

README:

1. The pseudocode describing main components of Chematica (execution of one retrosynthetic generation, evaluation of individual reactions, selection and expansion of subsequent retrosynthetic generations, construction and retrieval of complete synthetic pathways) is provided in file PSEUDOCODE_Aug2.pdf.
2. Example of one of 100,000+ reaction rules underlying Chematica is provided in file RULE.pdf
3. Chematica is virtualized using Docker containers and can be run on any multicore Linux server. For our experiments we used a Debian GNU/Linux 9.4 server with 62 Intel Core Processors with 2.2 GHz clock speed.
4. The program is owned and operated by Merck, KGaA who is managing all installation and maintenance on its internal servers. In the commercial version of Chematica, called Synthia™, licensed users log into Synthia's WebApp environment (akin to applications such as Reaxys or Sci Finder).

Merck KGaA provides webinars and online demos to organizations interested in licensing Chematica/Synthia. The Company distributes both commercial and academic licenses. For further information, please see

https://www.sigmaaldrich.com/chemistry/chemical-synthesis/synthesis-software.html?clid=Cj0KCQjwgo_5BRDuARIsADDEntS-8zJ6CK-gnb8HTjcaiE7pS4wCK7nL22prDc5g-yDtd12T7xi7QukaAqyUEALw_wcB

5. During use, the program takes as input a target molecule (either drawn in Chematica's structure editor or provided as SMILES) and constructs viable synthetic pathways leading to these targets. Isotopically-labelled targets are allowed (for details, see *Chem. Sci.* **10**, 9219–9232, 2019). Before search commences, the user specifies the conditions for the stop points (e.g., should the starting materials be commercially available and what are their highest allowable MW and price (in \$/g)? Are compounds with already-reported syntheses also allowed as stop-points? If so, what should be their maximal MW and/or synthetic popularity in the literature?). A default scoring function is typically applied (as in the examples described in the main text) although the user can also use advanced options to define his/her scoring own function. Optionally, the user can define multiple targets and search for the synthesis of an entire target library (*Chem. Sci.* **10**, 9219–9232, 2019). As another option, the user can specify which types of reactions or structural fragments to avoid during searches (this option is illustrated, for example, in Section S3.11 where alternative syntheses of Tacamonidine are constructed avoiding Pictet-Spengler cyclization; for algorithmic details of the functionality, see *Chem* **5**, 460–473, 2019).

The search is then initialized. As the program finds viable syntheses, it ranks them according to selection algorithms described in detail in *Chem. Sci.* **10**, 4640–4651, 2019. Pathways are returned and ranked (best-to-worse) and typically 50 top-scoring routes are displayed in the graph format as in **Figures 3-5** in the main text. Substance and reaction nodes within each pathway can be expanded to shown molecules' and reactions' details. As the algorithm finds more pathways, the

top-50 list is refreshed. Individual or multiple selected pathways from the list can be saved in graphml or pdf formats (the latter displaying all reaction details).

For medicinally-relevant small molecules, multiple viable pathways are typically found within minutes; for complex natural products such as those described in the main text, the searches take few hours to overnight.