Article

Computer-designed repurposing of chemical wastes into drugs

https://doi.org/10.1038/s41586-022-04503-9

Received: 26 June 2020

Accepted: 3 February 2022

Published online: 27 April 2022

Check for updates

Agnieszka Wołos[1,2], Dominik Koszelewski[2], Rafał Roszak[1], Sara Szymkuć[1], Martyna Moskal[1], Ryszard Ostaszewski[2], Brenden T. Herrera[3], Josef M. Maier[3], Gordon Brezicki[3], Jonathon Samuel[3], Justin A. M. Lummiss[3], D. Tyler McQuade[3], Luke Rogers[3] & Bartosz A. Grzybowski[1,2,4,5]✉

As the chemical industry continues to produce considerable quantities of waste chemicals[1,2], it is essential to devise 'circular chemistry'[3–8] schemes to productively back-convert at least a portion of these unwanted materials into useful products. Despite substantial progress in the degradation of some classes of harmful chemicals[9], work on 'closing the circle'—transforming waste substrates into valuable products—remains fragmented and focused on well known areas[10–15]. Comprehensive analyses of which valuable products are synthesizable from diverse chemical wastes are difficult because even small sets of waste substrates can, within few steps, generate millions of putative products, each synthesizable by multiple routes forming densely connected networks. Tracing all such syntheses and selecting those that also meet criteria of process and 'green' chemistries is, arguably, beyond the cognition of human chemists. Here we show how computers equipped with broad synthetic knowledge can help address this challenge. Using the forward-synthesis Allchemy platform[16], we generate giant synthetic networks emanating from approximately 200 waste chemicals recycled on commercial scales, retrieve from these networks tens of thousands of routes leading to approximately 300 important drugs and agrochemicals, and algorithmically rank these syntheses according to the accepted metrics of sustainable chemistry[17–19]. Several of these routes we validate by experiment, including an industrially realistic demonstration on a 'pharmacy on demand' flow-chemistry platform[20]. Wide adoption of computerized waste-to-valuable algorithms can accelerate productive reuse of chemicals that would otherwise incur storage or disposal costs, or even pose environmental hazards.

All our analyses are based on Allchemy's collection of approximately 10,000 generalized reaction transforms expert-coded based on the underlying reaction mechanism and including—but not limited to—robust reaction types common in chemical industries, especially pharmaceutical[21,22] but also agrochemical and flavour/fragrance. These reaction rules are much broader and also more accurate[23] than machine-extracted transforms[24], and the expert-coding approach has been validated by successful experimental execution of numerous computer-planned syntheses: such as in Allchemy for understanding the origins of life[16], and in Chematica/Synthia[25–29], which uses a different set of retrosynthetic rules, for the syntheses of drugs[26] and complex natural products[28].

Reaction rules

Each transform in Allchemy specifies the reaction type/class, scope of admissible substituents, structural motifs incompatible with a given reaction (approximately 400 groups are considered), typical conditions and reagents, suggested solvent (categorized as protic/aprotic and polar/nonpolar), temperature range (categorized as very low, less than −20 °C; low, −20 °C to +20 °C; room temperature (RT); high, +40 °C to +150 °C, and very high, greater than 150 °C), propensity of a given reaction to be performed in tandem with some other reaction(s), and more. Importantly, the programme also calculates a range of molecular properties (log*P*, polar surface areas and other structural descriptors, energies, pK_a values, and so on), flags problematic reagents (here, those in the US Environmental Protection Agency (EPA) List of Extremely Hazardous Substances[30], the European Union REACH regulation List of Substances of Very High Concern[31], and reagent guides from GlaxoSmithKline (GSK)[17,18]) and solvents, and uses environmental and health criteria[19] to suggest 'greener' alternatives, including the possibility of enzymatic reactions. For instance, Oxone instead of *meta*-chloroperoxybenzoic acid (*m*CPBA) is suggested in alkene epoxidation, thionyl chloride instead of triphenylphosphine

[1]Allchemy, Highland, IN, USA. [2]Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw, Poland. [3]On Demand Pharmaceuticals, Rockville, MD, USA. [4]Center for Soft and Living Matter, Institute for Basic Science (IBS), Ulsan, Republic of Korea. [5]Department of Chemistry, Ulsan Institute of Science and Technology (UNIST), Ulsan, Republic of Korea.
✉e-mail: nanogrzybowski@gmail.com

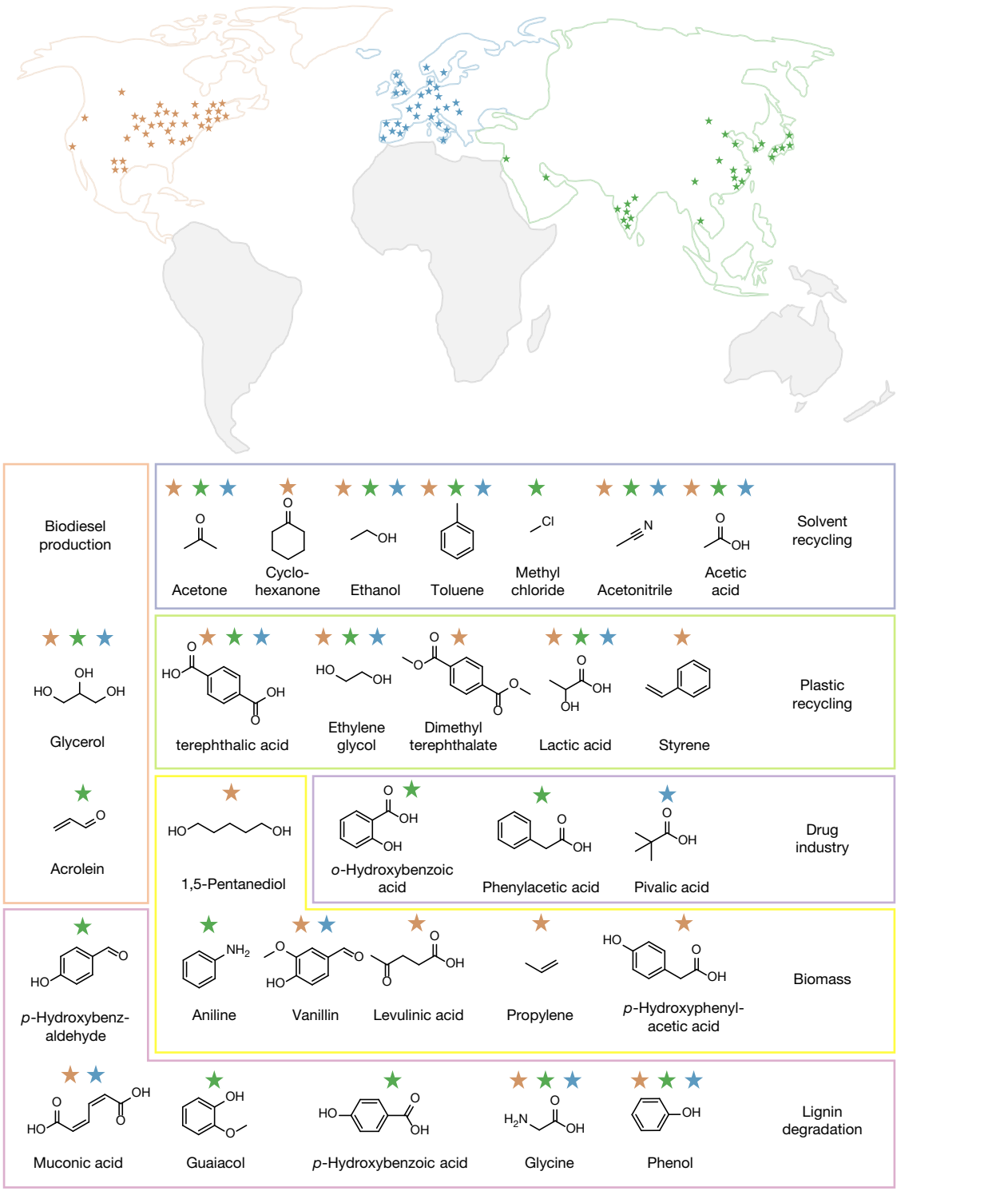


Fig. 1 | Examples of small molecules recycled from various types of industrial chemical waste. Coloured stars correspond to the map on top and indicate the geographical locations at which companies producing these substrates on large scales are located (for the list of companies, see Supplementary Table 1). We note that in addition to the recycling processes and industries indicated in the figure, the molecules shown here can also be desired targets of other processes (that is, not waste)—for instance, phenol can be produced in large quantities in the so-called cumene process, and terephthalic acid can be made via the Amoco process.

and tetrachloromethane in deoxychlorination; and *Pseudomonas* cell culture to carry out oxidation of lactic acid instead of Dess–Martin periodinane (see the synthesis of mirabegron in Fig. 3). In terms of solvents, dimethyl sulfoxide (DMSO) is given as an alternative to dimethylformamide (DMF) typically used in Williamson ether synthesis from phenols, and *t*-butyl ethyl ether rather than tetrahydrofuran (THF) is suggested for the reduction of ketones to alcohols, and so on. Also, for each reaction, the software calculates quantities such as atom economy[32] or reaction heats (using Benson's approach)[33]. Additionally, reaction sequences involving the same solvent in consecutive steps are promoted while reactions requiring very low or very high temperatures are penalized because of high energetic cost.

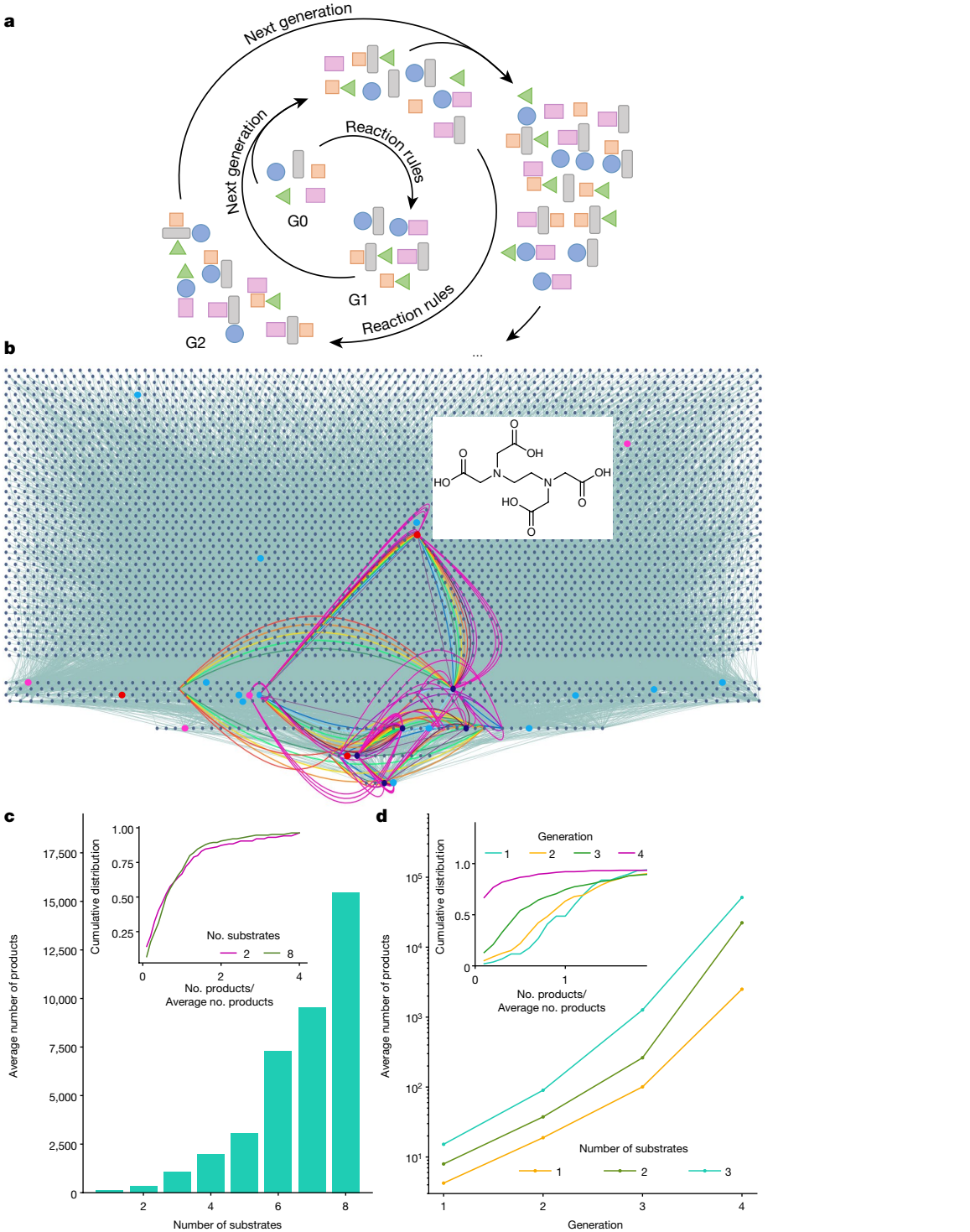


Fig. 2 | Generation and properties of 'forward' synthetic networks. **a**, Scheme of the iterative, forward-synthesis algorithm. Substrates in the zeroth generation, G0, are subject to reaction transforms to produce first generation of products, G1, which can be then pruned (generation G1′) by various structure-based or property-based filters. The G0 and G1′ molecules are then combined and the rules are again applied, this time creating molecules in generation G2 and, after pruning, G2′. The process continues until a user-defined limit of generations is reached. **b**, A screenshot from Allchemy illustrating how rapidly the network can expand from just few substrates (here, isopropanol, glycine, formaldehyde and 3,4-dihydroxyphenylglycol) in the absence of any pruning. Up to G4, this network contains 4,283 molecules (red nodes = drugs, blue = agrochemicals; pink = hazardous compounds). The colourful arcs trace 14 possible syntheses to edetic acid. **c**, Exponential growth of the average number of products obtained from a given number of 'waste' substrates after just three synthetic generations. Values for each histogram bar are averaged over $n = 189$ independent runs with different substrate sets drawn at random from the collection of 189 'waste' substrates. Inset, cumulative distributions of the number of products obtained in individual runs with different substrates (here, two and eight, normalized by the corresponding averages from the histogram). **d**, For a given number of substrates, the size of the network grows faster than exponentially with the number of synthetic generations. Each value in the plot is, as in **c**, an average over $n = 189$ independent runs with different substrate sets drawn at random from 189 'waste' substrates. Inset, changes of cumulative distributions for runs starting from two substrates analogous to those in **c** but for a different number of synthetic generations. As the number of generations increases, the fraction of runs with a below-average number of products also increases.

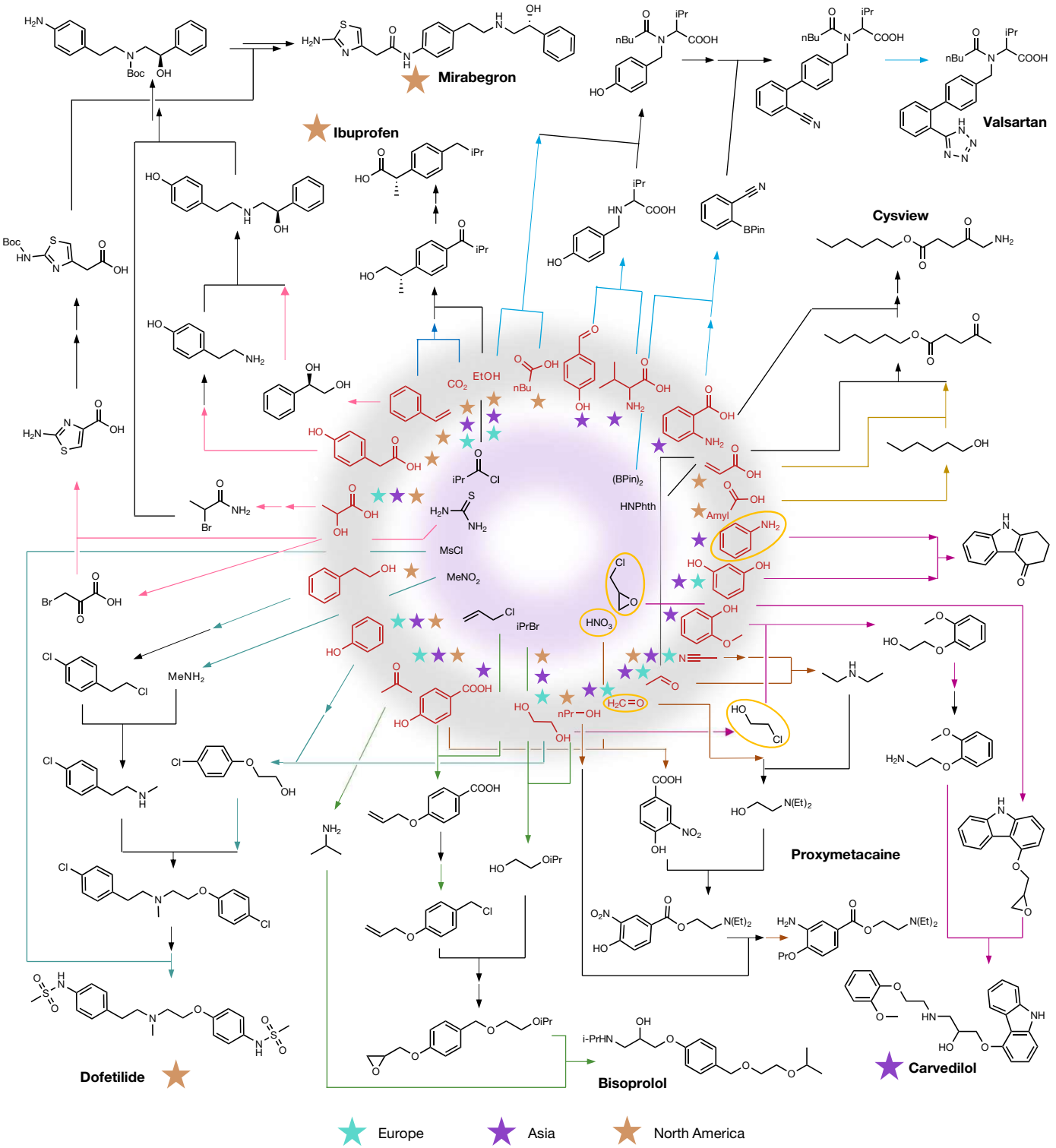


Fig. 3 | Examples of highly ranked syntheses of more advanced drugs starting from waste substrates and few simple, auxiliary molecules used frequently in organic synthesis. We show here only some intermediates along the routes; for more complete plans, also to some other drugs, see Extended Data Fig. 5. **a**, The auxiliary substrates are shown in the innermost circle. 'Waste' substrates are shown in red in the outer circle. Hazardous substances[30,31] are marked by yellow ellipses (for example, allylic alcohol). Small stars indicate geographical locations (Europe, Asia, North America; see Fig. 1 and Supplementary Information section 1) at which companies producing these substrates are located. Larger stars next to some of the drugs and agrochemicals indicate that they can be synthesized from 'same-star-colour' substrates available at the same geographical location (for example, ibuprofen can be made solely from waste substrates produced in North America). The synthetic pathways to different targets are differentiated by colours. Within each pathway, the reaction arrows for steps already reported in the literature are coloured, whereas those without literature precedent are in black. Details of all syntheses shown in this figure as well as other routes of each target are available at https://wasteresults.allchemy.net. HNPhth, phthalimide.

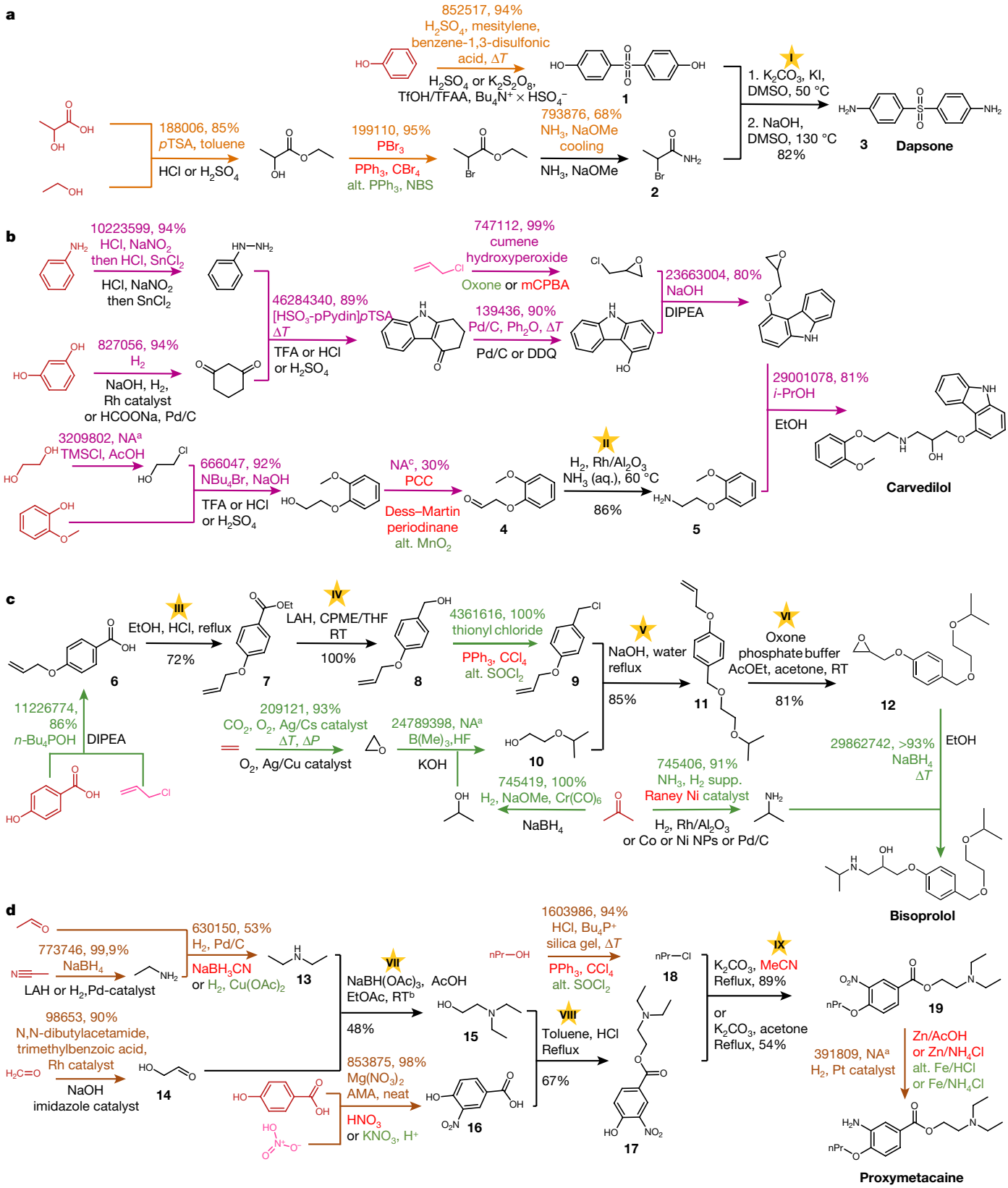


Fig. 4 | See next page for caption.

Fig. 4 | Experimental validation of selected, computer-designed pathways in laboratory scale syntheses. a–**d**, Allchemy-designed, waste-to-drug syntheses of dapsone (**a**), carvedilol (**b**), bisoprolol (**c**) and proxymetacaine (**d**). Steps lacking literature precedents and executed by us experimentally are marked by black arrows and with yellow stars above them (see Methods section 'Laboratory-scale validations'). Steps with existing literature precedent are indicated by coloured arrows. Above these arrows, text in the corresponding colour indicates the reaction ID from Reaxys, the literature-reported yield, and the conditions used (non-'green' conditions are given in red). For comparison, conditions suggested by Allchemy are provided below the arrows—'typical conditions' suggested by the programme are in black unless they involve harmful reagents (in red font); in the latter case, greener alternatives suggested by the programme are in green. In all pathways, 'waste' substrates are coloured red and commonly used chemicals are coloured in pink. [a]Information about the yield was not available in the source publication. [b]Dimer rather than very unstable monomer of glycolaldehyde was used in the reaction. [c]Reaction ID not available, literature precedent from SciFinder. CPME, cyclopentyl methyl ether; DDQ, 2,3-dichloro-5,6-dicyano-1,4-benzoquinone; DIPEA, *N*,*N*-diisopropylethylamine; DMSO, dimethyl sulfoxide; LAH, lithium aluminium hydride; RT, room temperature; THF, tetrahydrofuran; TFA, trifluoroacetic acid.

Construction of reaction networks

The reaction transforms are iteratively applied to the substrates of interest. Although the notion of 'chemical waste' may have different meanings, we consider here as substrates 189 small molecules that we identified to be waste by-products of large-scale industrial processes (see Methods, Fig. 1 and full list in Supplementary Information section 1). In the most basic version of the algorithm, the molecules produced in each synthetic generation, G, are combined with the products of preceding generations and with the original substrates, and the cycle is repeated until a user-defined limit of synthetic generations is reached (Fig. 2a). However, because the reaction networks[16,25,34,35] created by this method (Fig. 2b) expand very rapidly with the number of substrates and the number of generations (Fig. 2c, d), we bias network generation towards the syntheses of high-value molecules of interest (here, 2,466 approved drugs from DrugBank[36] and 1,647 agrochemicals subcategorized as pheromones, herbicides, insecticides and fungicides). In this approach, products made in each synthetic generation are retained for further calculations only if they are (i) small (molecular weight (MW) < 150) and can thus serve as useful building blocks for subsequent syntheses, or (ii) have 150 ≤ MW < 500 but above a certain fingerprint-based Tanimoto similarity[37] threshold to at least one of the 'target' drugs or agrochemicals. The similarity threshold is adjusted such that the total number of molecules retained in the network after each generation does not exceed a user-defined limit (the 'width' of the search, typically W = 10,000–100,000). In this way, we are able to propagate networks starting from large substrate collections (hundreds to greater than 1,000 molecules) up to generation 7 or even 8, and evaluate synthetic spaces spanning hundreds of millions of molecules. Such calculations take several days on a multicore workstation.

Pathway retrieval and scoring

Once the networks are generated, a breadth-first search algorithm is used to retrieve all syntheses connecting the wastes and valuable products. Because the shortest pathway(s) may not necessarily be optimal (for example, involving problematic conditions), we also consider routes longer by up to two steps. Owing to the high interconnectedness of the network, there are generally multiple syntheses of a given valuable product—from relatively few routes, illustrated by colourful arcs in Fig. 2b, to hundreds or even thousands in Extended Data Figs. 1, 2—and it is desirable to rank them with respect to various 'process' variables.

Here these variables are intended to 'add cost to' (penalize) the use of some undesired reaction conditions or properties and evaluate the overall pathway structure. With full definitions and details of rescaling to appropriate value ranges provided in Methods, these variables define: X_1 = a penalty on the use of harmful reagents based on GSK criteria[17,18]; X_2 = a penalty for problematic solvents as defined by GSK[19]; X_3 = a penalty for extreme temperatures; X_4 = a penalty proportional to the exothermicity or endothermicity of the reaction; X_5 = a set 'cost' for executing each reaction step (to promote shorter routes); X_6 = a penalty for low atom economy, defined in ref. [32]; X_7 = a penalty for pathways being linear rather than convergent and accounting for the position of the convergence point(s) and average yields[38]; X_8 = a 'geolocation' variable assigning penalty to pathways for which the waste substrates come from different continents (see Fig. 1, stars); X_9 = a penalty for pathways with high estimated cumulative process mass intensity (PMI), calculated based on a previous methodology[39] and using tables[40] of PMI values for individual reactions.

With the understanding that this selection may not be complete or accurate (notably, PMI values are only approximate, may entail substantial spread[39,40], and are not calibrated for flow conditions; information about large-scale pricing and batch-to-batch purity variations is currently unavailable to us, and so on), we use variables X_1–X_9 to define a simple 'cost' function, Cost = $[(w_7X_7 + \Sigma\text{StepCost})/X_8^{w_8}] \times X_9^{w_9}$, where StepCost = $\sum_{i=1}^{6} w_iX_i$. Within the Allchemy web application (https://waste.allchemy.net), the user can dynamically adjust the weights w_i of all variables and thus guide the selection of pathways meeting their process criteria. This is illustrated in Extended Data Fig. 3a, b, whereby without any penalties, the top-scoring synthesis of acetaminophen from phenol and acetic acid 'wastes' is four steps long but involves the use of thionyl chloride in toluene or dichloromethane (DCM) and $AlCl_3$ in DCM (of which DCM does not have any 'greener' replacement in the Friedel–Crafts acylation). When, however, penalties are assigned for harmful reagents and problematic solvents (Extended Data Fig. 3c), the programme prioritizes a one-step-longer pathway that avoids the acylation step and DCM. The new top-scored synthesis (Extended Data Fig. 3d) starts from *p*-hydroxybenzaldehyde and acetonitrile wastes. In the first step, an aldol reaction, the programme suggests lithium tetramethylpiperidide (LiTMP) as replacement for the harmful lithium diisopropylamide (LDA) typically used in this reaction. In subsequent alcohol oxidation, MnO_2—well rated for EHS (overall environment, health and safety score)[15]—is suggested as an alternative to the explosive Dess–Martin periodinane.

In all the examples discussed below, we ranked the pathways according to the cost function in which non-zero weights were assigned to all variables, although with the highest importance given to reagents, solvents and geolocation ($w_1 = w_2 = w_8 = 10$, $w_3 = w_4 = w_5 = w_6 = w_7 = w_9 = 1$).

Examples of synthetic networks

The first large-scale network was propagated from the 'basic' set of 189 waste substrates, with W = 10,000–30,000 and up to generation 7. Within some 300 million molecules comprising this network, the algorithm identified 69 drugs and 98 agrochemicals, suggesting 1–2,081 syntheses per target (on average, 216; see Extended Data Fig. 1a and all results stored at https://wasteresults.allchemy.net). Extended Data Fig. 4 highlights only some of the top-ranking pathways longer than three steps, with coloured arrows corresponding to steps previously reported in the literature. We observe that several targets can be made from waste available on the same continent (note the correspondence between large and small stars), and the vast majority of steps rely on benign conditions (an exception is the synthesis of eugenol in which the aromatic electrophilic alkylation

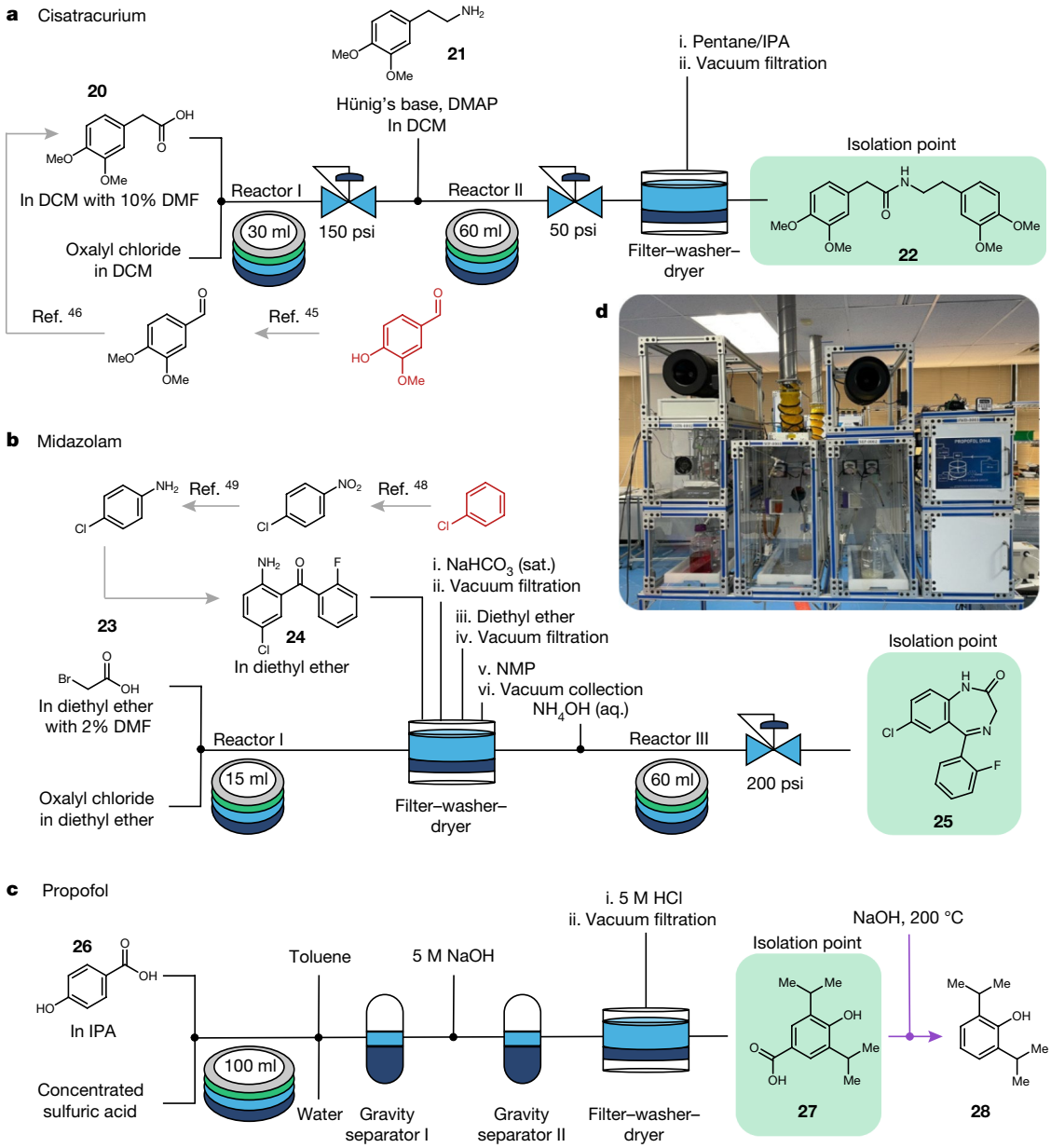


Fig. 5 | Syntheses of COVID-19 intensive care unit medications or their intermediates performed on an automated, modular ODP platform. **a**–**c**, Allchemy-designed, waste-to-drug syntheses of key intermediates of cisatracurium (**a**), and key intermediates of midazolam (**b**) and propofol (**c**; last violet arrow is offline decarboxylation). Manufacturing steps of stable hold points and potential purification points in the respective syntheses are depicted. Grey arrows indicate previously known, patented steps starting from waste molecules (red structures). Isolation points are highlighted in green. **d**, The process skid configured to manufacture the propofol. DCM, dichloromethane; DMF, dimethylformamide, DMAP, 4-dimethylaminopyridine; IPA, isopropyl alcohol; NMP, N-methyl-2-pyrrolidone.

step requires the use of 'non-replaceable' DCM). Given the simplicity of the targets, it is not surprising that most of the chemistries involved are straightforward, although not all approaches are necessarily obvious—for instance, synthesis of the dapsone antibiotic via a double Smiles rearrangement (see Fig. 4 and Methods for experimental validation).

Nevertheless, the 'wastes' alone clearly lack synthetic flexibility to build more complicated scaffolds—with this in mind, our second calculation augmented the set of waste substrates with the aforementioned 1,000 basic and popular reagents (https://github.com/rmrmg/wasteRepo/blob/main/popular_reagents.smi). Propagating the network with a more 'focused' width parameter, $W = 10{,}000$, and up to G8 generated a space of more than 160 million synthesizable compounds including 71 additional drugs and 20 agrochemicals.

These targets are more structurally complex than in the first network (for example, valsartan, mirabegron, dofetilide) and include some of the world's most prescribed medicines[41] (for example, salbutamol is ranked 7th, carvedilol is ranked 33th, and chlorhexidine is ranked 286th). Their syntheses (on average, 92 per target, see https://wasteresults.allchemy.net) are longer than those in Extended Data Fig. 4, and involve a higher proportion of steps that are, to our knowledge, previously unreported (black reaction arrows). Figure 3 and Extended Data Fig. 5 show some of the routes top-ranked according to the cost function: only in few cases they involve regulated intermediates (for example, aryl hydrazine in synthesis of carvedilol, oxirane in synthesis of bisopropol) or solvent and/or reagents for which the programme suggests no greener alternatives (for example, azide in synthesis of tetrazole ring in valsartan, diazomethane in mirabegron synthesis).

The interplay between various variables of the scoring scheme is further illustrated in Extended Data Fig. 6 for the synthesis of Cysview. Prioritizing only the pathway length yields a convergent route (blue reaction arrows) that starts from EPA-regulated[30] allyl alcohol and relies on the use of toxic and potentially carcinogenic diisopropyl azodicarboxylate (DIAD) and triphenyl phosphine in Mitsunobu reaction, and ozone in ozonolysis. A cost function penalizing the uses of harmful substances top-ranks a more linear pathway (violet arrows) in which, however, allyl alcohol is not used and the problematic steps are replaced by milder bromination (NH_4Br, Oxone conditions) and S_N2 reaction.

Finally, we considered a network to support a specific commercial operation, namely decentralized and fully automated production of pharmaceuticals and active pharmaceutical ingredients (APIs) by On Demand Pharmaceuticals (ODP)[20]. Propagating the network with a broad exploration width (W = 40,000–107,000) up to G5, generated a space of approximately 350 million molecules, including additional 27 drugs and 11 agrochemicals. Of particular and immediate interest, ODP identified drugs and/or their intermediates urgently sought[42] for ventilated COVID-19 patients: cisatracurium (a muscle relaxant), midazolam (a sedative), and propofol (an anaesthetic).

Experimental validations

Several routes traced by our algorithms within the abovementioned networks were committed to experimental validation. Initially, we performed laboratory-scale syntheses shown in Fig. 4 and intended merely to confirm the general correctness of computer-designed plans. These examples were chosen because the software either suggested some interesting transformations (for example, double Smiles rearrangement in the synthesis of dapsone) or because the proposed pathways lacked prior literature precedent for several steps (marked with yellow stars). The syntheses were generally straightforward and proceeded in good yields under benign conditions suggested by Allchemy (for details, see Methods and Supplementary Information section 5).

Next, we tested the applicability of computer-planned routes at larger scales and in realistic industrial settings, using ODP's flow chemistry platform[43] fed with adulterated waste streams (to mimic varying qualities of starting materials from various vendors at different locations). Specifically, in the continuous processes leading to strategic intermediates of cisatracurium, midazolam, and to propofol, ODP built strategic isolation points to ensure high product quality (Fig. 5a–c). For the cisatracurium intermediate (**22**), homoveratric acid (**20**) served as the starting material and a potential entry point for the second substrate, homoveratrylamine (**21**). Industrial-scale production of **20** from vanillin (recovered from biomass) and glycine (produced from lignin waste) has previously been described[44,45]. In our process, the acid chloride derivative of **20** was generated in the presence of 5 total mol% vanillic acid and guaiacol (both represent potential waste stream adulterants)[46] and subsequently reacted and isolated with no impact on product quality. Of note, **22** had a substantially different, Allchemy-calculated log*P* value (2.62) compared to either **20** (1.33) or **21** (1.21). This difference in partition coefficient led us to evaluate a binary solvent system (pentane/isopropyl alcohol), ultimately allowing for selective extraction of the more polar impurities from the process stream. With this purification in hand, a 12-h production run yielded greater than 1 kg of **22** with a liquid chromatogram area percent (LCAP) of more than 98% by high-performance liquid chromatography (HPLC) (Supplementary Table 3). The cumulative PMI for the process was 9 and compared to the theoretically predicted 24–84 range and 52 average (see Supplementary Tables 7, 10).

For midazolam, a benzodiazepine family member, ODP's entry points were bromoacetic acid (**23**) and the commercially available 2-amino-5-chloro-2′-fluorobenzophenone (**24**, log*P* = 3.29), which is available from approximately 97 suppliers and synthesizable from 4-chloroaniline, which is, in turn, manufactured on industrial scale from recycled chlorobenzene[47,48]. Batch experiments demonstrated minimal impact on conversion to the lactam when the benzophenone was contaminated with 10 mol% of both nitrobenzene and chlorobenzene. For the production run, the benzophenone was reacted with bromoacetyl chloride, which was also generated in-line from the corresponding acid and oxalyl chloride, to give 48 g of the acetamide over a 10-h run with a LCAP of 91.6% by HPLC (experimental cumulative PMI = 27 versus calculated range 24–84, with an average of 52; Supplementary Tables 8, 10). Whereas the acetamide (log*P* = 4.04) serves as a possible isolation and purification point, potential issues of stability and toxicity associated with the reactive acetamide functionality made further processing of this material to the corresponding lactam (**25**, log*P* = 3.27; see Supplementary Table 4) a more attractive hold point. This lactam serves as the main commercial entry point and currently has limited availability owing to the surge in demand as a result of COVID-19. Our process has the benefit of two distinct purification points that can purge potential impurities present in the waste stream.

Finally, propofol (**28**), a lipid-soluble anaesthetic, was manufactured from 4-hydroxybenzoic acid (**26**; produced from lignin waste) in isopropyl alcohol. To demonstrate the viability of this reaction as part of a circular economy, the feedstock (**26**, log*P* = 1.09) was adulterated with 10 total mol% vanillin and vanillic acid, two compounds that can also be obtained via lignin degradation[49]. Leveraging a distinct difference in log*P* values between this starting material and 3,5-diisopropyl-4-hydroxybenzoic acid intermediate (**27**, DIHA; log*P* = 3.37) enabled a continuous-stirred tank reactor coupled with concurrent gravity separations and precipitation, in the end providing multiple avenues for the successful purging of impurities present in the feed material. A 12-h production run through the ODP system yielded 150 g of DIHA with a LCAP of 99% by HPLC (Supplementary Table 5) and with cumulative PMI = 214, compared to Allchemy's calculated value in the 112–390 range with an average of 217 (see Supplementary Tables 9, 10). A portion of this material was forward-processed to deliver propofol API with a LCAP of 99.9% by HPLC (Supplementary Table 6). An additional purification unit operation such as a recrystallization step would enable an enhancement in purity specifications. All three processes, meant to be a part of a decentralized manufacturing, are designed to ultimately meet the supply of local hospital systems.

Conclusions

To summarize, we showed that computers equipped with comprehensive rules on chemical reactivity can rapidly trace and rank, to our knowledge, unprecedented numbers of circular syntheses establishing new, productive uses of industrial chemical waste. Naturally, we envision extensions and improvements to the schemes we described—in particular, if more accurate data with which to estimate PMI or *E* factors[40,50] and process scaling metrics became available, they should be updated in the cost function that ranks the candidate syntheses; when adequately broad substrate scopes for additional enzymes are delineated, these biocatalytic transformations[51,52] should be added to Allchemy's reaction knowledge base. In the fullness of time, applications such as Allchemy will be most impactful if adopted and shared across the chemical industry—for instance, with some companies inputting waste substrates they wish to dispose of, some indicating the products they would like to have synthesized, and some bidding to perform the waste-to-drug syntheses planned (or inspired) by the machine. In performing these tasks, we envision synergies between software such as Allchemy (to guide chemists to potential valorization opportunities) and distributed manufacturing networks such as ODP (to rapidly and cost-effectively deploy multiple production units utilizing locally available waste streams). Such an industry-wide system would help synchronize the circular-chemistry efforts but its implementation will probably require incentivization by administrative bodies overseeing chemical industry.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04503-9.

1. Yavrom, D. *An Overview of Hazardous Waste Generation (*EPA, accessed 28 April 2021); https://rcrapublic.epa.gov/rcra-public-web/action/posts/2
2. *Production-related Waste Managed by Chemical* (EPA, accessed 1 July 2021); https://www.epa.gov/trinationalanalysis/waste-managed-chemical-and-industry
3. Stahel, W. R. The circular economy. *Nature* **531**, 435–438 (2016).
4. Ellen MacArthur Foundation, World Economic Forum & McKinsey & Company. *The New Plastics Economy: Rethinking the Future of Plastics* (Ellen MacArthur Foundation, 2016).
5. Winans, K., Kendall, A. & Deng, H. The history and current applications of the circular economy concept. *Renew. Sust. Ener. Rev.* **68**, 825–833 (2017).
6. Keijer, T., Bakker, V. & Slootweg, J. C. Circular chemistry to enable a circular economy. *Nat. Chem.* **11**, 190–195 (2019).
7. Kümmerer, K., Clark, J. H. & Zuin, V. G. Rethinking chemistry for a circular economy. *Science* **367**, 369–370 (2020).
8. Kümmerer, K. Sustainable chemistry: a future guiding principle. *Angew. Chem. Int. Ed.* **56**, 16420–16421 (2017).
9. Tullo, A. H. Plastic has a problem; is chemistry the solution? *Chem. Eng. News* **97**, 29–34 (2019).
10. Zeng, H. & Li, C.-J. Conversion of lignin into high value chemical products. In Green Chemistry and Chemical Engineering (eds Han, B. & Wu, T.) 385–403 (Springer, 2018).
11. Sun, Z., Balint, F., de Santi, A., Saravanakumar, E. & Barta, K. Bright side of lignin depolymerization: toward new platform chemicals. *Chem. Rev.* **118**, 614–678 (2018).
12. Park, C. & Lee, J. Recent achievements in CO_2-assisted and CO_2-catalyzed biomass conversion reactions. *Green Chem.* **22**, 2628–2642 (2020).
13. Antonetti, C., Licursi, D., Fulignati, S., Valentini, G. & Raspolli Galletti, A. M. New frontiers in the catalytic synthesis of levulinic acid: from sugars to raw and waste biomass as starting feedstock. *Catalysts* **6**, 196 (2016).
14. Dabral, S. & Schaub, T. The use of carbon dioxide (CO_2) as a building block in organic synthesis from an industrial perspective. *Adv. Synth. Catal.* **361**, 223–246 (2018).
15. Zhang, F. et al. Polyethylene upcycling to long-chain alkylaromatics by tandem hydrogenolysis/aromatization. *Science* **370**, 437–441 (2020).
16. Wołos, A. et al. Synthetic connectivity, emergence, and autocatalysis in the network of prebiotic chemistry. *Science* **369**, eaaw1955 (2020).
17. Adams, J. P. et al. Development of GSK's reagent guides – embedding sustainability into reagent selection. *Green Chem.* **15**, 1542 (2013).
18. Henderson, R. K., Hill, A. P., Redman, A. M. & Sneddon, H. F. Development of GSK's acid and base selection guides. *Green Chem.* **17**, 945–949 (2015).
19. Henderson, R. K. et al. Expanding GSK's solvent selection guide – embedding sustainability into solvent selection starting at medicinal chemistry. *Green Chem.* **13**, 854 (2011).
20. Rogers, L. et al. Continuous production of five active pharmaceutical ingredients in flexible plug-and-play modules: a demonstration campaign. *Org. Process Res. Dev.* **24**, 2183–2196 (2020).
21. Brown, D. G. & Boström, J. Analysis of past and present synthetic methodologies on medicinal chemistry: where have all the new reactions gone? *J. Med. Chem.* **59**, 4443–4458 (2015).
22. Roughley, S. D. & Jordan, A. M. The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *J. Med. Chem.* **54**, 3451–3479 (2011).
23. Molga, K., Gajewska, E. P., Szymkuć, 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).
24. Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).
25. Szymkuć, S. et al. Computer-assisted synthetic planning: the end of the beginning. *Angew. Chem. Int. Ed.* **55**, 5904–5937 (2016).
26. Klucznik, T. et al. Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory. *Chem* **4**, 522–532 (2018).
27. Gajewska, E. P. et al. Algorithmic discovery of tactical combinations for advanced organic syntheses. *Chem* **6**, 280–293 (2020).
28. Mikulak-Klucznik, B. et al. Computational planning of the synthesis of complex natural products. *Nature* **588**, 83–88 (2020).
29. Molga, K., Dittwald, P. & Grzybowski, B. A. Computational design of syntheses leading to compound libraries or isotopically labelled targets. *Chem. Sci.* **10**, 9219–9232 (2019).
30. *Electronic Code of Federal Regulations (eCFR*, accessed 1 July 2021); https://www.ecfr.gov/cgi-bin/textidx?SID=2b4d2d375e73ebc5c93d8b2fe632cb6f&mc=true&node=pt40.28.355&rgn=div
31. *Candidate List of Substances of Very High Concern for Authorisation* (ECHA, accessed 1 September 2021); https://echa.europa.eu/candidate-list-table
32. Trost, B. M. Atom economy—a challenge for organic synthesis. *Angew. Chem. Int. Ed. Eng.* **34**, 259–281 (1995).
33. Benson, S. W. & Buss, J. H. Additivity rules for the estimation of molecular properties. Thermodynamic properties. *J. Chem. Phys.* **29**, 546–572 (1958).
34. Bishop, K. J. M., Klajn, R. & Grzybowski, B. A. The core and most useful molecules in organic chemistry. *Angew. Chem. Int. Ed.* **45**, 5348–5354 (2006).
35. Fialkowski, M., Bishop, K. J. M., Chubukov, V. A., Campbell, C. J. & Grzybowski, B. A. Architecture and evolution of organic chemistry. *Angew. Chem. Int. Ed.* **44**, 7263–7269 (2005).
36. Wishart, D. S. et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucl. Acids Res.* **36**, D901–D906 (2007).
37. Rogers, D. J. & Tanimoto, T. T. A computer program for classifying plants. *Science* **132**, 1115–1118 (1960).
38. Skoraczyński, G. et al. Predicting the outcomes of organic reactions via machine learning: are current descriptors sufficient? *Sci. Rep.* **7**, 3582 (2017).
39. Li, J., Albrecht, J., Borovika, A. & Eastgate, M. D. Evolving green chemistry metrics into predictive tools for decision making and benchmarking analytics. *ACS Sustain. Chem. Eng.* **6**, 1121–1132 (2017).
40. Borovika, A. et al. The PMI Predictor app to enable green-by-design chemical synthesis. *Nat. Sustain.* **2**, 1034–1040 (2019).
41. Kane, S. P. *The Top 300 of 2021* (ClinCalc, accessed 1 July 2021); https://clincalc.com/DrugStats/Top300Drugs.aspx
42. *Resilient Drug Supply Project* (Center for Infectious Disease Research and Policy, the University of Minnesota, accessed 1 July 2021); https://www.cidrap.umn.edu/sites/default/files/public/downloads/cidrap-rds-drug-shortages.pdf
43. Rogers, L., et. al. Continuous production of five active pharmaceutical ingredients in flexible plug-and-play modules: A demonstration campaign, *Org. Proc. Res. Dev.* **24**, 2183-2196 (2020).
44. Tengzhou Wutong Spice Co. Ltd. Reaction kettle device applicable to producing veratraldehyde and derivatives thereof. Chinese patent 203170325U (2013).
45. Guilin Teachers Technical College. Preparation method for aryl acetic acid derivative. Chinese patent 102070433A (2013).
46. Paterson, J. Poddutoori, P. & Romakh, V. Mechanism for production of biobased products from plant lignin. W.O. patent 2013/173316A1 (2013).
47. Dunn, R. O. Separation of chloronitrobenzene isomers by crystallization and fractionation. US patent 3311666A (1967).
48. Liaoning Shuntong Chemical Co. Ltd. A kind of preparation method of parachloroanilinum hydrochloride. Chinese patent 110467533A (2019).
49. Choi, W. J., Byun, J. W., Ahn, J. H., Ha, Y. W. & Seo, J.-H. Process of biologically producing a *p*-hydroxybenzoic acid. US patent 9206449B2 (2015).
50. Sheldon, R. A. The *E* factor 25 years on: the rise of green chemistry and sustainability. *Green Chem.* **19**, 18–43 (2017).
51. Turner, N. J. & O'Reilly, E. Biocatalytic retrosynthesis. *Nat. Chem. Biol.* **9**, 285–288 (2013).
52. Sheldon, R. A. & Woodley, J. M. Role of biocatalysis in sustainable chemistry. *Chem. Rev.* **118**, 801–838 (2018).

Methods

Retrosynthesis versus forward synthesis

Traditionally, chemists are accustomed to analysing how to make desired target molecules (retrosynthesis) rather than what molecules can be made from a given set of substrates (forward synthesis). However, a computerized retrosynthesis approach[25–29] is ill suited for our purpose because it is not a priori known which valuable products are synthesizable from the waste substrates: If retrosynthetic searches to these targets do not terminate after a long time, it is impossible to distinguish whether they simply need more iterations[28] or whether a given drug molecule cannot be navigated to waste precursors (and in this case, the searches will never terminate). By contrast, forward searches can exhaustively delineate the networks of molecules synthesizable from a given set of substrates including these (and only these) valuable products that are makeable from waste. Moreover, such networks are highly interconnected[16], ensuring that large numbers of possible synthetic solutions can be identified.

Choice of substrates

As 'chemical waste', we considered 189 small molecules which we identified to be waste by-products of large-scale industrial processes. Within this 'basic set', we further identified a 'commercial' subset of 56 molecules that are recycled from chemical waste or biomass, and are available commercially from companies located mostly in Asia, North America and Europe (see coloured star markers in Fig. 1 and full list in Supplementary Information section 1). For example, Chinese Jiangsu Kesheng Chemical Machinery company makes resorcinol as part of aramid fibre production process; USA-based BioCellection produces succinic, glutaric and adipic acids from plastic wastes, and European conglomerate Global Industrial Dynamics offers ethylene derived from waste biomass. All of these molecules are pre-loaded into the Allchemy software (https://waste.allchemy.net) and additional entities can be proposed via https://wastedb.allchemy.net portal (for details, see Supplementary Fig. 12). We note that although some of the 'wastes' are widely used as solvents, we are not interested in their uses as such—instead, they should be used as reaction substrates. In some searches, we also consider auxiliary sets—notably, 1,000 basic reagents most often used (as quantified in ref. [34]) in literature-reported syntheses and including molecules such as nitromethane, phthalimide and di-*tert*-butyl dicarbonate (for full list, see https://github.com/rmrmg/wasteRepo/blob/main/popular_reagents.smi).

Definitions of process variables X_1–X_9

Detailed definitions of the process variables discussed in the text are as follows.

X_1 is a penalty assigned to reactions using harmful reagents as defined by GSK criteria[17,18]. The GSK's original scores are rescaled to the range 0–1 (10 = most harmful). In most cases, alternative reagents are also suggested, and the final value is calculated as weighted average of the 'primary' and alternative conditions (0.3:0.7 weights).

X_2 penalizes problematic solvents as defined by GSK[19]. The specific value is assigned on the 0–10 scale as for X_1.

X_3 assigns a +10 penalty for extreme reaction temperatures below −20 °C or above 150 °C.

X_4 expresses a penalty that is linearly proportional to the exothermicity, $\Delta H/2$, or endothermicity, $\Delta H/5$, of reactions. The penalty is bounded to +10; ΔH is calculated using Benson's group contributions method and is expressed in kcal mol^{-1}.

X_5 assigns a +10 'cost' for executing each reaction step (this variable simply promotes shorter pathways). If consecutive steps can be performed in the same solvent (one pot), the penalty is reduced to 3.

X_6 penalizes reactions that are characterized for low atom economy, defined as in ref. [32], and takes into account both substrates and reagents. Its role is to promote reactions that produce the least amount of by-products and/or waste. Each reaction gets a score ranging from 0 to 10.

X_7 promotes convergent rather than linear pathways. This variable is defined to account for the position of the convergence point, and is expressed as an average of two terms, (linearity penalty + convergence location)/2. In this expression, the 'linearity penalty' is defined by the ratio of the longest linear sequence to the total number of reactions. The 'convergence location' term promotes routes in which convergence point(s) are closer to the final product, and is expressed as $1-\exp(-0.1\times\sum_i \mathrm{avgYield}^{-N_i})$, where avgYield is the average yield of a typical organic reaction (taken here as 75%)[38], N_i is a distance measured in synthetic steps from substrate i to the target, and the sum is over all substrates. The average of the two terms is multiplied by 10 to give a final score of a pathway between 0 and 10 (for examples of this scoring scheme for different pathway structures, see Supplementary Information section 4.3).

X_8 is a 'geolocation' variable that assigns a penalty to pathways in which the waste substrates come from different continents (see the stars in Fig. 1), implying increased transportation costs and/or longer delivery times. The overall pathway score is divided by a coefficient >1 if all 'waste' substrates are on the same continent. Here we promote such pathways by up to 20% (coefficient 1.25). If, for the substrates we considered, the location of production could not be determined, the geolocation was assigned to the company's country of origin (although, in the Allchemy web application, the variable can also be calculated for user-defined locations, see Supplementary Fig. 6).

X_9 penalizes pathways with high estimated cumulative PMI, calculated based on a previous methodology[39] and using tables[40] of PMI values for individual reactions. The raw value of cumulative PMI is rescaled to a range 1–1.5 based on the user-selected purification method. The overall pathway score is then multiplied by $X_9^{w_9}$, promoting pathways with the lowest cumulative PMI (for calculation details see Supplementary Information section 4.1).

Software details

Allchemy is a software platform for forward synthesis—that is, for iterative generation of synthetically plausible products and synthetic routes starting from arbitrary, user-defined substrates. The software can be run in either batch or web application modes; the web app can be used to visualize pathways obtained via both of these modalities. Allchemy's web-app is based on the Django (https://www.djangoproject.com) framework and uses the d3.js library (https://d3js.org) for graph representation. Substrates can be input as SMILES or drawn in Chemwriter (https://chemwriter.com). Results of synthetic calculations are stored using PostgreSQL (https://postgresql.org). Communication between the web app and Allchemy's backend is supported by Redis (https://redis.io) and RQ queue systems (https://python-rq.org).

The software has different modules focused on various aspects of forward synthesis: from the generation and exploration of networks created by prebiotic chemistries[16], to in silico combinatorial chemistry and scaffold optimization, to targeted searches towards specific molecules (here, drugs and agrochemicals). The prebiotic-chemistry module is based on ~600 reaction rules generally accepted as plausible under conditions of primitive Earth; other modules are based on ~10,000 rules covering reactions commonly used in pharmaceutical chemistry (including stereoselective ones) as well as those most capable of generating molecular diversity in as few synthetic generations as possible (multicomponent reactions, rearrangements). All rules are coded in the SMARTS notation and each has a much broader scope than any particular literature precedent underlying it (see section 'Reaction rules' and references[16,23,25,26]).

In the 'targeted' searches implemented in this work, at each synthetic generation (Fig. 2a, b), the rules are applied to the original substrates and to the subset of intermediates retained (that is, those that can still serve as useful building blocks and those above a certain similarity threshold to the 'target' molecules). A molecule is deemed suitable for a given reaction if it contains the core of at least one substrate as

defined by the reaction rule but, at the same time, does not contain any groups incompatible with the reaction. These matching conditions are evaluated using the 'GetSubstructMatches' function from the RDKit library (www.rdkit.org). Reactions are executed using the 'RunReactants' function from the ChemicalReaction class of the RDKit library with in-house enhancements to enforce proper stereochemistry and/or tautomeric forms. If a reaction template matches more than one locus on the substrate, RunReactants is executed at each and all of them. The products generated by RunReactants are filtered by algorithms developed in-house to recognize and eliminate chemically invalid molecules (for example, those violating Bredt's rules) as well as molecules that do not satisfy user-specified constraints (for example, those exceeding a certain allowed molecular mass). As the network of reactions is being generated, reaction paths leading to each molecule are stored as an ordered list of reaction steps, each of which is a tuple of reaction SMILES and reaction name.

Laboratory-scale validations

With reference to Fig. 4, we first considered synthesis of the antibiotic dapsone (Extended Data Fig. 4, bottom) from lactic acid and phenol. Unlike in a traditional route based on double aromatic nucleophilic substitution of 4-chloronitrobenzene with sodium sulfide, this synthesis relies on the Smiles rearrangement involving bisphenol S **1** and 2-bromopropionamide **2**, the latter prepared from lactic acid as described previously[53]. We validated this transformation, which is to our knowledge previously unreported, under benign conditions (K_2CO_3, KI, 50 °C in DMSO followed by NaOH, 130 °C in DMSO), achieving 82% yield (Fig. 4a, starred step I).

The second example was synthesis of carvedilol used to treat high blood pressure, congestive heart failure, and left ventricular dysfunction. Its proposed waste-to-drug synthesis (starting from aniline from biomass, guaiacol from lignin waste, and resorcinol from textile industry) features only one previously undescribed reaction, reductive amination of 2-(2-methoxyphenoxy)acetaldehyde **4**. We carried out this transformation, denoted by a star II in Fig. 4b in 86% yield using a previously proposed environmentally friendly approach[54] (Rh/Al_2O_3 catalyst and 25% aqueous solution of ammonia).

In the synthesis of a heart medication bisoprolol, four steps, denoted by stars III–VI in Fig. 4c, lacked direct literature precedent. Straightforward esterification of 4-(allyloxy)benzoic acid **6** (from 4-hydroxybenzoic acid recyclable from lignin processing) proceeded in 72% yield (star III), followed by quantitative reduction of ethyl 4-allyloxybenzoate **7** (star IV). Subsequent conversion of **8** to the corresponding 4-allyloxybenzyl chloride **9** was based on a published procedure and also proceeded in quantitative yield. This chloride was then alkylated with 2-isopropoxyethanol **10** (under phase transfer catalysis conditions with 50% $NaOH_{aq}$) to give allyl ether of 4-(2-isopropoxyethoxymethyl)-phenol **11** in 85% yield (star V). Finally, the unsaturated product was treated with Oxone in aqueous solution of phosphate buffer resulting in 4-(2-isopropoxy-ethoxymethyl)phenyl glycidyl **12** ether in 81% yield (star VI).

In the synthesis of the topical anaesthetic proxymetacaine (starting from *p*-hydroxybenzoic acid from lignin waste and four other waste substrates: propanol, formaldehyde, acetaldehyde and acetonitrile; see Supplementary Table 1), three steps required experimental validation. With reference to Fig. 4d, 2-(diethylamino)ethanol **15** was obtained from 1,4-dioxane-2,5-diol (dimer of **14**) and diethyl amine **13** in 48% yield (star VII) via reductive amination in ethyl acetate using $NaBH(OAc)_3$ as reducing agent. Esterification reaction between 2-(diethylamino)ethanol **15** and 4-hydroxy-3-nitrobenzoic acid **16** in dry toluene in the presence of catalytic amount of HCl followed to give 2-(diethylamino)ethyl 4-hydroxy-3-nitrobenzoate **17** in 67% yield (star VIII). Subsequently, this product engaged in alkylation reaction with *n*-propyl chloride **18** in acetonitrile providing 2-(diethylamino)ethyl 3-nitro-4-propoxybenzoate **19** in 89% yield or in 54% yield in greener acetone (star IX). Further synthetic details of this and other routes discussed in this section are provided in Supplementary Information section 5.

Regarding larger-scale validations, the processes for cisatracurium, midazolam, and propofol precursors were all conducted on ODP's reconfigurable platforms. Sub-kits utilized plug flow reactors with perfluoroalkoxy tubing flow paths, commercial continuous stirred tank reactors, and in-house designed filter–washer–dryers that have been described previously[20]. Reagents were purchased from their respective vendors and used as is without any need for additional purification. Simulated waste streams were created as described in Supplementary Information section 6, and analysis was carried out through HPLC versus a commercial standard.

Data availability

All data in support of the findings of this study are available within the Article and its Supplementary Information. Syntheses of all drugs and selected most interesting agrochemicals identified in large-scale network searches are available for analysis and re-ranking at https://wasteresults.allchemy.net. User manuals are available in Supplementary Information section 2.

Code availability

The interactive Allchemy web application is freely available for academic users at https://waste.allchemy.net (given server capacity, to five concurrent academic users on a rolling basis and two-week slots). Reaction rules and the source code of Allchemy are proprietary.

53. Oakes, F. T. & Leonard, N. J. Broadened scope of translocative rearrangements. Substituted 1,2,3-triazolo[1,5-*a*]-1,3,5-triazines. *J. Org. Chem.* **50**, 4986–4989 (1985).
54. Chatterjee, M., Ishizaka, T. & Kawanami, H. Reductive amination of furfural to furfurylamine using aqueous ammonia solution and molecular hydrogen: an environmentally friendly approach. *Green Chem.* **18**, 487–496 (2016).

Acknowledgements Development of the medicinal chemistry modules within the Allchemy platform (by A.W., R.R., S.S., M.M. and B.A.G.) has been supported by internal funds of Allchemy, Inc. D.K. and R.O. gratefully acknowledge funding from the National Science Centre, Poland (award 2016/23/B/ST5/03307) which supported the laboratory-scale syntheses described in this paper. Analysis of pathways and writing of the paper by B.A.G. was supported by the Institute for Basic Science, Korea (project code IBS-R020-D1). Development of the manufacturing processes for COVID-19 medications was supported by DARPA & CARES Act grant (HR0011-16-2-0029). The views, opinions and/or findings expressed are those of the author(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the US Government.

Author contributions A.W., R.R., S.S., M.M. and B.A.G. designed and developed the Allchemy platform and performed the analyses and calculations described in the paper. D.K. performed the syntheses described in Fig. 4, with supervision from R.O. B.T.H. and J.S. developed the process for cisatracurium intermediate; G.B., J.M.M. and B.T.H. developed the process for propofol; and J.A.M.L., D.T.M. and L.R. developed the process for midazolam; all these are described in Fig. 5. B.A.G. conceived and supervised the research and wrote the paper with help from other authors.

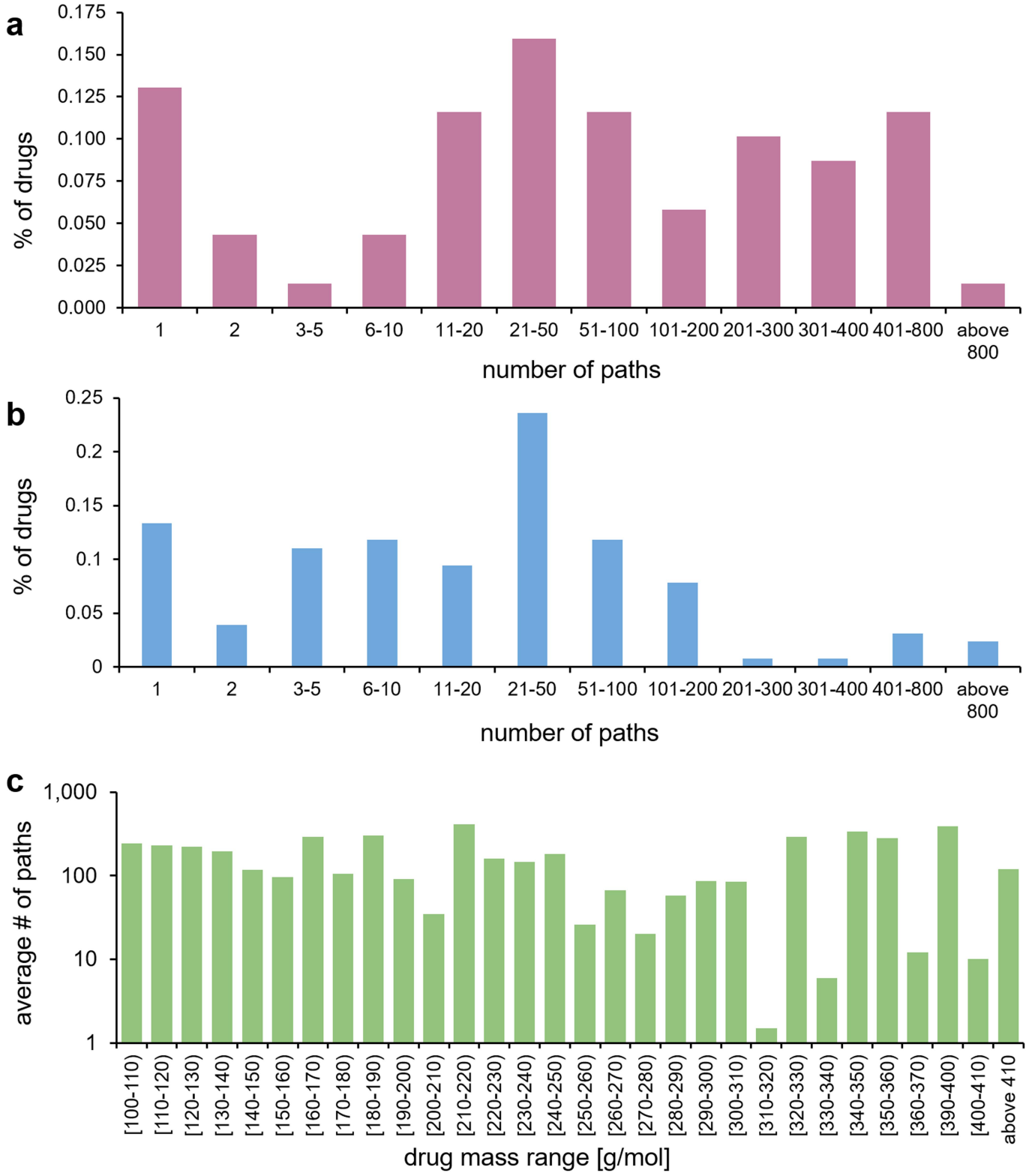
Competing interests A.W., R.R., S.S., M.M. and B.A.G. are consultants and/or stakeholders of Allchemy, Inc. Allchemy software is the property of Allchemy, Inc., USA. On Demand Pharmaceuticals (ODP) is planning to file for US Food and Drug Administration (FDA) approvals on the flow processes described in this study. All queries about access options to Allchemy, including academic collaborations, should be sent to saraszymkuc@allchemy.net. On Demand Pharmaceuticals inquiries can be sent to lrogers@ondemandpharma.com.

Additional information

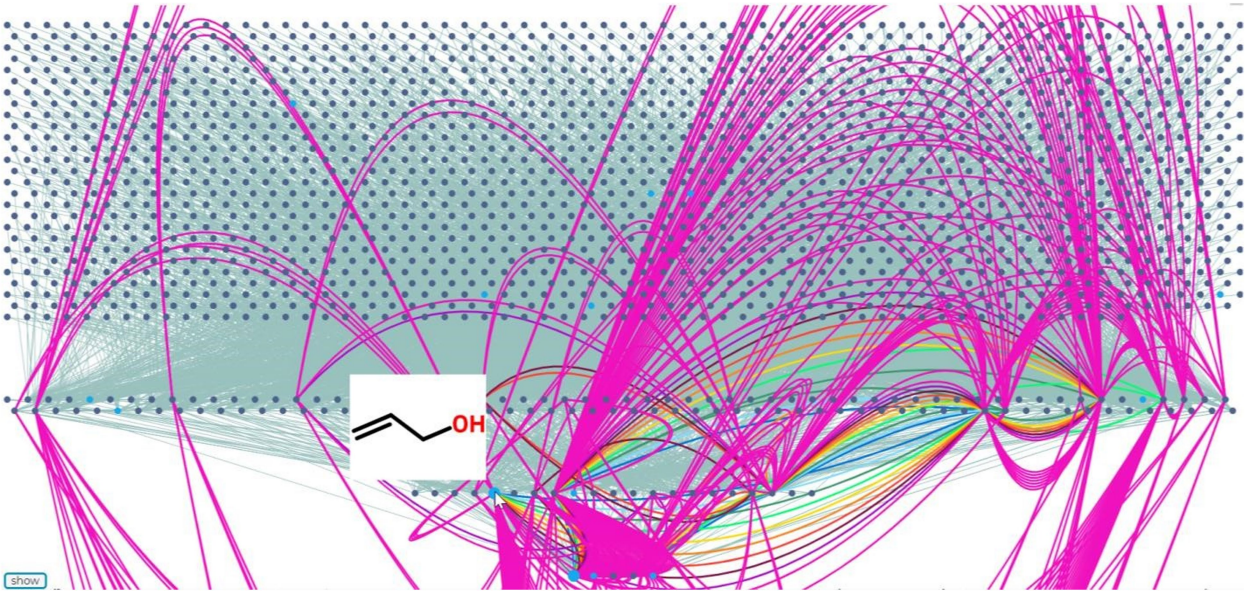
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04503-9.

Correspondence and requests for materials should be addressed to Bartosz A. Grzybowski.

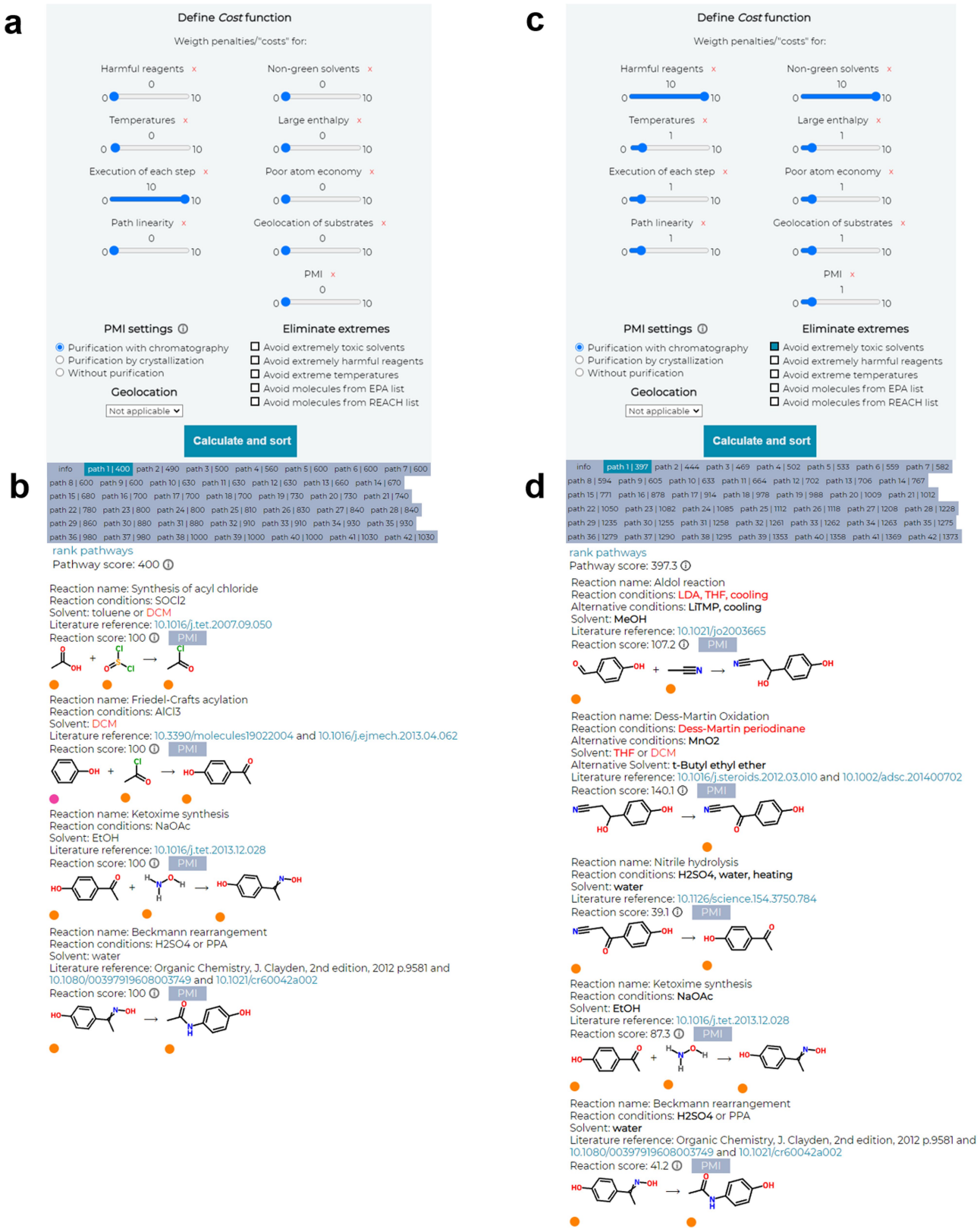
Peer review information *Nature* thanks Fabrice Gallou, Frank Roschanger and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



Extended Data Fig. 1 | Statistics of the numbers of synthetic pathways. **a**, Distribution of the numbers of synthetic paths for 69 drugs obtained from 189 waste molecules (as in Extended Data Fig. 4) within 7 synthetic generations. **b**, Distribution of number of synthetic paths for 127 drugs obtained from 189 waste molecules and 1,000 auxiliary, popular reagents (as in Fig. 3) within 8 synthetic generations. **c**, Histogram plotting the average number of pathways for drug targets of given molecular weights (mass ranges without drugs are omitted for clarity). Note the logarithmic vertical scale.

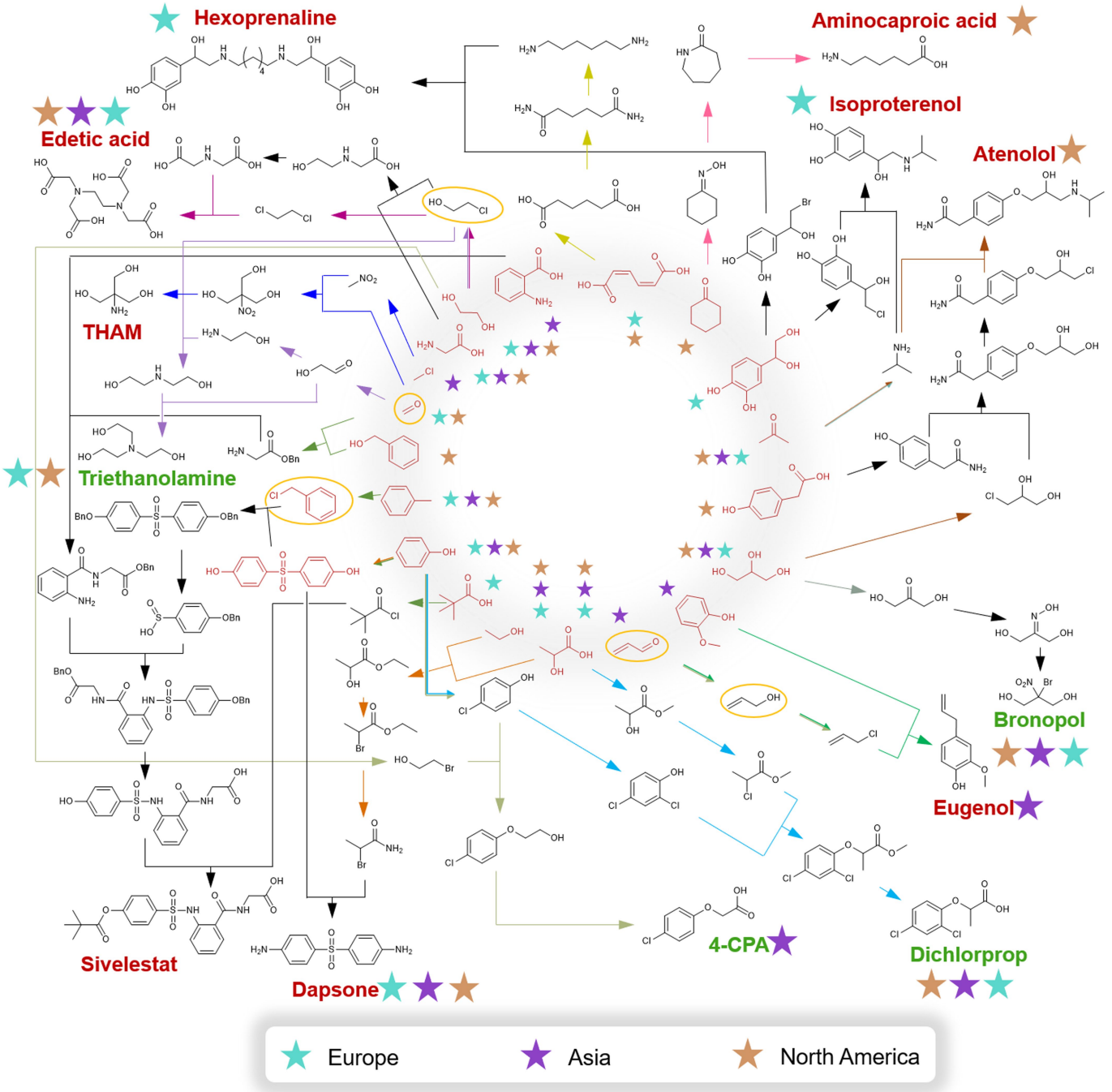


Extended Data Fig. 2 | Arcs tracing 52 syntheses of the allyl alcohol agrochemical. For large number of syntheses, the arc representation becomes messy (for example, some arcs extend beyond the window view)—nevertheless, this viewing modality enables rapid assessment of key intermediates (that is, hub nodes shared between different arcs).

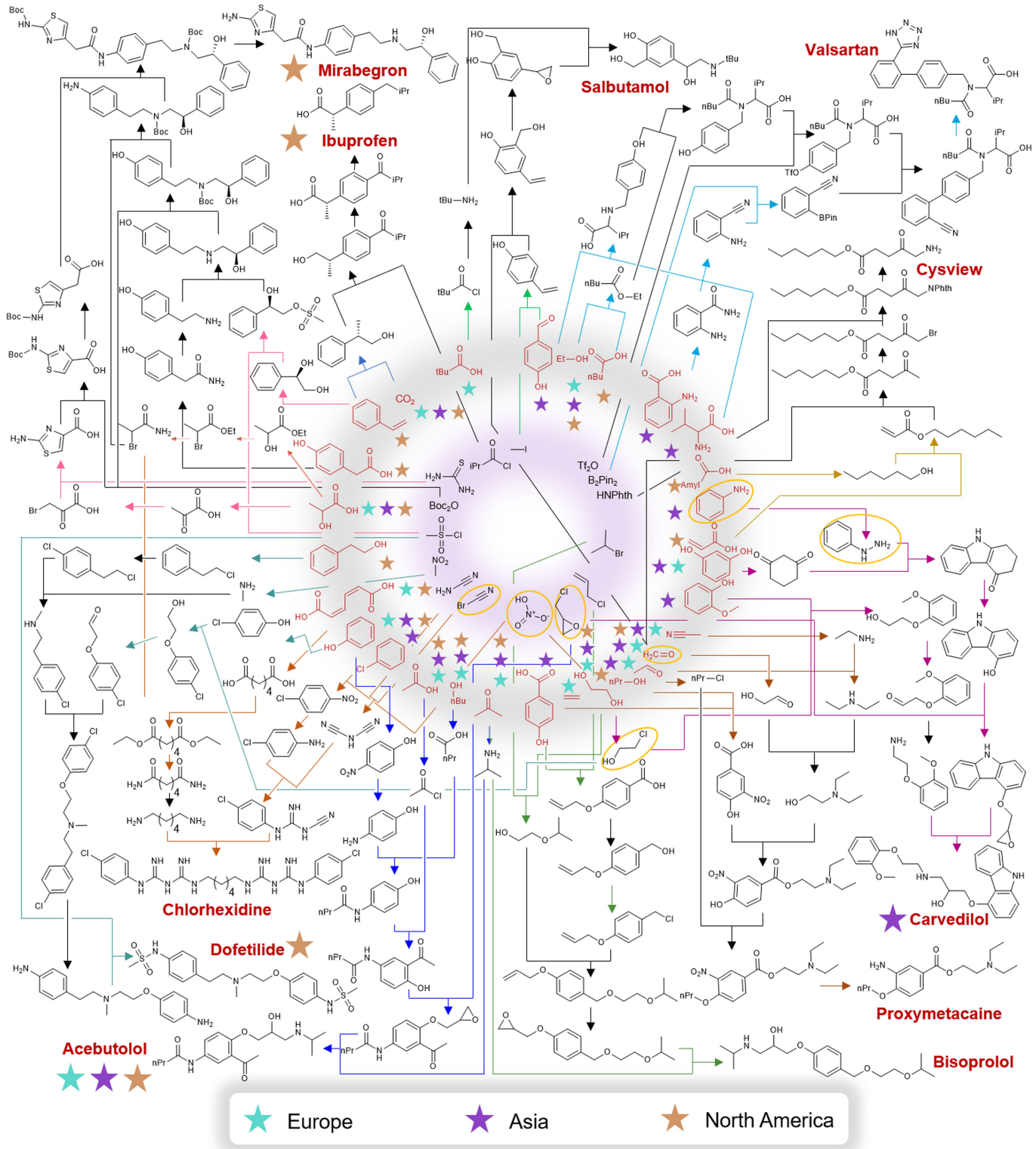


Extended Data Fig. 3 | See next page for caption.

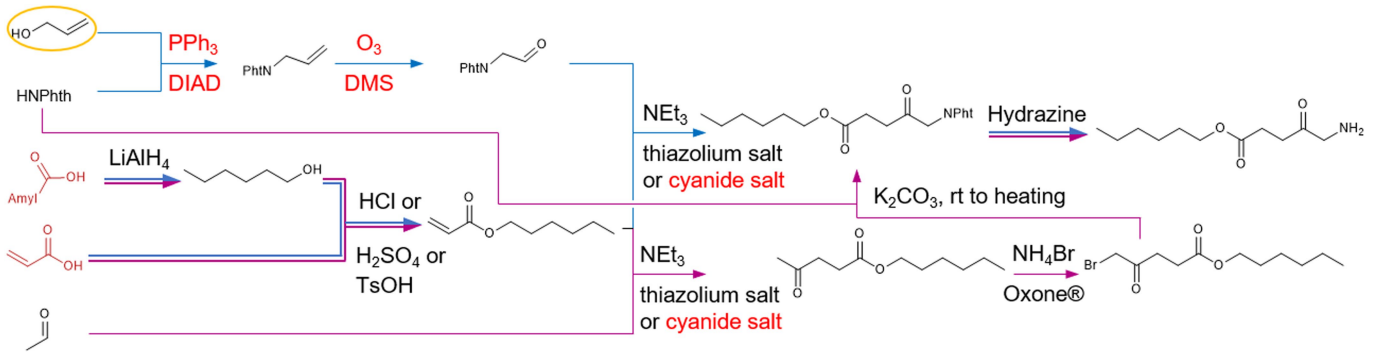
Extended Data Fig. 3 | Ranking of pathways according to process criteria. Allchemy identified 42 pathways, originating from 11 waste substrates and leading to acetaminophen. **a**, Screenshot shows the settings of the *Cost* function ranking these syntheses only for the overall length of synthesis (that is, assigning high cost for the execution of each step). **b**, The shortest route, but it entails Friedel–Crafts acylation using toxic and environmentally problematic DCM solvent (in red font) for which no 'greener' replacement is suggested. **c**, Another scoring scheme, this time assigning more penalty for problematic solvents and reagents (the checkboxes in the bottom assign even higher penalties, +200, to assure that any pathways violating these conditions will be at the very bottom of the rankings; here extremely toxic solvents are penalized). **d**, The top-ranking synthesis is one step longer but does not involve Friedel–Crafts acylation in DCM. Similar rankings are available for all syntheses generated by the web app (https://waste.allchemy.net) and for results of large-scale calculations from the text available at https://wasteresults.allchemy.net. For the use of these software programmes, please see the user manual in Supplementary Information section 2 and watch Supplementary Video 1.



Extended Data Fig. 4 | High-ranking syntheses of nine drugs and four agrochemicals entailing three or more steps and starting solely from recycled 'waste' molecules. The waste substrates are shown in red in the inner circle. Small stars indicate geographical locations (Europe, Asia, North America; see Fig. 1 and Supplementary Information section 1) at which companies producing and/or recycling these substrates are located. Larger stars next to some of the drugs and agrochemicals indicate that they can be synthesized from waste substrates available at the same geographical location. For instance, the drug hexoprenaline can be made from waste recycled solely in Europe (both of its substrates, muconic acid and 3,4-dihydroxyphenylglycol, are denoted by small light-blue stars) and is thus marked with a large light-blue star. The agrochemical triethanolamine is denoted by two large stars (light blue, Europe; orange, North America) because it can be made from the same-continent wastes either in Europe or in North America (one of the waste substrates, ethylene glycol, is recycled on all three continents and, accordingly, is denoted by three small stars; the other substrate, formaldehyde, is recycled at industrial scales only in Europe and North America and thus has two small stars). Synthetic pathways to different targets are differentiated by colours. Within each pathway, reaction arrows for steps already reported in the literature are coloured whereas those without literature precedent are in black. Hazardous substances[30,31] are marked by yellow ellipses (for example, formaldehyde). Details of all syntheses shown in this figure as well as other routes of each target are available at https://wasteresults.allchemy.net. Note that all pathways are top-scoring with exception of that for dapsone for which the software also found a two-step route starting from chlorobenzene waste, sulfuric acid and ammonia; however, this pathway is already known and patented and is not shown.



Extended Data Fig. 5 | Additional details of highly ranked syntheses of more advanced drugs starting from waste substrates and few simple, auxiliary molecules used frequently in organic synthesis. The figure accompanies and extends Fig. 3. All colour-coding schemes are the same. Details of all syntheses shown in this figure as well as other routes of each target are available at https://wasteresults.allchemy.net. HNPhth, phthalimide.



Extended Data Fig. 6 | Two syntheses of Cysview top-ranked by different *Cost* functions. Route traced by blue arrows was scored for overall length—it is concise but entails several harmful reagents (in red font) and intermediates (allyl alcohol in a yellow ellipse). Path traced by violet arrows is less convergent but uses greener reaction conditions (for details, see text).