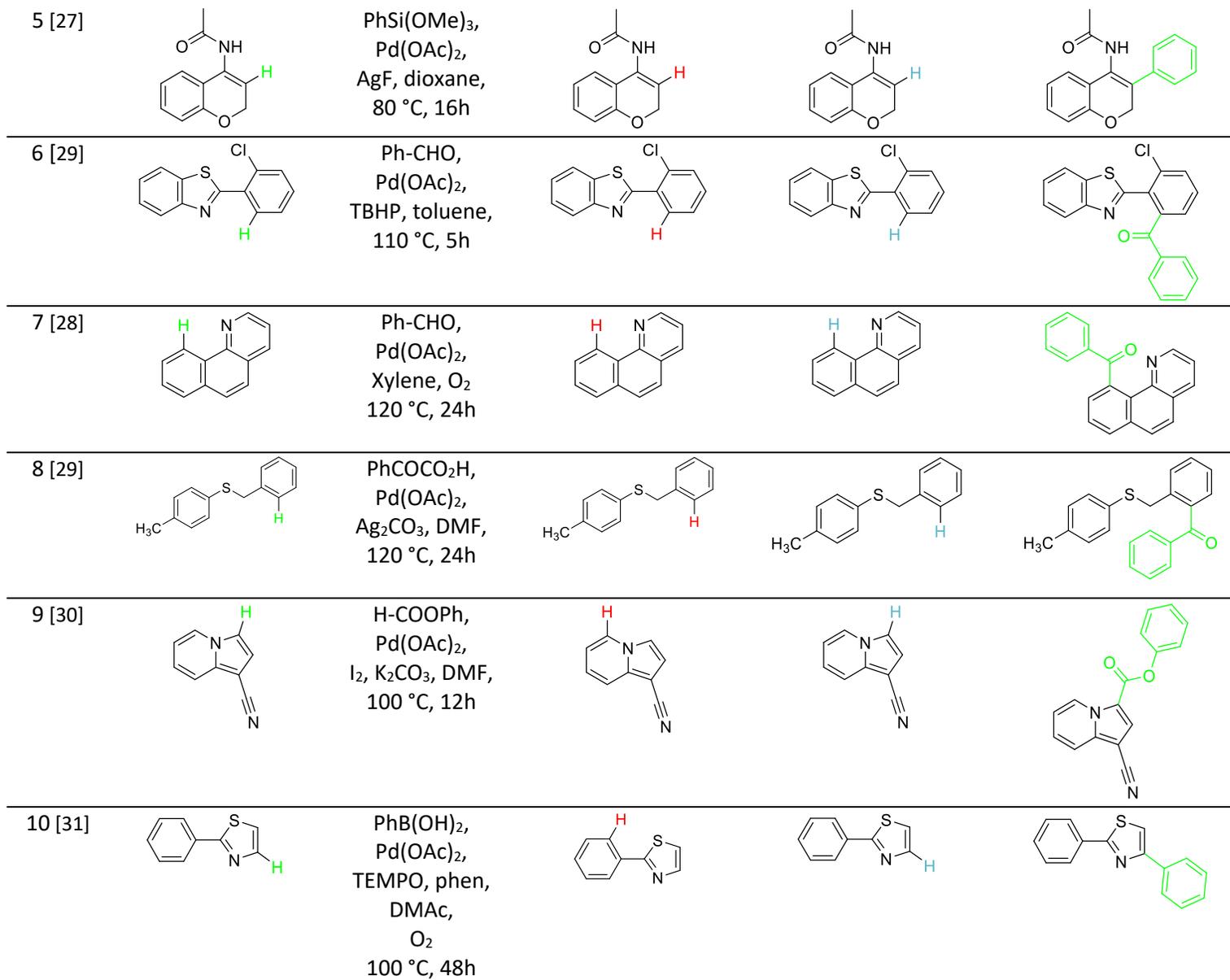
the selected reactions was devised and then compared against the previously reported experimental data.

For all the examples regioselectivity predicted by at least one mechanism matched the previously reported experimental results, see Table 1. In the cases where only one product was predicted it is expected to be isolated in high yield without the need of further purification from any other regioisomer. For the examples where formation of multiple products was expected due to the close energies of the respective reaction intermediates, the ratio of products was calculated from the relative energies of these intermediates using the Boltzmann distribution equation.

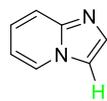
Table 1. Comparison of the published experimental results with the computational predictions for the Pd(OAc)₂-catalysed reactions.

Protons marked green are those that react under the conditions reported in the literature. Protons marked red and blue are the predicted active centers via the acidity and the electrophilic mechanisms respectively.

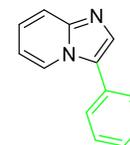
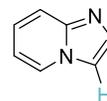
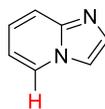
Entry [ref]	Starting molecule	Experimental conditions	Predicted active center		Experimentally isolated product
			Via acidity mechanism	Via electrophilic mechanism	
1 [23]		CO, EtOH, Pd(OAc) ₂ , Cu(OAc) ₂ , KOAc, DMF, KI, 100 °C, 13h		NO STABLE INTERMEDIATE	
2 [24]		CO, Pd(OAc) ₂ , Cu(OAc) ₂ , PivOH, mesitylene, 120 °C, 6h			
3 [25]		Cu(OAc) ₂ , Pd(OAc) ₂ , K ₂ CO ₃ , DMF, 60 °C, 0.6h			
4 [26]		PhCOCO ₂ H Pd(OAc) ₂ , K ₂ S ₂ O ₈ , MeCN, 25 °C, 16h			



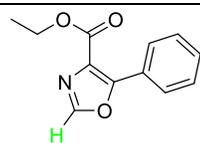
11 [32]



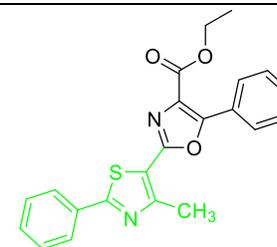
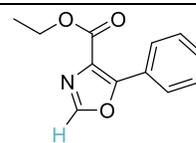
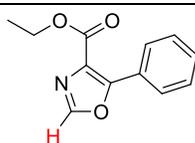
Benzene,
Pd(OAc)₂,
O₂, HOAc, DMA,
130 °C, 20h



12 [33]



CC(=O)Oc1nc(s1)C(=O)C
Pd(OAc)₂, CuCO₃,
dioxane, DMSO,
140 °C, 16h



Establishing the threshold between the two mechanisms

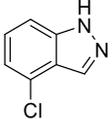
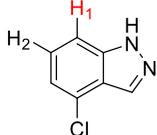
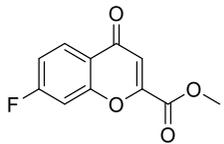
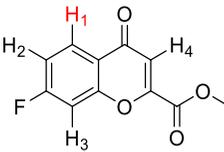
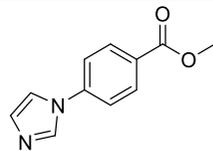
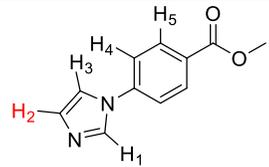
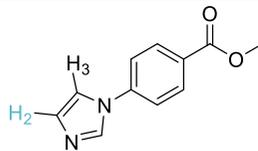
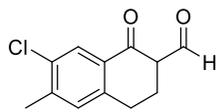
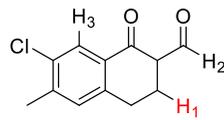
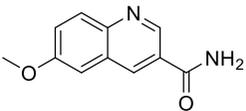
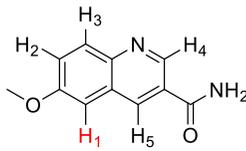
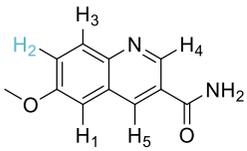
Although both, the proton-abstraction and the electrophilic aromatic-substitution, mechanisms are well established and described in the literature, it is not trivial to suggest the preferred mechanism for a given substrate based on a simple computational analysis. Through analysis of the results described above, the two-step evaluation algorithm was suggested.

Firstly, the optimized geometries were manually examined to ensure they represent the intermediates according to the particular mechanism. In particular, the bond length between the palladium atom and the corresponding carbon atom was given a maximum value of 2.4 Å to filter out inappropriate intermediates where there is no stable Pd-C bond [34].

Secondly, among the intermediates refined at the previous step, their relative Gibbs energies can be used to set a threshold establishing the likeliness of electrophilic aromatic substitution mechanism for C-H activation of a particular substrate. The more stable the ipso-complex between palladium acetate and the substrate is, the more likely the substrate is to follow the electrophilic mechanism. After performing the computational analysis of 12 examples which include five structures following the electrophilic mechanism, a threshold has been developed by choosing the example 6 as the reference, Table 1, and introducing the ipso-complex stability parameter. We define this parameter to be the energy difference between the most stable intermediate of the $S_{E}Ar$ mechanism and the one of the PA mechanism.

Table 2. Predicting C-H activation bond for heteroaromatic compound. Most probable intermediates for each mechanism are shown, and relative Gibbs free energy are given in kcal mol⁻¹. If only one possible intermediate is given, it means that either the other intermediates are unstable or the other intermediates have 10 more kcal mol⁻¹ Gibbs free energy than the most probable one. 'NO STABLE INTERMEDIATE' means instead of sitting on the corresponding carbon, the palladium sits on alternative atom. The predicted mechanism is given based on the threshold described in the previous section.

Entry	Starting molecule	Predicted mechanism	Computational prediction	
			Acidity mechanism	Electrophilic mechanism
1		S _E Ar		 H1:0.0 H2:0.1 H3:0.8 H4:1.0
2		S _E Ar		 H1:0.0 H2:0.5
3		PA	 H1:0.0 H2:0.7	 H2:0.0 H3:0.9
4		PA	 H1:1.9 H5:9.8 H6:0.0	 H1:0.0 H2:3.9 H3:2.9 H4:2.9 H5:4.2

Entry	Starting molecule	Predicted mechanism	Computational prediction	
			Acidity mechanism	Electrophilic mechanism
5		PA		H1:0.0 H2:4.6 NO STABLE INTERMEDIATE
6		PA		H1:0.0 H2:2.7 H3:9.7 H4:6.2 NO STABLE INTERMEDIATE
7		PA/ SeAr		H1:0.2 H2: 0.0 H3: 0.3 H4:0.9 H5:2.6 
8		PA		H1:0.0 H2:2.8 H3:10.0 NO STABLE INTERMEDIATE
9		PA		H1:0.1 H2:0.0 H3:14.9 H4:15.0 H5:15.8 

By comparing the computational results obtained to the literature experimental data, the two mechanisms can be segregated based on the following rules:

- 1) if the relative stability is below zero, the starting molecule will follow the proton abstraction mechanism.
- 2) If the relative stability is above five, the starting molecule will follow the electrophilic aromatic substitution mechanism.
- 3) If the relative stability is between zero and five, both mechanisms are regarded as plausible.

Although the rules set above seem rather approximate, they are consistent with the given examples, and further work aimed at increasing the accuracy and the scope of the algorithm is on-going. Based on the suggested rules, the predicted reactive centers for nine commercially available aromatic and heteroaromatic substrates as well as the most likely mechanisms are shown in the Table 2.

Conclusions

A computational algorithm rationalising the existing palladium catalysed C-H activation reactions has been developed. Computational threshold to distinguish between the two main mechanisms, proton abstraction (PA) and electrophilic aromatic substitution (S_EAr) mechanism, has been proposed and tested against the literature experimental data. This model can give not only the most probable reactive site and the appropriate mechanism, but also provides information for further kinetic studies and process development, thus contributing to the development of robust new chemical transformations.

Acknowledgements

This project was in part supported by EPSRC grant EP/K009494/1 (ML) and the National Research Foundation, Prime Minister's Office, Singapore under its CREATE programme, project "Cambridge Centre for Carbon Reduction in Chemical Technology" (LC and AL).

References

1. Grubbs, R. H.; Chang, S. *Tetrahedron* **1998**, *54* (18), 4413-4450.
2. Buchwald, S. L.; Mauger, C.; Mignani, G.; Scholz, U. *Adv. Synth. Catal.* **2006**, *348* (1-2), 23-39.
3. 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. *Science* **2015**, *347*, 49-53.
4. Echtermeyer, A.; Amar, Y.; Zakrzewski, J.; Lapkin, A. *Beilstein J. Org. Chem.* **2017**, *13*, 150-163; Holmes, N.; Akien, G. R.; Blacker, A. J.; Woodward, R. L.; Meadows, R. E.; Bourne, R. A. *React. Chem. Eng.* **2016**, *1* (4), 366-371, Article; Fitzpatrick, D. E.; Battilocchio, C.; Ley, S. V. *Org. Process Res. Dev.* **2016**, *20* (2), 386-394.
5. Segler, M. H. S.; Waller, M. P. *Chem. - A Eur. J.* **2017**, *23*, 6118-6128; Coley, C. W.; Jin, W. G.; Rogers, L.; Jamison, T. F.; Jaakkola, T. S.; Green, W. H.; Barzilay, R.; Jensen, K. F. *Chem. Sci.* **2019**, *10* (2), 370-377.
6. Schwaller, P.; Gaudin, T.; Lanyi, D.; Bekas, C.; Laino, T. *Chem. Sci.* **2018**, *9* (28), 6091-6098.
7. Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. *ACS Central Science* **2018**, *4* (11), 1465-1476.
8. Korwar, S.; Burkholder, M.; Gilliland, S. E.; Brinkley, K.; Gupton, B. F.; Ellis, K. C. *Chem. Commun.* **2017**, *53* (52), 7022-7025.
9. Roudesly, F.; Oble, J.; Poli, G. *J. Mol. Catal. A: Chem.* **2017**, *426*, 275-296.
10. Chen, X.; Engle, K. M.; Wang, D.-H.; Yu, J.-Q. *Angew. Chem. Int. Ed.* **2009**, *48* (28), 5094-5115.
11. McNally, A.; Haffemayer, B.; Collins, B. S. L.; Gaunt, M. J. *Nature* **2014**, *510* (7503), 129-133.
12. Balcells, D.; Clot, E.; Eisenstein, O. *Chem. Rev.* **2010**, *110* (2), 749-823.
13. Dedieu, A. *Chem. Rev.* **2000**, *100* (2), 543-600.
14. Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. *Comput. Phys. Commun.* **2010**, *181* (9), 1477-1489.
15. Kabeshov, M. A.; Sliwinski, E.; Fitzpatrick, D. E.; Musio, B.; Newby, J. A.; Blaylock, W. D. W.; Ley, S. V. *Chem. Commun.* **2015**, *51* (33), 7172-7175.
16. Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72* (1), 650-654.
17. Van Speybroeck, V.; Van Neck, D.; Waroquier, M.; Wauters, S.; Saeys, M.; Marin, G. B. *J. Phys. Chem. A* **2000**, *104* (46), 10939-10950.
18. Seregin, I. V.; Gevorgyan, V. *Chem. Soc. Rev.* **2007**, *36* (7), 1173-1193.
19. Sommai, P.-A.; Tetsuya, S.; Yoshiki, K.; Masahiro, M.; Masakatsu, N. *Bull. Chem. Soc. Japan* **1998**, *71* (2), 467-473.
20. García-Cuadrado, D.; de Mendoza, P.; Braga, A. A. C.; Maseras, F.; Echavarren, A. M. *J. Am. Chem. Soc.* **2007**, *129* (21), 6880-6886.
21. Hardinger, S. *Chemistry 14D thinkbook*; 2005.
22. Hammond, G. S. *J. Am. Chem. Soc.* **1955**, *77* (2), 334-338.
23. Chen, M.; Ren, Z.-H.; Wang, Y.-Y.; Guan, Z.-H. *J. Org. Chem.* **2015**, *80* (2), 1258-1263.
24. Luo, S.; Luo, F. X.; Zhang, X. S.; Shi, Z. J. *Angew. Chem. Int. Ed.* **2013**, *52*, 10598-10601.
25. Xia, J.-B.; Wang, X.-Q.; You, S.-L. *J. Org. Chem.* **2009**, *74* (1), 456-458.
26. Hossian, A.; Manna, M. K.; Manna, K.; Jana, R. *Org. Biomol. Chem.* **2017**, *15* (31), 6592-6603.
27. Zhou, H.; Xu, Y.-H.; Chung, W.-J.; Loh, T.-P. *Angew. Chem. Int. Ed.* **2009**, *48* (29), 5355-5357.
28. Jia, X.; Zhang, S.; Wang, W.; Luo, F.; Cheng, J. *Org. Lett.* **2009**, *11*, 3120-3123.
29. Xu, B.; Liu, W.; Kuang, C. *Eur. J. Org. Chem.* **2014**, *2014* (12), 2576-2583.
30. Wu, J.; Lan, J.; Guo, S.; You, J. *Organic Letters* **2014**, *16* (22), 5862-5865.

31. Tani, S.; Uehara, T. N.; Yamaguchi, J.; Itami, K. *Chemical Science* **2014**, 5 (1), 123-135, 10.1039/C3SC52199K.
32. Wang, S.; Liu, W.; Cen, J.; Liao, J.; Huang, J.; Zhan, H. *Tetrahedron Letters* **2014**, 55 (9), 1589-1592.
33. Zhang, F.; Greaney, M. F. *Angewandte Chemie International Edition* **2010**, 49 (15), 2768-2771.
34. Orpen, A. G.; Brammer, L.; Allen, F. H.; Watson, D. G.; Taylor, R. Typical interatomic distances: organometallic compounds and coordination complexes of the d- and f-block metals. In *International Tables for Crystallography Volume C: Mathematical, physical and chemical tables*; Prince, E., Ed.; Springer Netherlands: Dordrecht, 2004; pp 812-896.