Chem

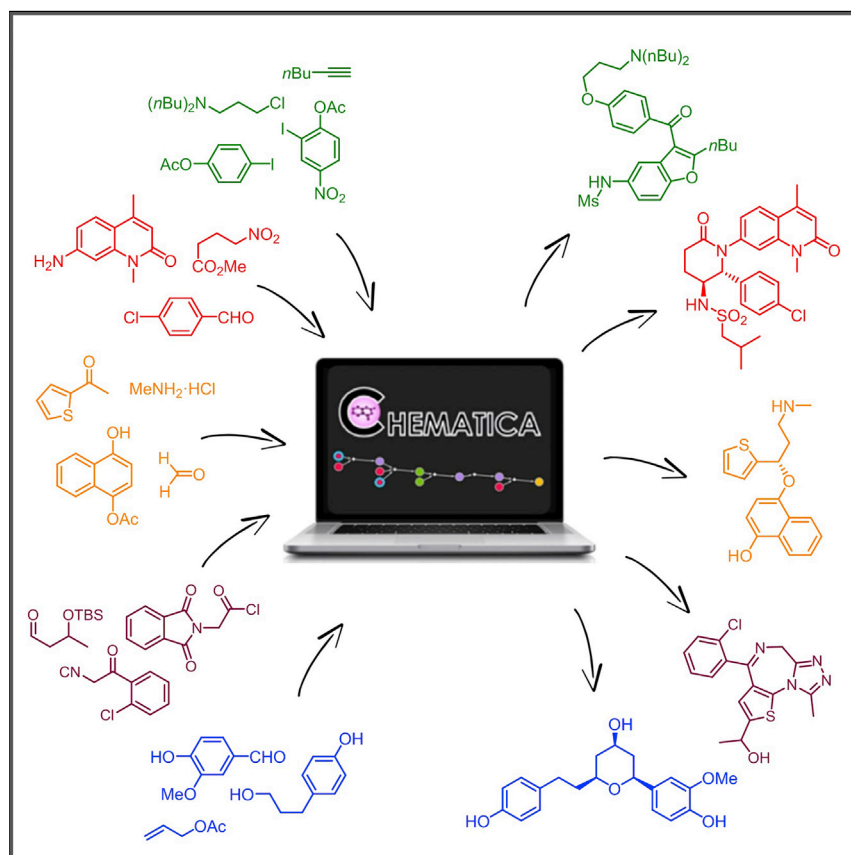**Article**

Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory

Tomasz Klucznik, Barbara Mikulak-Klucznik, Michael P. McCormack, ..., Milan Mrksich, Sarah L.J. Trice, Bartosz A. Grzybowski

milan.mrksich@northwestern.edu (M.M.)
sarah.trice@sial.com (S.L.J.T.)
nanogrzybowski@gmail.com (B.A.G.)

HIGHLIGHTS

Computer autonomously designs chemical syntheses of medicinally relevant molecules

The syntheses are successfully executed in the laboratory

The machine-designed routes improve on previous approaches

Multistep synthetic routes to eight structurally diverse and medicinally relevant targets were planned autonomously by the Chematica computer program, which combines expert chemical knowledge with network-search and artificial-intelligence algorithms. All of the proposed syntheses were successfully executed in the laboratory and offer substantial yield improvements and cost savings over previous approaches or provide the first documented route to a given target. These results provide the long-awaited validation of a computer program in practically relevant synthetic design.

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

Chem

CellPress

Article

Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory

Tomasz Klucznik,[1] Barbara Mikulak-Klucznik,[1] Michael P. McCormack,[2] Heather Lima,[2] Sara Szymkuć,[1] Manishabrata Bhowmick,[2] Karol Molga,[1] Yubai Zhou,[3] Lindsey Rickershauser,[2] Ewa P. Gajewska,[1] Alexei Toutchkine,[2] Piotr Dittwald,[1] Michał P. Startek,[4] Gregory J. Kirkovits,[2] Rafał Roszak,[1] Ariel Adamski,[1] Bianka Sieredzińska,[1] Milan Mrksich,[3,*] Sarah L.J. Trice,[2,*] and Bartosz A. Grzybowski[1,5,6,*]

SUMMARY

The Chematica program was used to autonomously design synthetic pathways to eight structurally diverse targets, including seven commercially valuable bioactive substances and one natural product. All of these computer-planned routes were successfully executed in the laboratory and offer significant yield improvements and cost savings over previous approaches, provide alternatives to patented routes, or produce targets that were not synthesized previously.

INTRODUCTION

Teaching the computer to plan chemical syntheses has been one of the outstanding challenges of modern-era organic chemistry. Despite decades of research and many ingenious approaches,[1–12] there have been no literature reports of complete synthetic pathways designed by the computer and then successfully executed in the laboratory.[11] The inadequacy of computer programs reflected, among other factors, their limited knowledge base of chemical transformations, their inability to navigate enormous "trees" of synthetic possibilities in an intelligent fashion, and the lack of higher-order logic prescribing how individual steps should be put together to produce elegant, or at least viable, pathways. Building on over a decade of research on chemical networks,[13–16] we have recently disclosed[10] a de novo retrosynthetic module within the Chematica platform (henceforth, simply Chematica) that unites network theory, modern high-power computing, artificial intelligence, and expert chemical knowledge to rapidly design synthetic pathways leading to arbitrary (i.e., previously made or never attempted) targets. Although Chematica has attracted considerable interest,[17,18] its predictions have not been verified experimentally until now. Here, we describe the results of a systematic evaluation in which synthetic pathways leading to eight structurally diverse and medicinally relevant targets were first designed by Chematica without any human supervision and subsequently executed in the laboratory. All of these syntheses were successful and either improved on previous approaches or provided the first documented route to a given target.

Starting in 2005,[13] we have published extensively on the algorithms and methods enabling representation of synthetic pathways as the so-called bipartite graphs, which can then be queried in the Chematica[10,16] platform according to different sets of criteria. In retrosynthesis, these criteria are rules describing various types of

The Bigger Picture

Although computers have demonstrated the ability to challenge humans in various games of strategy, their use in the automated planning of organic syntheses remains unprecedented. As a result of the impact that such a tool could have on the synthetic community, the past half century has seen numerous attempts to create in silico chemical intelligence. However, there has not been a successful demonstration of a synthetic route designed by machine and then executed in the laboratory. Here, we describe an experiment where the software program Chematica designed syntheses leading to eight commercially valuable and/or medicinally relevant targets; in each case tested, Chematica significantly improved on previous approaches or identified efficient routes to targets for which previous synthetic attempts had failed. These results indicate that now and in the future, chemists can finally benefit from having an "in silico colleague" that constantly learns, never forgets, and will never retire.

reactions. The rules are coded by expert chemists and to ensure applicability to arbitrary targets, they cover not only popular and simple transformations but also the advanced methodologies essential in the synthesis of complex targets. At the core of each of Chematica's ~50,000 rules is a decision tree such as the one shown in Figure 1A for double stereodifferentiating condensation of esters with aldehydes. The various conditions within the tree specify the range of admissible and also possible (i.e., not only those based on prior literature precedents) substituents or atom types. Importantly, all rules account for stereo- and regioselectivity and also for the "context" of the molecule; that is, for groups incompatible with the reaction or those to be protected (for these and other aspects of rule application including electronic and steric effects, see Supplemental Information, Sections S2–S5).

The reaction rules are only the basic "moves" from which the complete synthetic pathways ("games") are to be constructed. Because the number of choices at each retrosynthetic step[10] is ~100 (commensurate with the number of choices at each step in a chess game), the number of possibilities within n steps scales as 100^n. To search such an enormous synthetic space (Figures 1B and 1C; Supplemental Information, Section S6), intelligent algorithms are needed to truncate and revert from unpromising "branches" and channel the searches toward the most efficient and elegant sequences of steps. Chematica avoids unpromising routes by using numerous heuristics prohibiting unlikely structural motifs, penalizing reactions that are non-selective, or those that would have to proceed through very strained intermediates (Supplemental Information, Sections S5.2 and S6). The searches are then guided toward the most feasible solutions by the so-called scoring functions that evaluate (1) the sets of substrates made at each step and (2) the sequences of reactions that were used to reach any particular set. Importantly, to enable searches and scoring of substrate sets rather than individual molecules, the bipartite reaction graphs are transformed into so-called hypergraphs (with "supernodes" combining several individual substance nodes; Supplemental Information, Sections S6.1 and S6.2). The algorithm navigating these hypergraphs takes advantage of higher-order logic (concatenating individual steps into "strategic sequences," eliminating sequences of steps in which highly reactive groups are dragged along, etc.; see Supplemental Information, Section S7) and terminates when reaching commercially available (currently, over 200,000 chemicals from the Sigma-Aldrich catalog) or other synthetically popular substrates (ca. 7,000,000 molecules from literature and patents, each with the value of its connectivity within the Network of Chemistry[13–16]). Finally, because up to millions of viable pathways can be found for typical targets, dynamic linear programming algorithms are used to retrieve pathways that are not only best scoring but also significantly different from each other (Supplemental Information, Section S6.3). In the pathways presented to the user, each substance can be further inspected via built-in molecular mechanics tools, and each reaction comes with suggestions for reaction conditions, literature citation(s) illustrating this type of chemistry, information on which groups need to be protected (and with what protecting groups), examples of similar reactions reported in literature, and more (see Supplemental Information, Section S8).

RESULTS AND DISCUSSION

Choice of Targets

The algorithms and methods described above were used to design the syntheses of eight targets chosen as follows. The first six targets were provided by MilliporeSigma (MS, formerly Sigma-Aldrich) and were all biologically active compounds of high commercial value (>US$100/mg) for which previous (and in most cases numerous) synthetic attempts at MS were very low yielding, not scalable, or altogether failed. Here, the main objective for Chematica was to design routes improving over these

[1]Institute of Organic Chemistry, Polish Academy of Sciences, ul. Kasprzaka 44/52, Warsaw 01-224, Poland

[2]MilliporeSigma, 400 Summit Drive, Burlington, MA 01803, USA

[3]Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, IL 60201, USA

[4]Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, 02-097 Warsaw, Poland

[5]IBS Center for Soft and Living Matter and Department of Chemistry, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Eonyang-eup, Ulju-gun, Ulsan 689-798, South Korea

[6]Lead Contact

*Correspondence:
milan.mrksich@northwestern.edu (M.M.),
sarah.trice@sial.com (S.L.J.T.),
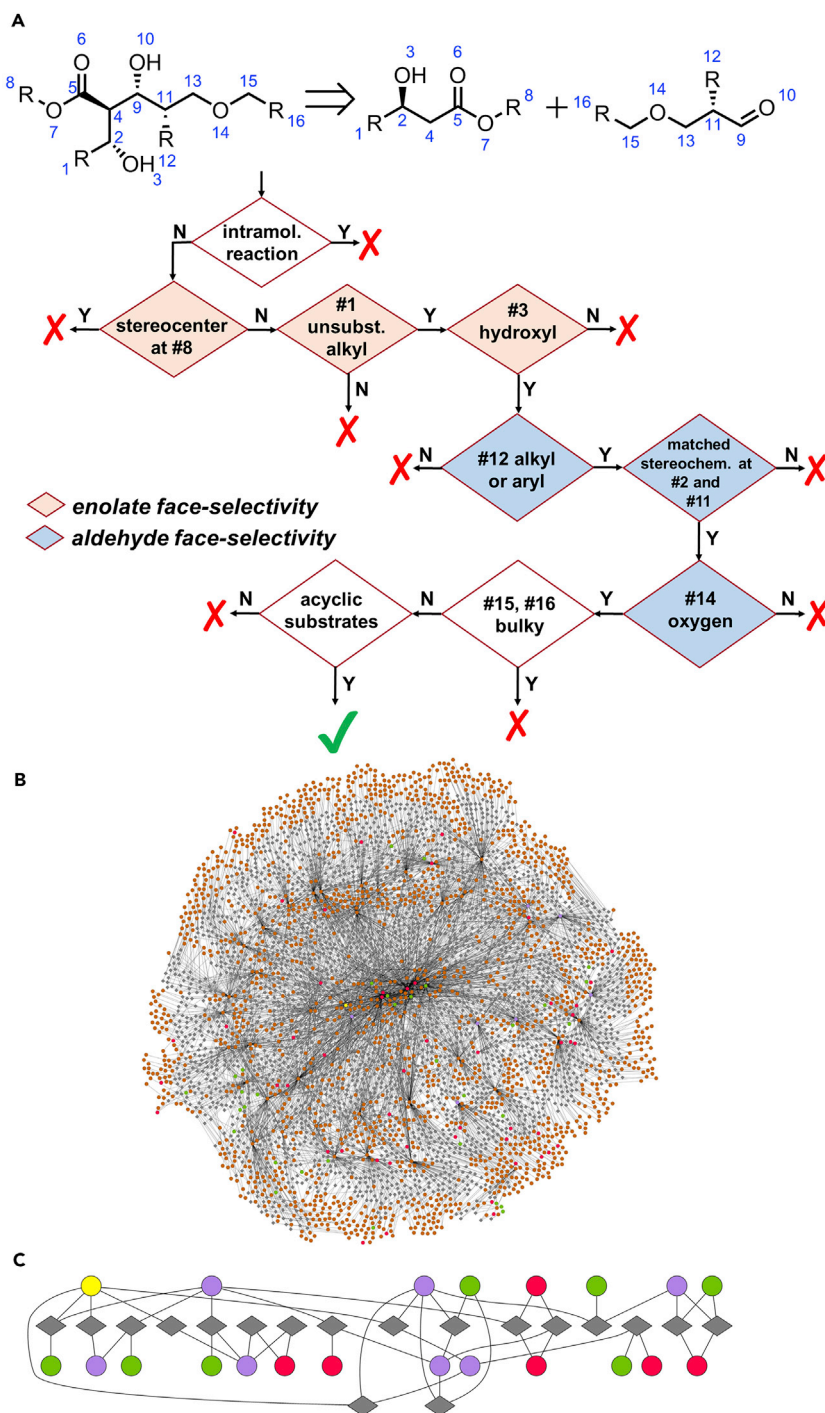nanogrzybowski@gmail.com (B.A.G.)

Figure 1. Reaction Rules and Reaction Networks Underlying Chematica

(A) An example of a decision tree for one of Chematica's ~50,000 reactions rules (double stereodifferentiating condensation of esters with aldehydes). The tree begins with a condition of the reaction being intermolecular. To ensure face selectivity of the enolate, conditions for the substituents at positions #8, #1, and #3 are considered. Conditions at positions #12, #2, #11, #14 follow and ensure proper face selectivity of the aldehyde. The last two conditions are common for both substrates. The substrates should be acyclic because cyclic structures might distort the aldehyde-titanium chelate conformation or face selectivity of the ester enolate. The other requirement concerns the consonant selectivity at both substrates that ensures the desired

Figure 1. _Continued_
diastereoselectivity. The mechanistic reasons for each condition as well as the translation of the tree into SMILES notation are discussed in detail in the Supplemental Information, Section S3.4.
(B) The rules such as the one described in (A) are used to explore the graphs of synthetic possibilities emanating from the target and growing with the number of search iterations. Each node corresponds to a set of substrates. The image shown here is for the early stage of planning the synthesis of the BRD7/9 inhibitor **8**. In a typical planning task, the program constructs and analyzes networks that are tens to thousands (sic!) of times larger than the one shown.
(C) A subgraph of (B) showing only the viable synthetic pathways terminating in commercially available (red nodes) or known (green nodes) substrates. Once such feasible syntheses are found, the program extracts them from its internal network representation and displays as in the actual synthetic routes in Figures 2 and 3 (also see Movie S1).

previous attempts in terms of overall cost. The seventh target was a blockbuster anti-arrythmic drug, dronedarone, from Sanofi-Aventis; this choice was made by the Grzybowski group and was motivated by the fact that tens of patents have been granted to protect dronedarone's synthesis (see list in the Supplemental Information, Section S16.1), raising the bar for the computer to find alternative pathways. The eighth target was selected by the Mrksich group to validate Chematica in the synthesis of a natural product, engelheptanoxide C, that was recently isolated but not yet synthesized.[19] Last but not least, the DARPA agency sponsoring most of our effort within the "Make-it" program—whose aim is to automate both synthetic planning and actual synthetic procedures—has been interested in whether a retro-synthetic software such as Chematica could "empower" less experienced chemists to perform synthetic work that would typically be carried out in classic synthetic laboratories. Accordingly, the first four targets (Figure 2) were made by MS chemists, whereas the last four (Figure 3) were synthesized by the Grzybowski and Mrksich students who are not experienced in multistep organic synthesis.

Execution Criteria and Expected Deliverables

The syntheses were planned completely autonomously by Chematica (running on a 64-core machine) within 15–20 min for all targets with the exception of dronedarone for which the search used an older and slower version of the software and was allowed to continue for several hours. The top-scoring pathway for each target was chosen provided that (1) it was significantly different from any previous approaches, and (2) the starting materials were immediately available. If the best-scoring pathway did not meet these criteria, the second-best was chosen (in three cases). No alterations in the proposed routes were allowed save for straightforward adjustments in reactions conditions (e.g., temperature, solvent, specific base, catalyst, etc.) for the sake of optimization. For the MS targets, the additional criteria—reflecting the constraints and demands of industrial reality—were to attempt each step no more than three to five times and to deliver at least a few hundred milligrams (and in most cases >1 g) of product within 8 weeks (no more than 70 hr of bench work) with a final high-performance liquid chromatography purity above 98% and no single impurity above 0.5% (except in the case of (S)-4-hydroxyduloxetine, which was obtained in ~95% purity). For the student-led projects, the time constraints were ca. 3–4 months with similar purity requirements.

Syntheses of Targets

The first target was the recently discovered and the first potent and selective inhibitor[20] of the so-called bromodomain-containing proteins BRD7 and BRD9 implicated in cancer. Previous attempts at MS to synthesize this quinolone-fused enantiomerically pure lactam (**8** in Figure 2A) according to the reported, eight-step literature procedure[20] resulted in very low isolated yields (overall, ~1%; see Supplemental Information, Section S10) and required the use of flash column chromatography (FCC) in all but one of the
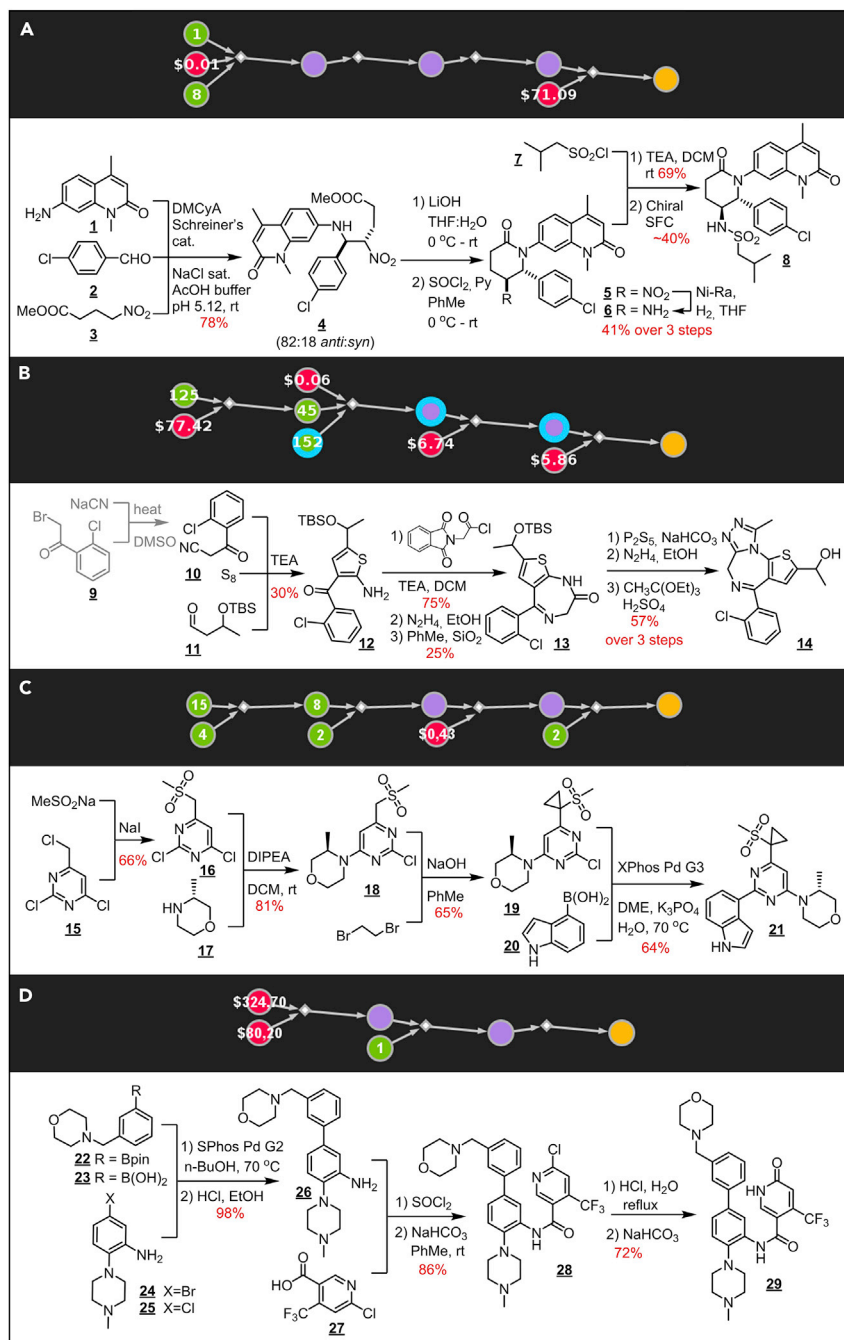
Figure 2. Syntheses of the First Four Targets Performed by the MS Chemists

BRD 7/9 inhibitor (A), α-hydroxyetizolam (B), ATR kinase inhibitor (C), and inhibitor of human acute-myeloid-leukemia cells (D). The images are screenshots of pathway "graphs" as displayed in Chematica (see also Movie S1). The positions of the structures in the chemical schemes below mirror, as much as possible, those in Chematica's graphs. Red nodes, commercial chemicals (with prices in US$/g); green nodes, known substances (with numerals denoting synthetic popularity); violet nodes, unknown substances; blue halo, protection needed.

steps. Chematica proposed a novel and shorter route starting with a concise, three-component aza-Henry reaction[21,22] of aryl amine **1**, aldehyde **2**, and nitroalkane **3**. This reaction gave the acyclic adduct **4** and its diastereomer in 78% yield and with
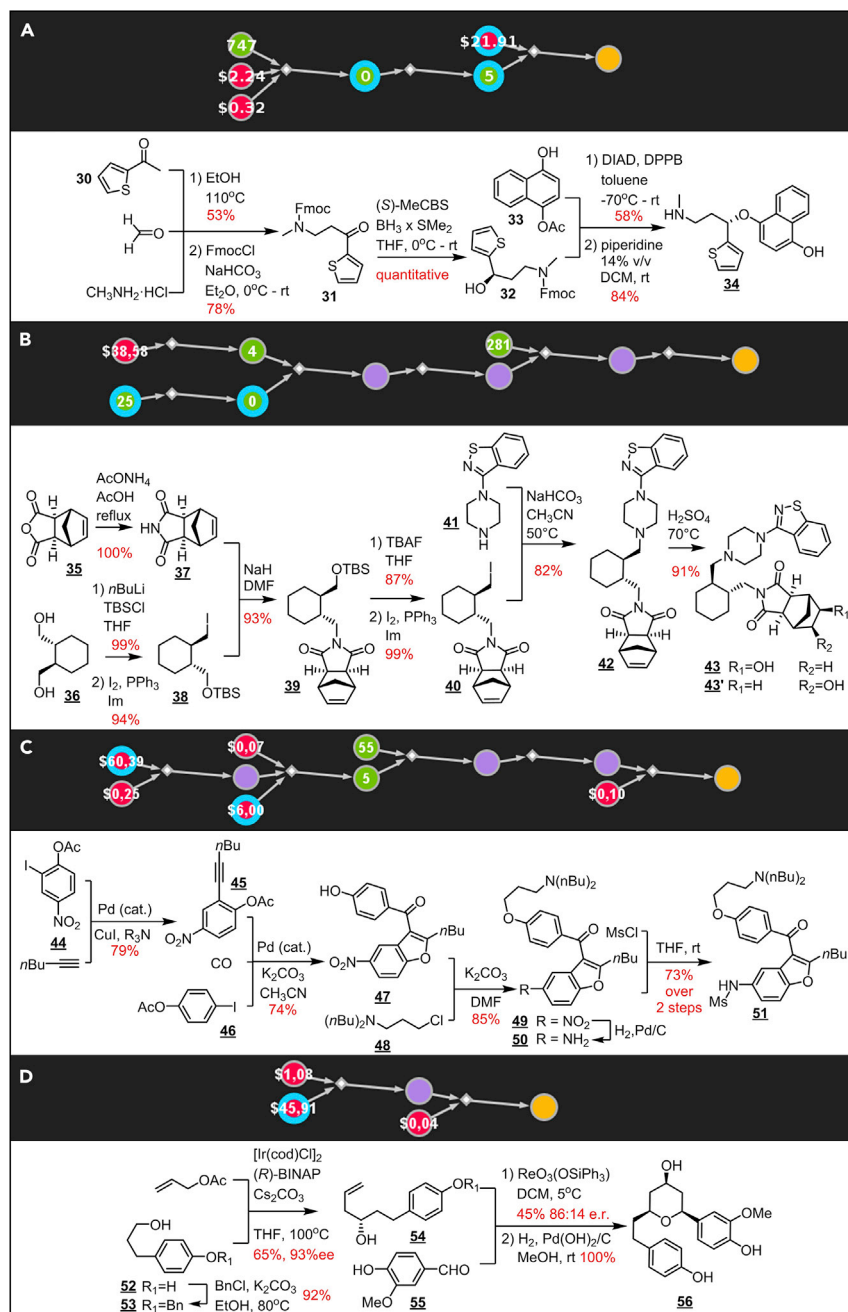
Figure 3. Syntheses of the Second Set of Four Targets Performed in the Grzybowski and Mrksich Laboratories

Figure360▷ For a Figure360 author presentation of Figure 3, see https://doi.org/10.1016/j.chempr.2018.02.002mmc3.

(S)-4-hydroxyduloxetine (A), 5β/6β-hydroxylurasidone (B), dronedarone (D), and engelheptanoxide C (D). Color coding of nodes is the same as in Figure 2.

82:18 distribution of *anti:syn* products. Intermediate **6** was then sulfonylated with **7** in 69% yield, and the desired enantiomer **8** was isolated by chiral supercritical fluid chromatography in 40% yield. Overall, Chematica's route improved the yield by six to eight times and required fewer (three versus five) FCC separations.

The second target was α-hydroxyetizolam (**14** in Figure 2B), which is one of two main metabolites used as indicators of etizolam usage in humans.[23] Although the synthesis of benzodiazepines is well established, making hydroxy metabolites presents more of a challenge and MS previously deemed **14** as "too risky" to prioritize. Chematica's proposal is unique in that it rests on a multicomponent Gewald reaction. Although this method has been used to prepare similar hydroxylated thiophene intermediates, this particular side chain has not been reported.[24] Chematica's route started with protected 3-hydroxybutanal **11** (need for protection, indicated in the software's plan by a blue halo) and nitrile **10** (purchased rather than made from bromoketone **9**). These substrates were allowed to react with sulfur under classic Gewald conditions to afford the desired thiophene **12**. The remaining steps followed classic benzodiazepine transformations to give the desired α-hydroxyetizolam **14** in 3.2% overall yield (see Supplemental Information, Section S11).

The third target was a potent and selective ATR kinase inhibitor denoted as **21** in Figure 2C. The synthesis of this target has been described previously[25] and involves seven steps (see Supplemental Information, Section S12). Although the overall literature-reported yield is about 20%, yields that MS were able to achieve on numerous attempts with the published route remained below 10%. In addition, the literature route is not readily amenable to scale-up. In this light, Chematica's short, four-step solution shown in Figure 2C appeared attractive, such that intermediate **18** was prepared in two rather than five published steps (cf. Supplemental Information, Section S12). This difference does not stem from using drastically different types of chemistries along the route but rather from a choice of a different starting material (as mentioned, Chematica "knows" some 200,000+ commercial substrates plus ca. 7,000,000 known molecules), which avoids some unnecessary interconversion of functional groups. When executed, Chematica's route offered an overall, reproducible, and gram-scalable yield of 20%–22%, and provided 30% time savings (45 versus 62 hr for the published protocol).

The fourth target, marked as **29** in Figure 2D, has been shown[26] to selectively inhibit proliferation and induced differentiation of human acute-myeloid-leukemia cells. The reported[26] preparation of **29** is, despite optimization efforts at MS, low yielding (1% overall; see Supplemental Information, Section S13), requires four chromatographic separations, and is not readily amenable to scale-up for commercialization. Chematica's route was attractive in that it proposed Suzuki coupling before forming the desired amide (which, in the literature pathway, was the lowest-yielding step, 10%). The route starting from the bromide **24** was chosen because it was commercially available. Conditions for the requisite Suzuki-Miyaura coupling were optimized with a KitAlysis Reaction Screening Kit.[27] Although several of the screening conditions were successful, the SPhos Pd G2 catalyst was chosen, and the desired biphenyl **26** could be purified without chromatography by recrystallizing the hydrochloride salt of the crude product. The hydrochloride salt was then converted to the amide **28** via Schotten-Baumann acylation, followed by aqueous hydrolysis of **28** and neutralization to afford **29** in 72% yield. In this step, the 2-Cl-pyridine **27** rather than the 2-OH derivative was used even though the latter could potentially avoid the hydrolysis step. Although Chematica identified both solutions, it decided to use 2-Cl on the basis of its higher synthetic popularity and significantly lower price. Overall, this approach allowed for minimal chromatographic purification steps, gave product in gram quantities with a significantly improved 60% yield, and offered time and cost savings of ~50%.

Turning to the syntheses performed by students in the Grzybowski and Mrksich laboratories, the fifth target was (*S*)-4-hydroxyduloxetine, **34** in Figure 3A, which is one

of the main metabolites of duloxetine (the blockbuster drug Cymbalta). This "simple-looking" target was chosen because of its importance in clinical testing, high market price (~$200/mg), and because MS had previously failed to reproduce the only literature route[28] involving a nucleophilic aromatic substitution between an alcohol and *O*-protected-4-fluoronaphthol but published without basic experimental details. Chematica suggested a much simpler route detecting no counterindications for the key Mitsunobu reaction between **32** and **33**. Indeed, this reaction proceeded neatly in 58% yield and was followed by simultaneous deprotection of Fmoc and OAc protecting groups in 84% yield. The overall yield of the entire pathway was 20% (see Supplemental Information, Section S14 for further details).

The sixth target was 5β/6β-hydroxylurasidone (**43**, **43′** in Figure 3B), which is the main active metabolite of lurasidone, a US Food and Drug Administration-approved atypical antipsychotic drug for the treatment of schizophrenia and bipolar disorder.[29] The objective for Chematica in this exercise was to design an efficient pathway that would avoid the only known but patented route[30] (see Supplemental Information, Section S15). The path designed by Chematica starts from anhydride **35** and diol **36**. The former was converted into imide **37** in quantitative yield, whereas the latter was protected (indicated by a blue halo in Chematica's plan) with *tert*-butyldimethylsilyl chloride and then iodinated into **38** in 94% yield. The imide was alkylated with the iodide to give **39** in 93% yield. This was followed by deprotection in 87% yield, activation to iodide **40** (99%), *N*-alkylation of 1,2-benzisothiazole **41** to produce **42** (82%), and finally hydroxylation of the double bond. We note that in the original patent,[30] hydroxylation was performed early on, before the potentially fragile 1,2-benzisothiazole moiety was installed. However, Chematica judged that this moiety would not be affected during hydroxylation, as indeed was verified experimentally and rewarded by an over 90% yield (versus 37% in the patented route). All in all, the route proved easily scalable to multigram quantities and gave an overall yield of ~55%; that is, two and a half times higher than reported in the patented route.

The ability to find routes significantly different from patented syntheses (see list of 46 patents in the Supplemental Information, Section S16.1) also motivated the synthesis of the seventh target, Sanofi-Aventis' dronedarone (**51** in Figure 3C). Chematica's route began with a Sonogashira coupling between aryl iodide **44** and 1-hexyne to give **45**. The next, key step was the palladium-catalyzed carbonylative annulation of alkyne **45**, aryl iodide **46**, and carbon monoxide.[31,32] This three-component reaction was not used in any previous syntheses of dronedarone, but we found it to proceed neatly in 76% yield to construct the entire central ring system of the target. Subsequent steps were straightforward (see Supplemental Information, Section S16), resulting in an overall pathway yield of 39.6%, which is comparable with the 41% yield of the Sanofi-Aventis synthesis;[33] the latter, however, starts from a nitrobenzofuran derivative (see Supplemental Information, Section S16.1), which is a more advanced intermediate than the simple 2-iodo-4-nitrophenol starting material Chematica selected (on the flipside of the coin, the iodide generated in the Sonogashira reaction might be costly to dispose of, which could be problematic if the synthesis was ever carried out on industrial scales).

Finally, the eighth target was engelheptanoxide C, **56**, which is a natural product recently isolated from stems of *Engelhardia roxburghiana* but not yet synthesized.[19] The main virtue of Chematica's pathway in Figure 3D is that the program was able to construct an elegant and convergent route by using chemistries most appropriate for this type of a scaffold. Specifically, a "modern" enantioselective allylation of a primary

alcohol **53** according to Krische's protocol[34] ([Ir(cod)Cl]$_2$, Cs$_2$CO$_3$, (R)-BINAP) proceeded in 65% yield and 93% ee and avoided a step that a "classic" aldehyde allylation would entail (requiring oxidation and involving fragile alkylaldehyde). This step set the first stereocenter. The remaining two stereocenters were created in one step via the Prins cyclization of the tetrahydropyran ring.[35,36] Depending on the conditions used, this reaction gave either 45% yield and 72% ee or 30% yield and 88% ee (see Supplemental Information, Section S17). In both cases, the synthesis of **56** was completed by quantitative hydrogenation carried over Pd(OH)$_2$/C.

Conclusion

In summary, after over a decade of laborious development, Chematica is finally capable of designing novel and experimentally efficient syntheses of medicinally and industrially relevant targets. Guided by its scoring functions promoting synthetic brevity and penalizing any reactivity conflicts or non-selectivities, the program finds solutions that might be hard to identify by a human user; for instance, four of the described syntheses rely on hard-to-spot multicomponent reactions and in at least one case (5β/6β-hydroxylurasidone), the program made a choice that a chemist might consider "risky." However, these approaches and choices are not a question of Chematica's luck but rather manifestation of its dexterity in analyzing and disconnecting complex graphs coupled with its ability to consider simultaneously large numbers of logically related criteria (of groups' cross-reactivity, selectivity, etc.). Looking forward, future development and wider dissemination of Chematica are critically reliant on the expansion of the underlying computer infrastructure (i.e., multiprocessor machines potentially linked into larger clusters) required for the exploration of extremely large retrosynthetic trees. With such expansion backed by MS and in close academic collaborations with several leading synthetic groups, Chematica's next and perhaps ultimate aim is to attack the syntheses of very complex targets at the forefront of modern synthesis.

EXPERIMENTAL PROCEDURES

Full experimental procedures are provided in the Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information includes algorithmic details of Chematica and experimental details of the syntheses, 132 figures, 8 schemes, and 1 movie and can be found with this article online at https://doi.org/10.1016/j.chempr.2018.02.002.

AUTHOR CONTRIBUTIONS

S.S., K.M., E.P.G., P.D., M.P.S., R.R., and T.K. were key developers of Chematica. M.B. synthesized the BRD7/9 inhibitor. H.L. synthesized α-hydroxyetizolam. A.T. synthesized ATR kinase inhibitor. M.P.M. synthesized the anti-leukemia drug

Chem

candidate and T.K. and B.M.-K. synthesized 5β/6β-hydroxylurasidone, dronedarone, and (S)-4-hydroxyduloxetine. A.A. and B.S. helped in the synthesis of (S)-4-hydroxyduloxetine, and L.R. and G.J.K. provided helpful advice; B.S. also helped prepare substrates for the synthesis of dronedarone. Y.Z. synthesized engelheptanoxide C under supervision of M.M. S.L.J.T. conceived and supervised the technology evaluation project on behalf of MilliporeSigma. B.A.G. conceived Chematica in graduate school and has directed its development ever since. All authors contributed to the writing of the manuscript.

DECLARATION OF INTERESTS

After the successful experiments outlined in this manuscript, Chematica was acquired by Merck KGaA, Darmstadt, Germany. For various access options, including academic collaborations, please contact S.L.J.T. Part of the current research on Chematica is sponsored by Merck KGaA.

REFERENCES AND NOTES

1. Corey, E.J., and Wipke, W.T. (1969). Computer-assisted design of complex organic syntheses. Science 166, 178–192.

2. Gelernter, H.L., Sanders, A.F., Larsen, D.L., Agarwal, K.K., Boivie, R.H., Spritzer, G.A., and Searleman, J.E. (1977). Empirical explorations of SYNCHEM. Science 197, 1041–1049.

3. Hanessian, S., Franco, J., and Larouche, B. (1990). The psychobiological basis of heuristic synthesis planning – man, machine and the Chiron approach. Pure Appl. Chem. 62, 1887–1910.

4. Hendrickson, J.B. (1977). Systematic synthesis design. 6. Yield analysis and convergency. J. Am. Chem. Soc. 99, 5439–5450.

5. Bauer, J., Herges, R., Fontain, E., and Ugi, I. (1985). IGOR and computer-assisted innovation in chemistry. Chimia 39, 43–53.

6. Ugi, I., Bauer, J., Bley, K., Dengler, A., Dietz, A., Fontain, E., Gruber, B., Herges, R., Knauer, M., Reitsam, K., et al. (1993). Computer-assisted solution of chemical problems – the historical development and the present state of the art of a new discipline of chemistry. Angew. Chem. Int. Ed. 32, 201–227.

7. Höllering, R., Gasteiger, J., Steinhauer, L., Schultz, K., and Herwig, A. (2000). Simulation of organic reactions: from the degradation of chemicals to combinatorial synthesis. J. Chem. Inf. Comput. Sci. 40, 482–494.

8. Ravitz, O. (2013). Data-driven computer aided synthesis design. Drug Discov. Today Technol. 10, e443–e449.

9. Bøgevig, A., Federsel, H.-J., Huerta, F., Hutchings, M.G., Kraut, H., Langer, T., Löw, P., Oppawsky, C., Rein, T., and Saller, H. (2015). Route design in the 21st century: the ICSYNTH software tool as an idea generator for synthesis prediction. Org. Process. Res. Dev. 19, 357–368.

10. Szymkuć, S., Gajewska, E.P., Klucznik, T., Molga, K., Dittwald, P., Startek, M., Bajczyk, M., and Grzybowski, B.A. (2016). Computer-assisted synthetic planning: the end of the beginning. Angew. Chem. Int. Ed. 55, 5904–5937.

11. The literature on computational predictions validated in the synthetic practice is very scarce. In our own work,[10] we showed that Chematica was able to suggest several synthetic pathways in which each individual step had been previously confirmed in experimental papers by other groups. Bøgevig et al.[9] used the ICSYNTH program to suggest several synthetic pathways, which, however, were not carried out because the funding of the project was cut (only one cyclization reaction leading to oxaspiroketone was performed). One other notable contribution was from Ivar Ugi, who used his bond-electron matrix formalism to propose and then execute several novel pericyclic reactions and a novel rearrangement of α-aminoalkylboranes.[6] More recently, Currie and Goodman[12] used a combination of reaction-template-matching and energy calculations to predict as feasible (under acidic conditions) and then execute experimentally (but under basic conditions) a retro-Claisen rearrangement of an intermediate hemiacetal in the multistep synthesis of (−)-dolabriferol. The remaining 18 steps of the synthesis were planned manually (i.e., not by the computer).

12. Currie, R.H., and Goodman, J.M. (2012). In silico inspired total synthesis of (−)-Dolabriferol. Angew. Chem. Int. Ed. 51, 4695–4697.

13. Fialkowski, M., Bishop, K.J., Chubukov, V.A., Campbell, C.J., and Grzybowski, B.A. (2005). Architecture and evolution of organic chemistry. Angew. Chem. Int. Ed. 44, 7263–7269.

14. Bishop, K.J.M., Klajn, R., and Grzybowski, B.A. (2006). The core and most useful molecules in organic chemistry. Angew. Chem. Int. Ed. 45, 5348–5354.

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

16. Kowalik, M., Gothard, C.M., Drews, A.M., Gothard, N.A., Wieckiewicz, A., Fuller, P.E., and Grzybowski, B.A. (2012). Parallel optimization of synthetic pathways within the network of organic chemistry. Angew. Chem. Int. Ed. 51, 7928–7932.

17. Gunther, M. (2016) Revolutionary computer program could change chemistry forever. Chemistry World. https://www.chemistryworld.com/news/software-could-revolutionise-chemistry/1017236.article.

18. Lowe, D. (2016) The algorithms are coming. http://blogs.sciencemag.org/pipeline/archives/2016/04/12/the-algorithms-are-coming.

19. Wu, H.C., Cheng, M.J., Peng, C.F., Yang, S.C., Chang, H.S., Lin, C.H., Wang, C.J., and Chen, I.S. (2012). Secondary metabolites from the stems of Engelhardia roxburghiana and their antitubercular activities. Phytochemistry 82, 118–127.

20. Clark, P.G.K., Vieira, L.C.C., Tallant, C., Fedorov, O., Singleton, D.C., Rogers, C.M., Monteiro, O.P., Benett, J.M., Baronio, R., Mueller, S., et al. (2015). LP99: discovery and synthesis of the first selective BRD7/9 bromodomain inhibitor. Angew. Chem. Int. Ed. 54, 6217–6221.

21. It is worth mentioning that the aza-Henry reaction required a large excess (~20 equiv) of methyl 4-nitrobutanoate, which contrasts with conditions reported in literature for other substrates.[22] Also, in the presence of strong acids or bases, 4 reverts back to starting amine

1 by a retro aza-Henry reaction. A small amount of the retro aza-Henry product was also observed during LiOH-mediated hydrolysis of **4**. Finally, we note that for this reaction, Chematica provided conditions from Cruz-Acosta et al.,[22] where either enantioselective or diastereoselective catalyst systems were reported. Because Schreiner's catalyst was commercially available (i.e., the enantioselective catalyst didn't need to be synthesized in house), the diastereoselective route was pursued.

22. Cruz-Acosta, F., de Armas, P., and Garcia-Tellado, F. (2013). Water-compatible hydrogen-bond activation: a scalable and organocatalytic model for the stereoselective multicomponent Aza-Henry reaction. Chem. Eur. J. *19*, 16550–16554.

23. Miki, A., Tatsuno, M., Katagi, M., Nishikawa, M., and Tsuchihashi, H. (2002). Simultaneous determination of eleven benzodiazepine hypnotics and eleven relevant metabolites in urine by column-switching liquid chromatography-mass spectrometry. J. Anal. Toxicol. *26*, 87–93.

24. Stransky, W., Weber, K., Walther, G., Harreus, A., Stenzel, J.C., Muacevic, G., and Bechtel, W.D. (1990). Thieno-1,4-diazepines. Patent US 4900729 A, filed October 11, 1988, and published February 13, 1990.

25. Foote, K.M., Blades, K., Cronin, A., Fillery, S., Guichard, S.S., Hassal, L., Hickson, I., Jacq, X., Jewsbury, P.J., McGuire, T.M., et al. (2013). Discovery of 4-{4-[(3R)-3-methylmorpholin-4-yl]-6-[1-(methylsulfonyl)cyclopropyl]pyrimidin-2-yl}-1H-indole (AZ20): a potent and selective inhibitor of ATR protein kinase with monotherapy in vivo antitumor activity. J. Med. Chem. *56*, 2125–2138.

26. Getlik, M., Smil, D., Zepeda-Velazquez, C., Bolshan, Y., Poda, G., Wu, H., Dong, A.P., Kuznetsova, E., Marcellus, R., Senisterra, G., et al. (2016). Structure-based optimization of a small molecule antagonist of the interaction between WD repeat-containing protein 5 (WDR5) and mixed-lineage leukemia 1 (MLL1). J. Med. Chem. *59*, 2478–2496.

27. Sigma-Aldrich. KitAlysis High-Throughput Screening Kits. https://www.sigmaaldrich.com/chemistry/chemical-synthesis/kitalysis-high-throughput-screening-kits.html.

28. Kuo, F., Gillespie, T.A., Kulanthaivel, P., Lanz, R.J., Ma, T.W., Nelson, D.L., Threlkeld, P.G., Wheeler, W.J., Yi, P., and Zmijewski, M. (2004). Synthesis and biological activity of some known and putative duloxetine metabolites. Bioorg. Med. Chem. Lett. *14*, 3481–3486.

29. Meyer, J.M., Loebel, A.D., and Schweizer, E. (2009). Lurasidone: a new drug in development for schizophrenia. Expert Opin. Investig. Drugs *18*, 1715–1726.

30. Kojima, A., Ono, Y., and Yoshigi, M. (1996). New imide derivative. Patent JP 08333368, filed June 9, 1995, and published December 17, 1996.

31. Hu, Y., Zhang, Y., Yang, Z., and Fathi, R.J. (2002). Palladium-catalyzed carbonylative annulation of o-alkynylphenols: syntheses of 2-substituted-3-aroyl-benzo[*b*]furans. J. Org. Chem. *67*, 2365–2368.

32. Nakamura, M., Ilies, L., Otsubo, S., and Nakamura, E. (2006). 2,3-Disubstituted benzofuran and indole by copper-mediated C−C bond extension reaction of 3-zinciobenzoheterole. Org. Lett. *8*, 2803–2805.

33. Gubin, J., Chatelain, P., Lucchetti, J., Rosseells, G., Inion H. (1993). Alkylaminoalkyl derivatives of benzofuran, benzothiophene, indole and indolizine, process for their preparation and compositions containing them. Patent US 5223510, filed July 26, 1991, and published June 29, 1993.

34. Kim, I.S., Ngai, M.-Y., and Krische, M.J. (2008). Enantioselective iridium-catalyzed carbonyl allylation from the alcohol or aldehyde oxidation level via transfer hydrogenative coupling of allyl acetate: departure from chirally modified allyl metal reagents in carbonyl addition. J. Am. Chem. Soc. *130*, 14891–14899.

35. Tadpetch, K., and Rychnovsky, S.D. (2008). Rhenium(VII) catalysis of Prins cyclization reactions. Org. Lett. *10*, 4839–4842.

36. Crosby, S.R., Harding, J.R., King, C.D., Parker, G.D., and Willis, C.L. (2002). Oxonia-Cope rearrangement and side-chain exchange in the Prins cyclization. Org. Lett. *4*, 577–580.