ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2025, **61**, 3528

Received 17th December 2024,
Accepted 30th January 2025

DOI: 10.1039/d4cc06605g

rsc.li/chemcomm

Phenotype-directed discovery of diverse, biologically-relevant molecular scaffolds†

Samuel D. Griggs,[abc] Amalia-Sofia Piticari,[c] Samuel Liver,[c] Chris Arter,[c] Sonja Sievers,[de] Stephen P. Marsden *[a] and Adam Nelson *[abc]

An array of reactions of diazo substrates with many possible outcomes was executed, and the biological relevance of the resulting products assessed in the cell painting assay. Reactions that had yielded bioactive products were scaled-up, and the products structurally elucidated. By bypassing the need to characterise all reaction products, this phenotype-directed approach enabled efficient discovery of functionally-distinctive molecules based on novel, structurally-diverse scaffolds.

The discovery of novel, biologically-relevant molecular scaffolds is a major challenge relevant to both medicinal chemistry and chemical biology. The historical exploration of molecular scaffolds has been remarkably uneven,[1] which has limited the diversity of scaffolds that have been exemplified in medicinal chemistry.[2] This uneven exploration stems, in large part, from the limited toolkit of robust reactions that dominates medicinal chemistry.[3] Diversity[4] and lead-oriented[5] synthetic approaches have been developed to explore diverse and novel chemical space efficiently, but do not explicitly target biologically-relevant regions. In contrast, biology-oriented synthesis takes inspiration from scaffolds based on natural products that have emerged on the basis of their function.[6]

In this paper, we describe a phenotype-directed approach for the identification of novel biologically-relevant chemical space (Fig. 1). Like activity-directed synthesis,[7] which has enabled discovery of diverse small molecules with specific biological functions, the approach is both structure-blind and function-driven. Initially, arrays of reactions with many possible outcomes are executed. The reaction products, whose structure are unknown at this stage, are then directly screened in the cell painting assay[8] which enables an assessment of biological relevance. Finally, the structures of (only) identified bioactive products are structurally elucidated. It was envisaged that the approach would enable discovery of diverse molecular scaffolds that were both synthetically-accessible and biologically-relevant.

We designed an array of 288 reactions in which all combinations of six diazo substrates, eight co-substrates and six reaction conditions were explored (Fig. 2). We chose to harness reactions of α-diazo esters because many different outcomes are possible (*e.g.* C–H, N–H and O–H insertion, cyclopropanation, ylid formation/reaction), and may be varied by changing the catalyst/conditions.[9] The selected α-diazo esters were the commercially-available α-unsubstituted substrate **S1**, the α-vinyl substrate **S2** (which can undergo cycloaddition with alkenes[10a]) and the diverse

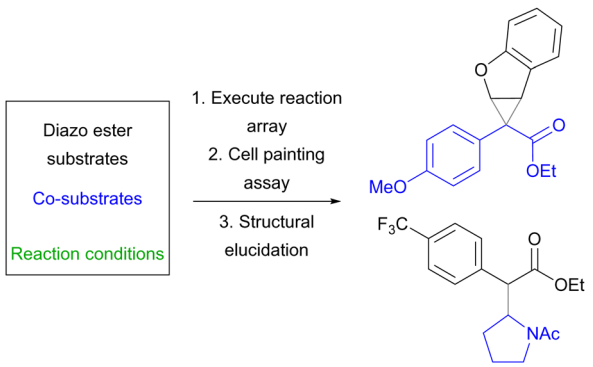

Fig. 1 Overview of phenotype-directed discovery of biologically-relevant molecular scaffolds. Reaction arrays are executed in which the substrate, co-substrate and reaction conditions are varied. The biological relevance of the purified products is assessed in the cell painting assay, and the structures of (only) bioactive products elucidated. The illustrated structures are hypothetical products that might be formed.

[a] *School of Chemistry, University of Leeds, Leeds, LS2 9JT, UK. E-mail: s.p.marsden@leeds.ac.uk, a.s.nelson@leeds.ac.uk*
[b] *Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK*
[c] *Rosalind Franklin Institute, Harwell Campus, Didcot, Oxfordshire, OX11 0QX, UK*
[d] *Max-Planck Institute of Molecular Physiology, Department of Chemical Biology, Otto-Hahn-Strasse 11, Dortmund, 44227, Germany*
[e] *Compound Management and Screening Center, Otto-Hahn-Strasse 11, Dortmund, 44227, Germany*

† Electronic supplementary information (ESI) available. CCDC 2410507–2410509. For ESI and crystallographic data in CIF or other electronic format see DOI: https://doi.org/10.1039/d4cc06605g

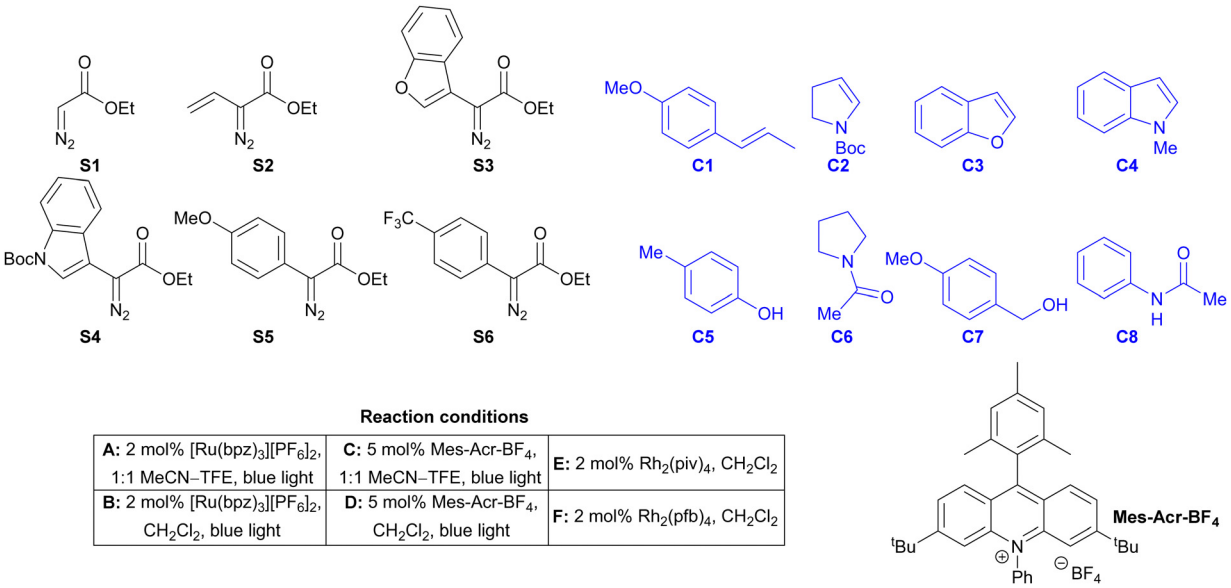

Fig. 2 Reaction array overview. The α-diazo substrates (black), co-substrates (blue) and reaction conditions are shown.

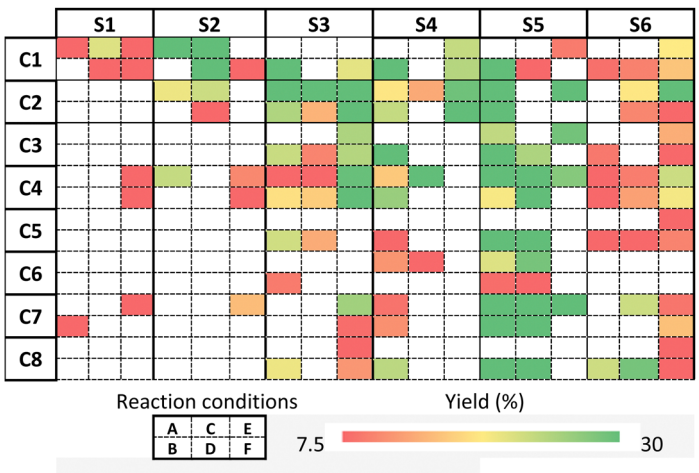

Fig. 3 Estimated yield of reaction array products determined by evaporative light-scattering detection. The conditions (A)–(F) are outlined in Fig. 2.

α-(het)arylated substrates **S3–S6.** The selected co-substrates **C1–C8** all contain at least one functional group that is precedented[9] to react with Rh carbenoids (*e.g.* alkene; potentially reactive C–H, N–H and O–H bond). The reaction conditions were selected on the basis of the potential for alternative reaction outcomes. $[Ru(bpz)_3][PF_6]_2$ and Mes-Acr-BF_4 are precedented photocatalysts for reactions of α-diazo esters,[10a,b] and photocatalysis can result in complementary outcomes to dirhodium catalysis.[10c] Rh_2piv_4 and Rh_2pfb_4 were selected as electronically- and sterically-distinct catalysts for metal carbenoid chemistry.[11]

The reactions were performed in vials in 96-well format on a 200 μl scale (with 20 μmol limiting substrate). The reactions were assembled from stock solutions (in 1 : 1 MeCN–TFE or CH_2Cl_2) using micropipettes. Each reaction involved a limiting co-substrate (**C1–8**, final concentration: 100 mM), an α-diazo ester (**S1–6**, final concentration: 150 mM) and a catalyst [$(Ru(bpz)_3][PF_6]_2$, final concentration: 2 mM; Mes-Acr-BF_4, final concentration: 5 mM; Rh_2piv_4, final concentration: 2 mM; or Rh_2pfb_4, final concentration: 2 mM). The reactions containing the photocatalysts catalysts were irradiated with two 40 W Kessil A160WE Tuna Blue lamps. After 16 h, the outcome of the reactions was determined by analytical UPLC/MS with, additionally, evaporative light-scattering detection[12] to enable determination of the approximate yield of each product (Fig. 3, panel A). To focus on intermolecular reaction products, we identified reactions that had yielded >7% of a unique product with molecular weight higher than the diazo substrate. For each of these reactions, the bulk of the crude reaction mixture was purified by mass-directed HPLC (see ESI†).

Table 1 Synthesis and morphological effects of resynthesized hit compounds

Substrates (method[a])	Product (yield)	Induction[b] /% (30 μM)	[c] (biosimilarity)
S4, **C2** (F)	**1** (86%)	73	WNK463 (88%)
S5, **C2** (E)	**2** (99%)	6[d]	Nicergoline (83%[d])
S4, **C8** (B)	**3** (25%)	14	[e]
S3, **C2** (D)	**4a** (5%[f])	14	Englerin A (83%)
	4b (27%[f])	13[d]	Aripiprazole (87%[d])
S4, **C7** (B)	**5** (26%)	35	I-CBP112 (87%)
S3, **C1** (F)	**6**[g] (44%)	19	[e]
S6, **C1** (F)	**7** (10%)	10	Caroverine (84%)
S4, **C3** (E)	**8** (25%)	6	PF-CBP1 (81%)

[a] See Fig. 2 for methods. [b] Percentage of significantly changed features in the cell painting assay (magnitude of median absolute deviation >3) relative to DMSO control. [c] Reference compound with highest biosimilarity. [d] At 50 μM. [e] No reference compound with >80% biosimilarity. [f] Yield over 3 steps after Boc removal and conversion into the corresponding Ts derivative. [g] 50 : 50 mixture of regioisomers.

The purified, but uncharacterised, products were screened in the cell painting assay which interrogates many biological pathways simultaneously, and enables an assessment of biological relevance.[8] In the assay, six dyes are used to stain different cellular compartments, and high-content imaging and automated image analysis enables determination of 579 features that correspond to specific morphological properties/characteristics of the cell. The biological profile of compounds may be described in terms of morphological fingerprints which capture changes in the features relative to DMSO control. The bioactivity of compounds, screened at 10, 30 and 50 μM, was assessed in terms of an induction value: the percentage of significantly changed features (magnitude of median absolute deviation >3) relative to DMSO control. The reactions that had yielded bioactive products (induction value >5%) were scaled up to enable structural elucidation and validation of their biological activity (Table 1). The reactions were performed on 0.1–0.4 mmol scale, and the resulting products were purified by

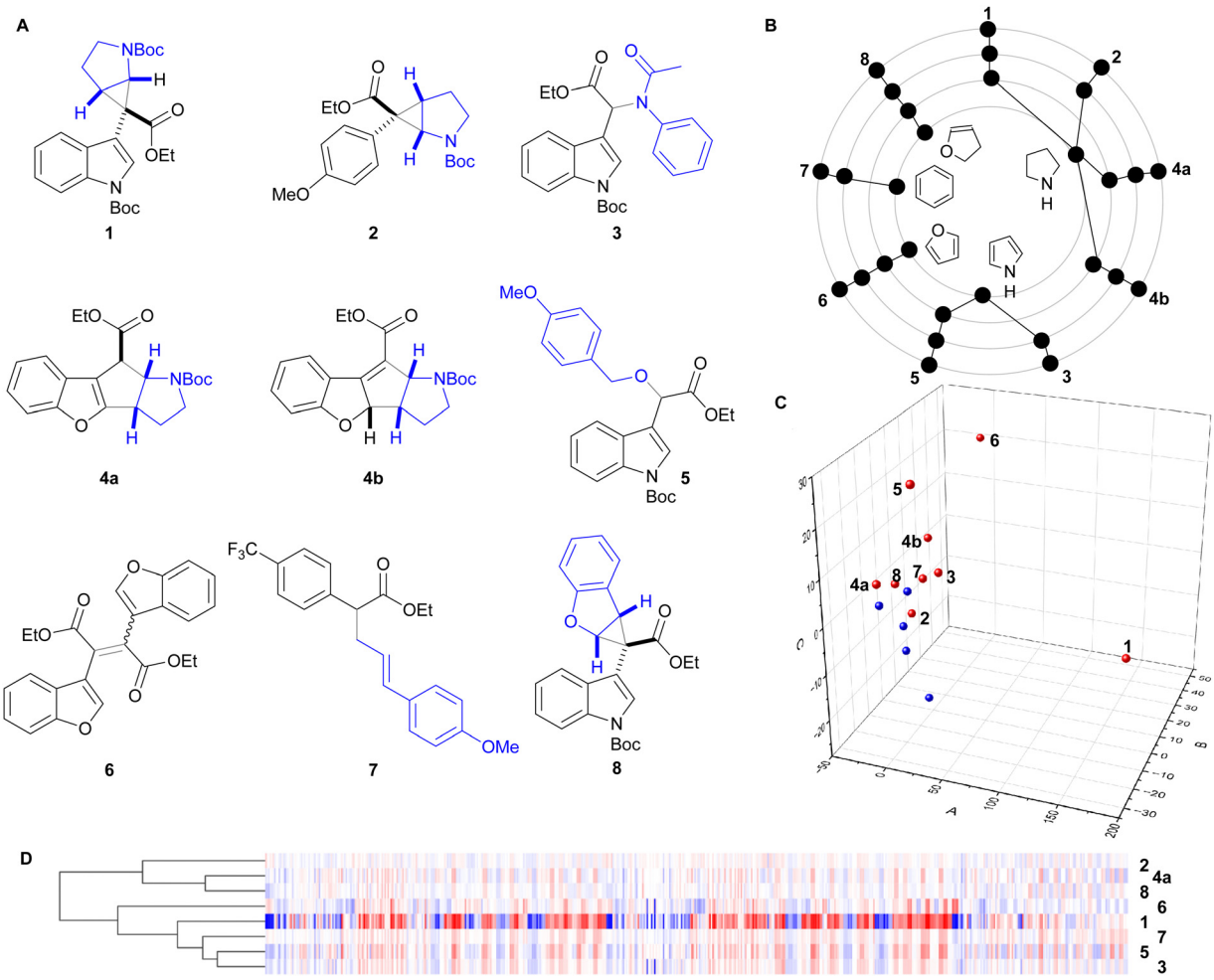

Fig. 4 Structures and activity of bioactive products. Panel A: products isolated from scaled-up reactions. Panel B: hierarchical tree showing the diversity of, and relationship between, bioactive products scaffolds. Black circles indicate the product (outer ring) and iteratively simplified (inner rings) frameworks. Panel C: dimension reduction analysis (PCA) of cell painting data (30 μM); explained variance: PC1 (A, 77%), PC2 (B, 8%), PC3 (C, 6%). The bioactive products (red) are shown with some compounds (blue) displaying >5% induction. Panel D: hierarchically-clustered (see ref. 8*d*) heat map of the morphological fingerprints of the bioactive products (30 μM).

flash column chromatography or mass-directed HPLC. In total, nine bioactive products (**1–3**, **4a–b** and **5–8**) were successfully purified and structurally elucidated using NMR spectroscopy and (for **4a** and **8**) X-ray crystallography (ESI) (Fig. 4, panel A).

The use of both metal- and photoredox-catalysed reactions of diazo compounds enabled formation of bioactive products *via* diverse reaction classes: enamide cyclopropanation (→**1**, **2** and **8**); insertion into an amide N–H bond (→**3**); insertion into allylic(→**7**) C–H bonds; insertion into an alcohol O–H bond (→**5**); and dimerization to yield alkenes (→**6**).[13] The fused tetracycles **4a** and **4b** were formed *via* a remarkable formal (3+2) cycloaddition between the α-benzofuran-3-yl α-diazo ester **S3** and the enamide **C2**; this transformation is reminiscent of a photocatalysed cyclopentene annulation of α-vinyl α-diazo esters.[14] Notably, **4a** and **4b** were separable regioisomers; however, their high (83%) biological similarity at 50 μM, a concentration at which they both have induction values >5%, suggests that they may have related mechanisms, or interconvert, in cells. The represented frameworks have significant structural diversity at all levels of hierarchy (Fig. 4, panel B).[15] The scaffolds (without α atoms) of **1**, **4a** and **4b** are novel, whilst that of **8** is found in just one patented compound.[16]

The morphological fingerprints of the bioactive products (Fig. 4, panels C and D) were compared to previously-identified functional clusters.[8c] Remarkably, only one compound, **1**, was >70% similar to any defined functional cluster (79% similarity to tubulin functional cluster at 30 μM). Furthermore, two of the compounds (**3** and **6**) did not have >80% biosimilarity to any of ~2300 reference compounds that have been previously evaluated in the cell painting assay.

In conclusion, our phenotype-directed approach enabled discovery of novel, structurally- and functionally-distinctive molecules. Critical to success was underpinning chemistry that enabled exploration of diverse chemical space, and an assay enabling broad assessment of biological relevance. Compounds with a range of distinct phenotypes were discovered using an array of 288 reactions that involved just 14 distinct starting materials.

We thank EPSRC (EP/W002914/1; EP/V011367/1; EP/N025652/1) for funding; and Dr Axel Pahl for discussions. Research at the Max Planck Institute of Molecular Physiology was supported by the Max Planck Society; and was co-funded *via* "Netzwerke 2021", an initiative of the Ministry of Culture and Science of the State of Northrhine Westphalia. The Compound Management and Screening Center (COMAS) is acknowledged for cell painting measurements.

Data availability

Supporting data are included in the ESI,† and deposited at CCDC (**https://www.ccdc.cam.ac.uk**) under accession numbers 2410507–2410509.†

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 A. H. Lipkus, Q. Yuan, K. A. Lucas, S. A. Funk, W. F. Bartelt, R. J. Schenck and A. J. Trippe, *J. Org. Chem.*, 2008, **73**, 4443.
2 S. R. Langdon, N. Brown and J. Blagg, *J. Chem. Inf. Model.*, 2011, **51**, 2174.
3 (*a*) D. G. Brown and J. Boström, *J. Med. Chem.*, 2016, **59**, 4443; (*b*) J. Boström, D. G. Brown, R. J. Young and G. M. Keserü, *Nat. Rev. Drug Discovery*, 2018, **17**, 709.
4 (*a*) M. D. Burke and S. L. Schreiber, *Angew. Chem., Int. Ed.*, 2004, **43**, 46; (*b*) S. L. Kidd, T. J. Osberger, N. Mateu, H. F. Sore and D. R. Spring, *Front. Chem.*, 2018, **6**, 460.
5 (*a*) A. Nadin, C. Hattotuwagama and I. Churcher, *Angew. Chem., Int. Ed.*, 2021, **51**, 1114; (*b*) R. Doveston, S. Marsden and A. Nelson, *Drug Discovery Today*, 2014, **19**, 813.
6 (*a*) H. van Hattum and H. Waldmann, *J. Am. Chem. Soc.*, 2014, **136**, 11853; (*b*) M. Grigalunas, S. Brakmann and H. Waldmann, *J. Am. Chem. Soc.*, 2022, **144**, 3314.
7 (*a*) G. Karageorgis, S. Warriner and A. Nelson, *Nat. Chem.*, 2014, **6**, 872; (*b*) G. Karageorgis and A. Nelson, *ChemMedChem*, 2020, **15**, 1776.
8 (*a*) J. C. Caicedo, S. Cooper, F. Heigwer, S. Warchal, P. Qiu, C. Molnar, A. S. Vasilevich, J. D. Barry, H. S. Bansal, O. Kraus, M. Wawer, L. Paavolainen, M. D. Herrmann, M. Rohban, J. Hung, H. Hennig, J. Concannon, I. Smith, P. A. Clemons, S. Singh, P. Rees, P. Horvath, R. G. Linington and A. E. Carpenter, *Nat. Methods*, 2017, **14**, 849; (*b*) S. Ziegler, S. Sievers and H. Waldmann, *Cell Chem. Biol.*, 2021, **28**, 300; (*c*) A. Pahl, B. Schölermann, P. Lampe, M. Rusch, M. Dow, C. Hedberg, A. Nelson, S. Sievers, H. Waldmann and S. Ziegler, *Cell Chem. Biol.*, 2023, **30**, 839; (*d*) T. Schneidewind, A. Brause, A. Pahl, A. Burhop, T. Mejuch, S. Sievers, H. Waldmann and S. Ziegler, *ChemBioChem*, 2020, **21**, 3197.
9 For examples, see: (*a*) H. M. Davies and D. Morton, *Chem. Soc. Rev.*, 2011, **40**, 1857; (*b*) A. Padwa and D. M. Weingarten, *Chem. Rec.*, 1996, **96**, 223; (*c*) M. P. Doyle and D. C. Forbes, *Chem. Rev.*, 1998, **98**, 911; (*d*) H. M. L. Davies and R. E. J. Beckwith, *Chem. Rev.*, 2003, **103**, 2861; (*e*) R. D. C. Gallo, G. Cariello, T. A. C. Goulart and I. D. Jurberg, *Chem. Commun.*, 2023, **59**, 7346; (*f*) Ł. W. Ciszewski, K. Rybicka-Jasińska and D. Gryko, *Org. Biomol. Chem.*, 2019, **17**, 432.
10 (*a*) F. J. Sarabia, Q. Li and E. M. Ferreira, *Angew. Chem., Int. Ed.*, 2018, **57**, 11015; (*b*) N. Holmberg-Douglas, N. P. R. Onuska and D. A. Nicewicz, *Angew. Chem., Int. Ed.*, 2020, **59**, 7425; (*c*) Ł. W. Ciszewski, J. Durka and D. Gryko, *Org. Lett.*, 2019, **21**, 7028.
11 A. I. Green, C. P. Tinworth, S. Warriner, A. Nelson and N. Fey, *Chem. Eur. J.*, 2021, **27**, 2402.
12 A. W. Squibb, M. R. Taylor, B. L. Parnas, G. Williams, R. Girdler, P. Waghorn, A. G. Wright and F. S. Pullen, *J. Chrom. A*, 2008, **1189**, 101.
13 For examples, see: (*a*) I. S. del Villar, A. Gradillas and J. Pérez-Castells, *Eur. J. Org. Chem.*, 2010, 5850–5862; (*b*) E. Aller, R. T. Buck, M. J. Drysdale, L. Ferris, D. Haigh, C. J. Moody, N. D. Pearson and J. B. Sanghera, *J. Chem. Soc. Perkin Trans. 1*, 1996, 2876–2884; (*c*) R.-T. Guo, Y.-L. Zhang, J.-J. Tian, K.-Y. Zhu and X.-C. Wang, *Org. Lett.*, 2020, **22**, 908–913; (*d*) H. M. L. Davies, M. G. Coleman and D. L. Ventura, *Org. Lett.*, 2007, **9**, 4971–4974; (*e*) G. G. Cox, D. J. Miller, C. J. Moody, E.-R. H. B. Sie and J. J. Kulagowski, *Tehrahedron*, 1994, **50**, 3195–3212; (*f*) M. P. Doyle, M. A. McKervey and T. Ye, *Modern Catalytic Methods for Organic Synthesis with Diazo Compounds: From Cyclopropanes to Ylides*, Wiley, New York, 1998.
14 F. J. Sarabia, Q. Li and E. M. Ferreira, *Angew. Chem., Int. Ed.*, 2018, **57**, 11015–11019.
15 A. Schuffenhauer, P. Ertl, S. Roggo, S. Wetzel, M. A. Koch and H. Waldmann, *J. Chem. Inf. Model.*, 2007, **47**, 47.
16 M. Bursavich*et al.*, *US Pat.*, **311**, 217, 2009.