Chinese Chemical Letters 34 (2023) 107936

Contents lists available at ScienceDirect

Chinese Chemical Letters

journal homepage: www.elsevier.com/locate/cclet

Research strategies in click chemistry: Measuring its cognitive contents and knowledge flow

Jing Zhu[a], Ming Chen[a], He Liang[b], Quanmin Li[a], Xuefeng Jiang[c,*]

[a] Department of Philosophy, East China Normal University, Shanghai 200241, China
[b] School of Journalism, Fudan University, Shanghai 200433, China
[c] Shanghai Key Laboratory of Green Chemistry and Chemical Process, Department of Chemistry, East China Normal University, Shanghai 200062, China

ARTICLE INFO

Article history:
Received 7 October 2022
Revised 12 October 2022
Accepted 13 October 2022
Available online 22 October 2022

Keywords:
Click chemistry
K. B. Sharpless
Research strategies
Connectivity
History and philosophy of chemistry

ABSTRACT

The issue about how outstanding scientists obtained innovative findings has drawn the interest of researchers in science, policy and scientometrics. Here, we attempt to address this question by using computational methods to measure the cognitive content and concepts of K. Barry Sharpless' research and estimate the knowledge flow of his click chemistry to other fields. First, we traced Sharpless' conceptual journey over time through topic modeling approach, mapping and clustering of the epistemic network from distant reading his publications. We find that connectivity and functions, the core features of click chemistry, are embodied in his constant search for simplicity. What makes simplicity possible is his continuous work with collaborators on reactivity and reaction mechanisms. Moreover, citation and link analysis show that click chemistry had a much richer impact on other research fields than what is generally acknowledged, and drew solutions to significant and practical questions back to chemistry from biology. Together with these findings, we propose that the click chemistry philosophy follows the way that values nature's principle. Chemistry has a clear-cut epistemic domain in modeling Nature. Thus, click chemistry as a concept on doing science beyond a connective technology goes across the boundaries between disciplines and impacts many other fields.

The issue about how outstanding scientists obtained innovative findings or the efficiency of their research strategies has drawn the interest of researchers in science, policy and scientometrics with data-driven technologies [1–3]. While bibliometric studies are concerned with Nobel laureates to understand their collaboration [4,5] or the growing interdisciplinarity of research [6,7], these works focused more on collaboration patterns or the knowledge flow among fields of science. Nonetheless, researchers are acutely aware that quantifying the cognitive extent of specific scientists or the knowledge flow of specific concepts can capture the features of elite scientists more comprehensively than the productivity-based measure [1,8–10].

Click chemistry is a term coined in 2001 by K. Barry Sharpless, winner of a second Nobel Prize in chemistry, for an approach to synthesis that prioritizes the use of a few key chemical reactions for molecule linkage containing particular chemical groups. It is considered a philosophy of not falling in love with complexity [11] and has rapidly created a prominent place in polymers, materials science and biomedicine. Why did Sharpless move his interest from asymmetric dihydroxylation, his first Prize work to click chemistry and propose his philosophy of "click", which led to extensive skepticism about its novelty and necessity at the beginning? To address the features of his research strategies, we investigate Sharpless' philosophy of click chemistry by using computational methods to measure the cognitive content and concepts of Sharpless' research and estimate the knowledge flow of his click chemistry to other fields. In our content-centered approach, we focus on the evolution of scientific thoughts and research strategies along with the expansion of interdisciplinary sciences.

We traced the conceptual journey of Sharpless over time through topic modeling [12], terms map and clustering technique [13] from distant reading his full-text publications in the five time periods from 1963 to 2020 with a working corpus containing 79,653 different lemmatized words and 792,074 word occurrences (Fig. S1 in Supporting information). VOSviewer and Citespace [14] were used to draw the term maps and the epistemic network in collaboration with four datasets we collected and processed. As these approaches are data-driven and unsupervised [12], they offer a quantitative and bottom-up view.

* Corresponding author.
E-mail address: xfjiang@chem.ecnu.edu.cn (X. Jiang).

https://doi.org/10.1016/j.cclet.2022.107936
1001-8417/

When comparing the topic coverage and relative frequency, the first result of topic model analysis is the importance of two most significant topics, concerning reaction mechanisms and research strategies or approaches with a combined occurrence weight of 36.7% in the corpus (Table S1 in Supporting information). Our analysis indicates that Sharpless put more weight on the methodology to explore new regions on reactivity rather than new reactions. The detailed topics in the category strategy show its diversity, covering reagent control strategy, ligand acceleration, slow addition, ligand-independent catalytic approach, on water strategy, and click chemistry. This is consistent with the findings of previous work that prize winners exhibit more diverse strategies [3]. The diachronic evolution of the topics (Table S2 in Supporting information) reveals that Sharpless' conceptual journey moved from experimental procedures, mechanisms to synthesis strategies, features of reactions, functions and applications. The fluctuating significance of these topics indicates that reaction mechanism concerning selectivity and reactivity is the most stable cluster and appears in all periods. Specifically, the pattern of research strategies, which changed over time but centered on the concept of simplicity, is important. The most prevalent terms of regioselectivity and enantioselectivity show that Sharpless moved to achieve selectivity through chemical synthesis rather than biosynthesis. Reagents and solvents that fall into the click chemistry category in the fourth period become simpler. Together, the research strategies orienting reactivity and then simplicity connect all other categories of topics. The evolution of Sharpless' focus on catalysis (Fig. S2 in Supporting information), which are the engine driving Sharpless' research [15], further illustrates how strategies are centered on simplicity. The privileged reactivity led to his involvement with metal and ligand-accelerated catalysis in his early career. Following more effective but simpler catalysts invented, especially the new ligand-independent catalysis process became simpler, the methods to fulfill his concept of reactivity and then simplicity to forge more functional molecules in mild conditions became available in experiments. Comparing the details of metal-free catalysts in the 1960s and 2010s further reflects the reaction conditions changing from extreme to mild.

Furthermore, findings from link analysis for Sharpless' co-occurrence terms map network emphasize connectivity and modulate function are the unique features of click chemistry. Five clusters formed by the strongest links (Fig. 1a) reflect the same conceptual journey trends revealed in the topic model analysis. When the number of strongest links increases to 200, four clusters link together by four terms (Fig. 1b). A common characteristic of these terms is heteroatoms, in which Sharpless gets the appealing modulate functional groups as shown by terms with the strongest links (Fig. 1a) connecting yellow and sky blue clusters. The term selectivity connects with protein and nAChR. These links demonstrate that click chemistry is a strategic method for the efficient combination of small modules together through heteroatom links rather than new types of reactions. Fig. 1c shows a striking result. The isolated green cluster finally connects with the other four until the term water appears. Water connects all five clusters because water is the simplest solvent, and water-tolerant reactions can employ functional groups that have biocompatibility or can easily have wide applications. This shows that the simplicity and connectivity help function address diverse applications and needs. In other words, click chemistry is beyond a connective technology and places equal emphasis on the structure and function of the diverse array of linkers [16].

To further explore how the research strategies emerged and evolved, we integrate the cognitive content into the patterns of co-authorship networks through social network analysis. The limited but continuous cognitive contents during different periods (Fig. 2), such as amino alcohols and inhibitors, show Sharpless' focus on

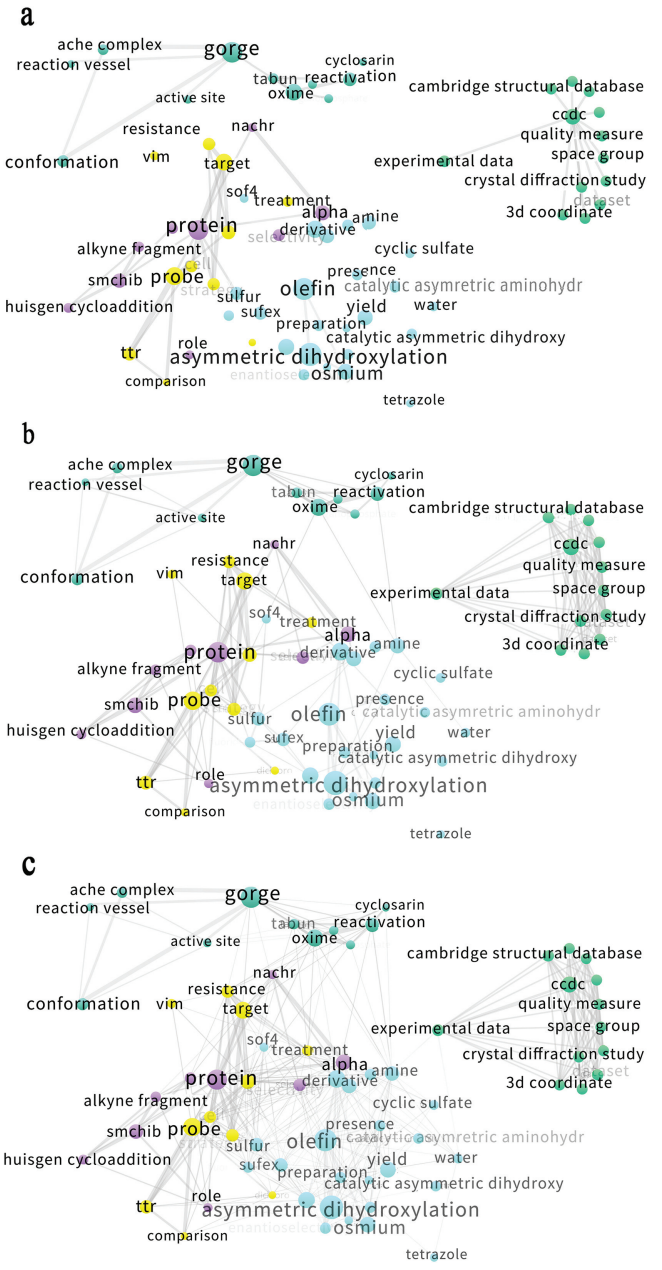

Fig. 1. Sharpless' term co-occurrence network showing the 50 (a), 200 (b), 400 (c) strongest links. The size of the nodes represent the probability of the terms. The thickness and the shade of the lines represent the co-occurrence frequencies of terms. (a) Five clusters were formed regarding reactivity and stereoselectivity (turquoise), experimental procedure (green), features of click chemistry (yellow), types of reactions (sky blue), and application of click chemistry in medicine (purple). (b) Turquoise and sky blue clusters link together through two links by amine, oxime and SOF_4. The latter also connects yellow and sky blue clusters, together with a stronger link *via* sulfur. Sky blue and yellow clusters are connected by terms related to modulate functional groups, such as protein, probe, target, strategy, cell. (c) The terms water and amine connects all five clusters.

functions instead of the diversity of molecules. Initially, our results, on its face, may seem to be in tension with earlier work that new discoveries in biomedical chemistry always link chemicals never linked before [3]. However, a tie between two chemicals may be novel and previously linked chemicals are now linked in a new way. Sharpless and co-authors developed simple and small chemicals (amino alcohols) with the most useful functional groups as ligands to selectively bind complex molecules of biological significance (inhibitor, infectious diseases, bacterial resistance) in a new

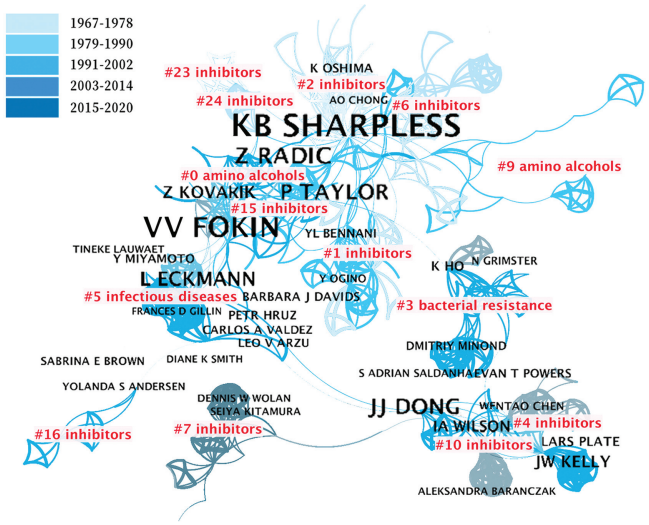

Fig. 2. Sharpless' epistemic network map in collaboration and its evolution. The colors represent the time periods. The names of the coauthors appear as the nodes, in which the size of the nodes represents the collaboration strength. The degree threshold with collaboration times is 12. The weights of links represent the strength of the research subjects between coauthors. The order of the numbers before the research subjects represents the order of the size of research subjects through spectral clustering.

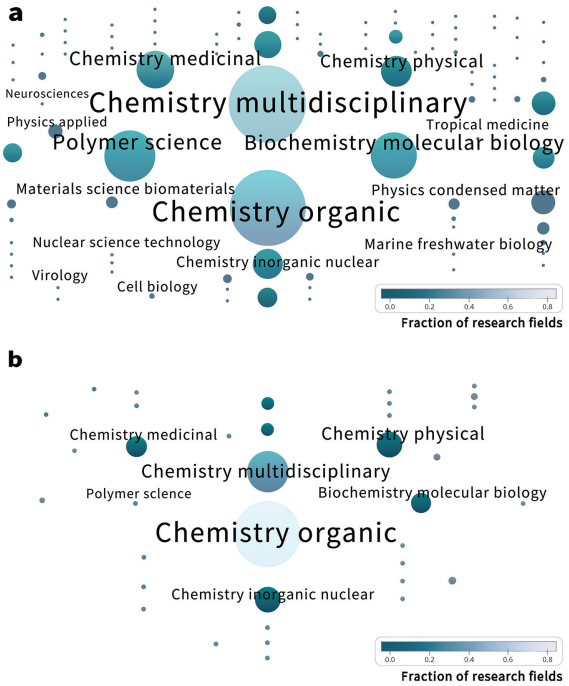

Fig. 3. The research fields distribution map of citing publications. Each node represents one research field. The full names of research fields are shown in Fig. S6 (Supporting information). (a) Publications that cited three representative papers in click chemistry are scattered in 103 research areas. Chemistry multidisciplinary is the largest field (41.02%), while the fraction of chemistry organic is 38.34%. Polymer science (17.67%), chemical medicinal (11.7%), biomedical research methods (3.83%), and nanotechnology (3.14%) are the top four new emerging fields. (b) Publications that cited Nobel Prize paper on asymmetric dihydroxylation scatter in 37 research areas. Chemistry organic is the largest field (67.14%), while the fraction of chemistry multidisciplinary is 23.89%.

way through an array of reactions from asymmetric synthesis to the connective technology in click chemistry. Moreover, his interest was rooted in his early work on amino alcohols in bioactive natural product synthesis, and it is highly disciplinary. When considering the co-authors, it seems a highly heterogeneous group in collaborative work (Fig. S3 in Supporting information). However, a few core co-authors for connecting different clusters show their importance in subjects' jump and connections with new research strategies. The continuing efforts of the Sharpless Lab to fully understand reaction mechanisms helped to lay the foundation for searching useful new reactivity and general methods. These core co-authors with high density (Fig. 2) have constantly collaborated with Sharpless before and after his reception of the first Nobel Prize.

We explore the breadth of impact of click chemistry and underlying beliefs through citation and link analysis on citing publications. It is astonishing that the citing publications on click chemistry are scattered much more fields beyond organic chemistry than publications on Sharpless' first Nobel Prize work (Fig. 3). The details of research field distribution demonstrate that citing publications on click chemistry are seen not only all aspects of drug discovery anticipated by Sharpless, materials science, but also most of fields in natural sciences and social sciences (Fig. 3a). Click chemistry had a much greater impact on a set of completely different disciplines, such as physics condensed matter, biophysics, computer science information systems, quantum science technology in natural sciences, and education, history, and philosophy of science in social sciences (Table S3 in Supporting information). The fluctuation of fractions of citations shows the increasing influence of click chemistry on more research areas (Fig. S5 in Supporting information). These results highlight a diversity of research fields that are inspired by click chemistry, which is much richer than what is generally acknowledged. Our exploration of the underlying beliefs on citing from reading the contents of these publications reveals that the methods, techniques, concepts, and insights of click chemistry inspire many other research fields.

A strong awareness of the patterns in the co-occurrence terms map formed by link analysis on citing publications shows that click chemistry not only allows particular new materials or drugs to be prepared based on modularity and efficiency [17], it clicks nearly all research fields with its philosophy valuing simplicity, connectivity and function (Fig. S4 in Supporting information). The citing publications clustered the same five parts as Sharpless' work. The analysis of the changes on catalysis (Fig. S2) highlights specific trend that enzymes appear on the two sides of his lifespan research. This implies a big turn from biology-centered to chemistry-centered research in medicine and life sciences. More specifically, click chemistry draws solutions to significant and practical questions back to chemistry from biology, echoing with the development trend of organic chemistry and biochemistry [18]. Click chemistry is, therefore, a prime example of a disciplinary discovery that has a significant impact on several disciplines.

In summary, with combining the theories of philosophy of science into data-driven technologies, we find that the concept of Sharpless' click chemistry derives from his continuous efforts on reactivity, which aims to simplicity, connectivity and practicality. Taking nature's cue for inspiration, Sharpless specializes in the seemingly useless carbon-heteroatom links, which nature uses sparingly, to develop new transformations satisfying the criteria of a click reaction. This strategy follows the same way that values nature's principle. Sharpless' continuous efforts on reaction mechanisms, which are unique features of chemistry and deserve a renaissance [19], may make the broad impact of click chemistry possible, because of the richer knowledge about a reaction pathway centered on chemical bonds [20,21], the higher the chances of connectivity control. Chemistry has a clear-cut epistemic domain when it is on the way to model nature. The success of click chemistry inspires us to rethink the value of chemistry. While interdisciplinarity now spans many fields across the natural and life sciences, the hard core in chemistry and the independence of chemistry should not be ignored.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Research Funds of Happiness Flower ECNU (No. 2019ECNU-XFZH006), the National Social Science Fund of China (No. 19BZX041), the National Natural Science Foundation of China (Nos. 22125103 and 21971065) and the Science Technology Commission Shanghai Municipality (Nos. 22JC1401000, 20JC1416800 and 20XD1421500).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cclet.2022.107936.

References

[1] A. Clauset, D.B. Larremore, R. Sinatra, Science 355 (2017) 477–480.
[2] S. Fortunato, C.T. Bergstrom, K. Borner, et al., Science 359 (2018) eaao0185.
[3] A. Rzhetsky, J.G. Foster, I.T. Foster, et al., Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 14569–14574.
[4] S. Tong, P. Ahlgren, Scientometrics 112 (2017) 75–90.
[5] H.F. Chan, A.S. Onder, B. Torgler, Scientometrics 105 (2015) 2215–2235.
[6] A.J. Gates, Q. Ke, O. Varol, et al., Nature 575 (2019) 32–34.
[7] M. Szell, Y. Ma, R. Sinatra, Nat. Phys. 14 (2018) 1075–1078.
[8] J. Murdock, C. Allen, S. DeDeo, Cognition 159 (2017) 117–126.
[9] S. Milojevic, J. Informetr. 9 (2015) 962–973.
[10] Y.X. Liu, E. Biskup, Y.Q. Wang, et al., Scientometrics 125 (2020) 1213–1228.
[11] R.F. Service, Science 320 (2008) 868–869.
[12] D.M. Blei, Commun. ACM 55 (2012) 77–84.
[13] L. Waltman, N.J. van Eck, E.C.M. Noyons, J. Informetr. 4 (2010) 629–635.
[14] C.M. Chen, J. Am. Soc. Inf. Sci. Technol. 57 (2006) 359–377.
[15] K.B. Sharpless, Tetrahedron 50 (1994) 4235–4258.
[16] C.J. Smedley, G.C. Li, A.S. Barrow, et al., Angew Chem. Int. Ed. 59 (2020) 12460–12469.
[17] C.J. Hawker, V.V. Fokin, M.G. Finn, et al., Aust. J. Chem. 60 (2007) 381–383.
[18] E.I. Negishi, Angew. Chem. Int. Ed. 50 (2011) 6738–6764.
[19] J.I. Seeman, G. Restrepo, Angew. Chem. Int. Ed. 59 (2020) 12250–12266.
[20] H. Chang, Ambix 64 (2017) 360–374.
[21] M. Weisberg, Philos. Sci. 75 (2008) 932–946.