Deep metabolome annotation and the necessity of a thought experiment

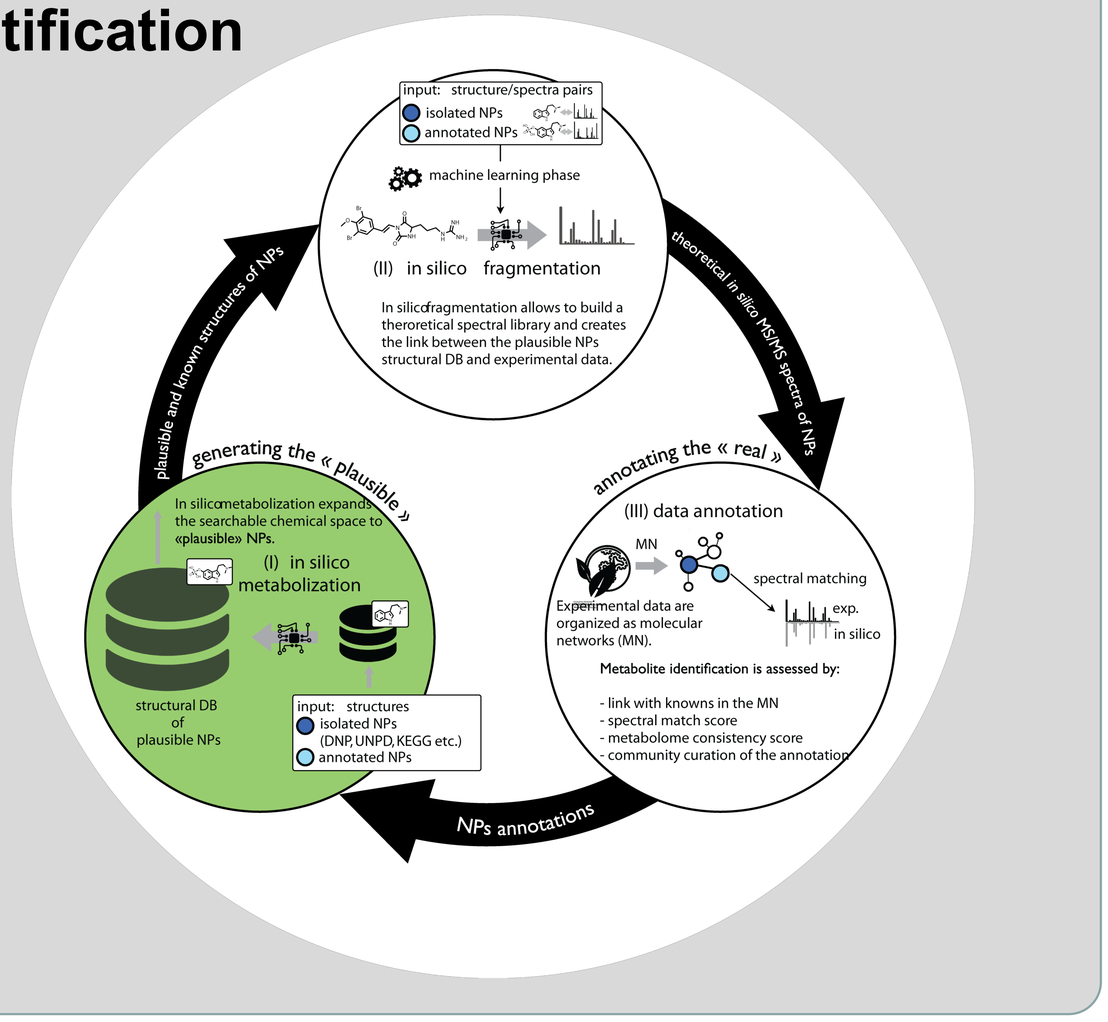
Pierre-Marie Allard[a], Jonathan Bisson[b], Alexandre Da Silva[c], Jean-Luc Wolfender[a], Olivier P. Thomas[d], Mehdi Beniddir[e] and Grégory Genta-Jouve[c]

[a] *Université de Genève (UNIGE), Ecole de pharmacie Genève-Lausanne (EPGL) – Rue Michel-Servet 1 - 1211 Genève 4 - Suisse*
[b] *College of Pharmacy - University of Illinois at Chicago, 833 S. Wood Street (781) Chicago, IL 60612 - États-Unis*
[c] *Laboratoire de toxicologie cellulaire et analytique (C-TAC), U.M.R./C.N.R.S. 8638 COMETE, Faculté des Sciences Pharmaceutiques et Biologiques, Université Paris Descartes, Sorbonne Paris Cité, 4 Avenue de l'Observatoire, F-75006 Paris, France*
[d] *National University of Ireland Galway, Marine Biodiscovery lab, University road, Galway, Irlande*
[e] *Université Paris Sud CNRS : UMR8076, Faculté de Pharmacie Université Paris-Sud 5, rue J-B Clément 92296 Châtenay-Malabry cedex - France*

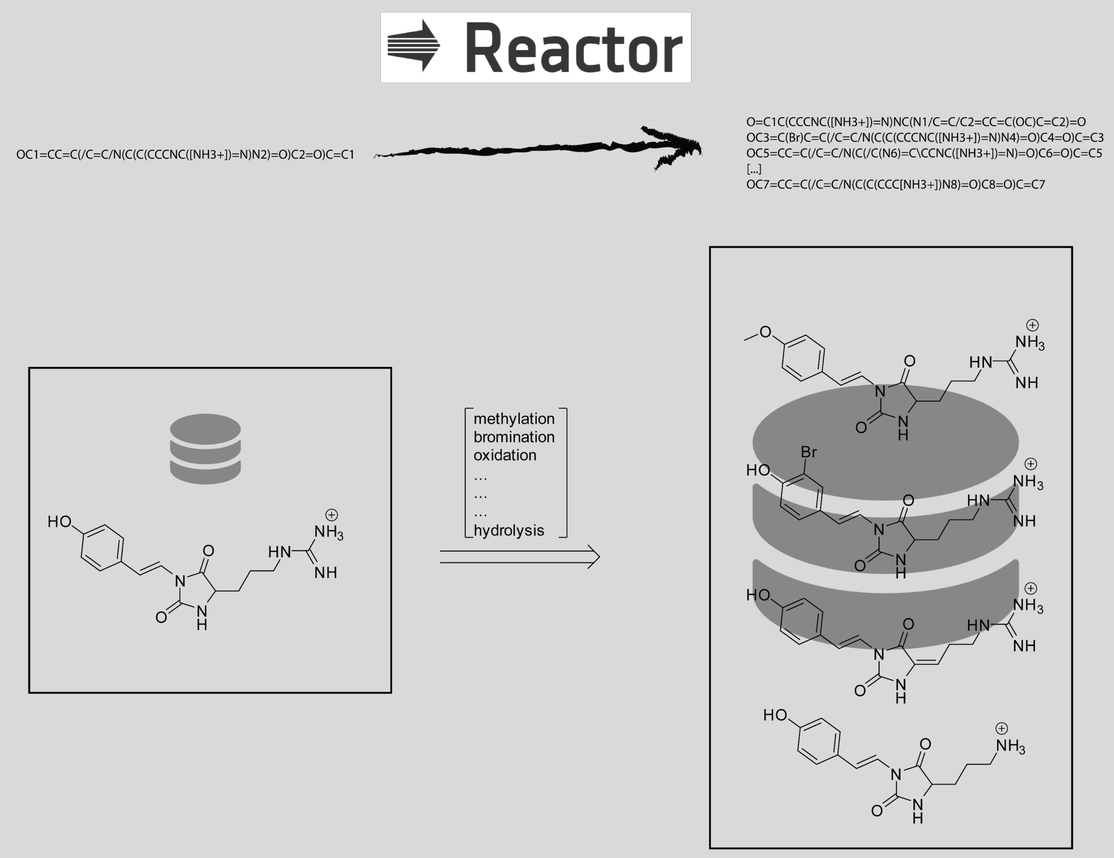
Introduction

The metabolome of an organism consists in the totality of the small molecules present in this organism at a given time. Unlike genome, transcriptome or proteome, it represents the real-time content of the living organism at the metabolic level. In order to understand complex biological interactions between cells, tissues, organs or individuals in a specific ecosystem, it is fundamental to first identify all the components, i.e. all the molecules that can be encountered. Over the last decade, identification of small biologically generated compounds, the metabolites, has emerged as the main issue in the field of metabolomics and despite all the great efforts of the community during the last years only a relatively small number of compounds have been referenced in accessible databases, often with only hypothetical structures. Herein we propose to discuss the philosophical aspects of the important question "**what is a known natural product**?" in order to move forward to a new paradigm in metabolite identification: **identifying the real from the possible**.

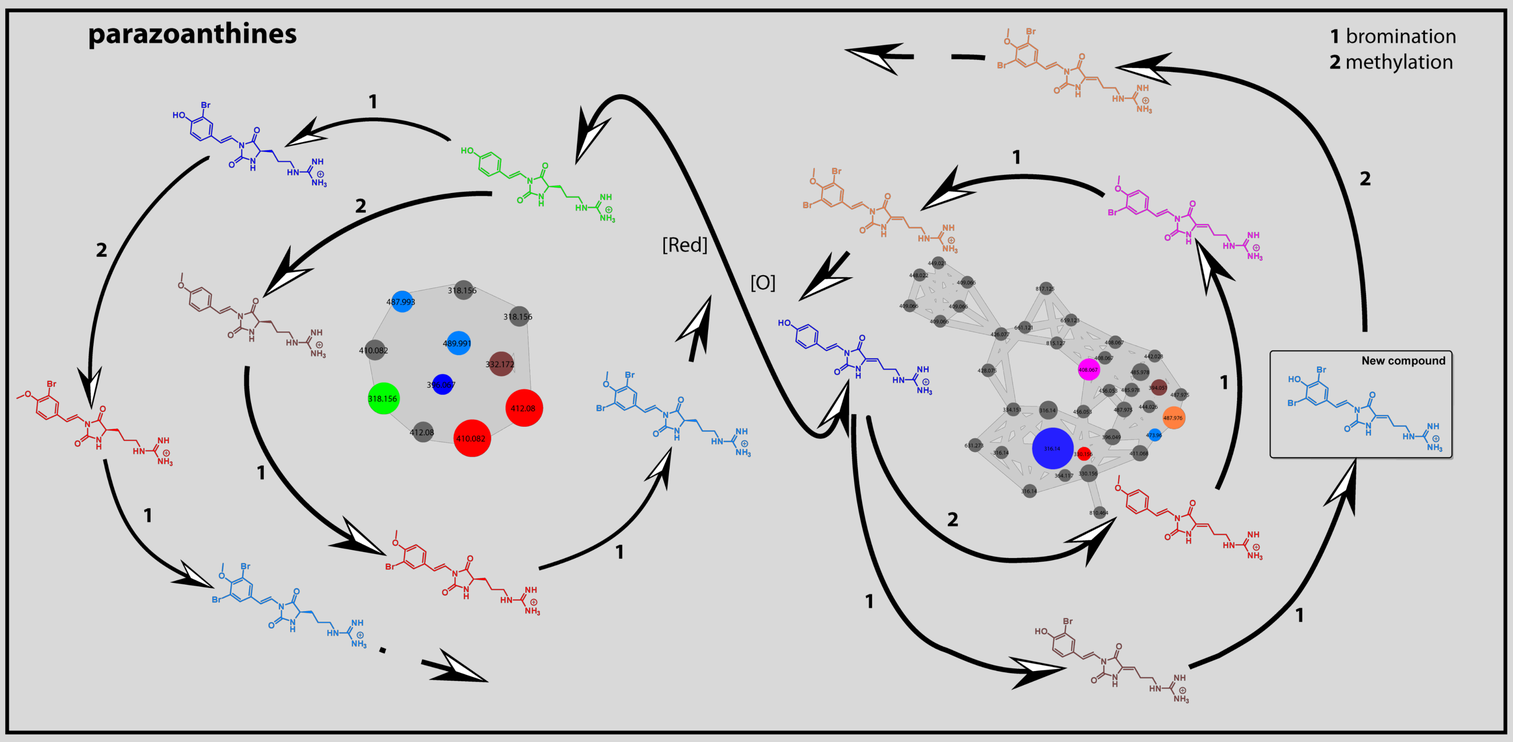
The virtuous cycle of metabolite identification

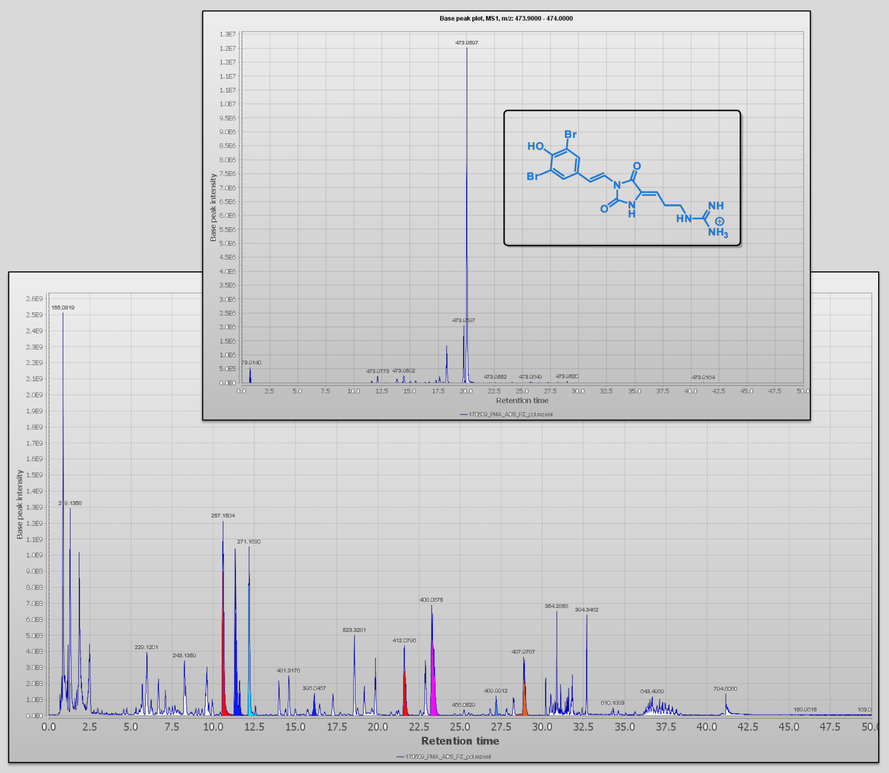


In-silico metabolization



Results





Conclusion

Here we present the first fully automated identification of a new natural product using the metabolome consistency principle [1] and the virtuous cycle of identification [2]. The structure was indeed proposed by a modification of a known substance (parazoanthine B) using chemical and biochemical reactions implemented in the Chemaxon Reactor software [3]. After structure generation, *in silico* spectra prediction was performed using CFM-ID [4]. The automated mapping of the molecular network (generated using GNPS [5]) was performed using the ISDB script [6].

[1] Audoin et al. Metabolites 2014, 4(2), 421-432; doi:10.3390/metabo4020421
[2] Allard et. Curr. Opin. Chem. Biol. 2017, 36, 40-49 doi:10.1016/j.cbpa.2016.12.022
[3] https://www.chemaxon.com/products/reactor/
[4] Allen et al. Nucleic Acids Res. 2014, 42, W94-9. doi: 10.1093/nar/gku436.
[5] Wang et al. Nature Biotechnology 34, 828–837 (2016) doi:10.1038/nbt.3597
[6] Allard et al. Anal. Chem., 2016, 88 (6), pp 3317–3323