ChemComm

Accepted Manuscript

This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the author guidelines.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the ethical guidelines, outlined in our author and reviewer resource centre, still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemcomm

Please do not adjust margins

Journal Name

COMMUNICATION

A new water soluble copper N-heterocyclic carbene complex delivers mild O^{6}G-selective RNA alkylation

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Dnyaneshwar Rasale,[a] Kiran Patil,[a] Basilius Sauter,[a] Stefanie Geigle, Saule Zhanybekova,[a] and Dennis Gillingham*[a]

We show here that copper carbenes generated from diazo acetamides alkylate single RNAs, mRNAs, or pools of total transcriptome RNA, delivering exclusively alkylation at the O^6 position in guanine (O^{6}G). Although the reaction is effective with free copper some RNA fragmentation occurs, a problem we resolve by developing a novel water-stable copper N-heterocyclic carbene complex. Carboxymethyl adducts at O^{6}G are known mutagenic lesions in DNA but their relevance in RNA biochemistry is unknown. As a case-in-point we re-examine an old controversy regarding whether O^{6}G damage in RNA is susceptible to direct RNA repair.

Modifications of RNA play a critical role in biology and biotechnology. More and more we see that post-transcriptional RNA modification is an integral element of RNA regulation and gene control.[1, 2] Chemists use modifications to improve the stability, activity, and membrane permeability of RNA therapeutics.[3] Despite the rapid growth in research on RNA modifications,[1] many remain difficult to synthesize. Although commercial phosphoramidites are available for the more common modifications, rarer or unnatural ones are typically unavailable. We have recently shown that copper-carbenes generated from diazo compounds can alkylate the O^{6}G in DNA with high chemoselectivity.[4] Although divalent metals can degrade RNA, we wondered whether the copper-carbene transfer to O^{6}G would also occur in RNA and whether it could be used to prepare modified RNA derivatives. The occurrence and consequences of RNA alkylation damage is poorly understood,[5-8] although a number of papers have shown that some AlkB-type iron dioxygenases can repair alkylation damage in RNA.[9-12] We can find only four papers that specifically investigate O^{6}G damage in RNA – all dealing with methylation. Two of these are recent, with one looking at the effects of O^{6}G methylation on the fidelity of ribosomal translation[13] and the other at the next generation sequencing (NGS) behaviour of O^{6}G lesions.[14] There are also two papers from the eighties with conflicting conclusions about the susceptibility of methylations at the O^{6}G position to direct RNA repair.[15, 16] Whereas these earlier repair studies used unselectively methylated tRNA (*i.e.* a mixture of lesion types), the technique we report here (see Fig. 1) gives us access to pristine O^{6}G lesions in RNA, facilitating a precise reanalysis of whether methyl guanine methyltransferase can directly repair RNA.

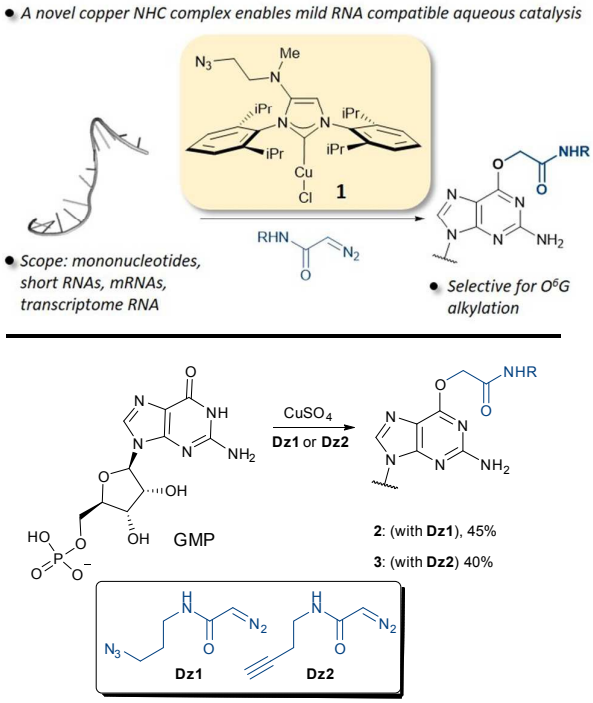

Fig. 1. Top: Overview: Diazoacetamides alkylate diverse RNA substrates with a single type of alkylation; **Bottom**: The 2'hydroxyl does not interfere with O^{6}G alkylation.

[a] *Dr. D. Rasale, Dr. K. Patil, MSc B. Sauter, Dr. S. Geigle, Dr. S. Zhanybekova, Prof. Dr. D. Gillingham*
Department of Chemistry
University of Basel
St. Johanns-Ring 19
CH-4056, Basel
E-mail: dennis.gillingham@unibas.ch

Electronic Supplementary Information (ESI) available: See DOI: 10.1039/x0xx00000x

ChemComm Accepted Manuscript

Please do not adjust margins

Direct modification of canonical RNAs is ideal for chemical tailoring because RNA substrates from both chemical synthesis and transcription can be used directly. With current technology, however, it is also the most difficult approach. The state-of the art in this so-called post-synthetic tailoring is to repurpose RNA modification enzymes to accept unnatural substrates[17-21] or to use DNAzymes[22] to introduce 2'-hydroxyl tags. While these efforts put a strong foot forward in resolving the problems in RNA synthesis, more work is needed. For example modifications in the Watson-Crick face can still only be accessed by total chemical synthesis. Here we contribute the first example of a direct chemical reaction that is chemoselective for O^6G RNA lesions.

Our previous work has shown that alkylation of DNA with diazo compounds is efficient and selective for O^6G;[4] our first step here was to determine whether this chemoselectivity was maintained in the more functionalized and less stable RNA. To test the impact of the 2'-hydroxyl on the copper catalysed reaction we prepared two diazoacetamides bearing biorthogonal chemical tags (**Dz1** and **Dz2**). Reaction of these compounds with GMP using copper sulfate as a catalyst delivered clean conversion to a single modified product and we could confirm O^6G selectivity by both NMR and LC-MS fragmentation analysis (see bottom of Fig. **1**, S12-13 and scheme S5-7).

With reactivity in RNA substrates established we next examined whether larger RNAs were alkylated and whether the O^6G selectivity was maintained. Indeed, in both a small oligonucleotide (see **ssRNA1** in **Fig. 2A**), or in a whole transcriptome sample (see **Fig. 2B**) the copper catalysed alkylation delivered the modified product. The reaction of **ssRNA1** is a first-in-class selective RNA modification – this type of direct internal RNA modification has only been achieved previously with enzymatic transformations or deoxyribozymes.[19, 22-25] Although RNA degradation is observed (see HPLC of crude reaction in **Fig. 2C,** S17) the product can be

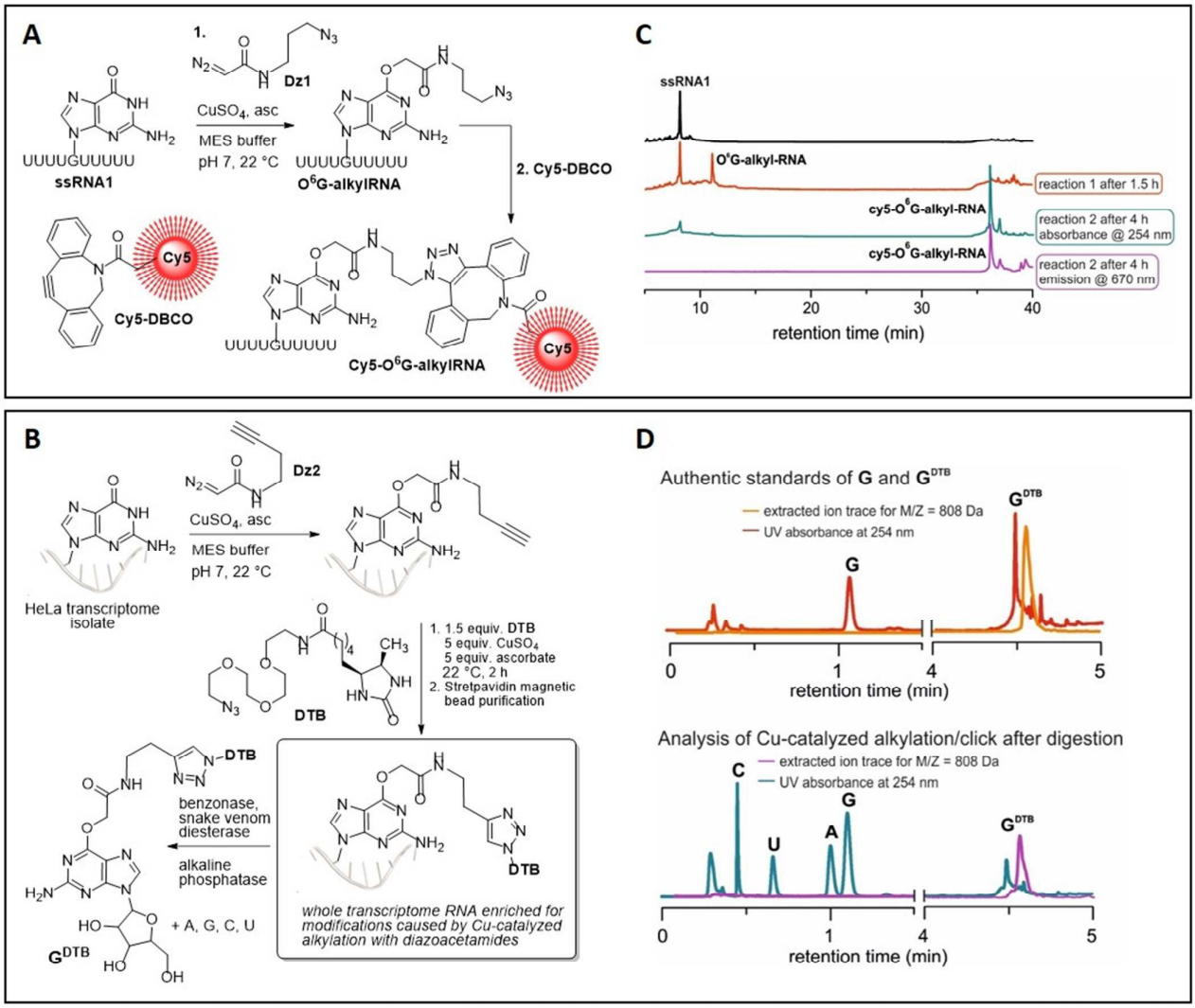

Fig. 2 Copper-catalyzed RNA alkylation with diazo compounds lead to selective O^6G modification in both a short oligo (**A**), and in HeLa whole transcriptome RNA (**B**). The ssRNA with one guanine delivers largely a single product (see the red HPLC trace in **C**). Complete RNA digestion (see last step in **B**) can be used to identify modified bases after alkylation (**D**). Authentic standards (described in the ESI) allow a confident assignment of the modified G. The only modified base seen in the HPLC and ESI MS after alkylated samples are digested and analysed is G (**D**); no masses corresponding to modification of other bases are observed.

ChemComm Accepted Manuscript

This journal is © The Royal Society of Chemistry 20xx

obtained in high purity by preparative HPLC. Derivatization with the strain promoted azide-alkyne cycloaddition (SPAAC) to attach a Cy5 fluorophore, offers further confirmation of the alkylation product (see **Fig. 2A** & **C**). The method outlined in Fig. 2 provides a simple one-step protocol to O^6G modified RNAs that uses readily available materials and is easy to execute.

We chose the RNA 10-mer **ssRNA1** as our first substrate in larger RNAs to simplify product analysis. Even in the complex transcriptome samples, however, all alkylations were O^6G-type. To confirm this we first appended desthiobiotin (**DTB**) to the alkylated transcriptome samples (see Fig. 2B) through a copper-catalyzed azide-alkyne cycloaddition. Streptavidin magnetic bead selection then delivered an RNA library highly enriched in alkylated sequences. From here we could confirm the expected high selectivity for O^6G alkylation through complete digestion,[26] followed by LC-MS of the resulting mixture of nucleosides. The masses observed were consistent only with O^6G alkylation products (comparison with an authentic standard prepared as described in Scheme S8-S9 and Figure S8-S9 in the ESI) and not with any of the other bases.

Although the copper catalysed alkylation is highly selective for O^6G modification, the Lewis acidity and redox activity of copper leads to some RNA degradation with large RNAs such as, for example, mRNAs (see the ESI Figure S4). We therefore set out to identify a ligand for copper that led to a stable, yet active complex that did not degrade the RNA. Chelating nitrogen ligands were quickly ruled out because they completely suppressed catalysis. A commercially available copper N-heterocyclic carbene (CAS#: 578743-87-0) showed activity, but its poor water solubility made it impractical. After thorough ligand screening[27] of both known[28-30] and new water-soluble NHC ligands we arrived at the electron rich NHC ligand **4** as one that gave a stable and catalytically active copper complex. The dimethylamine variant of NHC ligand **4** was first introduced by Cesar and Lavigne for use in palladium catalysis,[31] but its capacity to bind copper as well as its activity in copper catalysis has never been explored. We made an azide-bearing version of the Cesar-Lavigne ligand in three steps in 7% overall yield (see the Schemes S1-S2 in the ESI for complete synthetic details). The formation of the corresponding copper complex (**1**) was readily achieved via the silver complex in 19% yield over two steps. Crystals of **1** could be grown and the structure confirmed by X-ray analysis (see blue box in Fig. 3). Test reactions of GMP with diazoacetamide **Dz3** confirms that **1** is active in mononucleotide alkylation (see Fig. 3B). Although we found the parent complex **1** was reasonably water-soluble, the azide tag was a precautionary inclusion in case we needed to include further solubilizing substituents. As an example, a click reaction with water soluble alkyne **5** cleanly delivered complex **6**, which was completely water soluble and active in **GMP** alkylation (see Fig. 3B). Nevertheless, since the precursor complex **1** was active and water-soluble, it was used in all further reactions. Importantly a degradation study with a 290 nucleotide mRNA prepared by in vitro transcription demonstrates that while copper sulfate is detrimental to RNA, the copper complex leads to no substantial degradation even at a 100-fold excess (see the ESI figure S4).

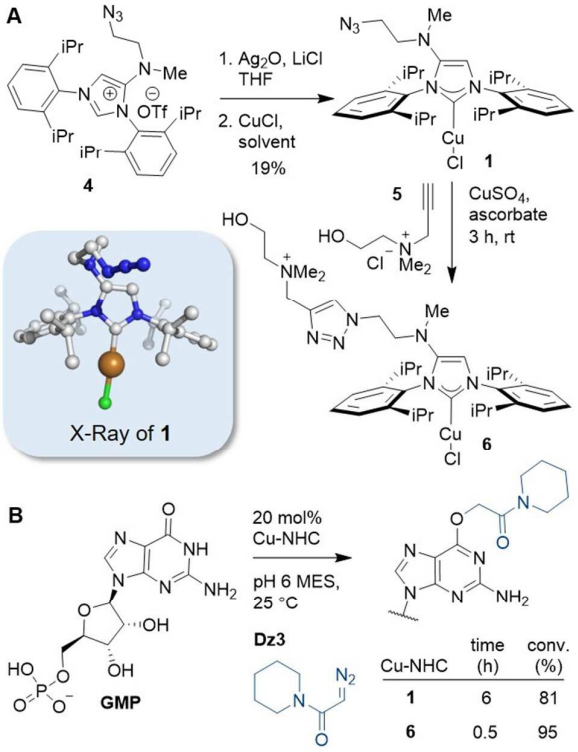

Fig. 3 A An electron rich N-heterocyclic carbene ligand forms a stable complex with copper, which can be further derivatized. **B** Both the parent complex **1** and the derivatized complex **6** are active in guanine alkylation.

To determine whether the copper catalysed alkylation demonstrated any specific sequence or structural preferences we performed next generation sequencing (NGS) on alkylated transcriptome samples, using an adaptation of a protocol we recently developed for profiling anticancer agents.[32] In contrast to our earlier protocol, here we prepared the sequencing libraries using RNA samples enriched in alkylation (prepared as described earlier in Fig. 2B, and the complete protocol is described in detail in the ESI). Reverse transcription of the alkylated RNAs delivers cDNAs with stops and mutating read-throughs at sites of alkylation that we use to pinpoint putative sites of alkylation. With sites of alkylation identified we asked three questions of the sequencing data: Is there any sequence preference in the alkylation? Is there any structural preference in the alkylation? Is the guanine selectivity seen in the in vitro samples reflected in the NGS data? The lack of information in sequence logo plots showed there was no strong preference for alkylating any specific sequence, while examining significant positions across different structural types shows a near random distribution (i.e. the alkylation can occur at any G, see the ESI Figure S1-S2). Analysis of significant stops and mutations, however, show a strong preference for guanine. The fact that guanine residues dominate the significant mutations jibes with recent evidence that read-

ChemComm Accepted Manuscript

throughs in reverse transcription at sites of O^6G alkylation are highly mutagenic.[14]

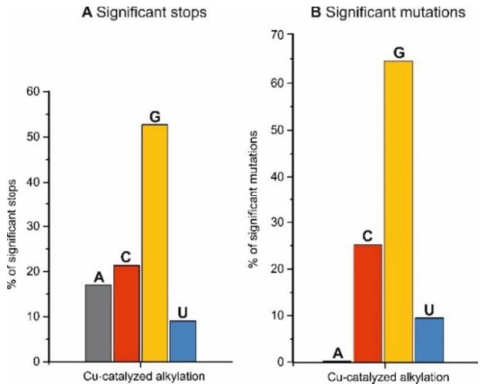

Fig. 4 The copper catalysed RNA alkylation shows a preference for guanine modification as measured by both the number of significant stops (**A**), or the observed mutations (**B**).

In summary we provide a new technique for the direct modification of diverse RNA types. Although the method is effective with copper sulfate, the new Cu-NHC complex **1** leads to less RNA degradation, maintains alkylation activity, and has a chemical handle for further derivatization. More broadly speaking, the stability and water compatibility of **1** give it great potential for other types of aqueous copper catalysis. The RNA modification is selective for O^6G, even in large RNAs such as mRNAs and in complex samples such as transcriptome isolates. Now that we have developed a nucleobase selective reaction our next goal is to use the modularity of the ligand to find sequence specific RNA alkylation catalysts.

Acknowledgements

Sequencing data analysis was performed at sciCORE (https://scicore.unibas.ch/), the scientific computing core facility of the University of Basel. D.G.G acknowledges the Swiss National Science Foundation (grant #s:200021_172521 and 200021_149805) for financial support of this work.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1. T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs and M. Wagner, *Angew. Chem. Int. Ed.*, 2012, **51**, 7110-7131.
2. J. Mauer, X. Luo, A. Blanjoie, X. Jiao, A. V. Grozhik, D. P. Patil, B. Linder, B. F. Pickering, J.-J. Vasseur, Q. Chen, S. S. Gross, O. Elemento, F. Debart, M. Kiledjian and S. R. Jaffrey, *Nature*, 2017, **541**, 371-375.
3. S. F. Dowdy, *Nat. Biotechnol.*, 2017, **35**, 222.
4. S. N. Geigle, L. A. Wyss, S. J. Sturla and D. G. Gillingham, *Chem. Sci.*, 2017, **8**, 499-506.
5. C. L. Simms and H. S. Zaher, *Cell. Mol. Life Sci.*, 2016, **73**, 3639-3653.
6. A. Calabretta, P. A. Küpfer and C. J. Leumann, *Nucleic Acids Res.*, 2015, **43**, 4713-4720.
7. Carrie L. Simms, Benjamin H. Hudson, John W. Mosior, Ali S. Rangwala and Hani S. Zaher, *Cell Reports*, 2014, **9**, 1256-1264.
8. E. J. Wurtmann and S. L. Wolin, *Critical Reviews in Biochemistry and Molecular Biology*, 2009, **44**, 34-49.
9. G. Zheng, John A. Dahl, Y. Niu, P. Fedorcsak, C.-M. Huang, Charles J. Li, Cathrine B. Vågbø, Y. Shi, W.-L. Wang, S.-H. Song, Z. Lu, Ralph P. G. Bosmans, Q. Dai, Y.-J. Hao, X. Yang, W.-M. Zhao, W.-M. Tong, X.-J. Wang, F. Bogdan, K. Furu, Y. Fu, G. Jia, X. Zhao, J. Liu, Hans E. Krokan, A. Klungland, Y.-G. Yang and C. He, *Mol. Cell*, 2013, **49**, 18-29.
10. M. Westbye, E. Feyzi, P. Aas, C. Vågbø, V. Talstad, B. Kavli, L. Hagen, O. Sundheim, M. Akbari, N.-B. Liabakk, G. Slupphaug, M. Otterlei and H. Krokan, *J. Biol. Chem.*, 2008, **283**, 25046-25056.
11. F. Drabløs, E. Feyzi, P. A. Aas, C. B. Vaagbø, B. Kavli, M. S. Bratlie, J. Peña-Diaz, M. Otterlei, G. Slupphaug and H. E. Krokan, *DNA Repair*, 2004, **3**, 1389-1407.
12. P. A. Aas, M. Otterlei, P. O. Falnes, C. B. Vagbo, F. Skorpen, M. Akbari, O. Sundheim, M. Bjoras, G. Slupphaug, E. Seeberg and H. E. Krokan, *Nature*, 2003, **421**, 859-863.
13. B. H. Hudson and H. S. Zaher, *RNA*, 2015, **21**, 1648-1659.
14. A. M. Kietrys, W. A. Velema and E. T. Kool, *J. Am. Chem. Soc.*, 2017, **139**, 17074-17081.
15. A. E. Pegg, K. Morimoto and M. E. Dolan, *Chem. Biol. Interact.*, 1988, **65**, 275-281.
16. P. Karran, *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 5285-5289.
17. J. M. Holstein, L. Anhauser and A. Rentmeister, *Angew. Chem. Int. Ed.*, 2016, **55**, 10899-10903.
18. J. M. Holstein, D. Schulz and A. Rentmeister, *Chem. Commun.*, 2014, **50**, 4478-4481.
19. D. Schulz, J. M. Holstein and A. Rentmeister, *Angew. Chem. Int. Ed.*, 2013, **52**, 7874-7878.
20. Y. Motorin, J. Burhenne, R. Teimer, K. Koynov, S. Willnow, E. Weinhold and M. Helm, *Nucleic Acids Res.*, 2011, **39**, 1943-1952.
21. M. Tomkuvienė, B. Clouet-d'Orval, I. Černiauskas, E. Weinhold and S. Klimašauskas, *Nucleic Acids Res.*, 2012, **40**, 6765-6773.
22. L. Büttner, F. Javadi-Zarnaghi and C. Höbartner, *J. Am. Chem. Soc.*, 2014, **136**, 8131-8137.
23. D. Gillingham and R. Shahid, *Curr. Op. Chem. Biol.*, 2015, **25**, 110-114.
24. J. M. Holstein, L. Anhäuser and A. Rentmeister, *Angew. Chem. Int. Ed.*, 2016, **55**, 10899-10903.
25. J. M. Holstein, D. Stummer and A. Rentmeister, *Chem. Sci.*, 2015, **6**, 1362-1369.
26. E. P. Quinlivan and J. F. Gregory, *Anal. Biochem.*, 2008, **373**, 383-385.
27. S. N. Geigle, Doctor of Philosophy, University of Basel, 2017.
28. H. Ibrahim, C. Gibard, C. Hesling, R. Guillot, L. Morel, A. Gautier and F. Cisnetti, *Dalton Trans.*, 2014, **43**, 6981-6989.
29. C. Gibard, H. Ibrahim, A. Gautier and F. Cisnetti, *Organometallics*, 2013, **32**, 4279-4283.
30. C. Gibard, D. Avignant, F. Cisnetti and A. Gautier, *Organometallics*, 2012, **31**, 7902-7908.
31. Y. Zhang, V. César, G. Storch, N. Lugan and G. Lavigne, *Angew. Chem. Int. Ed.*, 2014, **53**, 6482-6486.
32. B. Sauter and D. Gillingham, *ChemBioChem*, doi:10.1002/cbic.201800235.

ChemComm Accepted Manuscript

This journal is © The Royal Society of Chemistry 20xx