Recent trends in catalytic \(sp^3\) C–H functionalization of heterocycles†

Milanpreet Kaur and Jeffrey F. Van Humbeck

Heterocycles are a ubiquitous substructure in organic small molecules designed for use in materials and medicines. Recent work in catalysis has focused on enabling access to new heterocycle structures by \(sp^3\) C–H functionalization on alkyl side-chain substituents—especially at the heterobenzylic position—with more than two hundred manuscripts published just within the last ten years. Rather than describing in detail each of these reports, in this mini-review we attempt to highlight gaps in existing techniques. A semi-quantitative overview of ongoing work strongly suggests that several specific heterocycle types and bond formations outside of C–C, C–N, and C–O have been almost completely overlooked.

Introduction

Scope and goals of this mini-review

Herein, we present an overview of 206 recent papers (2010 or later) that encompass 4904 discrete catalytic reactions.† The reader may be curious to know what inspired this undertaking, and why this particular set of publications was chosen. One area of our laboratory investigates copper-8 and iron-catalyzed \(sp^3\) C–H functionalization of azaheterocycle substrates.

Milanpreet Kaur was born in Yamuna Nagar (Haryana), India. She completed both her B.Sc. (2015) and M.Sc. (2017) in Chemistry at Hindu College, University of Delhi. Currently, she is completing her graduate studies under the supervision of Prof. Jeffrey Van Humbeck, Department of Chemistry, University of Calgary, Canada. Her research interest is focused on designing a new catalytic system for site selective C–H functionalization.

Jeffrey F. Van Humbeck is a native of Calgary, Alberta, Canada, where he completed his undergraduate education at the University of Calgary before venturing south of the 49th parallel. His PhD was conducted under the supervision of Prof. David W. C. MacMillan at Princeton University, studying the mechanism of interaction between organocatalytic reactive intermediates and transition metals. Next, he changed fields and coasts, joining the laboratory of Prof. Jeffrey R. Long at the University of California, Berkeley, where he investigated the use of porous organic polymers in gas purification, nuclear fuel processing, and as ion-conducting solid electrolytes. Jeff began his independent career at MIT in July 2014, where his group made their first discoveries in site-selective C–H functionalization. In January 2018, Jeff returned to Canada to take up a position at the University of Calgary.
After establishing our particular mechanistic approach with some initial success, the natural question arose: what next?

To help guide our future work, we wanted a semi-quantitative overview of ongoing work in the area of sp³ C–H functionalization of heterocycles, broadly defined. We were most interested to learn if specific reaction types or substrate classes are currently poorly served by existing technology. While numerous high-quality reviews on C–H functionalization have been published recently,207–214 such resources are naturally focused on what has been achieved and not what is missing. These are important resources, but did not answer our key question.

Our strategy to create an overview that would highlight current gaps was this: first, we divided the major topics of interest—reaction mechanism, bond formation, and heterocycle type—into predefined categories that are listed within. Next, we set out to find a broadly representative set of publications we could use to roughly enumerate how thoroughly any particular [mechanism, bond formation, and heterocycle] combination had been investigated. This was sensibly achieved with a broad SciFinder™ search, followed by reasonable filtering of the results as described in Fig. 1. We first looked for any C–H functionalization of an alkyl group attached to a heterocycle that fit three criteria: (i) having a publication year or 2010 or later, (ii) in an academic journal, and which (iii) has a catalyst as recognized by the search tools. Our search initially focused on only carbon substituents at the position adjacent to the heterocycle, to avoid massive quantities of trivial reactions such as the oxidation of an alcohol to an aldehyde. However, if a manuscript featured a mixture of examples, where in some cases all ‘R’ substituents were carbon and in other cases some ‘R’ substituents were heteroatoms, both were included in the statistics we detail below.

This search delivered a number of publications that was well beyond our ability to analyse. Upon slightly closer inspection, though, there were a number of further filters that could be applied so that the focus would truly rest on recent catalyst development in C–H functionalization. First, we excluded transformations that are much better described as another reaction type. For example, we decided not to consider aldol reactions of ketones that simply happen to have heterocycle substitution. Similarly, we felt that publications reporting traditional radical reactions (e.g. halogenation using N-bromosuccinimide) using common radical initiators as catalysts (e.g. AIBN; azobisisobutyrylnitrile) do not reflect ongoing work in catalyst development in this area. Next, we also excluded reactions that involved stoichiometric in situ activation of the heterocycle (e.g. such as deprotonation with a strong base) prior to a catalytic transformation taking place.

Finally, we also restricted our analysis to manuscripts that had a significant heterocycle focus and/or scope. This excluded target-oriented synthesis that might feature a single reaction of the type we were interested in, and most mechanistic studies of reactions in this class were also excluded. More difficult decisions had to be made around broadly applicable C–H functionalization chemistry. Should a paper that features fifty all-carbon alkyl aromatic substrates and only five alkyl heterocycles be considered? Although a rather qualitative filter, we gave our honest best effort to answer that exact question: does this publication have a significant heterocycle focus and/or scope. These features—focus and scope—do not necessarily occur together. In some cases, the focus of the paper was on other substrate classes, but the sheer number of heterocycles reported justified its inclusion. In others, while only a handful of substrates were reported, the focus was evident as every example was a heterocycle.

What results is a set of 206 papers that is by no means exhaustive and was not chosen by random sampling. Chemical reaction indexes such as SciFinder™ also do not perfectly categorize reactions. We can say this with certainty, as we discovered that our own publications did not appear in our initial searches. As such, we do not suggest taking any of the conclusions below as being decimal-point accurate.

By any reasonable count, however, the 4904 unique reactions contained in these papers represent a significant fraction of academic work done in the last decade. In any place where we think that our publication-selection strategy has falsely resulted in interesting trends, we have tried to alert the reader. Most importantly, we believe that some of the trends observed...
when analysing these data are so suggestive that they are almost unquestionably significant. There are clearly areas of heterocycle space that need much better service from catalysis. We hope that this mini-review will serve to encourage laboratories other than our own to address some of these challenges.

Results and discussion

Manuscript-level analysis: reaction type

The most important deconstruction of the data we gathered is that based on the heterocycle structure, which is presented later on. To appreciate some of our structure-based analysis, a starting set of expectations is necessary. First, we present some manuscript-level analysis of the paper set that can provide the necessary context for structure-based analysis.

As shown in Fig. 2, we classified papers based on the bond formation pursued, and further sub-divided based on the structure of the reaction partner. Carbon–carbon bond formations predominate, comprising 145 of 206 reports. Carbon–nitrogen and carbon–oxygen bonds are described at a relatively similar rate—27 and 22 times, respectively—with both being reported more often than all other types of bonds combined (i.e. ‘C-Other’).

To us, the relative ranking of C–C, C–N, and C–O bonds makes sense for a specific reason. The sub-division based on the reaction partner shows that there is simply a much greater diversity of possible coupling partners that can be used to make C–C bonds, as compared to C–N or C–O. Carbon–carbon multiple bonds, whether of an alkene, alkyne, or activated with an oxygen-based leaving group (both benzylic and allylic), represent almost a quarter of the papers in our set (53 of 206). The carbon–heteroatom double bonds of carbonyls and imines have been explored to nearly the same extent (44 of 206). Carbon–bromine bonds are ideal substrates for transition-metal catalysed pathways (22 of 206), with a variety of other partners each contributing a handful of further examples. Reagents for carbon–nitrogen and carbon–oxygen bond formation are dominated by the element/element multiple bonded species (i.e. N=N and O=O, 20 of 49) and the element-hydrogen bond (i.e. R,N–H and RO–H, 18 of 49).

Arguably, then, the relative lack of other bond forming processes reported is surprising. The diversity of possible coupling partners to make C–X bonds is certainly greater than that incorporated into our data set.

Manuscript-level analysis: reaction mechanism

For a similar reason, having an idea of the distribution of mechanisms that were reported within our data set is important for understanding structure-based analysis. Our definitions for each mechanism type are presented in Fig. 3, and the results of this analysis are presented in Fig. 4. We define mechanisms according to the following definitions. ‘Organometallic’ feature metal–carbon bonds, either on the heterocycle or on the reaction partner it encounters. ‘Lewis Acid’ and ‘Brønsted Acid’ encompasses heteroatom coordination of either the heterocycle, or its reaction partner, with the difference between them being the traditional definition of a Brønsted acid being a proton exclusively. ‘Redox’ catalysts can engage in single electron transfer with either the heterocycle or its reaction partner, or can include two-electron redox processes, as would be demonstrated by the halogen redox cycle shown in Fig. 3.

We used ‘Organocatalyst’ to encompass any organic catalyst not acting by a classical Brønsted acid or base pathway (e.g. enamine/iminium catalysis, counterion catalysis, etc.). ‘Radical’ catalysts act to create a carbon-centered radical on the sp³ functionalized carbon of the heterocycle. Brønsted base catalysts are as classically defined. A handful of manuscripts leveraged multiple/synergistic catalyst pathways and could not be assigned to only one of these groups.

Organometallic mechanisms represent nearly half of our set. We ascribe this behaviour to two factors. First, as we will show below, one major class of heterocycle substrates is ser-
vicied almost entirely by organometallic pathways, which certainly affects the manuscript-level analysis. Second, organometallic pathways can be seen as being the most versatile mechanism class. Five of the other seven possible mechanisms (i.e. Lewis Acid, Brønsted Acid, Brønsted Base, Organocatalyst, and Multiple) deliver predominantly carbon–carbon bonds. 69 of 75 manuscripts using these mechanisms deliver C–C bonds. Radical mechanisms, conversely, deliver almost exclusively carbon–heteroatom bonds (20 of 21). Organometallic mechanisms have delivered essentially any type of desired new bond, across a wide range of substrate types, and hence have contributed the greatest share of current work.

Structure-level analysis: reaction distribution by structure type

Before any counting began, we divided all heterocycles into eight groups, with no expectation as to how common they would be in the data (Fig. 5). The first division was by the ring size and aromatic nature: 6-membered aromatic vs. 5-membered aromatic vs. all other aromatics and saturated. Both 6- and 5-membered rings were sub-divided into 3 sections each. 6-Membered rings were divided based on whether the C–H bond that is broken is in ‘resonant’ communication with the heteroatoms in the ring system. That is: if an arbitrary reactive intermediate (i.e. anion, cation, radical) were placed on the sp³ side-chain, could resonance structures be drawn that place this reactive intermediate on any of the heteroatoms in the ring system? By this definition, 4-alkylpyridines are ‘6-Resonant’, while 3-alkylpyridines are ‘6-Non-Resonant’. We considered extended heterocycles to be in the same category as their parent heterocycle: 2-alkylpyridines and 2-alkylquinolines are both considered ‘6-Resonant’. We expected that 8-substituted quinolines (and related structures) specifically deserved their own category, given their widespread use as model substrates for method development. A similar split was made for 5-membered rings. 2-Alkylimidazole would be considered ‘5-Resonant’, while 3-alkylimidazole would be considered ‘5-Non-Resonant’. We expected that five-membered rings featuring a single heteroatom (e.g. furan, thiophene, indole, etc.) would...
have unique chemistry and deserved their own category. If the C–H bond being broken is found on the all-carbon ring of an extended heterocycle (e.g. 5-, 6-, or 7-alkylquinolines), then that was to be considered an example of ‘Remote’ C–H functionalization. All other heterocycles encountered were grouped into a final category.

We found the results to be striking. Three of our eight preset categories were essentially absent from 206 manuscripts and 4904 reactions. Zero examples of five-membered rings functionalized at a non-resonant position were observed, other than reactions listed as ‘trace’ yield, which were not included. Only one single example of remote functionalization was observed as a major product (i.e. excluding the minor component of product mixtures) and is shown in Fig. 6.

Non-resonant 6-membered substrates were nearly rare, accounting for only ~0.5% of reported reactions. A further investigation of the eight papers which report these substrates illustrates just how unusual it is for a catalytic reaction to be specific for these positions.16,47,49,88,103,130,134,135,140 One strategy to activate these positions is demonstrated by four publications from the Walsh laboratory (ref. 16, 103, 134 and 135), and describe the production of di- and tri-arylmethane products through palladium or nickel catalysis. In three of those four reports, reactive positions adjacent to all-carbon aromatics are also demonstrated as competent coupling partners (Fig. 7). Given that basic heterocycles such as pyridine often inhibit transition metal catalysis, it is a very noteworthy positive feature of these systems that heterocycles are well-tolerated. From a mechanistic perspective, we would describe such reactions as being heterocycle compatible, but not necessarily heterocycle selective. Four other manuscripts report only heterocycle substrates, with a mixture of resonant and non-res-
onant positions activated. These reports may, indeed, be het-
erocycle selective—but are not necessarily selective for non-res-
onant positions. In fact, we found only one manuscript where
the paper focused exclusively on non-resonant positions. In
that specific case, Höke, Herdtweck, and Bach reported the
use of a highly specific hydrogen-bonding interaction between
a rhodium paddlewheel catalyst flanked with chiral lactams
and a 3-alkylquinolone substrate as a mechanism to access
these positions selectively. 47

At first glance, ‘All Others’ seems to be underrepresented,
and here we will highlight a very significant effect of our litera-
ture filter described in the introduction. The actual search
terms entered into SciFinder™ do not include sp3 functionali-
ization inside the heterocycle ring itself. Yet, substrates such as
N-Boc-pyrrolidine and xanthene make up a significant fraction
of the reactions classified as ‘All Others’. These reactions were
discovered from their inclusion in manuscripts that also
feature substrates which fit our search terms. So, this number
is unquestionably under reported, and would have been much
greater if we had exhaustively searched for sp3 C–H functionali-
izations inside heterocycle rings. We did not pursue that strat-
egy, as we felt as though it was essentially a different review
topic. The types of strategies pursued for functionalization
adjacent to an aromatic heterocycle and within a saturated het-
erocycle ring are likely very different. We did not exclude these
reactions from our analysis though—thought it might
provide some qualitative context, as long as the readers were
properly warned to take those numbers to be very rough
indeed.

Even with some prior expectations, we were surprised by
the high number of 8-alkylquinolines reported. We had
expected to see this well-established tool for organometallic
C–H activation chemistry well-represented. To support that
assumption, we next detailed the distribution of reaction
mechanisms used for each substrate class that had a signifi-
cant number of examples (Fig. 8).

Three observations seem most important to us from this
analysis. First, it is indeed true that 8-alkylquinoline is used
almost exclusively to investigate organometallic C–H activation
pathways. Second, there was a potentially significant difference
in the number of organocatalytic pathways reported for 5- and
6-membered resonant substrates. Third, and perhaps natu-
rally, single-heteroatom 5-membered substrates (e.g. indole,
thiophene, furan, etc.) are more often subject to redox catalysis
than resonant/quinoline substrates. For the reasons men-
tioned just above, we did not further analyse the ‘All Other’
group, but it is included here for the curious reader. The high
number of reports using 8-substituted quinolines is driven pri-
marily by rhodium and palladium catalysis, which make up 27
of the 32 papers that feature organometallic mechanisms, with
cobalt, ruthenium and iridium making up the remainder
(Fig. 9).

With regard to the issue of the relatively high number of
organocatalytic pathways reported for 5-membered resonant
heterocycles, that result seems to be a somewhat coincidental
matter of two particular substrate structures. Four of the five
papers that report these mechanisms use the same—highly
acidic—heterocycle, a 4-nitroisoxazole. The final report also
features a highly acidified substrate, a cyclic sulfonilimine.
The highly acidic nature of these substrates allows for
sufficient tautomerization and/or deprotonation to occur in
the absence of a strong base so that iminium and phase
transfer catalysis can be used. With regards to the use of
single-heteroatom 5-membered substrates, again, our
expectations were met. The ability of these species to gen-
rate benzylic cation equivalents—with various precise struc-
tures—that can couple with nucleophiles is not shared by
the relatively electron poor substrates found in other
classes.
Conclusions

Missing chemistries: do they really matter?

Without question, in our eyes the most important result of this analysis is the near-complete dearth of catalytic chemistry that is being targeted at non-resonant and remote aromatic positions. Whatever systematic effects our publication filter strategy may have, it is unquestionably true that reactions on these substrates are orders-of-magnitude less common than the other structure types. Such a statement leads to a natural question, however: does it really matter? We think that compelling arguments can be made on either side, and we will give our balanced thoughts here. The reader can decide for themselves whether more effort is needed in these areas.

Is this a chicken-or-the-egg issue? Are remotely alkylated quinolines particularly difficult to make, and so are not available substrates? Are thiazoles substituted at non-resonant positions not useful for pharmaceutical or materials applications, and so synthetic routes to form them are not valuable? Also, one could point to the fact that we excluded very well established radical initiators like AIBN from our search. Perhaps traditional reaction platforms serve these substrates perfectly well.

While we lack the omnipotence to answer these questions authoritatively, we can again look at the data (Fig. 10). We chose 5-, 6-, and 7-substituted quinolines and 4-, and 5-substituted thiazoles as representative heterocycles. We used the single C–H functionalization of a methylene carbon atom as a representative transformation. In this case, we had no restriction on the publication year or the nature of the catalyst used, which should thus include traditional chemistries like AIBN/NBS bromination.

It is unquestionably true that these alkylheterocycles are relatively rare. They have been reported as product structures in just over 13 000 reactions published in academic journals. Following from this, they have been reported as the substrate of a catalytic reaction only 1500 times. Even still, of these 1500 reactions only four involved C–H functionalization—excluding two reactions where the desired products were the minor component of a regioisomeric mixture. All four reactions were radical brominations. The absence of evidence is not evidence of absence, however. We do not see any fundamental reason why many existing reaction platforms could not be adapted for these systems.

Even still, we argue in favour of increased investigation in these areas for two reasons. Fundamentally, it seems hard to argue against investigation of a reasonable substrate class where no single modern catalytic approach has been applied. It may be the case that absolutely nothing novel or interesting is encountered, and a catalyst that is able to react with 2-ethylpyridine also suffices for 6-ethylquinoline and 4-ethylthiazole. It is hard to put too much weight on this argument, however, given that it would also be an argument for ending research in synthetic methodology nearly entirely.

Second, we believe that as target-oriented research attempts to use C–H functionalization in more and more complex set-
tings, issues of site-selectivity are going to become increasingly relevant.\textsuperscript{215–226} Consider complex small molecules investigated in pharmaceutical applications, such as verucerfont or nizatidine (Fig. 11). If C–H functionalization for the synthesis of potential metabolites or for attachment to a polymer nanoparticle was desired, how could the site of reaction be controlled? Developing such technology for the synthetic community will require accessing non-resonant and remote positions, as well as positions adjacent to heterocycles that are currently not well explored. This is an area where traditional chemistry will also come up short. While some particular feature of the substrate structure might guide NBS bromination to a single position, that will only provide access to the inherently most reactive site. It does not seem at all apparent how inherently less reactive positions could be selectively accessed without the use of catalysis. To our eye, this seems to be a problem well suited for the application of supramolecular chemistry. Could heterocycle coordination and C–H activation occur at two different locations in a catalyst scaffold—held in precise relative positions—to break bonds that are inaccessible to current catalysts? Or, will the development of a successful system avoid heterocycle coordination entirely? Perhaps other non-covalent interactions could guide the reaction towards unusual sites. We hope that research groups active in this area of C–H functionalization will consider investigating more substrates that are dramatically underreported in their ongoing work.

**Conflicts of interest**

There are no conflicts to declare.

**Acknowledgements**

Funding for this work was provided by the NSERC of Canada through the Discovery Grant program, and the University of Calgary through start-up funds.

**Notes and references**