Computer-assisted planning of hydroxychloroquine's syntheses commencing from inexpensive substrates and bypassing patented routes.

Sara Szymkuć[1+], Ewa P. Gajewska[1+], Karol Molga[1+], Agnieszka Wołos[1], Rafał Roszak[1], Wiktor Beker[1], Martyna Moskal[1], Piotr Dittwald[1] & Bartosz A. Grzybowski[1,2,3]*

[1] Institute of Organic Chemistry, Polish Academy of Sciences, ul. Kasprzaka 44/52, Warsaw 02-224, Poland

[2] IBS Center for Soft and Living Matter and

[3] Department of Chemistry, UNIST, 50, UNIST-gil, Eonyang-eup, Ulju-gun, Ulsan, 689-798,

[+]Authors contributed equally

*Correspondence to: nanogrzybowski@gmail.com

Abstract: A computer program for retrosynthetic planning helps develop multiple "synthetic contingency" plans for hydroxychloroquine, a promising but yet unproven medication against COVID-19. These plans are designed to navigate, as much as possible, around known and patented routes and to commence from inexpensive and diverse starting materials, such as to ensure supply in case of anticipated market shortages of the commonly used substrates. Looking beyond current COVID-19 pandemics, development of similar contingency syntheses is advocated for other already-approved medications, in case such medications become urgently needed in mass quantities to face other public-health emergencies.

Keywords: COVID-19 pandemic, hydroxychloroquine, economically feasible syntheses, computer-assisted retrosynthesis.

Faced with the eruption of the coronavirus pandemic, individual academic and clinical laboratories, funding agencies, and entire governments are intensifying efforts to develop and deploy safe and effective vaccines and/or antiviral medications. Whereas vaccines may become available within a year or so, development and approval of a brand new drug will, most likely, require a significantly longer time, not relevant to the current exigency. Accordingly, much of the ongoing effort has been focused on drugs that are already approved and could be re-purposed against COVID-19. In particular, first reports have been emerging in the scientific literature[1,2] that chloroquine (CQ) and hydroxychloroquine (HCQ) – vintage drugs to treat malaria as well as some autoimmune diseases – efficiently inhibit SARS-CoV-2 infection in vitro by slowing down entry of viruses into the cell and by blocking their transport from early endosomes to endolysosomes[2,3], causing noticeable enlargement of the former and affecting the pH levels[4] within the endolysosomal tract. Since HCQ is less toxic than CQ[5,6] and given current lack of viable alternatives, the use of this relatively safe drug against the COVID-19 pandemic appears imminent, even in the absence of comprehensive clinical data – in fact, Novartis has just announced[7] that it intends to donate for experimental treatments up to 130 million 200 mg doses by the end of May, 2020, including all of its current stock[7]. Still, should HCQ prove effective, the demand might soon surpass supply. Moreover, the key synthetic methods leading to HCQ are very often protected by patents, including some very recent ones (see **Figure 1**), and we cannot exclude the possibility that monetary, corporate interests would interfere with humanitarian inspirations. In addition, the failure of the worldwide logistics and supply chains that accompanies COVID-19 pandemic might render some key substrates temporarily unavailable, in effect delaying execution of the proven synthetic routes and calling for alternative synthetic solutions. Anticipating such complications, we harnessed the power of

Chematica[8-16] – an experimentally-tested[9,10] platform for computer-assisted retrosynthesis of both known and unknown target molecules – to design syntheses of HCQ that would (1) commence from various inexpensive and popular starting materials (so that the syntheses minimize the abovementioned supply problems); (2) circumvent patented methodologies whenever possible[16]; and (3) minimize the use of expensive methodologies and/or reagents. In the following, we briefly outline the computational methods underlying *Chematica*'s retrosynthetic searches, summarize the known syntheses of HCQ, and then describe novel ones identified by *Chematica* to meet conditions (1)-(3). We hope that at least some of these syntheses can become useful in streamlining economically feasible and widely accessible production of HCQ. We remain open to performing – on a *pro bono* basis – similar synthetic analyses for organizations considering production and unrestricted (both geographically and economically) distribution of other potential anti-COVID-19 agents, should such agents become available in the near future.

Chematica is a sophisticated platform for fully automated design of pathways leading to arbitrary (i.e., both known and new) targets. The software combines elements of network theory[16,17] with an expert knowledge-base of synthetic transformations as well as multiple reaction-evaluation routines (based on machine learning,[11,12] quantum mechanics,[8,9] and molecular dynamics[9,13]) to search over vast trees of synthetic possibilities. The reaction transforms (currently, ~ 100,000) are expert-coded based on the underlying reaction mechanisms and are broader than any specific literature precedents (for comparison with machine extraction of rules from reaction repositories, see [13]). Each rule specifies the scope of admissible substituents, accounts for stereo- and regiochemistry requirements, recognizes groups that must be protected under given reaction conditions, and identifies functionalities that are outright incompatible. The searches are guided by combinations of functions (either heuristic[8,9] or best-in-class AI-based[12]) that score both synthetic

positions as well as costs of individual reactions. The pathways identified by the program terminate in either commercially available chemicals (here, more than 200,000 molecules from Sigma-Aldrich catalogs, each with price per unit quantity; also see below for price re-scaling) or those already known in the literature (ca. 6 million substances, each accompanied by a measure of synthetic popularity[8,16], i.e., how many times a given substance was used in prior syntheses). Since the program typically identifies a large number of possible routes, the network of viable syntheses already found is queried by dynamic linear programming algorithms to select pathways with the lowest cost (propagated recursively from substrates to products with the consideration of estimated yields), and that offer diverse retrosynthetic strategies.[14] In setting up a particular search, the user can specify parameters influencing the economy of the solutions, notably, the upper price threshold and/or the minimal synthetic popularity of the starting materials, the relative cost of performing a reaction operation, or the desired estimated yield. The user can also eliminate certain types of transformations or unwanted reagents (e.g., expensive catalysts). He/she is also able to "lock" certain bonds or fragments in the target such that they are not disconnected along the synthetic plan – as described in detail in [15], this functionality is useful in navigating around patented routes. Depending on the number of imposed constraints, a typical search for a drug-like molecule takes from few to tens of minutes and within this time inspects tens to hundreds of thousands of reaction candidates. Ultimately, a user-specified number of top-scoring pathways (typically 50-100) are returned and displayed as bipartite graphs with nodes that are expandable to display molecular structures, suggested reaction conditions typical to a given reaction class, and more.[8,9]

The results described in the following come from various searches executed by our team over the course of two days and using three machines, each with 64 cores. Multiple searches were performed on the newest version of the program (not yet transitioned onto the commercial

Synthia™ platform owned and distributed by Sigma-Aldrich/Merck) with various parameters to reflect different economic scenarios of the desired syntheses and with different types of abovementioned constraints. In all, these searches considered on the order of millions of potential intermediates and synthetic plans. The common feature of the searches was the desire to offer alternatives to existing syntheses and to suggest multiple synthetic plans using diverse but inexpensive starting materials. In considering the prices of the starting materials, we naturally realized that catalog prices from a specialty-chemicals retailer such as Sigma-Aldrich, S-A, are significantly higher than from whole-sale producers. Still, substrates inexpensive in S-A are even less inexpensive from larger-scale suppliers, as evidenced by the correlation shown in **Figure 2** and spanning substrates of the new syntheses of HCQ we identified. We will discuss these issues in more detail along with specific routes.

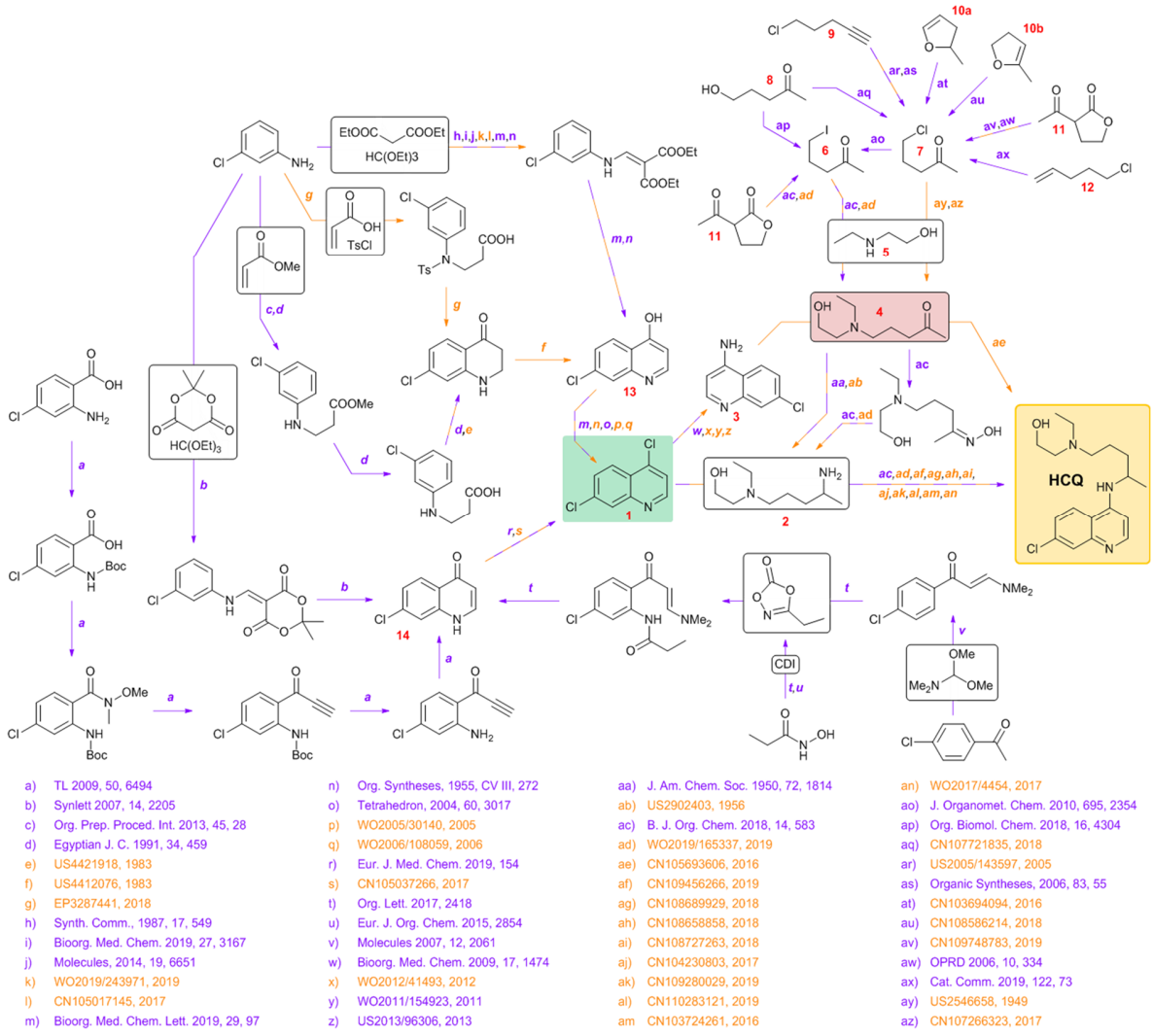

Figure 1. Synthetic network summarizing known syntheses of hydroxychloroquine, HCQ, along with the pertinent literature (patents in orange, scientific publications in violet). The two key intermediates are highlighted by colored boxes: in terms of availability and price, **1** (*green*) is significantly less problematic than **4** (*red*).

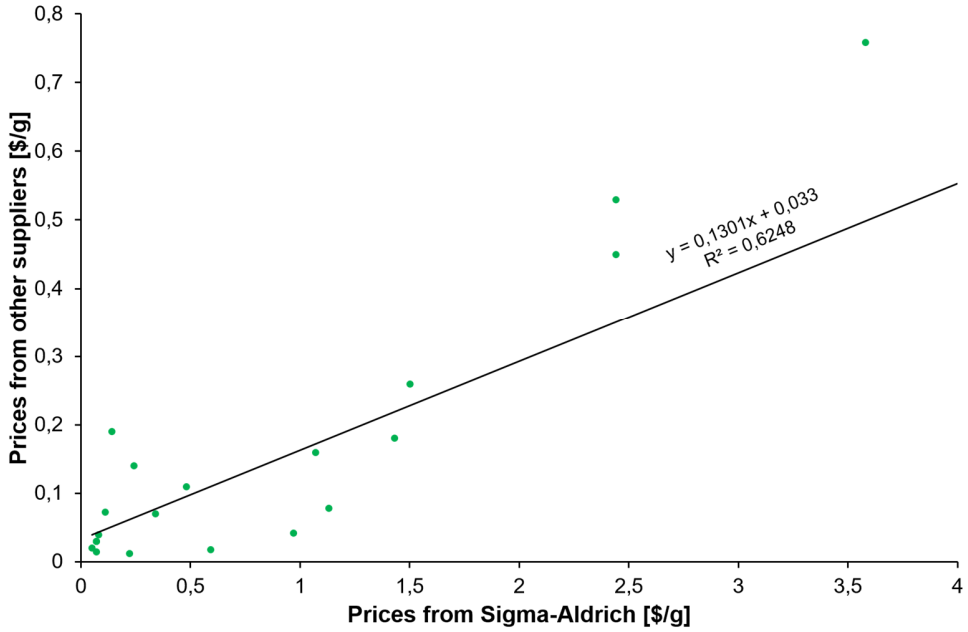

Figure 2. Correlation between prices (scaled per gram) of the substrates of new syntheses of HCQ (see **Figure 3**) from (*x-axis*) Sigma-Aldrich' catalog interfaced with *Chematica* and (*y-axis*) the least inexpensive options we were able to identify from larger-scale manufacturers. Not surprisingly, the latter are, on average, twelve times less expensive (median = 4.67).

To begin with, we surveyed the available literature to construct a synthetic network summarizing currently known syntheses of HCQ (**Figure 1**). Somewhat remarkably, although HCQ has been off-patent for decades, a large proportion of methods involved have been patented, sometimes quite recently, substantiating our concern of potential IP complications in case of emergency production by independent agents. These solutions hinge at the late stage attachment of the side chain performed either via (i) nucleophilic aromatic substitution of dichloroquinoline **1** and amine **2**, or (ii) reductive amination of aminochloroquinoline **3** (itself derived from **1**) and ketone **4**, the latter being the starting material for the preparation of **2**. The two "hubs" of the

network are, obviously, **1** or **4** though they are quite different from the economic and logistic points of view. The heterocyclic part of HCQ, **1**, is rather inexpensive (1.50 $/g from S-A, 0.26 $/g from Biosynth Carbosynth) and in case of supply problems, can be sourced (in 94% yield, via chlorination using $POCl_3$) from hydroxychloroquinoline **13** which, in turn, can be made in ~ 40% yields in two steps either from 3-chloroaniline, diethyl malonate and ethyl orthoformate (respectively, 9.51 $/g from S-A, 0.05 $/g from Oakwood Chemical, OC; 0.04 $/g S-A, 0.015 $/g OC; 0.12 $/g S-A, 0.03 $/g OC) or from 3-chloroaniline, acrylic acid (or methyl acrylate) and tosyl chloride (respectively, 9.51 $/g S-A, 0.05 $/g OC; 1.48 $/g S-A, 0.02$/g from Gelest Inc.; 1.43 $/g S-A, 0.03 $/g from Alfa Aesar ; 0.02 $/g S-A, 0.04 $/g from Alfa Aesar). Alternatively, **1** can be obtained from chloroquinolinone **14**, available via a similar two-step sequence starting from 3-chloroaniline, Meldrum's acid (1.75 $/g S-A, 0.07 $/g from AbaChemScene) and ethyl orthoformate. Some more recent approaches for the preparation of **14** hinge on different starting materials (4-chloroacetophenone or 2-amino-4-chlorobenzoic acid) but require at least four steps. Additionally, C-H activation of enaminone derived from chloroacetophenone requires expensive bimetallic catalyst ($Cp^*Co(CO)I_2/AgSbF_6$).

In contrast, ketone **4** is not easily sourced (no prices listed on e-molecules) and is likely the production bottleneck. This intermediate can be prepared via alkylation of aminoalcohol **5** (0.11 $/g SA, 0.04 $/g from Arcos Organics) with haloketones **6**/**7**, which in turn can be derived from hydroxyketone **8**, chloroalkyne **9**, enol ethers **10a**/**10b**, lactone **11** or chloroalkene **12**. These substrates, except from **8** and **11** (both available for less than 0.5 $/g from suppliers like Combi-Blocks or ChemScene) are relatively expensive (from 4 $/g to even 585 $/g) so these methodologies are probably unsuitable for industrial up-scaling.

Without any search constraints, *Chematica* generally identified many of these known solutions (or their very close analogs, differing in insignificant details). The program began to find substantially different pathways especially upon application of restrictive thresholds for the prices/popularities of the starting materials. **Figure 3** summarizes 17 routes we found most economically viable and concise (see Supplementary Information for enlarged views). In addition to routes relying on nucleophilic aromatic substitution of dichloroquinoline and reductive amination of aminochloroquinoline, the software was able to avoid these steps, replacing them with methodologies such as A3-coupling (path **15**), Cu-cat. coupling between heteroaryl iodide and amine (paths **2** and **3**), three-component reaction between amine, aldehyde and halide under Barbier-type conditions (path **14**), or alkylation of aromatic amine with alkyl iodide (path **11**). Other innovative aspects of *Chematica*'s plans are manifest in the routes to prepare the side-chain of the HCQ which, as we saw before, is the major factor driving availability/cost of the overall synthesis. The machine's proposals include, for example, opening of a lactam with Grignard reagent (paths **1** and **16**), or alkylation of a lactone followed by ring-opening to install a primary iodide functionality – which is a very convenient group for subsequent alkylation (path **13**). Other interesting approaches use multicomponent Mannich reaction. In pathway **2** this reaction is combined with subsequent Henry reaction, and in pathway **10** it follows a Curtius rearrangement. Both Henry reaction and Curtius rearrangement are interesting alternatives to reductive amination or reduction of oxime used for the introduction of the nitrogen atom. As already mentioned, all of these proposed routes avoid expensive catalysts and commence from inexpensive starting materials, readily available in large quantities (e.g., ethylamine at 0.018 $/g, 2-bromoethyl acetate at 0.22 $/g, 5-chloro-2-pentanone at 0.08 $/g, or ethanolamine at 0.012 $/g). Only few of these substrates were used in the previously published/patented syntheses. In **Figure 3**, their prices are indicated in red font.

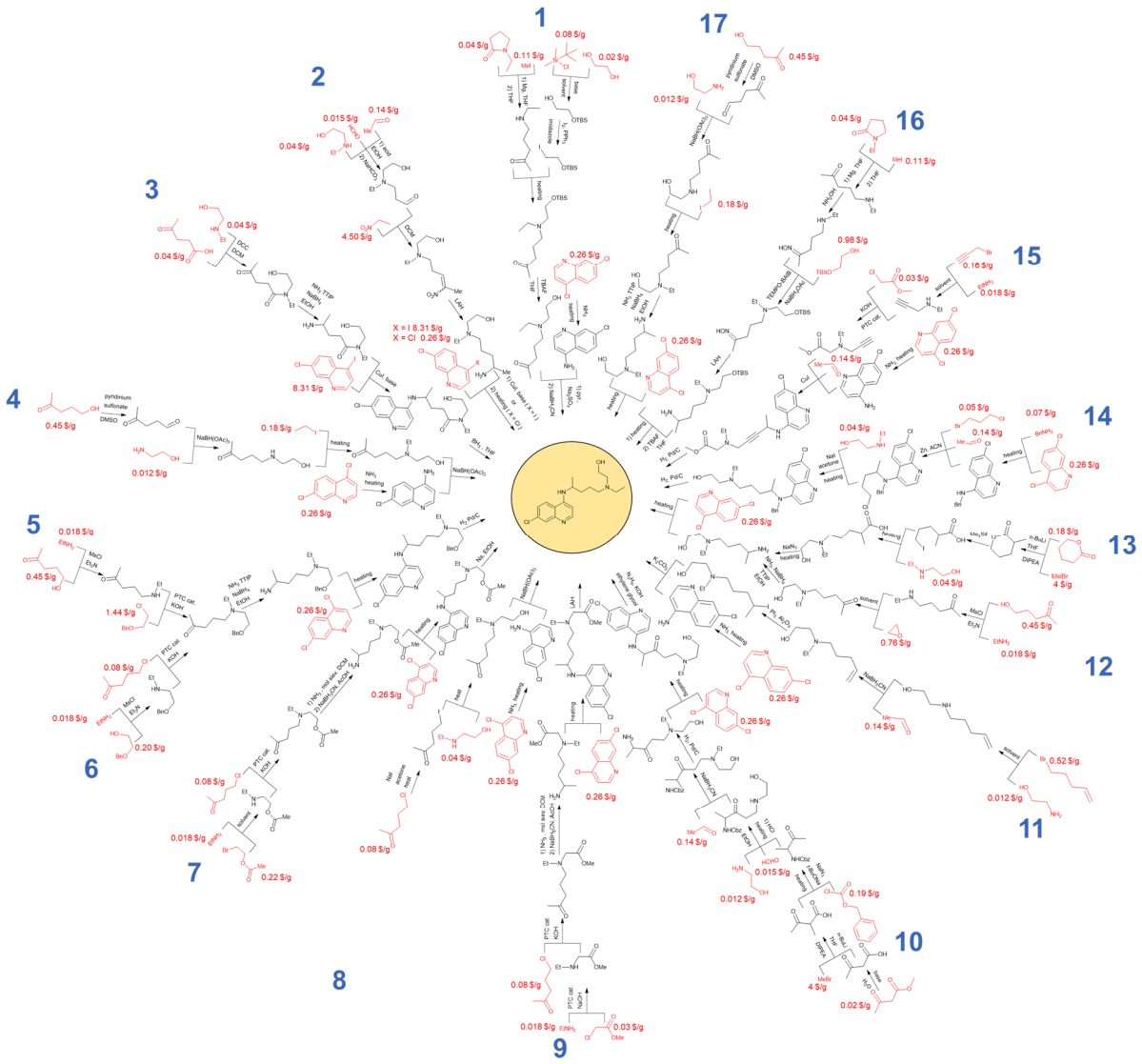

Figure 3. Novel syntheses of hydroxychloroquine (HCQ) designed automatically by *Chematica*. Substrates and their prices (the lowest ones we were able to identify) scaled to $/g are colored in red.

In summary, we capitalized on the speed and chemical accuracy of modern computer-assisted synthetic planning to develop alternative and economical "contingency" plans for the synthesis of HCQ. Although these syntheses could, without doubt, be also identified by human experts alone, tracing them to inexpensive substrates while minimizing the use of previously-

described methodologies might be a rather tedious and time-consuming enterprise, incompatible with the COVID-19 emergency at hand. In a broader context, this exercise made us realize that the current system of chemical/pharmaceutical production is heavily reliant on efficient but far-and-between methods – while this approach works at "peacetime," it might be very vulnerable to the disruption of global supply chains of key starting materials, effectively leaving us without alternative means of production. Consequently, we advocate development of contingency plans for all other approved drugs in case they are needed in large quantities on a short call. It seems to us it is time to transition the planning of national/global chemical production of key therapeutics from Napoleonic improvisation ("*I have never had a plan of operations*") to von Moltke's far-sighted calculation ("*Strategy is a system of expedients*").

Author contributions

K.M., E.P.G, S.S., P.D., W.B., M.M. have been the key developers of *Chematica*. K.M., E.P.G, S.S. performed the synthetic analyses described in the paper. A.W. and R.R. helped with pricing and synthetic data. B.A.G. conceived *Chematica* in graduate school and has directed its development ever since. All authors contributed to the writing of the manuscript.

Acknowledgements

This work was performed on a volunteer basis in authors' free time (over the weekend of March 21/22, 2020). This being said, we are grateful to the U.S. DARPA who, for many ears, sustained development of Chematica under the "Make-It" Award, 69461-CH-DRP #W911NF1610384. We also thank MilliporeSigma/Merck KGaA for the use of their computer resources. B.A.G. also gratefully acknowledges personal support from the Institute for Basic Science Korea, Project Code

IBS-R020-D1. A.W. gratefully acknowledges personal support from the National Science Center, NCN, Poland under the Symfonia Award (#2014/12/W/ST5/00592).

Conflict of interest

The authors declare no financial interest in this work. While Chematica was originally developed and owned by B.A.G.'s Grzybowski Scientific Inventions, LLC, neither he nor the co-authors currently hold any stock in this company, which is now property of Merck KGaA, Darmstadt, Germany. Most of the authors have, until recently, collaborated with Merck KGaA, Darmstadt, on Chematica's development within the DARPA "Make-It" award. All queries about access options to Chematica (now rebranded as Synthia^{TM}), including academic collaborations, should be directed to Dr. Sarah Trice at sarah.trice@sial.com.

References:

1. Wang, M. et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* **30**, 269–271 (2020).

2. J. Liu, R. Cao, M. Xu, Xi Wang, H. Zhuang, H. Hu, Y. Li, Z. Hu., W. Zhong, M. Wang, Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro. *Cell Discov.* **6**, 16 (2020). https://doi.org/10.1038/s41421-020-0156-0

3. Mauthe, M. et al. Chloroquine inhibits autophagic flux by decreasing autophagosome-lysosome fusion. *Autophagy* 14, 1435–1455 (2018).

4. Borkowska, M., Siek, M., Kolygina, D. et al. Targeted crystallization of mixed-charge nanoparticles in lysosomes induces selective death of cancer cells. *Nat. Nanotechnol.* (2020). https://doi.org/10.1038/s41565-020-0643-3

5. Weniger, H. Review of side effects and toxicity of chloroquine. *Bull. World Health* **79**, 906 (1979).

6. McChesney, E. W. Animal toxicity and pharmacokinetics of hydroxy-chloroquine sulfate. *Am. J. Med.* **75**, 11–18 (1983).

7. https://www.novartis.com/news/media-releases/novartis-commits-donate-130-million-doses-hydroxychloroquine-support-global-covid-19-pandemic-response

8. Szymkuć, S. *et al.* Computer-assisted synthetic planning: The end of the beginning. *Angew. Chem. Int. Ed.* **55**, 5904–5937 (2016).

9. Klucznik, T. *et al.* Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory. *Chem* **4**, 522–532 (2018).

10. Gajewska, E.P., Szymkuc, S., Dittwald, P., Startek, M., Popik, O., Mlynarski, J., Grzybowski, B.A. Algorithmic discovery of tactical combinations for advanced organic syntheses. *Chem* **6**, 280–293 (2020).

11. Beker, W., Gajewska, E.P., Badowski, T., Grzybowski, B.A. Prediction of major regio-, site-, and diastereoisomers in Diels-Alder reactions by using machine-learning: The importance of physically meaningful descriptors *Angew. Chem. Int. Ed.* **58**, 4515-4519 (2019).

12. Badowski, T., Gajewska, E.P., Molga, K., Grzybowski, B.A. Synergy between expert and machine-learning approaches allows for improved retrosynthetic planning. *Angew. Chem. Int. Ed.* **59**, 725-730 (2020).

13. Molga, K., Gajewska, E.P., Szymkuc, S., Grzybowski, B.A. The logic of translating chemical knowledge into machine-processable forms: a modern playground for physical-organic chemistry. *React. Chem. Eng.* **4**, 1506-1521 (2019).

14. Badowski, T., Molga, K., Grzybowski, B.A. Selection of cost-effective yet chemically diverse pathways from the networks of computer-generated retrosynthetic plans. *Chem. Sci.* **10**, 4640-4651 (2019).

15. Molga, K., Dittwald, P., Grzybowski, B.A. Navigating around patented routes by preserving specific motifs along computer-planned retrosynthetic pathways. *Chem* **5**, 460–473 (2019).

16. Grzybowski, B. A., Bishop, K. J. M., Kowalczyk, B. & Wilmer, C. E. The 'wired' universe of organic chemistry. *Nat. Chem.* **1**, 31–36 (2009).

17. Kowalik, M. *et al.* Parallel optimization of synthetic pathways within the Network of Organic Chemistry. *Angew. Chem. Int. Ed.* **51**, 7928–7932 (2012).