Chemistry
A European
Journal

Chemistry Europe
European Chemical Societies Publishing

Accepted Article

Title: In silico anticipation of metabolic pathways extended to organic chemistry reactions: a case study with caffeine alkaline hydrolysis and the origin of camellimidazoles

Authors: Victor Turpin, Mehdi A. Beniddir, Grégory Genta-Jouve, Adam Skiredj, Jean-François Gallard, Karine Leblanc, Pierre Le Pogam, and Erwan Poupon

To be cited as: *Chem. Eur. J.* 10.1002/chem.202002196

Link to VoR: https://doi.org/10.1002/chem.202002196

WILEY-VCH

10.1002/chem.202002196

WILEY-VCH

COMMUNICATION

In silico anticipation of metabolic pathways extended to organic chemistry reactions: a case study with caffeine alkaline hydrolysis and the origin of camellimidazoles

Victor Turpin,[a] Mehdi A. Beniddir,[a] Grégory Genta-Jouve, [b,c] Adam Skiredj,[a] Jean-François Gallard,[d] Karine Leblanc, [a] Pierre Le Pogam, *[a] and Erwan Poupon* [a]

[a] Victor Turpin, Dr. Mehdi A. Beniddir, Dr. Adam Skiredj, Karine Leblanc, Dr. Pierre Le Pogam, Prof. Dr. Erwan Poupon
Chimie des Substances Naturelles, BioCIS, Université Paris-Saclay, CNRS
92290 Châtenay-Malabry (France)
* E-mail: erwan.poupon@universite-paris-saclay.fr; pierre.le-pogam-alluard@universite-paris-saclay.fr
[b] Dr. Grégory Genta-Jouve
Laboratoire de Chimie-Toxicologie Analytique et Cellulaire (C-TAC), UMR CNRS 8038 CiTCoM, Université de Paris,
4, Avenue de l'Observatoire, 75006, Paris (France)
[c] Dr. Grégory Genta-Jouve
Laboratoire Écologie, Évolution, Interactions des Systèmes Amazoniens (LEEISA), USR 3456, Université de Guyane, CNRS Guyane
275, route de Montabo, 97334 Cayenne (French Guiana)
[d] J.-F. Gallard
Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Université Paris-Saclay
1, avenue de la Terrasse, 91198 Gif-sur-Yvette (France)

Supporting information for this article is given via a link at the end of the document.

Abstract: Camellimidazoles A–C were recently reported as new natural substances in Keemun black tea. Although a "biosynthetic" route to these intriguing imidazole dimers was proposed from caffeine by the authors in this seminal report, we envisioned that a artefactual scenario, consisting of alkaline hydrolysis of caffeine and spontaneous cascade reactions with a methylene donor such as formaldehyde or methylene chloride, could also have led to their formation. To capture the diversity of molecules obtained under these conditions (*i. e.* alkaline treatment of caffeine/formaldehyde), an *in silico* MetWork-based pipeline was implemented, highlighting the sought-after camellimidazoles B and C. A wealth of further compounds were also tagged, notably comprising the herein newly described and unnatural camellimidazoles D−F that were subsequently confirmed as anticipated *in silico* upon extensive spectroscopic analyses. Likewise, camellimidazoles B and C could also be obtained using methylene chloride as an alternative methylene donor which may also have occurred in the initial phytochemical pipeline that implied this solvent. The current investigation emphasizes the fitness of MetWork-tagging to extend the logic of *in silico* anticipation of metabolic pathways to organic chemistry reactions.

Caffeine (**1**) content of tea leaves is notably stable all along fermentation process (*i.e.* "green" to "black" tea)[1,2] in opposition to polyphenols which are allowed to oxidize under controlled humidity and temperatures (theaflavin and thearubigin formation for example).[3,4] In that context, camellimidazoles A–C (**2**–**4**) were recently reported as new, seemingly natural, methylene-bridged dimeric imidazoles from Keemun black tea.[5] These structures, along with some trimethylallantoine derivatives later described by the same research group,[6] were proposed to be generated through an alternative caffeine catabolic pathway. Even though a detailed "biosynthetic" pathway was proposed by the authors, a few simple chemical reactions from caffeine can be deemed to afford camellimidazoles, hinting their plausible unnatural origin.[5] The present work aimed at assessing whether camellimidazoles could be obtained from caffeine, having in mind the specific conditions used by Wang et al. to isolate these compounds (*vide infra*), rather than being the end products of the new catabolic pathway of caffeine, as proposed by the authors in the seminal report of camellimidazoles. Accordingly, two different purely chemical scenarios (involving formaldehyde or methylene chloride as a methylene donor) might have led to camellimidazoles in the specific work reported by Wang et al.

At first, a retro(bio)synthetic analysis of camellimidazoles reveals, indeed, the need for four building blocks, *i.e.*:
- imidazoles **5** (aka caffeidine) and **6** that may both arise from hydrolytic decarboxylation of caffeine (**1**);
- methylamine that may be provided through hydrolysis of imidazoles **5** and **6**;
- and a bridging methylene source such as formaldehyde or methylene chloride, the origin of which will be discussed in the following.

Isohypsic cascades of Mannich-type reactions and involving cross condensations of **5** and **6** may then easily explain the formation of camellimidazoles A-C (Figure 1).

Taking into account the basic conditions of some extraction steps of the seminal paper, a mere chemical reactivity can rather be envisaged to afford the newly reported natural substances, especially when having in mind the huge quantity of starting material (200 kg). This chemical reactivity-driven scenario would rely on repeatedly described purine alkaline hydrolysis reactions.[7–9] Quite strikingly, the putative biosynthetic scheme proposed by the authors relies on a surprisingly similar way as it proceeds *via* a caffeidine (**5**) intermediate which was formerly chemically obtained from caffeine (**1**) upon alkaline hydrolysis.

10.1002/chem.202002196

COMMUNICATION

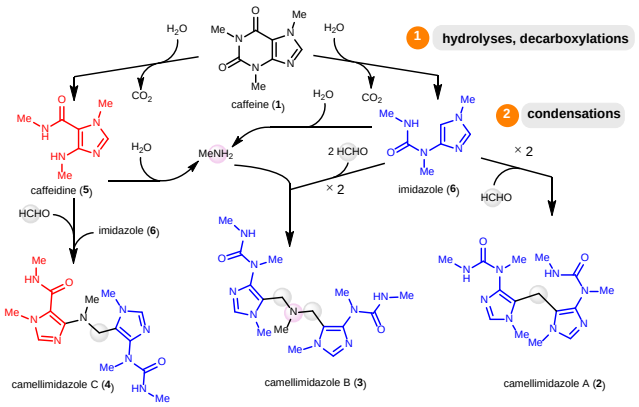


Figure 1. The (bio)chemical logic of camellimidazoles formation.

The current investigation aims at assessing the chemical diversity generated under accelerated caffeine hydrolytic degradation conditions that were deemed suitable for camellimidazoles formation (*i.e.* in the presence of formaldehyde in protic medium). Reaction crude mixtures can, in this particular situation, be seen as "artificial metabolomes" that are of very limited complexity by comparison to natural organism extracts, so that they may advantageously enter a prioritization rational workflow. For this purpose, untargeted mass spectrometric data of the crude reaction batches were generated by LC-HRMS/MS prior to being organized as a molecular network (MN).[10] The resulting MN was then further labelled by in silico-generated molecular structures obtained with MetWork.[11] From an applied standpoint, these putative structures were obtained based on foreseen chemical reactions that were first implemented into MetWork server and their MS/MS spectra were further predicted by CFM-ID (Competitive Fragmentation Modeling for Metabolite Identification) .[12] As a last step, a similarity comparison between theoretical and experimental MS/MS spectra is performed to annotate the MN (Figure 2).

10.1002/chem.202002196

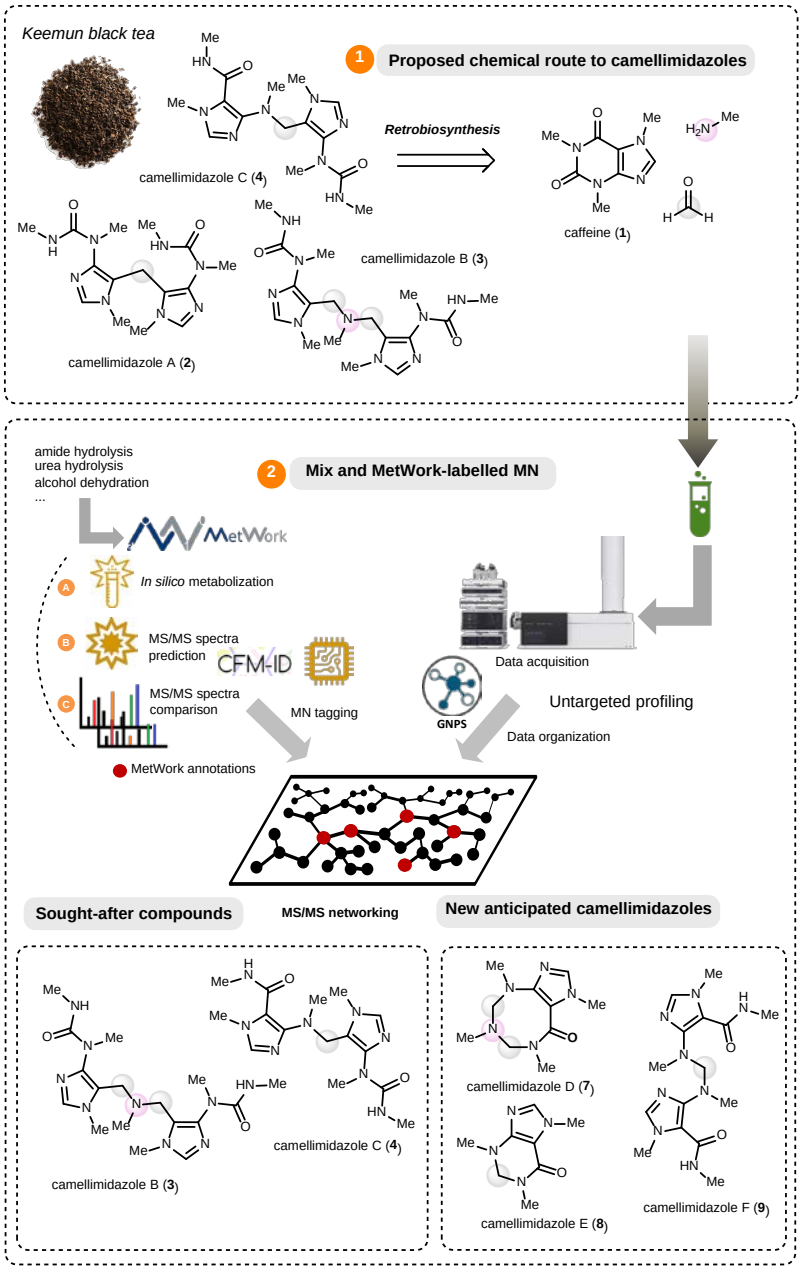


Figure 2. Overview of the analytical workflow developed in this study along with the structures of the obtained compounds.

The Wang *et al.* "biosynthetic" hypothesis was mimicked at the bench. Caffeine (**1**) was reacted under basic conditions (NaOH) and hydrolysis was accelerated by heating (80 °C). Mild acidic conditions were then imposed to favor the key decarboxylating steps presumably leading to imidazoles **5** and **6** (Figure 1). Finally, formaldehyde was added in the mixture which was allowed to slowly react at room temperature for a dozen days. To get a wide insight into the generated species, the crude batch was profiled by HPLC-HRMS/MS to be later organized as a molecular network.[10,13] The later was further labelled by in silico generated compounds obtained through MetWork pipeline as described in the following section. As an implementation of the metabolome consistency concept,[14] MetWork was developed to anticipate new natural products from untargeted tandem mass spectrometric analyses organized as molecular networks.[10] For this purpose, MetWork web server features (i) a collaborative library of (bio)chemical transformations and (ii) a CFM-ID based MS/MS spectra prediction module. From a spectral library match annotation in the molecular network, in silico metabolization algorithms generate putative structures prior to predicting the associated MS/MS spectra. Lastly, a similarity comparison between the MS/MS spectra is performed to annotate the node. Based on aforementioned literature reports on purines and imidazoles reactivity, the foreseen chemical reactions were uploaded into MetWork

10.1002/chem.202002196

web server: amide hydrolysis, urea hydrolysis, alcohol dehydration, amine hydroxymethylation, enamine hydroxymethylation, carbamic acid decarboxylation, imidazole decarboxylation, iminium amination, hemiaminal dehydration, aza-Michael addition and Michael addition with enamine (Figure 3A) and are made available to users. This reaction panel was applied *in silico* on caffeine (**1**), methylamine, and formaldehyde. This analytical pipeline instantly tagged the seminal camellimidazoles B and C (**3** and **4**) along with most intermediates proposed to intercede in camellimidazole formations and some putative unprecedented structures such as methylene-bridged imidazole derivatives (Figure 3B).

10.1002/chem.202002196
WILEY-VCH

COMMUNICATION

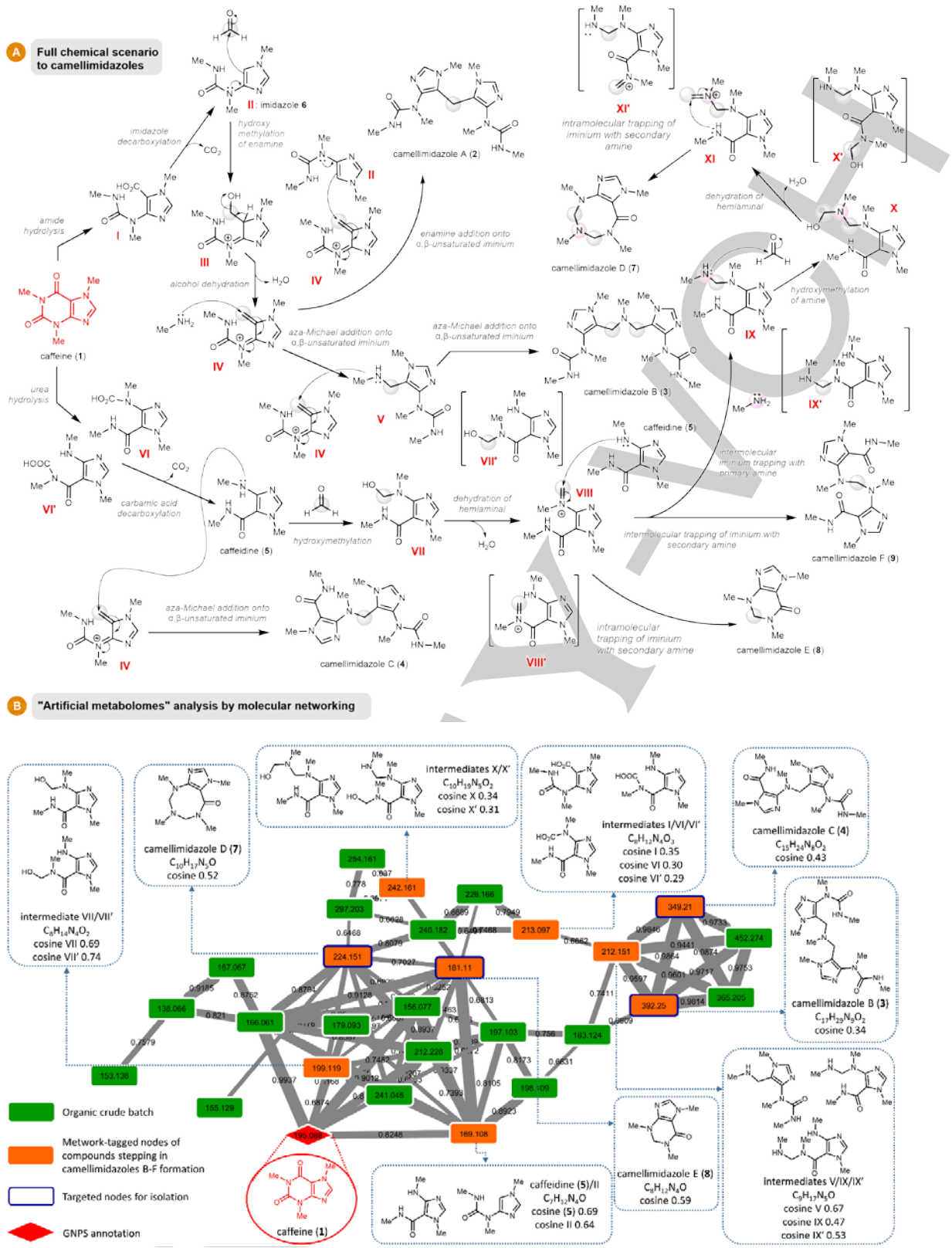


Figure 3. A: Putative chemical scenario to camellimidazoles with detailed mechanisms. **B**: MetWork-tagged cluster of the organic crude batch highlighting inferred intermediates to the camellimidazole series and some anticipated new structures, highlighting prioritized ions for isolation and NMR structure elucidation.

Camellimidazoles B and C, and caffeidine (**3** and **4**) were instantly tagged upon MetWork data processing, with this latter being already proposed as a "biosynthetic" pillar *en route* to the former. Mass spectrometric-streamlined isolation of these compounds validated their tentative identities through the thorough analysis of the complete set of NMR spectra.[5]

Under our conditions, camellimidazole A (**2**) seems not to have been obtained, the methylene bridge formation in this latter may require a series of slower reactions (enamine *versus* nitrogen nucleophilic attack to anchor the second imidazole unit). Besides having pinpointed these expected structures, MetWork facilitated the annotation of a further dozen ions. Some of these putative structures were prioritized for isolation based on i) the degree of scaffold originality ii) the possibility to be isolated for subsequent NMR structure elucidation. Three methylene-containing, unnatural compounds were indeed isolated and characterized, they were subsequently named camellimidazoles D–F (**7**–**9**). The structure elucidation of these new compounds is detailed in the Supporting Information and further validate the Computer Assisted Natural Products Anticipation (CANPA) approach proposed last year.[15,16] MS/MS spectra related to the different isolated camellimidazoles were uploaded to the GNPS spectral libraries, with their individual identifiers being provided in the supporting material.

As mentioned before, camellimidazoles were isolated in minute amounts from tea (8 to 12 mg from 200 kg of tea) following a convoluted phytochemical workflow which could have triggered chemical transformations of caffeine towards camellimidazoles through a succession of basic steps and exposure to methylene chloride.[5] Efforts were made by the first authors to address this query, but the analytical evidence to support the claimed genuineness of camellimidazoles A–C as natural products (*viz*. UPLC-HR-ESI-MS of fresh leaves and processed materials) are poorly convincing (level of confidence 4/5 according to Schymanski *et al. i.e.* "insufficient evidence to propose possible structure"[17]). To benefit from a higher degree of analytical confidence, the extraction process having led to the dubious characterization of camellimidazoles A−C in this seminal report were repeated by us from two different Keemun black tea samples for subsequent HPLC-HRMS^2 chemical profiling. The obtained tandem-mass spectrometric data were later organized as MNs and annotated against (i) the experimental MS/MS spectra of all isolated camellimidazole derivatives and (ii) the CFM-ID predicted fragmentation patterns of all intermediates formerly shown to step in camellimidazole synthesis. The lack of any annotation as camellimidazoles in the obtained MNs further supported the alkaline hydrolysis scenario as the (sole) plausible route to reach the camellimidazole series (Figures S2-S5, Supporting Information).

A last point to mitigate was the nature of the mechanisms having led to camellimidazole frameworks in the specific phytochemical workflow followed by Wang and coll., that did not use formaldehyde nor methylamine. The *in situ* generation of methylamine can be related to the alkaline hydrolysis of caffeine,[7] as reasonably assumed by the authors. The claimed endogenous content of HCHO in *Camellia sinensis* is, conversely questionable, especially when taking into account black tea manufacturing process. This led us to infer that methylene chloride should have served as the real methylene donor instead.[18,19] With this in mind, liquid-liquid extraction conditions were mimicked by stirring caffeine (**1**) in a biphasic system constituted of a basic aqueous phase (NaOH, 2 eq.) and dichloromethane at room temperature for 3 days. Untargeted LC-MS^2 analysis of the crude batch and subsequent MN revealed the occurrence of the camellimidazole derivatives reported from our former reactive conditions (Figure S6, Supporting Information), strengthening the possibility of an artefactual origin.

Besides unveiling in a straightforward manner the likely occurrence of formerly reported camellimidazoles and its postulated "biosynthetic" intermediates, the MetWork pipeline also shed light on some unprecedented appendages. Mass spectrometric approaches offer interesting perspectives owing to their high sensitivity. As such, tandem mass spectrometry-based dereplicative strategies lie at the forefront of modern natural product discovery research and proved their ability to provide a sharp glance into the chemical space encompassed by an analyzed sample to further pinpoint and streamline new products from various natural sources.[20] As of 2020, the field of organic synthesis does not yet benefit from such advanced mass spectrometric platforms and the degree of structural information conveyed by these techniques remains most often limited to a poorly informative molecular formula. To overcome this bottleneck, the CFM-ID predicted mass spectrometric fragmentation of *in silico* derivatized reactants is herein demonstrated as a valuable strategy to straightforwardly pinpoint tentative structures from the crude batch without the need for any standard nor even any former literature report. The specific example of camellimidazoles might be ideally fitted to retrieve the full extent of advantages offered by the MetWork pipeline since it involves (i) a set of limited and simple organic molecules and (ii) a reduced number of possible chemical processes to afford a diversity of more complex compounds through iterative metabolization steps. This workflow is highly dependent on the predictive value of CFM-ID which, due to its machine-learning functioning, has uneven performances across structural series. As such, structures close to those comprised in the training set of CFM-ID have higher chances of having their tandem mass spectrum correctly predicted than structures having no such close derivative. In the present case, we were delighted to observe that the predicted MS^2 spectra of our various camellimidazoles were of good quality notwithstanding their original structures. The rather unique scaffold of camellimidazoles could indeed be correctly predicted with cosine values ranging from 0.3 to 0.6 which are very good scores when dereplicating against in silico MS^2 spectra. Accordingly, MetWork workflow can be deemed of peculiar interest where chemical diversity stems from such « minimalistic » conditions, *e.g.* bio-inspired cascade reactions[21] or prebiotic chemistry-related area.[22]

Acknowledgements

We thank the French "Agence Nationale pour la Recherche" (grant ANR-19-CE07-0002-01) for funding this study. Laurent Evanno is thanked as the Ph. D. director of Victor Turpin.

Keywords: Natural products • Mass spectrometry • Molecular networking • MetWork platform • Total synthesis

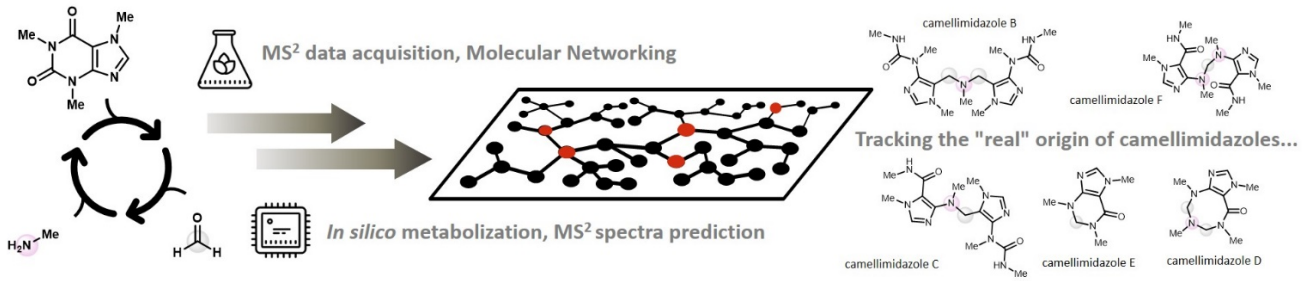
[1] S. Sang, J. D. Lambert, C.-T. Ho, C. S. Yang, *Pharmacol. Res.* **2011**, *64*, 87–99.
[2] S. Li, C.-Y. Lo, M.-H. Pan, C.-S. Lai, C.-T. Ho, *Food Funct.* **2013**, *4*, 10–18.
[3] A. Robertson, D. S. Bendall, *Phytochemistry* **1983**, *22*, 883–887.
[4] Y. Matsuo, T. Tanaka, I. Kouno, *Tetrahedron Lett.* **2009**, *50*, 1348–1351.
[5] W. Wang, X. Tang, F. Hua, T.-J. Ling, X.-C. Wan, G.-H. Bao, *Org. Lett.* **2018**, *20*, 2672–2675.
[6] W. Wang, B.-Y. Zhu, P. Wang, P. Zhang, W.-W. Deng, F.-H. Wu, C.-T. Ho, T.-J. Ling, Z.-Z. Zhang, X.-C. Wan, G.-H. Bao, *Org. Lett.* **2019**, *21*, 5147–5151.
[7] K. Kigasawa, K. Ohkubo, H. Shimizu, T. Kohagizawa, R. Shoji, *Chem. Pharm. Bull.* **1974**, *22*, 2448–2451.
[8] T. Itaya, K. Ogawa, *Chem. Pharm. Bull.* **1985**, *33*, 1906–1913.
[9] T. Itaya, H. Matsumoto, *Chem. Pharm. Bull.* **1985**, *33*, 2213–2219.
[10] J. Y. Yang, L. M. Sanchez, C. M. Rath, X. Liu, P. D. Boudreau, N. Bruns, E. Glukhov, A. Wodtke, R. de Felicio, A. Fenner, W. R. Wong, R. G. Linington, L. Zhang, H. M. Debonsi, W. H. Gerwick, P. C. Dorrestein, *J. Nat. Prod.* **2013**, *76*, 1686–1699.
[11] Y. Beauxis, G. Genta-Jouve, *Bioinformatics* **2018**, DOI 10.1093/bioinformatics/bty864.
[12] F. Allen, R. Greiner, D. Wishart, *Metabolomics* **2015**, *11*, 98–110.
[13] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. B. P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Ø. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein, N. Bandeira, *Nat. Biotechnol.* **2016**, *34*, 828–837.
[14] C. Audoin, V. Cocandeau, O. Thomas, A. Bruschini, S. Holderith, G. Genta-Jouve, *Metabolites* **2014**, *4*, 421–432.
[15] A. E. Fox Ramos, C. Pavesi, M. Litaudon, V. Dumontet, E. Poupon, P. Champy, G. Genta-Jouve, M. A. Beniddir, *Anal. Chem.* **2019**, *91*, 11247–11252.
[16] L. A. Shaala, D. T. Youssef, J. M. Badr, S. M. Harakeh, G. Genta-Jouve, *Mar. Drugs* **2019**, *17*, 584.
[17] E. L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H. P. Singer, J. Hollender, *Environ. Sci. Technol.* **2014**, *48*, 2097–2098.
[18] R. Besselièvre, N. Langlois, P. Potier, *Bull. Soc. Chim. Fr* **1972**, *4*, 1477.
[19] R. J. Capon, *Nat. Prod. Rep.* **2020**, *37*, 55–79.
[20] A. E. Fox Ramos, L. Evanno, E. Poupon, P. Champy, M. A. Beniddir, *Natural product reports* **2019**, *36*, 960–980.
[21] E. Poupon, E. Gravel, *Chem. Eur. J.* **2015**, *21*, 10604–10615.
[22] K. Ruiz-Mirazo, C. Briones, A. de la Escosura, *Chem. Rev.* **2014**, *114*, 285–366.

10.1002/chem.202002196

WILEY-VCH

COMMUNICATION

Entry for the Table of Contents



Surmising an artefactual nature for camellimidazoles recently isolated from black tea, their synthesis was undertaken from caffeine. A molecular-networking based workflow revealed the occurrence of the targeted compounds and *anticipated* some unprecedented derivatives, benefitting from a virtual metabolization platform. The value of *in silico* prediction in organic synthesis is here emphasized as a way to answer critical biosynthetic issues.