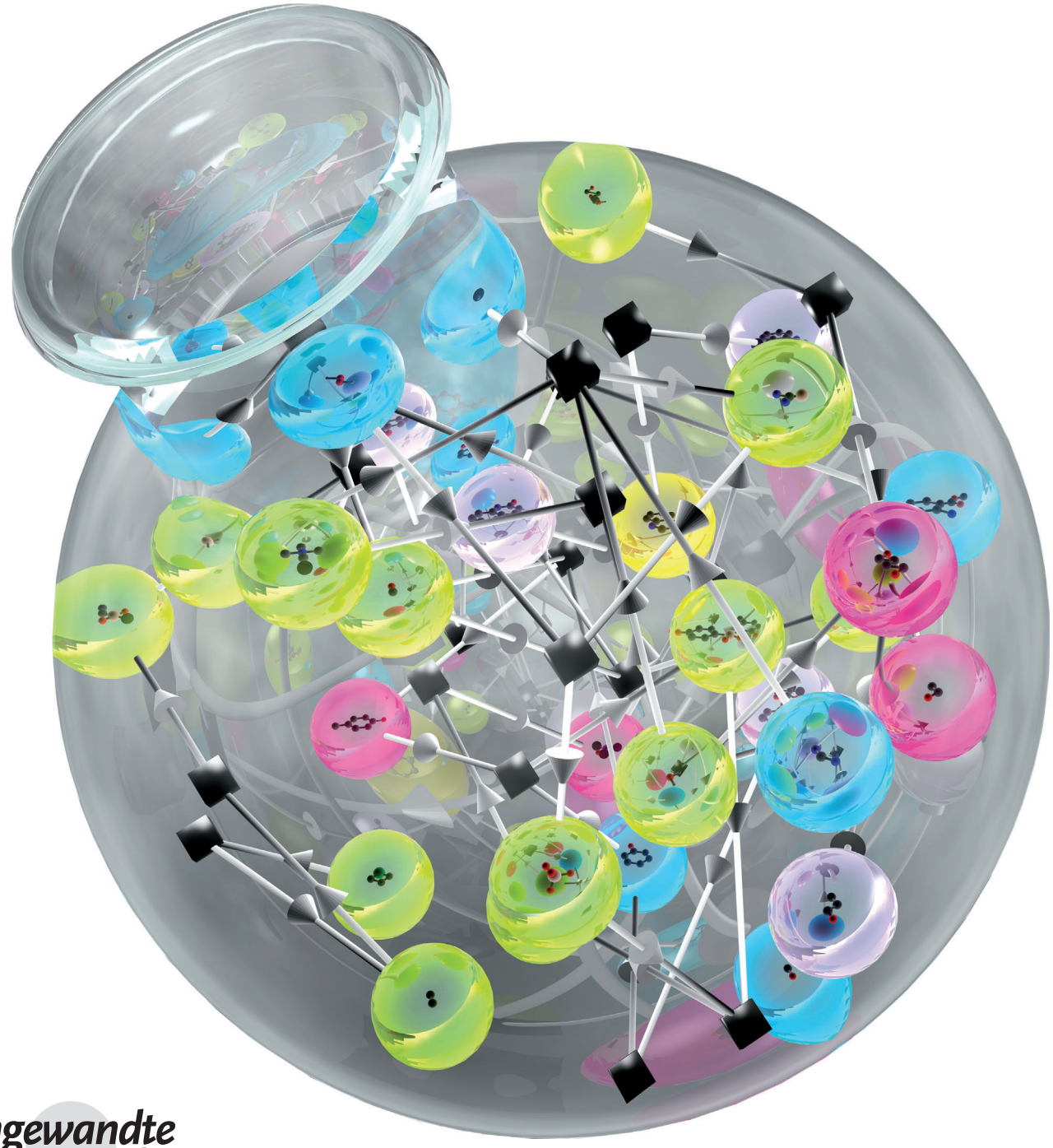
Chemical Networks (1)

DOI: 10.1002/anie.201202155

Rewiring Chemistry: Algorithmic Discovery and Experimental Validation of One-Pot Reactions in the Network of Organic Chemistry**

*Chris M. Gothard, Siowling Soh, Nosheen A. Gothard, Bartlomiej Kowalczyk, Yanhu Wei, Bilge Baytekin, and Bartosz A. Grzybowski**



In 2005 and 2006, we published the first reports[1a,b] on the representation of all synthetic knowledge as a giant network in which molecule "nodes" are connected by reaction "arrows" (Figure 1). In these early works, we focused on the topological structure and evolution of this network and

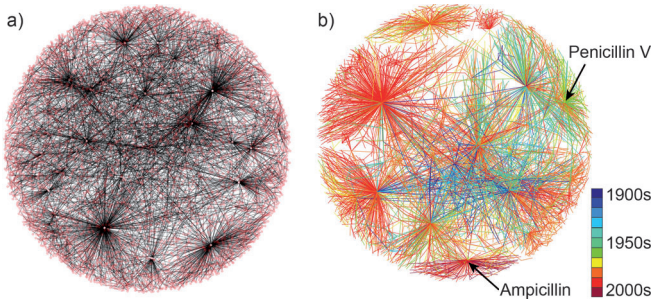


Figure 1. a) A small (ca. 5500 nodes, ca. 0.1% of the total) fragment of the network of organic chemistry (NOC), where individual nodes represent the molecules and arrows represent reactions. The representation in b) has the reaction arrows colored by the times these reactions were first reported. This representation emphasizes the fact that NOC, by itself, is not a "coherent" giant chemical system but only a repository of reactions discovered separately, without regard for their mutual compatibility. At best, it can be said that there was a "coherent" interest in certain areas of chemistry (for example, synthetic activity around the Penicillin V node in the 1960s, following the first total synthesis).

demonstrated the scale-free network topology, existence of hub molecules central to organic synthesis, exponential growth of the network in time, correlations between molecular masses,[1a] trends in reactivity based on network connectivity,[1c,d] and more. While our analyses had little applicability to the everyday synthetic practice, we envisioned[1e] that such a junction between network theory and synthesis would one day be achieved. Now, we are reporting, in three consecutive communications,[14] the extension of chemical-network concepts into methods directly relevant to experimental chemistry: 1) discovery of one-pot reactions; 2) optimization of multiple reaction pathways, and 3) the detection and blocking of synthetic pathways leading to dangerous chemicals. The first communication in this series addresses one of the most important challenges in organic chemistry: namely, how to "wire" individual reactions into sequences that could be performed in one pot. One-pot reactions[2–5] save resources and time by avoiding isolation, purification, characterization, and production of chemical waste after each synthetic step. Sometimes, such reactions are identified by chance or, more often, by careful inspection of individual steps that are to be wired together; even this latter process, however, is invariably subjective and depends on the knowledge and intuition of any individual chemist (or group of chemists) involved. Herein, we show that the discovery of one-pot reactions can be facilitated by computational methods. We first describe algorithms that identify possible one-pot reactions within the network of all known synthetic knowledge and then demonstrate that the computationally predicted sequences can indeed be carried out experimentally in good overall yields. The experimental examples are chosen to "rewire" small networks of reactions around practically important chemicals: quinoline scaffolds, quinoline-based enzyme inhibitors, and thiophene derivatives. In this way, we replace individual synthetic connections with two-, three-, or even four-step one-pot sequences.

The network of organic chemistry (NOC; Figure 1) is constructed from reactions reported in the chemical literature since 1779 and nowadays stored in chemical databases. Pruning the raw data[1a,b] to remove catalysts, solvents, substances that do not participate in reactions, and duplicate or incomplete reactions, leaves about 7 million reactions and about 7 million substances on which further analyses are based. This dataset is translated into a network by representing chemical substances as network nodes, and the reactions as arrows directed from the reaction's substrates to products.

At first glance, this giant network of chemistry might look akin to the metabolic networks of biochemical reactions. In reality, however, metabolic networks are true chemical systems comprising reactions that can, in most cases, occur concurrently within the same reaction medium (that is, in a cell); in contrast, the network of organic chemistry is a collection of individual reactions performed by different chemists at different times (Figure 1b), typically under different conditions and in different solvents. Our goal is to find within NOC the combinations of reactions that can be wired into the simplest systems, that is, the linear, one-pot sequences.

To do so, the possible reaction sequences within NOC are evaluated for one-pot-compatibility by several computational criteria. The initial, trivial check is whether for a given sequence (for example, a two-step A→B→C), there already exists within NOC a direct A→C connection; if so, this sequence is no longer considered. Assuming no A→C connection is known, the algorithm applies various filters to determine the compatibility of the individual reactions to be wired together. Filter **#1** checks for the compatibility of functional groups on all molecules participating in a putative sequence. Specifically, a house-written program is first used to unambiguously partition each of the molecules into functional groups taken from a list of 322 common chemical functionalities (Figure 2a; Supporting Information, Section 1). The constituent groups are then compared against a 322×322 "master" matrix where all possible group combinations are classified[6] as mutually unreactive (that is, compatible; gray

[*] Dr. C. Gothard,[+] Dr. S. Soh,[+] N. Gothard, Dr. B. Kowalczyk, Dr. Y. Wei, Dr. B. Baytekin, Prof. B. A. Grzybowski
Department of Chemistry, Northwestern University
2145 Sheridan Rd., Evanston, IL 60208 (USA)
E-mail: grzybor@northwestern.edu
Homepage: http://dysa.northwestern.edu

[+] These authors contributed equally to this work.

[**] This work was supported by the Non-equilibrium Energy Research Center (NERC)—an Energy Frontier Research Center funded by the U.S. Department of Energy under Award Number DE-SC0000989. B.A.G. would like to thank his former students Dr. C. Campbell (3M), Prof. M. Fialkowski (Polish Academy of Sciences), Prof. R. Klajn (Weizmann Institute), and Dr. S. Smoukov (Cambridge University) for their help in the early development of NOC.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201202155.

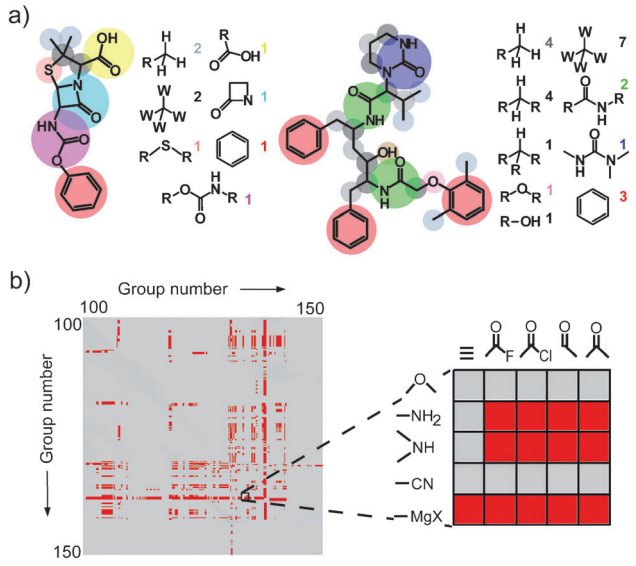


Figure 2. Illustration of the group-compatibility filter (**#1**). a) Examples of algorithmic partitioning of molecules into specific functional groups. The full list of possible 322 groups is included in the Supporting Information, Section 1. b) A large fragment (left) and further magnification (right) of the group-compatibility 322×322 master matrix used to determine the compatibility or incompatibility of groups involved in a putative one-pot sequence. The classifications are made assuming typical reaction conditions (see [11]). The zoomed fragment contains some familiar group combinations and illustrates their well-known reactivity trends (for example, ethers are poor nucleophiles and generally unreactive, primary and secondary amines, on the other hand, are reactive towards all kinds of electrophiles, and so on).

entries in Figure 2b) or reactive (incompatible; red entries). If incompatibilities are found, the algorithm suggests the order of addition to avoid conflicts (for instance, in a two-step A→B→C sequence, if A reacts with C, then A must be consumed before the reagents to make C are added). Filter **#2** verifies whether reaction conditions required in each step would cause unintended reactions of functional groups in other steps. These rules are summarized in the form of a table comprising 97 typical reaction types/conditions versus 322 functional groups (Supporting Information, Section 2). Filter **#3** checks for the compatibility of solvents using a digitized version of the well-known solvent miscibility tables (although reactions can occur at solvent interfaces, it is not efficient for one-pot reactions to combine solvents that are not miscible). Filters **#4**–**#8** are based on the compatibility of 600 common reagents and are summarized in the Supporting Information, Section 3. For instance, filter **#4** checks for anhydrous versus aqueous conditions (for example, in Gattermann reactions, which install aldehyde groups in aryl systems under aqueous conditions, subsequent one-pot steps cannot involve water-sensitive reactants or reagents, such as Grignard compounds, alkali metal hydrides, or organolithium reagents). In a similar way, filter **#5** checks for oxidizing versus reducing conditions, which are incompatible unless the oxidizing or reducing reagents can be converted into unreactive spectator species. Filter **#6** determines acid–base compatibility and alerts the user as to whether acidic, basic, or neutral conditions are incompatible among reaction steps, such that the addition of acids/bases at specific times, needs to be planned. Finally, filter **#7** checks for the incompatibilities in terms of hydride/proton sources, and filter **#8** checks for the compatibility of chemical groups on the reagents (akin to filter **#1** for substrates/products). Although the rules stored in the filter tables comprise over 86000 chemical criteria to evaluate candidate one-pot sequences, the entire analysis takes only a small fraction of a second on a typical desktop computer. Also, when suitable reactions are identified and potential conflicts resolved by proper reaction timing (that is, order of addition), an optional step in the algorithm is to check for the commercial availability of the substrates and reagents (in the current version of our software, against the list of ca. 20000 chemicals, mostly from Sigma Aldrich).

Of course, the true value of any theoretical–chemical algorithm is in experimental validation. In principle, the method can be tested to identify one-pot reactions from among any of the possible 1.8 billion two-step sequences present within the NOC. While our algorithm has already identified over a million (and counting!) possible sequences, such randomly chosen reactions might be of no real-world interest, and so herein we chose to illustrate the performance of the method by "wiring" reaction sequences within classes of compounds that are of popular interest and/or practical importance. As the examples in Figure 3 and 4 span 27 one-pot syntheses (14 two-step, 12 three-step, and 1 four-step sequences), the main text focuses on their key aspects (full experimental procedures and structural characterization of all compounds are included in the Supporting Information, Section 7). Finally, we emphasize that all the sequences and yields reported below are based on the one-pot procedures as suggested by the algorithm; that is, without any human "tinkering" to optimize the yields, and so on (for comparison of one-pot versus sequential reactions yields, see the Supporting Information, Section 4).

Figure 3 has the examples of one-pot reactions involving quinoline-based molecules. We chose these molecules because 1) quinolines are core structural units in medicinal plant alkaloids, drug therapeutics[7] (notably, antimalarial drugs[7b]), dyes, and bioactive materials;[7b] and 2) the predicted sequences illustrate well the performance of the algorithm in the detection and avoidance of cross-reactivity conflicts. Initiating the searches in the network vicinity of 2,4-dihyroxyquinoline (**1** in Figure 3a), which is a common and commercially available precursor for many C2- and C4-substituted quinolines, our algorithm predicted seven viable two-step one-pot reactions (red dashed arrows) and two three-step sequences (blue dashed arrows). These plausible sequences involved the chlorination of **1**, followed by arylation, alkynylation, or amination at the C2 and/or C4 positions and were subsequently carried out with overall experimental yields indicated next to the arrows. In evaluating these syntheses, the functional group compatibility filters **#1** and **#8** (compare with Figure 2b) detected a reactivity conflict between aryl chlorides and alkylating reagents (especially, boronic acids for Suzuki couplings and alkynes for Sonogashira reaction, which can both alkynylate or arylate the C2 position) and between some of the reagents or reaction

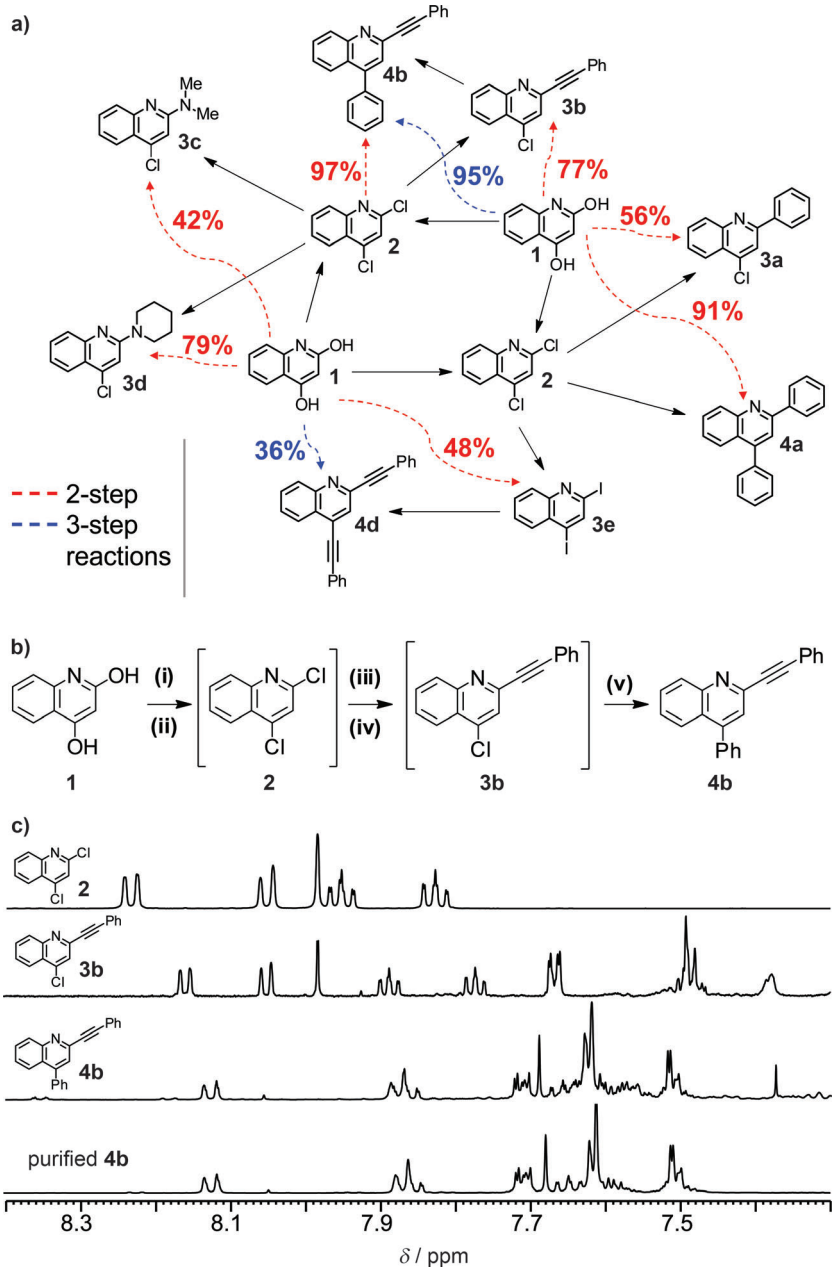


Figure 3. a) Individual reactions reported in NOC (black arrows) and the "wired" one-pot sequences predicted and carried out with the indicated yields (for all experimental details, see the Supporting Information, Section 7). Note that **1** and **2** are shown twice to make the scheme look less congested. b) An example of a three-step one-pot sequence involving i) chlorination of 2,4-dihydroxyquinoline (**1**) with $POCl_3$ to generate the corresponding 2,4-dichloroquinoline (**2**), ii) evaporation of phosphoryl chloride ($POCl_3$)/HCl and iii) basification with Cs_2CO_3 (aq), iv) alkylation with phenylacetylene at the C2 position to provide **3b** and, finally, v) alkylation at the C4 position with phenylboronic acid to give product **4b**. The ^{1}H NMR spectra in c) illustrate reaction progress and the purity of the final product **4b** (see also the Supporting Information, Section 7).

conditions (filters **#2**, **#4**, and **#6**); together, these algorithmic guidelines translated into the proper timing of reagent's addition. Figure 3b provides one illustrative example where a three-step sequence involving chlorination, Sonogashira, and Suzuki couplings was carried out cleanly, without isolating **2** and **3b**, with 95% overall yield. In this example, chlorination with $POCl_3$ (water-sensitive, carried out under acidic conditions) was found by filters **#4** and **#6** to be incompatible with the basic and aqueous conditions common in many Pd-catalyzed cross-coupling reactions. Accordingly, with 2,4-dichloroquinoline (**2**) generated almost quantitatively in situ, the reaction was basified before proceeding to C2 alkynylation and then, under the same conditions, to the final Suzuki coupling. The ^{1}H NMR spectra in Figure 3c showcase the clean transitions among the steps.

To further challenge the detection capabilities of our program, we focused on quinoline platforms that use acyclic precursors for the ring construction. This approach allows for a greater variety of substituents and access to more types of therapeutic targets such as the inhibitors of phosphoinositide 3-kinase delta (PI3Kδ), a key enzyme in the signaling pathway involved in airway inflammation.[8,9] Figure 4a shows the network of syntheses of several PI3Kδ inhibitors, inhibitor precursors, or closely related compounds with four two-step, eight three-step, and one four-step one pot sequences predicted and then validated experimentally. Here, the algorithmic compatibility filters gave predictions/suggestions similar to those we discussed in the context of Figure 3 (that is, changing to basic conditions after the cyclization and chlorination steps followed by alkynylation, arylation, or amination). Furthermore, the algorithm correctly identified that there were no group/reaction condition incompatibilities in the initial cyclization and chlorination steps. Perhaps the most striking example of our method's effectiveness is the prediction, and then execution, of a four-step one-pot sequence combining cyclization, chlorination, alkynylation, and arylation. This sequence, indicated by a violet arrow in Figure 4a and highlighted in Figure 4b, was carried out with an overall yield of 49%. We note that the algorithm-identified sequences in Figure 4a provide an attractive approach for large-scale preparations, as they allow for flexible and regioselective introduction of substituents using acyclic precursor **5** and a substituted aniline **6c**.

The final set of examples (Figure 4c) deals with the one-pot syntheses of thiophenes, which are of widespread interest owing to their multifunctional properties, adaptability, chemical robustness, and applications in electronics and photovoltaics.[10] Central to our thiophene network is 3-bromothiophene, which was detected by network searches as a synthetic

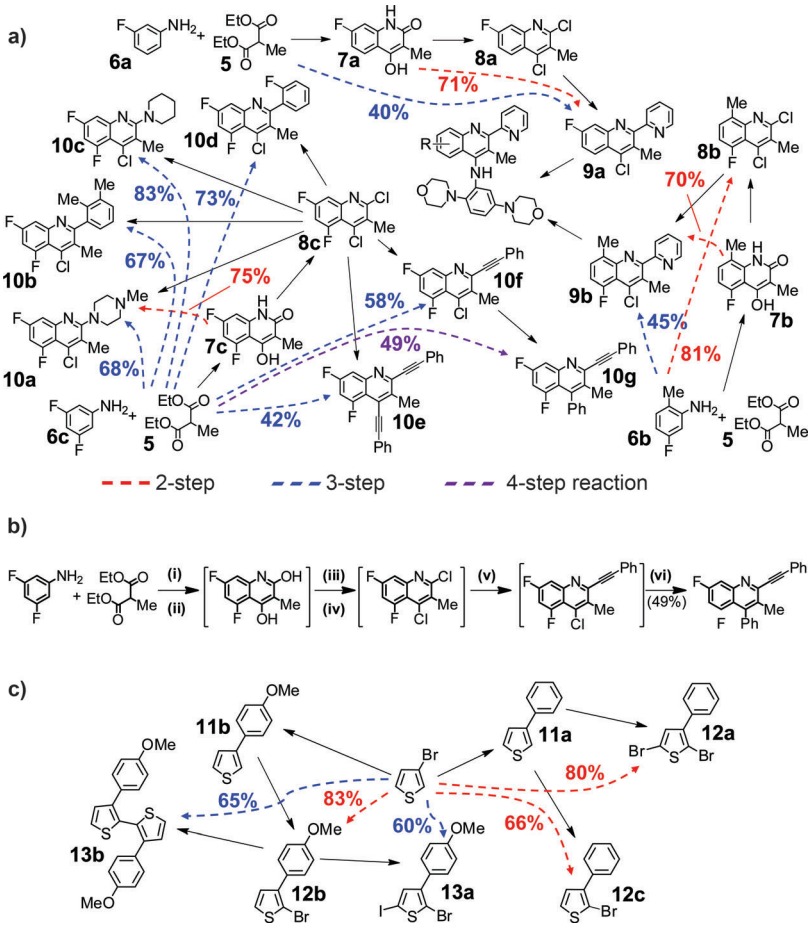


Figure 4. Rewiring reaction networks of individual reactions involving PI3Kδ inhibitors and substituted thiophenes. Previously reported individual reactions correspond to black arrows; two-step sequences are represented by red arrows, three-step sequences by blue arrows, and a four-step sequence is denoted by a purple arrow. a) Network of PI3Kδ inhibitors or precursors and closely related compounds. b) Four-step one-pot synthesis of **10g**: i), ii) cyclization, iii), iv) chlorination, v) Sonogashira reaction, and vi) Suzuki coupling. c) A small reaction network centered on 3-bromothiophene. For experimental details, see the Supporting Information, Section 5.

"hub" and thus a key starting material within our thiophene network. This small, "rewired" network comprises three two-step and two three-step one-pot sequences involving four types of reactions: arylation (Suzuki cross coupling), bromination, iodination, and dimerization. In the pathways going through **11a** to **12a** or **12c**, the program deemed the reactions compatible with the proviso that the basic conditions of the Suzuki coupling be changed to acidic for the bromination step; this was achieved using acetic acid, which is a common medium for NBS bromination reactions[11] and can scavenge remaining metal catalyst. In the pathways going through **11b** (prepared using Grignard reagent and $[NiCl_2(dppp)]$ catalyst), regioselective bromination led to **12b**, the iodination of which then completed the three-step sequence to give **13a**. In a similar fashion, after forming **12b** by a Grignard bromination sequence, a second Grignard reaction was carried out to provide homodimerization product **13b**.

Having discussed the success cases, it is important to outline the pitfalls of the method. While our algorithm has so far generated over a million structurally diverse one-pot sequences, it is clearly impossible to validate all of them experimentally. Instead, we estimated the likelihood of false-positive predictions by closely inspecting about 500 predicted sequences and cross-checking them against the original research describing the constituent/individual reactions. In few percent of cases, the predicted sequences turned out to be unfeasible because the underlying chemical databases did not report, or reported incorrectly, the key reagents or reaction conditions present in the original reports. This result underscores the need for faithful translation of the literature data into chemical database content. A much less frequent source of errors (only few cases we encountered so far) is the algorithm's incomplete "knowledge" of the mechanistic details of the reactions to be wired. One illustrative example is included in the Supporting Information, Section 5, where a predicted sequence failed experimentally because of an unforeseen transformation of Lawesson's reagent into species reactive toward one of the intermediates. We recognize that there is an ongoing need to improve the filters/rules that our algorithm uses; the goal is that such improvements will ultimately render the algorithm on a par with the detailed synthetic knowledge of experienced organic chemists (including stereoselective syntheses; see the next communication, Figure S12).[14a]

In summary, we have described one-pot reactions that were first "blindly" discovered by a computational method (that is, without any human guidance during algorithm execution) and then validated experimentally. The current work is to the best of our knowledge the first experimentally validated demonstration of a computer-driven discovery of one-pot sequences. A software product for NOC searches, one-pot reaction analyses, optimization of synthetic pathways, and other functions described in the following two communications[14] will be commercially available later in 2012.[12] In the meantime, our immediate goal is to extend the algorithm to the discovery of multicomponent reactions such as the famous Ugi reaction.[13a] Incidentally, it was the late Ivar Ugi who first considered the use of computers for optimizing and rewiring chemical syntheses.[13b,c] Three decades later, we finally have computers and algorithms powerful enough to rewire chemistry into an optimal and coherent chemical systems. As wiring individual computers together gave rise to the internet revolution, so will the wiring of chemical reactions into systems give rise to the revolution of the "chemical internet". Our aim with this and the subsequent two communications in the series[14] is to illustrate how powerful a tool the chemical internet can

become in connecting molecules through synthetically advantageous or even optimal pathways.

Received: March 19, 2012
Revised: May 18, 2012
Published online: July 13, 2012

Keywords: algorithms · chemical networks · one-pot reactions · synthetic methods

[1] a) M. Fialkowski, K. J. M. Bishop, V. A. Chubukov, C. J. Campbell, B. A. Grzybowski, *Angew. Chem.* **2005**, *117*, 7429 – 7435; *Angew. Chem. Int. Ed.* **2005**, *44*, 7263 – 7269; b) K. J. M. Bishop, R. Klajn, B. A. Grzybowski, *Angew. Chem.* **2006**, *118*, 5474 – 5480; *Angew. Chem. Int. Ed.* **2006**, *45*, 5348 – 5354; c) B. Kowalczyk, K. J. M. Bishop, S. Smoukov, B. A. Grzybowski, *J. Phys. Org. Chem.* **2009**, *22*, 897 – 902; d) S. Soh, Y. Wei, B. Kowalczyk, C. M. Gothard, B. Baytekin, M. Branicki, B. A. Grzybowski, *Chem. Sci.* **2012**, *3*, 1497 – 1502; e) B. A. Grzybowski, K. J. M. Bishop, B. Kowalczyk, C. E. Wilmer, *Nat. Chem.* **2009**, *1*, 31 – 36.

[2] a) S. Kobayashi, *Chem. Soc. Rev.* **1999**, *28*, 1 – 15; b) L. Li, G. Zhang, A. Zhu, L. Zhang, *J. Org. Chem.* **2008**, *73*, 3630 – 3633; c) D. Sureshkumar, S. M. Koutha, S. Chandrasekaran, *J. Am. Chem. Soc.* **2005**, *127*, 12760 – 12761; d) J.-C. Wasilke, S. J. Obrey, R. T. Baker, C. G. Bazan, *Chem. Rev.* **2005**, *105*, 1001 – 1020; e) R. Mossetti, T. Pirali, D. Saggiorato, C. G. Tron, *Chem. Commun.* **2011**, *47*, 6966 – 6968; f) A. Dömling, *Chem. Rev.* **2006**, *106*, 17 – 89; g) D. Enders, C. Wang, J. W. Bats, *Angew. Chem.* **2008**, *120*, 7649 – 7653; *Angew. Chem. Int. Ed.* **2008**, *47*, 7539 – 7542.

[3] For optimally timed addition of reagents/solvents in one-pot reactions, see, for example: a) M. Galibert, O. Renaudet, P. Dumy, D. Boturyn, *Angew. Chem.* **2011**, *123*, 1941 – 1944; *Angew. Chem. Int. Ed.* **2011**, *50*, 1901 – 1904; b) L. Albrecht, L. K. Ransborg, B. Gschwend, K. A. Jorgensen, *J. Am. Chem. Soc.* **2010**, *132*, 17886 – 17893; c) A. Grirrane, A. Corma, H. Garcia, *Science* **2008**, *322*, 1661 – 1664; d) C. Vaxelaire, P. Winter, M. Christmann, *Angew. Chem.* **2011**, *123*, 3685 – 3687; *Angew. Chem. Int. Ed.* **2011**, *50*, 3605 – 3607; e) E. W. Dunn, G. W. Coates, *J. Am. Chem. Soc.* **2010**, *132*, 11412 – 11413.

[4] For discussion of one-pot reactions maximizing "atom economy" and "step economy", see: a) N. Z. Burns, P. S. Baran, R. W. Hoffmann, *Angew. Chem.* **2009**, *121*, 2896 – 2910; *Angew. Chem. Int. Ed.* **2009**, *48*, 2854 – 2867; b) P. A. Wender, V. A. Verma, T. J. Paxton, T. H. Pillow, *Acc. Chem. Res.* **2008**, *41*, 40 – 49; c) T. Newhouse, P. S. Baran, R. W. Hoffmann, *Chem. Soc. Rev.* **2009**, *38*, 3010 – 3021; d) T. Gaich, P. S. Baran, *J. Org. Chem.* **2010**, *75*, 4657 – 4673; e) B. M. Trost, *Science* **1991**, *254*, 1471 – 1477.

[5] For examples of one-pot reactions where both the order and time in which each reagent is added are important, see: a) L. F. Tietze, U. Beifuss, *Angew. Chem.* **1993**, *105*, 137 – 170; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 131 – 163; b) A. M. Jacobine, G. H. Posner, *J. Org. Chem.* **2011**, *76*, 8121 – 8125.

[6] This (and other filter tables in the Supporting Information) was constructed by independent analyses by five Ph.D.-level organic chemists (the authors); the reactive/unreactive classifications of all entries were then reached by consensus and assuming typical reaction conditions.

[7] a) B. S. Moon, B. S. Lee, D. Y. Chi, *Bioorg. Med. Chem.* **2005**, *13*, 4952 – 4959; b) J. P. Michael, *Nat. Prod. Rep.* **2007**, *24*, 223 – 246.

[8] a) Amgen Inc. Patent US2010/331293A1, **2010**; b) Amgen Inc. Patent WO2010/151737A2, **2010**.

[9] S. J. Park, K. H. Min, C. L. Lee, *Respirology* **2008**, *13*, 764 – 771.

[10] M. Muccini, *Nat. Mater.* **2006**, *5*, 605 – 613.

[11] Z. Li, J. Ding, N. Song, J. Lu, Y. Tao, *J. Am. Chem. Soc.* **2010**, *132*, 1360 – 1361.

[12] Disclosure: B.A.G. is the founder and President of the GSI company that will market this software; he is also the C.S.O of ProChimia Surfaces (see analyses in the next Communication). See also the movie posted at: http://dysa.northwestern.edu/chematica-demonstration.html, and also the Supporting Information, Section 6).

[13] a) I. Ugi, R. Meyr, U. Fetzer, C. Steinbrückner, *Angew. Chem.* **1959**, *71*, 386; b) I. Ugi, et al., *Pure Appl. Chem.* **1988**, *60*, 1573 – 1586; c) I. Ugi, et al., *Angew. Chem.* **1993**, *105*, 210 – 239; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 201 – 227.

[14] a) M. Kowalik, C. Gothard, A. M. Drews, N. Gothard, A. Wieckiewicz, P. E. Fuller, B. A. Grzybowski, K. J. M. Bishop, *Angew. Chem.* **2012**, *124*, 8052; *Angew. Chem. Int. Ed.* **2012**, *51*, 7928; b) P. E. Fuller, C. Gothard, N. Gothard, A. Wieckiewicz, B. A. Grzybowski, *Angew. Chem.* **2012**, *124*, 8057; *Angew. Chem. Int. Ed.* **2012**, *51*, 7933.

