COMPUTATIONAL FRAMEWORK FOR THE IDENTIFICATION OF BIOPRIVILEGED MOLECULES

Xiaowei Zhou,[*, a] Zachary J. Brentzel,[b] George A. Kraus,[c] Peter L. Keeling,[c]
James A. Dumesic,[b] Brent H. Shanks,[c] and Linda J. Broadbelt[*, a]
[a] Northwestern University, Evanston, IL 60208.
[b] University of Wisconsin-Madison, Madison, WI 53706.
[c] Iowa State University, Ames, IA 50011.

Abstract

Bridging chemical and biological catalysis by bioprivileged molecules provides a useful and flexible new paradigm for producing biobased chemicals. However, the discovery of bioprivileged molecules has been demonstrated to require extensive experimental effort over a long period of time. This work developed a computational framework for systematically screening and identifying C6 bioprivileged molecule candidates and associated feasible chemical pathways to biobased chemicals including novel compounds, thereby guiding experimentation and computational analysis. Analysis of the 100 bioprivileged molecule candidates identified from 29,252 C6 compounds creates a paradigm for establishing a rational structure for the design of bioprivileged molecules in general. The computational framework can be extended to identification of other biological candidates including C4s and C5s.

Keywords

Bioprivileged molecules, Biobased chemicals, automated network generation, reactivity.

Introduction

The integration of biological catalysis and chemical catalysis is emerging as an effective strategy to produce biobased chemicals from biomass carbohydrates, where biological catalysis can be used to defunctionalize sugars to yield platform species that can be catalytically upgraded into final products using chemical catalysis. (Schwartz et al., 2014, 2016) Moreover, the use of platform molecules allows for the more easily diversified production of a variety of species including both commodity chemicals and high-value building blocks.

Very recently, the concept of bioprivileged molecules was introduced by Shanks and Keeling (2017). as a useful new paradigm for bridging the biological and chemical catalysis gap to develop biobased chemicals and create value from biomass. Bioprivileged molecules are defined as "biology-derived chemical intermediates that can be efficiently converted to a diversity of chemical products including both novel molecules and drop-in replacements".

Unfortunately, there is no current strategy for the direct identification and development of bioprivileged molecules. While a small number of bioprivileged molecules such as muconic acid, 5-hydroxymethylfurfural (5-HMF), and triacetic acid lactone (TAL) have been identified, the process has been somewhat serendipitous and slow. Moreover, there are a number of key challenges that prevent the development of bioprivileged molecules: Where is the optimal handoff between biology and chemical catalysis? Can the identification of bioprivileged molecules be accelerated using computational approaches?

[*] To whom all correspondence should be addressed, L. J. Broadbelt at broadbelt@northwestern.edu; tel: +1-847-467-1751 and X. Zhou at xiaowei.zhou2@chemours.com; tel:+1-302-353-5936.

In this work, we developed a computational framework for identification of all possible C6HxOy molecules (29,252) that can be honed down to a manageable number of candidate bioprivileged molecules based on analysis of structural features, reactive moieties and reactivity of species, and the evaluation of the reaction network and resulting products based on automated network generation (Broadbelt et al., 1994).

Development of the computational platform

As shown in Fig. 1, there are two major modules in the computational platform for the identification of bioprivileged molecules. The first module is a simple filtering of the candidate C6 oxygenates as potential bioprivileged molecules based on the analysis of structure and reactivity as captured through the reactivity indices. The second module is to screen and evaluate candidate molecules using detailed network generation with a clear demarcation between the two modules.

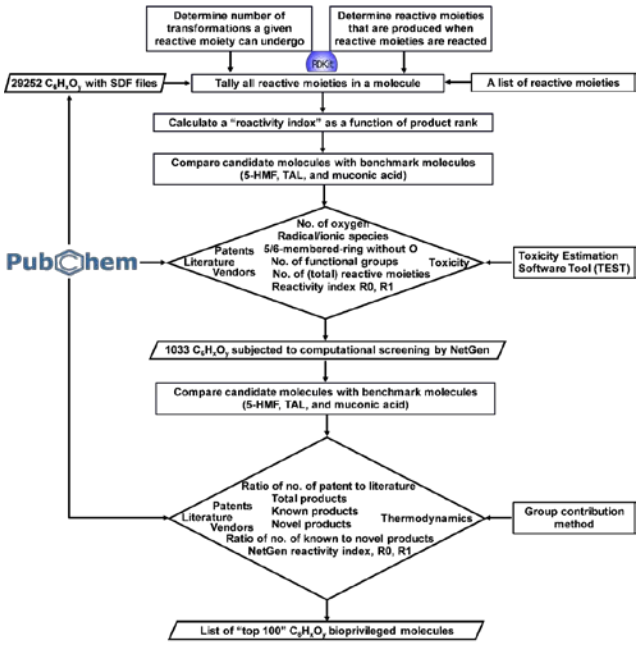

Figure 1. Flowchart of identification of C6 bioprivileged molecules

Required input is the structure data format (SDF) of the starting molecules and the reaction rules. On-the-fly estimation of thermodynamics by a group contribution method is introduced as a screening criterion to identify the feasibility of reactions and pathways. Generated species are dynamically linked to the PubChem database for identification of novel products and evaluation of the known products as attractive candidates. Application of the proposed computational framework in screening 29,252 C6 species and identifying a list of 100 C6HxOy bioprivileged molecule candidates is presented. Each of the 100 candidate molecules falls into one of nine broad compound classes and is typically composed of carbon atoms with a different chemical environment and, as a result, distinct reactivity patterns. Sensitivity analysis of the parameters used in the filtering steps leading to the bioprivileged molecules that were identified is discussed, and analysis of favorable structural features, reactive moieties, and functionalities of C6HxOy candidate bioprivileged molecules is performed.

Conclusions

We present here a computational framework that combines evaluation of structural moieties, reactivity, and selectivity in order to rank 29,252 C6HxOy molecules, honing the list down to 100 bioprivileged molecule candidates. The framework is sufficiently general that it can be applied to a variety of different carbon number substrates and refined to include other constraints suggested by experiments and theory. The framework also can be run in reverse if a specific final biobased product were desired in order to find the intermediates that are candidates for its precursor.

Acknowledgments

The approach advocated in this opinion is reflected in the core mission of the National Science Foundation Engineering Research Center for Biorenewable Chemicals (CBiRC). CbiRC is funded by the National Science Foundation under Award no. EEC-0813570. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily represent the views of the National Science Foundation. The authors thank Dr. Lindsay Oakley, Rob Brydon, and Dr. James Jeffryes at Northwestern University, Pradeep Natarajan at Indian Institute of Technology Madras, Professor Matthew Neurock at University of Minnesota, Professor Bob Davis at University of Virginia, Dr. Wolf-Dietrich Ihlenfeldt at Xemistry GmbH in Germany, and Dr. Paul A. Thiessen and Dr. Bo Yu at NCBI / NLM / NIH for fruitful discussions and useful suggestions.

References

Schwartz, T. J, O'Neill, B. J., Shanks, B. H., Dumesic, J. A. (2014). Bridging the Chemical and Biological Catalysis Gap: Challenges and Outlooks for Producing Sustainable Chemicals. *ACS Catal.*, 4, 2060.

Schwartz, T. J., Shanks, B. H., Dumesic, J. A. (2016). Coupling chemical and biological catalysis: a flexible paradigm for producing biobased chemicals. *Curr. Opin. Biotechnol.*, 38, 54.

Shanks, B. H. Keeling, P. L. (2017). Bioprivileged molecules: creating value from biomass. Green Chem., 19, 3177.

Broadbelt, L. J., Stark, S. M., Klein, M. T. (1994). Computer Generated Pyrolysis Modeling: On-the-Fly Generation of Species, Reactions, and Rates. *Ind. Eng. Chem. Res.*, 33, 790.