Journal Name

COMMUNICATION

BINOPtimal: A Web Tool for Optimal Chiral Phosphoric Acid Catalyst Selection

Jolene P. Reid,[a,b] Kristaps Ermanis [a] and Jonthan M. Goodman [a*]

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

DOI: 10.1039/x0xx00000x

www.rsc.org/

A catalyst selection program, BINOPtimal, has been developed. This interactive web tool selects the best performing chiral phosphoric acid catalysts from analysis of the starting materials, imine and nucleophile, on the basis of rules derived from the transformations within its database. This procedure has been applied to an example transformation demonstrating the potential to assist reaction design. The tool is available at www-mmm.ch.cam.ac.uk.

BINOL-derived chiral phosphoric acid catalysed addition of nucleophiles to imines is an established, highly general method enabling the formation of multiple types of bonds, (C-C, C-O, C-S, C-N, C-P and C-H) in asymmetric fashion.[1-4] However, the optimal catalyst and other reaction parameters can vary for different substrate combinations, contributing to the difficulty in selecting the best conditions.[5-7] We have developed a rule-based computational toolkit, BINOPtimal (BINOL-derived chiral phosphoric acid optimizer) which by automatic analysis of the reagent structures aids the best selection of catalysts for this reaction type.

Since the development of the first programs for computer aided organic synthesis, many have been developed to assist chemists in planning synthesis,[8-11] predicting reaction outcomes,[12-16] and structure determination.[17-18] Our work has been focused on the development of computational tools for the prediction and analysis of catalyst selectivity. Recently, we reported mechanistic principles for this reaction type summarized in a unified stereochemical model (Figure 1).[7] By considering the steric demands of the different parts of the imine and nucleophile, we can determine how they fit into the chiral cavity and which of the competing pathways is favored. These details can be applied to guide the correct choice of chiral catalyst via the simplified decision tree shown in Figure 1. Although useful, the simultaneous analysis of multiple molecular features on both reactants may limit application. For example, a key decision in the tree is whether the *N*-substituent is small or large, but this generally depends on the nucleophile, and so requires a table lookup. An automated tool would make this decision tree quick and easy to use.

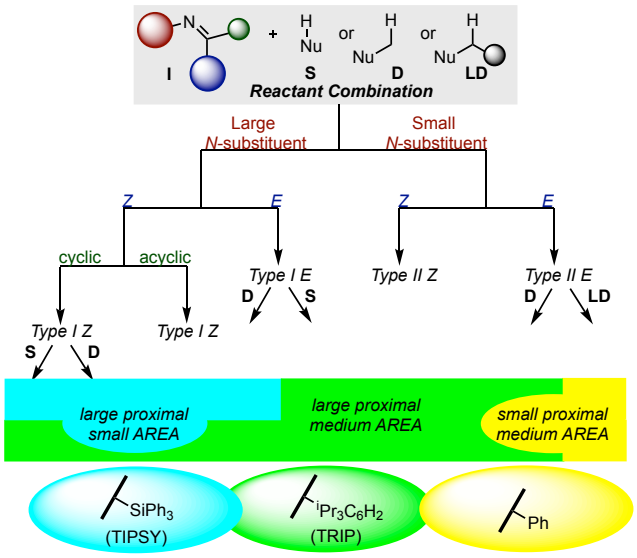

Figure 1. Unified stereochemical model. The transition state (TS) pathway is dependent on the structure of the imine (I) and nucleophile (S, D, LD).

Here, we introduce a new interactive web tool for chiral phosphoric acid catalyst selection. BINOPtimal, has been developed to allow chemists to anticipate, analyze and understand their experimental results of catalyst selectivity.

The development of a catalyst selection tool required formalizing our developed stereochemical model in even greater detail and carefully considering the limits of its generality. We chose to implement the program as a webtool, which could be used from any web browser. The program applies the series of carefully considered, coded rules to determine the best catalysts, making decisions on the important aspects of the reactants determining selectivity using selected organic transformations (Table 1). The decision making process is fully automated and the script is responsible for determining the structure of the reactants and analyzing the results (Figure 2). The structures are drawn in a molecular editor integrated in the website[19] – both the imine and nucleophile are entered in their active form.

[a] *Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail jmg11@cam.ac.uk*
[b] *Department of Chemistry, University of Utah, 315 S 1400 E, Salt Lake City, UT 84112, USA.*

†Electronic Supplementary Information (ESI) available: Computational details, complete list of authors in the Gaussian09 reference, Cartesian coordinates, energies, and values of imaginary frequencies of all the transition state structures.

Table 1. Contents of the decision tree block.

Type of transformation
Strecker reaction[23]
peroxidation of imines[24]
hydrophosphonylation of imines[25]
addition of thiols,[26] alcohols,[27] amines,[28] imides[29]
Friedel-Crafts reaction[30]
addition of enols,[31-32] enamides[33] and enamines[34]
reductive amination[35]
transfer hydrogenation[36]
addition of diazoesters,[37] diazoamides,[38] diazophosphonates[39]

The application logic was coded in Python 2.7,[20] using RDKit for the chemoinformatic functionality.[21] The results from the Python application are returned as text and images, integrated in a dynamic webpage and served by Apache webserver.[22] The site is available at **http://www-mmm.ch.cam.ac.uk**.

The program first matches the supplied 2D inputs against a series of SMARTS strings.[40] Each SMARTS string encodes a pattern matching a certain class of imines or nucleophiles. Based on which SMARTS strings return a positive match, imine and nucleophile type can be determined. The SMARTS strings have been coded to recognize the following variables: *N*-substituent size (large or small), imine type (cyclic or acyclic) and nucleophile type (symmetrical, displaced or large displaced). Due to the interrelatedness of certain structural features not all variables are easily assigned. For instance, the size of the *N*-substituent is dependent on the nucleophile type and so both must be considered. For this purpose, we have analyzed a series of literature reactions for different combinations and developed general rules for when the *N*-substituent is small and when it can be considered large. A few examples are shown in Table 2 and a complete set can be found in the SI. The imine configuration (*E* or *Z*) is determined straight from the structural input. The correct catalyst for the given reagent combination assessed on these four variables is then determined by the *decision tree* block. It contains rules based on relevant contributions from calculations previously reported and an expansive library of chemical reactions of this type, including the latest examples (Table 1).[7,41] The program automatically navigates the decision tree and presents the path taken to the user, providing a graphical means to assist in interpretation of the outcome.

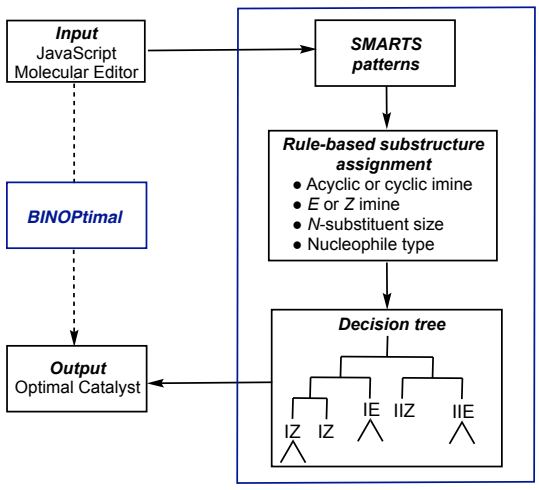

Figure 2. BINOPtimal algorithm.

Table 2. Imine-Nucleophile steric relationship table.

Imine *N*-substituent	Nucleophile	Size of *N*-substituent
Acetyl	Hantzsch ester	Large
Acetyl	Indole	Small
Aromatic	Diazoesters	Large
Aromatic	Enamide/Enol	Small
Benzoyl	Alcohol/Enol	Large

The top catalysts are presented in order of certainty and based on extensive literature precedent of reactions of this type. Ideally, BINOPtimal would provide 100% confidence for a single catalyst, but all certainty beyond a random selection is useful, and a high-probability that the correct choice is 1 of 3 or 4 is helpful.

Often more than a single catalyst can be effective for a reaction. BINOPtimal therefore should be of practical use in the planning of organic synthesis. A list of the general imines and nucleophiles recognized by the program has been compiled and any combination of these is predictable (see SI). The tool also indicates our confidence in the predictions for a particular combination. This ranges from certain (literature precedent), confident, doubtful or in small number of cases – unable to make a prediction because of particularly complex combination of factors. As the final part of the development of the tool, it underwent extensive testing on 70 literature reactions (over 1000 transformations) and successfully predicted the correct catalyst structure for each.

Applications of BINOPtimal is shown in the examples below outlining the capabilities of the program. We first focus on case studies in which the optimal catalyst has been found and show that these are correctly identified by our program (Figure 3). The transfer hydrogenation of imines by Hantzsch esters was an early example of a transformation catalyzed by phosphoric acids. Investigations by Rueping,[42] MacMillan[35] and List[36] all use slightly different conditions but overall the transformation is the same. We allow BINOPtimal to consider the reactants, the imine is entered in the *Z* form. SMART patterns are matched against the inputs and determine that the electrophile is an acyclic *Z* imine with a large *N*-substituent and the nucleophile symmetrical (Figure 3A). The program considers these factors that contribute to the reaction pathway in the *decision tree block* and concludes the *Type I Z* mechanism to be most likely with TIPSY (3, 3'=$SiPh_3$) and TRIP (3, 3'=2,4,6-iPrC_6H_2) to be top optimal catalyst choices, in agreement with experiment.

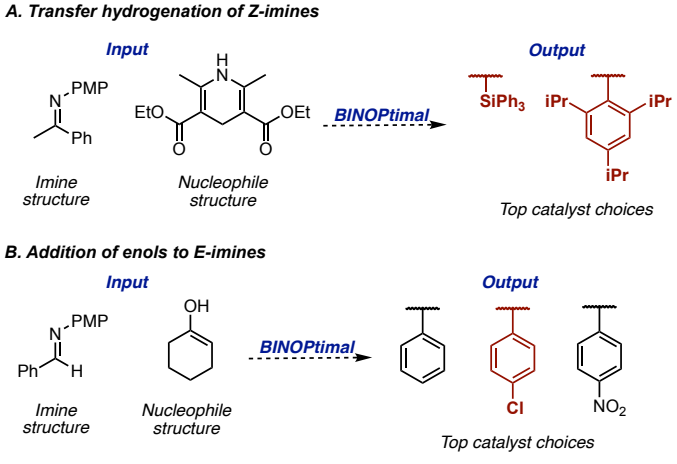

Figure 3. BINOPtimal identifies the correct catalyst 3,3' substituents for the two examples shown. The correct choice(s) are highlighted in red.

For most nucleophiles, such as Hantszch esters, the H-bond that holds the nucleophile to the catalyst is in line with the nucleophilic site. However, in some cases, the bond may be displaced to one side. This can promote a *Type II* process but only if the *N*-substituent is considered to be small. The catalyst requirements for these reactions can be different (Figure 1). As an example, we allow BINOPtimal to consider the addition of cyclohexanone to PMP-aldimines (Figure 3B).[32] The reactants are inputted in their reactive forms, *E* imine and the enol tautomer of cyclohexanone. Through the same process the program considers the electrophile as an acyclic *E* imine with a small *N*-substituent and the nucleophile as large displaced. Considering all this information the *decision tree block* correctly determines a *Type II E* pathway to be most likely and 4-ClC_6H_4 to be the optimal catalyst substituents.

To explore the reaction design capabilities of the program we applied it to a reaction that had yet to be rendered highly enantioselective. The hydrophosphorylation of imines catalysed by chiral phosphoric acids was originally reported by Akiyama in 2005.[25] At that time little was known about the catalyst structural requirements for high enantioselectivity, as such highly versatile and selective catalysts had yet to be identified. Akiyama *et al.* showed a subset of α-amino phosphonates could be obtained in moderate selectivities using a chiral phosphoric acid catalyst, CPA-1 (Figure 4). To analyse the possibility of improving enantioselectivity through changing the substituent structure at the 3 and 3' positions, BINOPtimal considered the combination of reactants. In this example we use the *E* imine and the phosphite tautomer as the active forms of reactants. According to the *rule-based substructure assignment* block the program considers the electrophile as an acyclic *E* imine with a large *N*-substituent and the nucleophile as symmetrical. Next BINOPtimal applies its rules on these structures, evaluates the TS pathway and features, and concludes that a *Type I E* reaction pathway is most likely. Figure 4 shows a selection of top catalysts. To analyse the program suggestions, we consider the TRIP catalysed reaction using ONIOM (B3LYP/6-31G**:UFF) calculations and compare the relative corrected energies between TS leading to competing products (Figure 5).

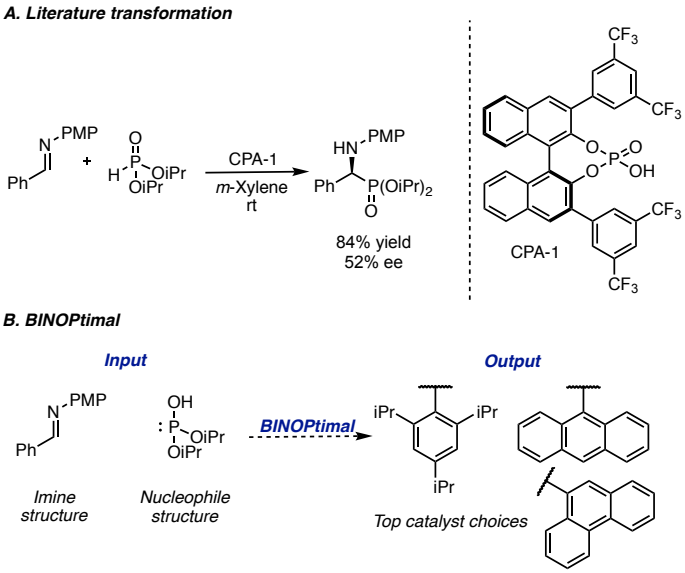

Figure 4. Hydrophosphorylation of imines catalysed by CPA-1 and application of BINOPtimal to aid the identification of better performing catalysts.

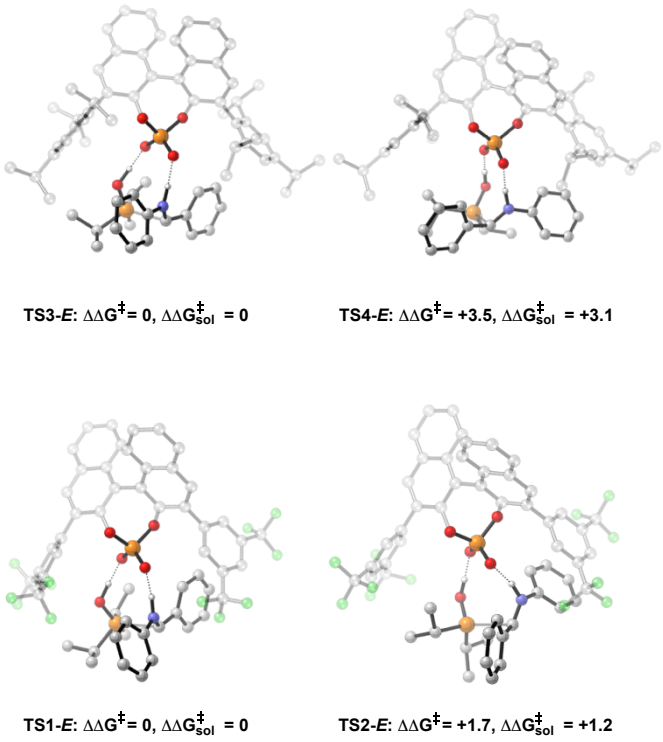

Figure 5. Competing TS for 3,5-$CF_3C_6H_4$ and TRIP derived phosphoric acid catalysed phosphorylation of imines. ONIOM (B3LYP/6-31G**:UFF), single-point energy M06-2X/6-31G**. Grayed-out regions were treated with UFF, and the full-colour regions were treated B3LYP/6-31G**. All energies in kcal mol^{-1}. Structures generated using CYLview.[43]

Slightly simplified molecules were used in the calculations: PMP group was replaced for a Ph. To further test the validity of the calculations the results were compared with those from CPA-1. Consistent with BINOPtimal, the energy differences between the pathways leading to the competing products increased with the TRIP catalyst. For the CPA-1 catalysed reaction the lowest energy calculated TS is **TS1-E** (*Type I E*), which is in good agreement with experiment and BINOPtimal. The *Z* pathways are higher in energy due to the internal substrate steric interactions, structures and energies of these TS can be found in the SI. The directly competing one was that of the opposite imine orientation (*Type II E*). The reduced steric profile of the catalyst substituents creates more space in the chiral cavity, thus, allows the imine to be placed parallel to the 3,3' substituents. This tilted position of the imine greatly reduces the steric interactions with the catalyst lowering the energy of the *Type II E* pathway, **TS2-E,** accounting for the small free energy difference, 1.2 kcal mol^{-1}, between the pathways. Increasing the steric bulk of the 3,3' groups increases the steric interactions between the reagents and the catalyst substituents. The nucleophile adopts a different position within the catalyst cavity and this shifting distorts the parallel arrangement of the imine leading to greater steric interactions between the catalyst, resulting in, raising the energy of the *Type II E,* **TS3-E**, relative to the *Type I E,* **TS4-E**. In order to test the ONIOM method, key TS were optimized using the B3LYP functional and 6-31G* basis set followed by single point corrections as before. These values compare well, and similarly to the ONIOM method, TRIP is predicted to lead to a better performance (see SI). We expect the

greater energetic preference, validated using two computational methods, for *Type I E* with the BINOPtimal suggested catalyst, TRIP, to translate to experiment.

As the next step we applied BINOPtimal to new reactions and demonstrate that it accurately accounts for the outcomes in these cases.[44-45] This shows BINOPtimal's power is not limited to classifying the vast literature but it can also be used to analyse new reactions and predict optimal catalysts for new synthetic strategies.

In summary, building on previous computational work, we have created BINOPtimal, a web tool for optimal chiral phosphoric acid selection. BINOPtimal can be easily used by non-experts to aid chemical synthesis and to provide new understandings to experimental results for 70 literature reactions reporting over 1000 transformations in total. The program is able to analyze reactions that have yet to be tested and to suggest the optimal catalyst. It is our aim to further extend the tool to more reactions of this type.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

J.P.R thanks the EPSRC for financial support. K.E thanks the Leverhulme and Isaac Newton trust for the financial support.

References

1 T. Akiyama, *Chem. Rev.* 2007, **107**, 5744.
2 M. Terada, *Synthesis* 2010, **2010**, 1929.
3 M. Rueping, A. Kuenkel and I. Atodiresei, *Chem. Soc. Rev.* 2011, **40**, 4539.
4 D. Parmar, E. Sugiono, S. Raja and M. Rueping, *Chem. Rev.* 2014, **114**, 9047.
5 J. P. Reid, L. Simón and J. M. Goodman, *Acc. Chem. Res.* 2016, **49**, 1029.
6 J. P. Reid and J. M. Goodman, *J. Am. Chem. Soc.* 2016, **138**, 7910.
7 J. P. Reid and J. M. Goodman, *Chem. Eur. J.* 2017, **23**, 14248.
8 (a) E. J. Corey, *Pure Appl. Chem.* 1967, **14**, 19. (b) A. P. Johnson, C. Marshal, and P. N. Judson, *J. Chem. Inf. Comput. Sci.* 1992, **32**, 411. (c) A. P. Johnson and C. Marshal, C. *J. Chem. Inf. Comput. Sci.* 1992, **32**, 418. (d) A. P. Johnson and C. Marshal, *Chem. Inf. Comput. Sci.* 1992, **32**, 426.
9 C. W. Coley, W. H. Green and K. F. Jensen, *Acc. Chem Res.* 2018, **51**, 1281.
10 K. Funatsu and S. Sasaki Tetrahedron *Comput. Methodol.* 1988, **1**, 27.
11 K. Satoh and K. Funatsu *J. Chem. Inf. Comput. Sci.* 1999, **39**, 316.
12 W. L. Jorgensen and E. R. Laird *Pure Appl. Chem.* 1990, **62**, 1921.
13 W. D. Ihlenfeldt and J. Gasteiger, *Angew. Chem., Int. Ed. Engl.* 1995, **34**, 2613.
14 R. Höllering, J. Gasteiger, L. Steinhauer, K. Schulz and A. J. Herwig, *Chem. Inf. Comput. Sci.* 2000, **40**, 482.
15 H. Satoh and K. J. Funatsu, *Chem. Inf. Comput. Sci.* 1996, **36**, 173.
16 (a) I. M. Socorro, K. Taylor and J. M. Goodman, *Org. Lett.* 2005, **7**, 3541. (b) I. M. Socorro and J. M. Goodman, *J. Chem. Inf. Model,* 2006, **46**, 606.
17 M. W. Lodewyk, M. R. Siebert and D. J. Tantillo, *Chem. Rev.*, 2012, **112**, 1839.
18 (a) S. G. Smith and J. M. Goodman, *J. Am. Chem. Soc.*, 2010, **132**, 12946. (b) N. Grimblat, M. M. Zanardi and A. M. Sarotti, *J. Org. Chem.*, 2015, **80**, 12526. (c) K. Ermanis, K. E. B. Parkes, T. Agback and J. M. Goodman, *Org. Biomol. Chem.* 2017, **15**, 8998.
19 B. Bienfait and P. Ertl. JSME: a Free Molecule Editor in JavaScript. *J. Cheminf.* 2013, **5**, 24.
20 https://www.python.org/downloads/release/python-2715/
21 RDKit: Cheminformatics and Machine Learning Software. 2013, [http://www.rdkit.org].
22 https://httpd.apache.org/
23 M. Rueping, E. Sugiono and C. Azap, *Angew. Chem., Int. Ed.* 2006, **45**, 2617.
24 W. Zheng, L. Wojtas and J. C. Antilla, Angew. *Chem., Int. Ed.* 2010, **49**, 6589.
25 T. Akiyama, H. Morita, J. Ito and K. Fuchibe, *Org. Lett.* 2005, **7**, 2583.
26 G. K. Ingle, M. G. Mormino, L. Wojtas and J. C. Antilla, *Org. Lett.* 2011, **13**, 4822.
27 G. Li, F. R. Fronczek and J. C. Antilla, *J. Am. Chem. Soc.* 2008, **130**, 12216.
28 G. B. Rowland, H. Zhang, E. B. Rowland, S. Chennamadhavuni, Y. Wang and J. C. Antilla, *J. Am. Chem. Soc.* 2005, **127**, 15696.
29 Y. Liang, E. B. Rowland, G. B. Rowland, J. A. Perman and J. C. Antilla, *Chem. Commun.* 2007, 4477.
30 Y.-X. Jia, J. Zhong, S.-F.Zhu, C.-M. Zhang and Q. L. Zhou, *Angew. Chem., Int. Ed.* 2007, **46**, 5565.
31 D. Uraguchi and M. Terada, *J. Am. Chem. Soc.* 2004, **126**, 5356.
32 Q.-X. Guo, H. Liu, C. Guo, S.-W. Luo, Y. Gu and L.-Z. Gong, *J. Am. Chem. Soc.* 2007, **129**, 3790.
33 M. Terada, K. Machioka, K. Sorimachi, *Angew. Chem. Int. Ed.* 2006, **45**, 2254.
34 G. Dagousset, W. Erb, J. Zhu and G. Masson, *Org. Lett.* 2014, **16**, 2554.
35 R. I. Storer, D. E. Carrera, Y. Ni and D. W. C. MacMillan, *J. Am. Chem. Soc.* 2006, **128**, 84.
36 S. Hoffmann, A. M. Seayad and B. List, *Angew. Chem. Int., Ed.* 2005, **45**, 7424.
37 D. Uraguchi, K. Sorimachi and M. Terada, *J. Am. Chem. Soc.* 2005, **127**, 9360.
38 X. Zeng,X. Zeng,Z. Xu, M. Lu and G. Zhong, *Org. Lett.* 2009, **11**, 3036.
39 H. Zhang, H. Wen, L. Gan and Y. Peng, *Org. Lett.* 2012, **14**, 2126.
40 http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html (SMARTS is a trademark of Daylight Chemical Information Systems)
41 J. P. Reid and J. M. Goodman, *Org. Biomol. Chem*, 2017, **15**, 6943.
42 M. Rueping, E.Sugiono, C. Azap, T. Theissmann and M. Bolte, Org.Lett. 2005, 7, 3781.
43 Legault, C. Y. CYLview, 1.0b; Université de Sherbrooke: Sherbrooke, Canada, 2009 (http://www.cylview.org).
44 A. A. Rexit, S. Luo and M. Mailikezati *J. Org. Chem.* 2016, **81**, 11384.
45 C M.-W. Chen, Q. Yang, Z. Deng, Y. Zhou, Q. Ding and Y. Peng, *J. Org. Chem.* 2018, **83**, 8688.

