EurJOC
European Journal of Organic Chemistry

Review
doi.org/10.1002/ejoc.202200575

Chemistry Europe
European Chemical Societies Publishing

www.eurjoc.org

VIP *Very Important Paper*

Special Collection

Diversity-Oriented Synthesis and Chemoinformatics: A Fruitful Synergy towards Better Chemical Libraries

Elena Lenci*[a] and Andrea Trabocchi[a]

In these two decades, Diversity-Oriented Synthesis (DOS) has changed significantly: from the historical focus on structural diversity, new approaches have been developed in order to increase the biological outcome of the molecules within the library, such as the privileged-based DOS or the Diversity-Oriented Fluorescence Library Approach (DOFLA). In this context, chemoinformatics can assist organic chemists in identifying biologically-relevant regions of the chemical space not yet explored. Also, chemoinformatics tools can be used to graphically visualize and/or predict the structural diversity and complexity of the compounds within the library. This review highlights the improvement that DOS has received from chemoinformatics, presenting recent articles (published between 2016 and 2021) in which DOS libraries are analysed by chemoinformatics tools.

1. Introduction

From the definition given by Schreiber in 2000,[1] Diversity Oriented Syntesis (DOS) has evolved significantly. At the beginning, the principal goal in developing DOS strategies was to obtain compounds with the highest structural diversity possible[2] subdivided in terms of appendages, functional groups, stereochemical and skeletal diversity,[3] by using no more than five synthetic steps. Starting materials were chosen to maximise the diverse reactivity of the intermediates, in order to balance the need for building diverse and complex compounds with the requirement of keeping synthetic routes short and efficient. Several synthetic strategies have been developed for this purpose, such as the widely known Build/Couple/Pair approach[4] or the Relay Catalytic Branching Cascade (RCBC) approach.[5]

Then, a shift in the focus of DOS library design was observed, moving from structural diversity to biological relevance, and more recently, to functional relevance.[6] Since it is estimated that the number of small molecules populating the chemical universe is more than 10^{60},[7] it is more convenient to drive the synthetic efforts towards those areas of the chemical space that are enriched with biologically relevant compounds, possibly underpinned ones, instead of exploring large areas of the chemical space.[8] DOS libraries with compounds containing bioactive elements have led to the identification of small molecule modulators of non-traditional drug targets,[9] even when the size of the library was limited, such as in the case of the identification of the antimitotic dosabulin from a small DOS collection of only 35 compounds.[10] Following the principles of the recently developed performance-directed synthesis[11,12] and activity-directed synthesis,[13,14] in most recent papers regarding DOS, libraries have been designed with the aim of installing on final compounds the chemical features that are traditionally challenging to incorporate and that are found in natural products or biologically active compounds, such as in privileged-based DOS or in the Diversity-Oriented Fluorescence Library Approach (DOFLA). In this context, chemoinformatics[15,16] represent a real opportunity for the identification of biologically relevant areas of the chemical space,[17] suggesting scaffolds that have been preserved by Nature in evolution, as in the case of the so-called Structural Classification of Natural Products (SCONP) approach[18] devoleped by Waldmann for Biology-Oriented Synthesis (BIOS).[19,20] Later, BIOS concepts have been extended beyond natural products, to the screening of all those databases that include bioactivity data, such as ChEMBL and PubChem, in order to drive the synthetic efforts towards areas of the chemical space related to difficult targets.[21] Also, chemoinformatics tools can be used to measure the structural diversity and complexity of the compounds within the library, providing a visual and easily interpretable representation of the chemical space covered by DOS libraries, in order to communicate their relevance to medicinal chemists and chemical biologists involved in the screening.[22] The diversity and complexity of the molecules in a chemical library can be evaluated in multiple ways, mainly depending on the goals of the study and the molecular representation chosen for the analysis.[23] Each method has some advantages and some limitations and choosing the right chemoinformatics tool can be trivial, as the results may be subjected to multiple interpretations. Table 1 shows the tools most widely used for analysing DOS libraries,[24] that are recalled throughout the text in selected examples of DOS works published between 2016 and 2021. The discussion about these papers has been subdivided into the three main approaches currently used to obtain biologically-relevant libraries: (i) DOS of

[a] *Dr. E. Lenci, Prof. A. Trabocchi*
Department of Chemistry "Ugo Schiff",
University of Florence
Via della Lastruccia 13, 50019 Sesto Fiorentino, Florence, Italy
E-mail: elena.lenci@unifi.it

Special Collection *Part of the "DCO-SCI Prize and Medal Winners 2020/2021" Special Collection.*

Table 1. Chemoinformatics tools used to analyse chemical diversity and complexity of DOS libraries that are reported in this review paper.

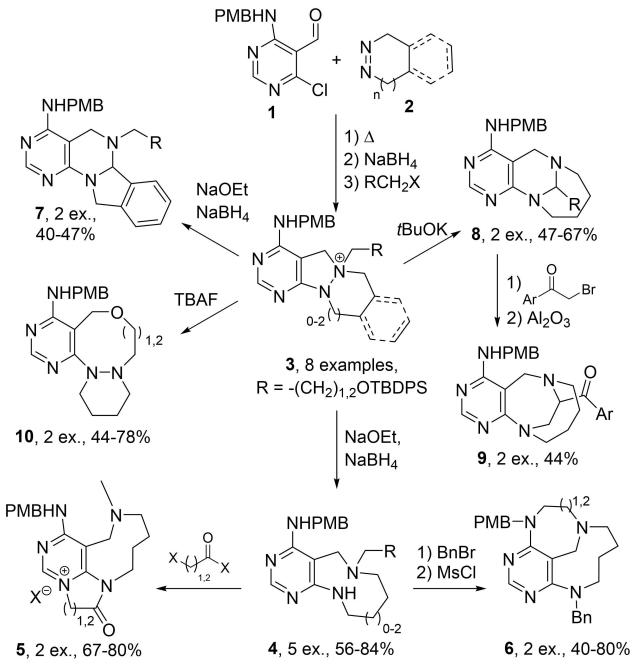
Tool	Utility	Case study
Alignment analysis	Offers the 3D superposition of structures to analyse scaffold and stereochemical diversity	Figure 1 Ref. [29]
Structural fingerprints	Calculates structural diversity between compounds. Very useful for the analysis of appendage and functional group diversity, even though, especially for large libraries, it can be difficult to be visualized	Figure 2 Ref. [33]
Principal Moment of Inertia (PMI)	Provides a distribution of the compounds in a triangular graph depending on shape diversity and three-dimensional complexity	Figure 3 Ref. [36]
Principal Component Analysis (PCA)	Provides a distribution of the compounds in a 2D or 3D plot based on a set of physico-chemical properties	Figure 4 Ref. [41]
Molecular complexity graph	Provides a distribution of the compounds in a 2D plot based on molecular complexity index (calculated in different ways)	Figure 5 Ref. [66]
Multidimensional scaling (MDS)	Provides a distribution of the compounds in a 2D plot based on molecular fingerprint similarity	Figure 6 Ref. 70
Lead-Likeness and Molecular Analysis (LLAMA)	Provides the virtual generation of a library around a defined scaffold and the distribution of the compounds obtained in the lead-like and drug-like chemical space, with the assignment of "lead-likeness penalty" values, depending on Congreve's rules	Figure 7 Ref. [74]
Consensus Diversity Plot (CDP)	Provides a 2D plot in which four measures of diversity are condensed: structural fingerprints, scaffold diversity, whole molecular properties and size of the databases	Figure 10 Ref. [86]

privileged structures-based libraries, (ii) DOS of libraries of macrocycles, (iii) DOS of sp^3-rich compounds.

2. DOS and Chemoinformatics of Privileged Structures-Based Libraries

With the purpose of increasing the chance of exploring biologically-relevant chemical space, one of the most recently used strategies in the field of DOS is the privileged structure-based DOS (pDOS),[25–28] in which the synthesis is planned to obtain compounds around a common scaffold, usually the skeleton of a natural product or a pharmacopore that is found in a high number of drug candidates. In this context, Bum Park and coworkers developed a pDOS strategy starting from pyrimidines **1** and cyclic hydrazines **2**, that were assembled to give a complex pyrimido-pyrazolo-pyridazine intermediate **3**, used for the achievement of seven skeletally bridged or medium-sized azacycles around pyrimidine.[29]

As shown in Scheme 1, the cleavage of the N–N bond of **3** with NaOEt and $NaBH_4$ gave 9-membered diazonane-fused pyrimidine **4**, that was further transformed into scaffold **5**, using dielectrophiles such as bromoacetyl bromide or chloropropionyl chloride, and into scaffold **6**, through an intramolecular



Scheme 1. pDOS of pyrimidine-containing medium-sized azacycles.

Elena Lenci received the Doctor Europaeus PhD title in Organic Chemistry from the University of Florence, Italy, in 2017, after spending a research period in the group of Prof. Dixon at the University of Oxford, UK. She is currently working at the University of Florence, in the field of diversity-oriented synthesis, chemoinformatics and medicinal chemistry of anticancer agents. In 2021, she received the Organic Chemistry for Life Sciences Award Junior from the Organic Chemistry Division of the Italian Chemical Society.

Andrea Trabocchi is Associate Professor of Organic Chemistry at the University of Florence since 2017. He received training on peptide chemistry at Imperial College, London, and holds a PhD in Chemical Sciences. Current research interests and projects under several research collaborations are in the fields of heterocyclic and peptidomimetic chemistry applied to oncology and infectious diseases, particularly focusing on enzyme inhibitors.

substitution reaction using benzyl bromide and mesyl chloride. When intermediate **3** contains an aromatic ring fused to the pyridazine nucleus, the treatment with NaOEt and $NaBH_4$ gave preferentially the ring-rearranged tetracyclic product **7**, instead of **4**, because of the presence of more than one benzylic carbon adjacent to the quaternized nitrogen. Also, when the basic treatment was performed in the absence of a hydride source, an intramolecular reaction involving the migration of the N–N–C bond to an N–C–N bond was observed with the achievement of the diazabicyclo[4.3.1]decane-pyrimidine **8**, whose structural complexity was improved into the scaffold **9** by adding bromoacetophenone. Finally, when the R group of **3** was a silyloxy derivative, the treatment with tetrabutylammonium fluoride allowed for the achievement of 8- or 9-membered oxadiazacycles **10** through intramolecular nucleophilic attack of the alkoxide moiety to the C–N bond.

To assess and graphically visualize the structural diversity of the compounds, the authors determined the energy minimized conformer of each scaffold and overlayed their calculated 3D structures[30] using as the pyrimidine substructure as the focal point (Figure 1).[29]

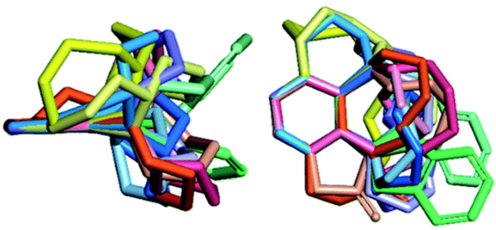
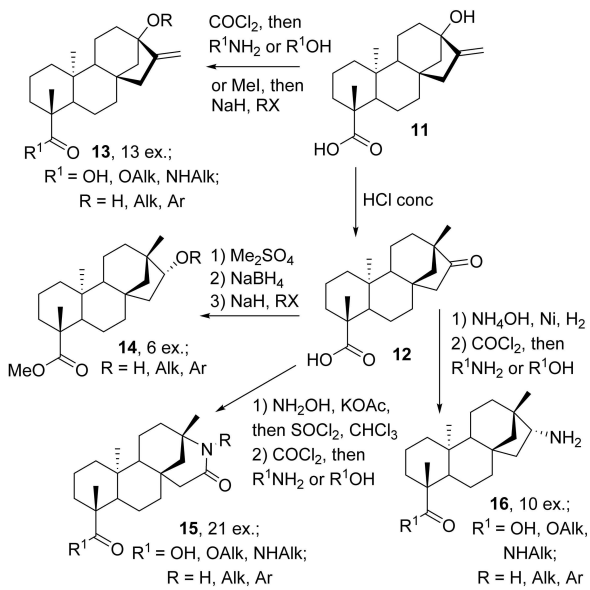


Figure 1. Structural alignment of the scaffolds **4**–**9** in their energy minimized conformers. Adapted from Ref. [29] published by the Royal Society of Chemistry.



Scheme 2. Solid-phase pDOS of steviol derivatives. Yields were not reported by the authors.

This tool is useful to have an intuitive and fast image of the scaffold and stereochemical diversity of the compounds, hovewer it does not offer a quantitative measurement of the structural similarity. To do that, the authors calculated the average Tanimoto similarity coefficent and performed a PMI analysis, comparing their DOS library with a collection of natural products containing benzannulated medium-sized rings, thus showing that the three-dimensional complexity of the DOS products was comparable to those of natural compounds.

Similarly, Georg and coworkers reported a pDOS around the diterpene core structure of stevioside, mainly using solid-phase synthetic protocols.[31] As shown in Scheme 2, a small library of over 90 compounds was obtained by derivatizing the hydroxyl and carboxylic acid moieties of steviol **11** and isosteviol **12**, obtained from **11** through a Wagner-Meerwein rearrangement,[32] and by transformation of the D-ring of **12** into the alcohol **14** by reduction, into the lactam **15** by Beckmann rearrangement and into amine-containing compounds **16** by reductive amination.

The structural diversity of these compounds was assessed by a fingerprints analysis using Tanimoto Extended Connectivity Fringer Print (ECFP) similarity.[33] Tanimoto similarity is calculated with the expression: $T(A,B) = c/(a+b-c)$ where T(A,B) is any value from 0 to 1, being 1 for the highest similarity and 0 for the lowest similarity. For a database with n compounds, $n(n-1)/2$ pairwise comparisons are to be computed and can be stored in a similarity matrix, such as a heatmap. In the paper reported by Georg and coworkers,[31] the similarity values were graphed in a matrix where the calculated similarity for each pair of compounds was coloured in blue for values less than 0.45 (that was found to be the average value for the library) and with red for high scores of similarity (more than 0.8) (Figure 2). This is a very illustrative method to show the differences between compounds, as for the DOS library of steviol no compounds have a similarity ratio of more than 0.7. However, for large libraries containing thousands of compounds this method is not useful, as the visualization becomes more difficult, and in general, the matrix can be too large. One way to visualize these results and, furthermore, conduct quantitative and statistical analysis with other compound databases is by analyzing the off-values of the diagonal of the similarity matrix, for example by analysing the cumulative distribution function for each data set. The authors also performed a PMI analysis and evaluated some physicochemical parameters, including Fsp^3, cLogP and number of stereocenters, to compare the average values obtained for their library with those obtained by two commercially-available libraries, namely the ChemBridge CombiSet and the Maybridge Library. This analysis showed that DOS steviol compounds have higher average values of cLogP and Fsp^3, which means that these compounds are more lipophilic and more structurally complex. These compounds were then assayed with successful results against several isolated enzymes and receptors, including neuropilin-1 (NRP-1), tyrosyl-DNA phosphodiesterase 1 (TDP1), glucagon-like peptide-1 (GLP1) receptor, and even in several phenotypic assays against entire microorganisms, including Hepatitis C Virus (HCV), Hepatitis B Virus (HBV) and *Plasmodium falciparum*, thus

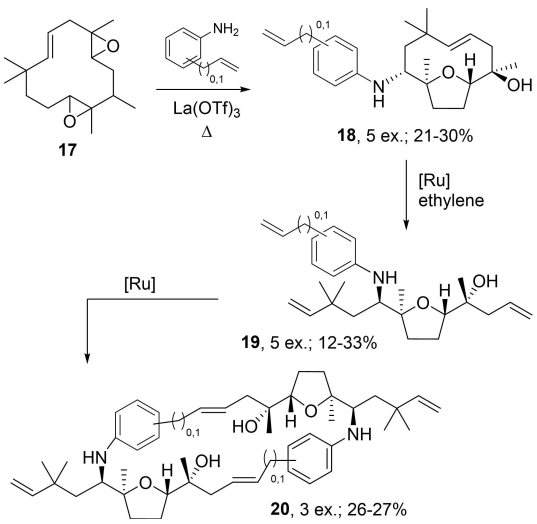
Cmpd	13a	13b	16a	15a	15b	16b	15c	13d	16a	14a	15d	15e	14b	14c	15f
13a	1.000	0.456	0.406	0.389	0.403	0.346	0.333	0.677	0.329	0.463	0.472	0.507	0.438	0.456	0.384
13b	0.456	1.000	0.695	0.661	0.657	0.440	0.423	0.427	0.425	0.329	0.346	0.368	0.333	0.342	0.651
16a	0.406	0.695	1.000	0.712	0.453	0.631	0.453	0.382	0.437	0.338	0.372	0.397	0.342	0.351	0.700
15a	0.389	0.661	0.712	1.000	0.436	0.453	0.647	0.367	0.419	0.325	0.358	0.382	0.329	0.338	0.667
15b	0.403	0.657	0.453	0.436	1.000	0.721	0.690	0.450	0.487	0.294	0.341	0.329	0.345	0.306	0.430
16b	0.346	0.440	0.631	0.453	0.721	1.000	0.746	0.395	0.507	0.305	0.369	0.358	0.357	0.317	0.447
15c	0.333	0.423	0.453	0.647	0.690	0.746	1.000	0.381	0.487	0.294	0.356	0.345	0.345	0.306	0.430
13d	0.677	0.427	0.382	0.367	0.450	0.395	0.381	1.000	0.434	0.432	0.500	0.473	0.468	0.427	0.363
16a	0.329	0.425	0.437	0.419	0.487	0.507	0.487	0.434	1.000	0.411	0.354	0.325	0.507	0.425	0.413
14a	0.463	0.329	0.338	0.325	0.294	0.305	0.294	0.432	0.411	1.000	0.421	0.537	0.529	0.683	0.321
15d	0.472	0.346	0.372	0.358	0.341	0.369	0.356	0.500	0.354	0.421	1.000	0.682	0.575	0.416	0.461
15e	0.507	0.368	0.397	0.382	0.329	0.358	0.345	0.473	0.325	0.537	0.682	1.000	0.429	0.529	0.493
14b	0.438	0.333	0.342	0.329	0.345	0.357	0.345	0.468	0.507	0.529	0.575	0.429	1.000	0.662	0.325
14c	0.456	0.342	0.351	0.338	0.306	0.317	0.306	0.427	0.425	0.683	0.416	0.529	0.662	1.000	0.333
15f	0.384	0.651	0.700	0.667	0.430	0.447	0.430	0.363	0.413	0.321	0.461	0.493	0.325	0.333	1.000

Figure 2. Representative chart of the Tanimoto pairwise similarity for the DOS steviol library compounds. Red indicates similarity coefficient of 1.0 (identical) and blue indicates similarity coefficient less than 0.45 (average similiarity value for the DOS library).

confirming the biological relevance of the DOS library based on steviol privileged structures.

A similar DOS approach was used by Oshima and coworkers who reported a library of terpenoid compounds with a high ratio of sp^3 carbon atoms, exploiting a Lewis-acid-catalyzed transannulation of the sesquiterpenoid humulene.[34] Starting from humulene diepoxide **17**, the ring-opening reactions of the two epoxides by vinylanilines or allylanilines and their subsequent transannulation allowed for the synthesis of 12-oxabicyclo[7.2.1]dodecanes **18**, that were further rearranged by metathesis using Stewart-Grubbs' catalyst into monocyclic tetrahydropyrans **19** (Scheme 3). Finally, cross-methathesis in the presence of a second-generation Grubbs' catalyst allowed for the generation of macrocyclic scaffolds **20**. The chemical diversity of this small library of terpenoid compounds was evaluated by Principal Moments of Inertia (PMI) analysis in comparison with a reference set of top-selling drugs[35] and a small collection of 47 natural products. PMI analysis employs normalized shape-based descriptors to position minimum energy conformation of each library member in a triangular plot, where the vertices represent a perfect rod (acetylene), disc (benzene) and sphere (adamantane).[36] PMI plots can be obtained not only by using standard chemoinformatics packages such as MOE or KNIME, but also from the open source Indigo.[37]

As shown by the PMI graph reported in Figure 3, the terpenoid pDOS library, despite the relative small number of members, shows a high level of shape diversity, with a higher tendency towards the sphere-disc axes, which means a higher



Scheme 3. pDOS of terpenoid derivatives.

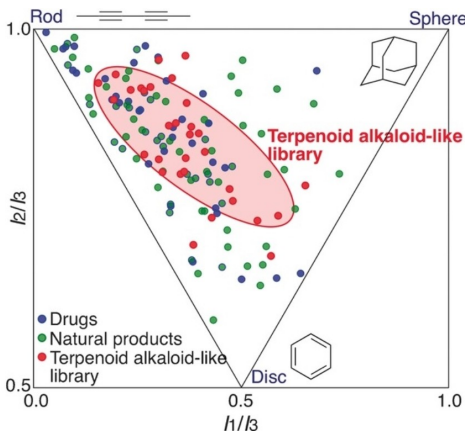
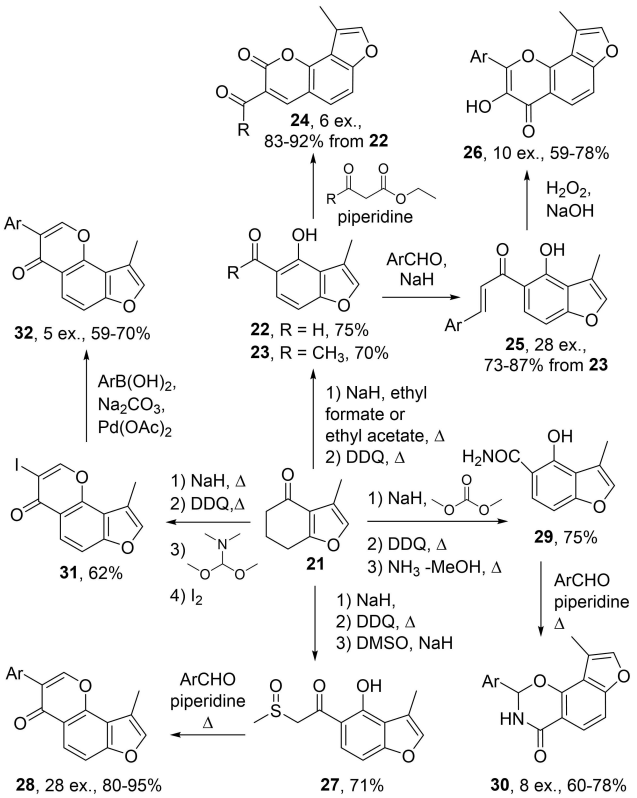


Figure 3. PMI plot of the pDOS library of terpenoids in comparison with 60 top-selling drugs and 47 natural products. Reproduced from Ref. [34], Copyright 2016, Wiley-VCH.

three-dimensional character of the molecules. Finally, the biological potential of these compounds was demonstrated by applying them in a high-throughput assay against two molecular targets involved in lipid metabolism, namely Peroxisome-Proliferator-Activated Receptor alpha (PPARα) and Carnitine Palmitoyltransferase 1 (CPT-1),[38] which results in the identification of a compound acting on these two targets more potently than the reference antihypolipidemic drug bezafibrate.

Another pDOS library was reported by Wu and coworkers around the 3-methylfuran core,[39] a framework that is found in a variety of natural products, including menthofuran and tanshinone. In their approach, cyclohexane-1,3-diketone **21** was transformed into a panel of different intermediates (**22**, **23**, **25**, **27**, **29**, **31**) using dehydrogenation-aromatization reactions mediated by sodium hydride (NaH) and 2,3-dichloro-5,6-dicyano-1,4-benzoquinone (DDQ) in combination with other reagents (Scheme 4). These intermediates were further transformed into 3-methylfuran containing compounds, by applying different reaction conditions. For example, coumarins **24** were obtained from **22** by reaction with methylene compounds, whereas furanochalcones **25** were obtained from **23** by Claisen-Schmidt condensation, that were further transformed into furanoflavonols **26** exploiting Algar-Flynn-Oyamada reaction mediated by NaOH and H_2O_2.[40] β-ketosulfoxide intermediate **27** was used to obtain furanoflavones **28** after treatment with aromatic aldehydes and piperidine, whereas furanoisoquinolinones **30** and furanoisoflavones **32** were obtained, respectively, from amide intermediate **29** and from 3-iodochromone **31** through intramolecular cyclization with aromatic aldehydes or cross-coupling reactions with boronic acids.

To computationally assess the structural diversity of the 3-methylfuran pDOS library, Principal Component Analysis (PCA) was selected. In this approach, a set of physicochemical descriptors is calculated for each compound and condensed into single dimensional numerical values that are plotted into 2D or 3D graphs, positioning DOS libraries in the chemical space, with respect to set of known compounds, such as drugs or natural products.[41] Although visualization graphs can be significantly different depending on the descriptors chosen, PCA is a reliable tool to compare the ability of different compound sets in exploring the chemical space, even though it is not a method to quantitatively assess the chemical diversity.[42] Principal Component can be easily calculated by using the web-based public tool ChemGPS-NP or Chem-Des,[43] that calculate most common physicochemical descriptors, such as molecular weight, LogP values, number of H-bond acceptors and donors, number of rotatable bonds, relative negative and positive charge, topological polar surface area (tPSA), van der Waals volumes, and so on. All these chemical properties are then condensed into principal components, which contain more than one chemical property. For example, in the graph showed in the paper by Wu (Figure 4), the major contributions of principal component one (PC1) are the number of all atoms, molecular surface areas and solvent-accessible surface areas, whereas the logP and tPSA are the key factors associated with principal component two (PC2), and relative negatively and positively charge surface with principal component three (PC3). The library of furan compounds developed by Wu was compared with the set of blockbuster drugs defined by Tan[35] and with a collection of 20 coumarin and flavonoid natural products, in order to show that the DOS library perfectly overlaps midway between the chemical space covered by drugs and those explored by natural products. This library was then submitted to a phenotypic screening to evaluate their potential in conferring host resistance against plant invaders,[44] finding that some of the 3-methylfuranochalcones derivatives **25** were able to induce plants in developing resistance to *Nilaparvata lugens* nymphs (also called brown planthopper).



Scheme 4. pDOS of 2-methylbenzofuran derivatives.

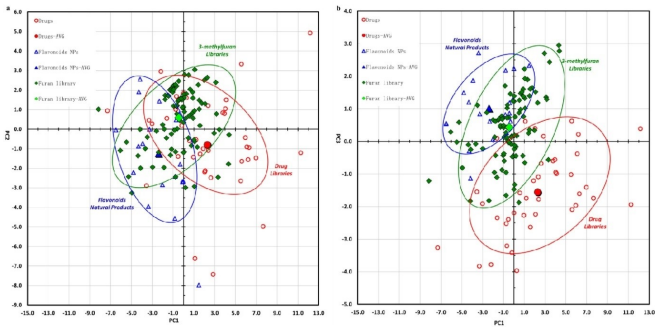
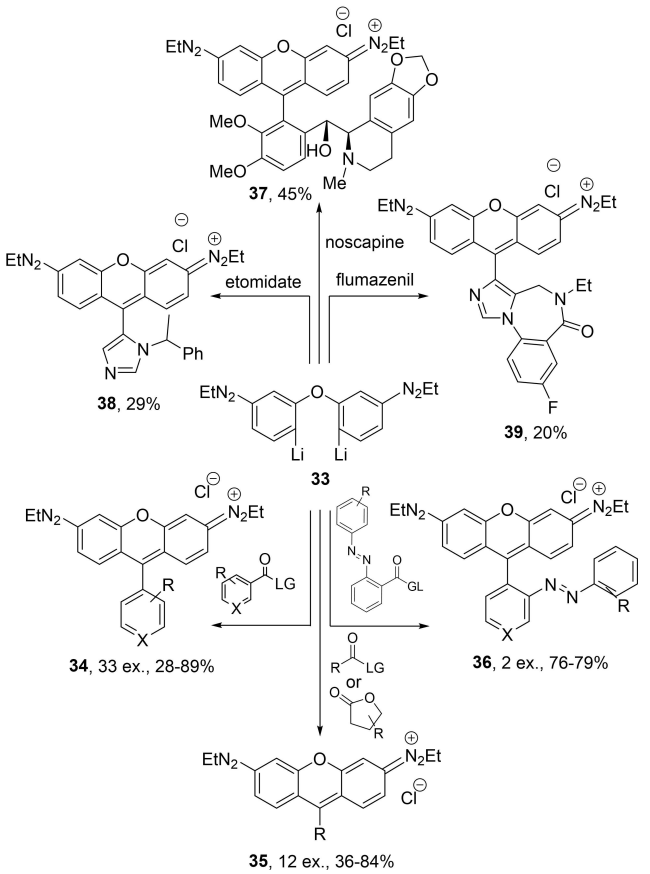


Figure 4. Comparative PCA plots reporting (a) PC1 versus PC2 and (b) PC1 versus PC3 of the pDOS library of 3-methy-furan derivatives (green), block-buster drugs (red) and a collection of coumarins and flavonoid natural products (blue). Reproduced from Ref. [39], Copyright 2017, Wiley-VCH.

In the context of privileged-based DOS, a strategy that proved to be interesting for the discovery of bioactive molecules is the Diversity-Oriented Fluorescence Library Approach (DOFLA), that consists in synthesizing a large number of fluorescent molecules using divergent synthetic strategies.[45] As the molecules that are obtained contain fluorophores, this library can be easily applied in phenotypic assays on complex systems, such as human cancer tissues, by using automated fluorescence microscopy. However, DOFLA libraries often do not show a high scaffold diversity, due to the required two-dimensional conjugated structures of fluorescent probes.[46] In this context, Yang and coworkers reported a mild one-step synthesis of a DOFLA library around rhodamine, which permits the late-stage introduction of natural products, pharmaceuticals, and bioactive compounds, thus increasing significantly the scaffold diversity of the fluorescent compounds generated.[47] Late stage fucntionalization is a very important tool for DOS approaches as it offers the possibility to use directly unfunctionalized C–H bonds as points of diversification, introducing diversity into a complex scaffold.[48–50] In this way, Yang and coworkers reported the construction of a library of 70 different compounds around rhodamine, via nucleophilic condensation of the dilithium reagent **33** with readily available esters and anhydrides, thus obtaining derivatives **34**–**36** in which the C-9 of rhodamine was functionalized with (hetero)aromatic rings, alkyl chains and aliphatic polycyclic structures (Scheme 5). Also, they proved that this strategy could be used to couple rhodamine with more complex substrates, including the drugs noscapine (upon reaction of its dihydrofuran-2(3*H*)-one ring) and flumazenil and etomidate (upon reaction of their ethyl carboxylate functions) to give the corresponding compounds **37**–**39** in good yields.

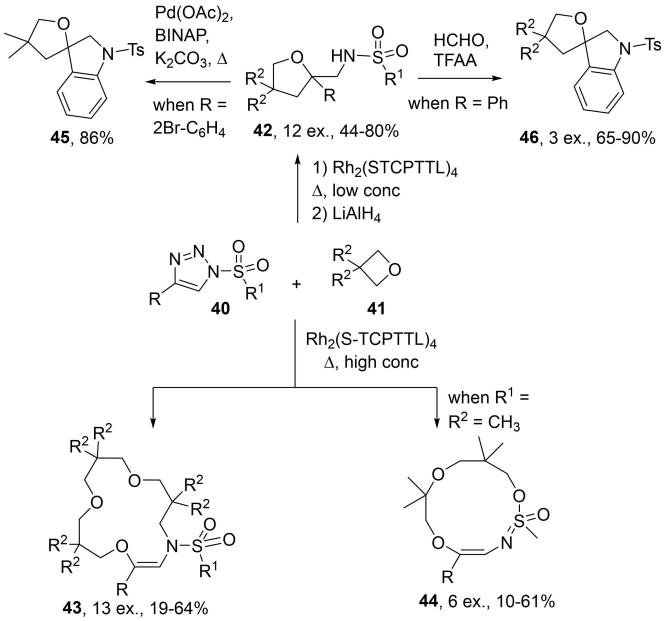
The structural diversity of the DOFLA rhodamine collection was analysed through a newly introduced index, Diversity Number Finger Print Features (DNFPF) value, defined by the authors as the total number of fingerprint features divided by the number of molecules within the library. Specifically, they calculated fingerprint similarity using Tanimoto ECFP-6 for each molecule, counted the total number of unique fingerprint features collected, and divided this number by the number of molecules. A DNFPF value of 18.875 was calculated for the library of rhodamine, a number significantly higher as compared with previously reported DOFLA libraries. Phenotypic screening against two methicillin-resistant bacteria, *Staphylococcus aureus* and *Acinetobacter baumannii*, resulted in the identification of two compounds (containing an adamantane and a 1,2-diphenyldiazene at C-9 position) with MIC values around 2–10 $\mu g\,mL^{-1}$, that did not induce the quick emergence of resistance, as seen for other antibiotics, such as vancomycin, linezolid and daptomycin.



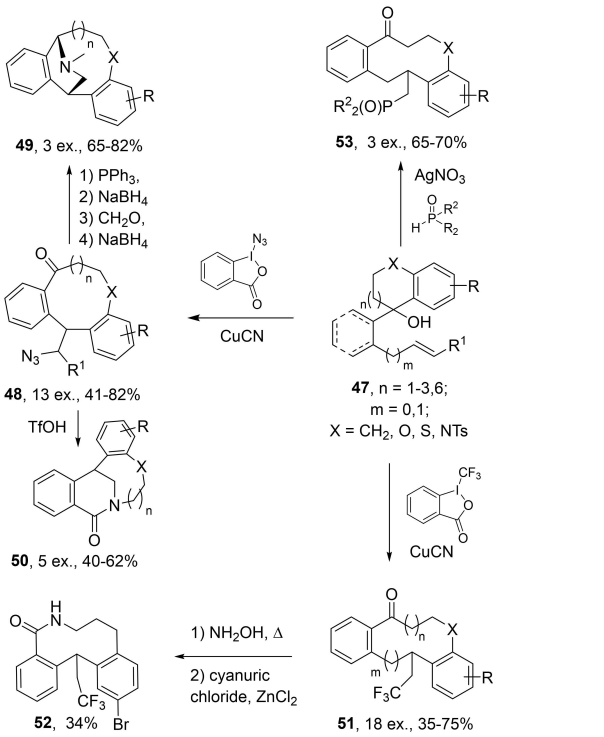
Scheme 5. Diversity-Oriented Fluorescence Library Approach (DOFLA) for the synthesis of rhodamine derivatives.

3. DOS and Chemoinformatics of Libraries of Macrocycles

Macrocycles and medium-sized rings are receiving increased attention in small-molecule drug discovery, especially for targeting protein-protein interactions (PPI) and "undraggable" targets, because such skeletons can populate underpinned area of the chemical space.[51] They are still quite rare in DOS libraries, mainly because the unfavourable enthalpic and entropic reasons in creating macrocycles contrast with the efficiency criteria required for DOS.[52] However, novel chemistries and ring-expansion approaches have proven to be a fruitful source for the development of DOS libraries of macrocycles and/or medium-size rings with a high skeletal diversity,[53,54] as proven by chemoinformatics analysis. In this context, Lacour and coworkers reported the DOS of a macrocycles library by subjecting *N*-sulfonyl-1,2,3-triazoles **40** and cyclic ethers **41** to different reaction conditions (Scheme 6).[55] In particular, different products were observed depending on the concentrations of the oxetane **41**. Using regular concentration conditions (0.1 M), 2-iminotetrahydrofurans **42** were formed by a [1 + 4] condensation mechanism, whereas under high concentration conditions (1.0 M), macrocycles **43** or **44** were observed by a [3 + 4 + 4 + 4] polycondensation mechanism, with **44**



Scheme 6. DOS of hetero- and macrocycles from oxetanes and α-iminocarbenes.



Scheme 7. DOS of benzannulated medium-sized rings from cyclic benzyl alcohols.

obtained starting from mesyl triazoles and 3,3-dimethyloxetane. Finally, 2-iminotetrahydrofurans **42** were further transformed into spiro-indoline **45** and spiro-tetrahydroquinoline **46** through Buchwald-Hartwig and Pictet-Spengler cyclizations, respectively. The structural diversity of the copmpounds was assessed by PCA and PMI analysis. In particular, the PMI analysis showed that the DOS library of macrocycles exhibited a broad shape distribution, with macrocycles **43** and spirocycle **45** populating the right-hand side of the plot.

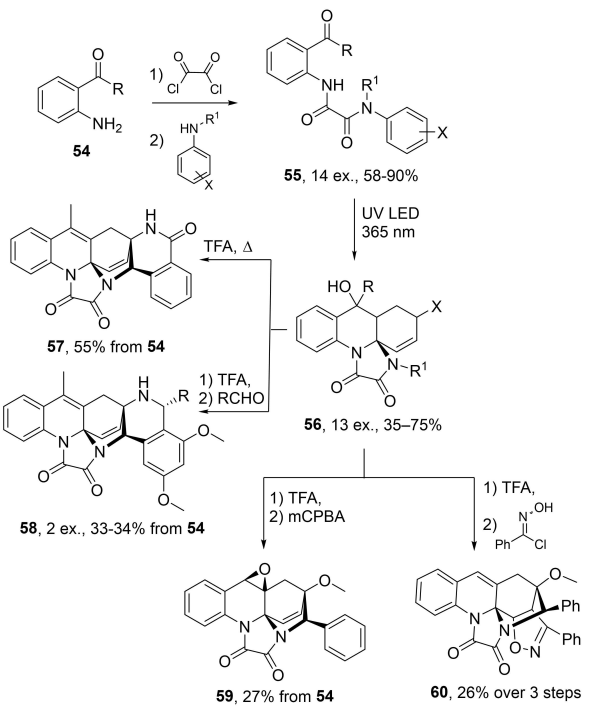
Similarly, Liu and coworkers reported the DOS of 37 distinct benzannulated medium-sized scaffolds, starting from polyfunctional cyclic benzyl alcohol **47** through a sequential 1,4- or 1,5-aryl migration/ring expansion sequence (Scheme 7).[56] For example, the radical azidation promoted by azidoiodinane and CuCN and subsequent 1,4-aryl migration gave benzannulated cyclic ketones **48** possessing 9, 10 or even 14 atoms. These macrocyclic compounds were easily transformed into isopavine analogues **49** in three simple synthetic steps, namely a intramolecular Staudinger/aza-Wittig reaction, followed by subsequent $NaBH_4$ reduction and reductive amination with formaldehyde. Also, the application of intramolecular Schmidt-Aubé reaction conditions (i.e. the presence of a Brønsted acid such as TfOH) allowed to transform **48** into medium-bridged lactam derivatives **50**. Similarly, the reaction of **47** with the Togni's reagent and CuCN led to trifluoromethylated macrocyclic ketones **51**, that were further derivatized into ten-membered lactam **52** by Beckmann reaction. Finally, the reaction of **47** with phosphine oxide in the presence of $AgNO_3$ opened the way of an aryl migration/ring expansion process in concomitance with phosphonylation, which gave ten-membered scaffolds **53**. PCA analysis was used to evaluate the diversity of the 52 compounds prepared in the DOS library in comparison with a set of 27 benzannulated medium-sized ring natural products, a set of 47 brand-name small molecule drugs, and a set of 20 commercially available drug-like compounds present in the Molecular Libraries Small Molecule Repository. The PCA was based on 19 structural and physicochemical parameters and the results revealed distinct chemical spaces occupied by the DOS library of benzannulated medium-sized rings with those explored by natural products containing similar benzannulated medium-sized rings.

4. DOS and Chemoinformatics of sp^3-Rich Compounds Libraries

A complementary strategy for producing biologically-relevant chemical libraries is the use of complexity-generating transformations that can give access to compounds with topographically complex cyclic frameworks with a high ratio of sp^3 carbon atoms.[8,57] In fact, increasing the three-dimensional character of a molecule is generally associated with a more successful outcome in drug discovery.[58,59] This approach, called by Hergenrother and coworkers as "complexity-to-diversity (CtD)",[60] involves the production of complex libraries via the controlled application of complexity generating reactions, such as ring distortion transformations, cycloadditions,[61] domino process and stereospecific C–H activation methodologies.[62–64] Also, photochemistry has allowed to perform arduous reactions, such as cycloadditions, in which the ground ground state

reactions are not possible, under mild conditions. In this context, Kutateladze and coworkers reported a Photoassisted Diversity-Oriented Synthesis of complex sp^3-rich polyheterocyclic scaffolds starting from the oxalyl-amidobenzaldehyde photoprecursor **55**, obtained in two synthetic steps from anylines **54** and oxalyl chlorides.[65] As shown in Scheme 8, the photoprecursors **55**, once activated by UV light irradiation, were able to undergo a photoinduced cascade process leading to the formation of complex scaffolds **56**, containing a tetrahydroquinoline core fused to a imidazolidine-4,5-dione and a cyclohexadiene ring.

These intermediates were further diversified into complex polyheterocyclic scaffolds by standard chemical modifications, such as into the lactams **57** by treatment with TFA, into the tetrahydroisoquinoline **58** through Pictet – Spengler reaction, into the epoxide **59** by reaction with mCPBA and into isoxazoline **60** by 1,3-dipolar cycloaddition of benzonitrile-oxide with the azabicyclo[2.2.2]octene moiety of **56**.



Scheme 8. Photoassisted DOS of sp^3-rich polyheterocyclic scaffolds.

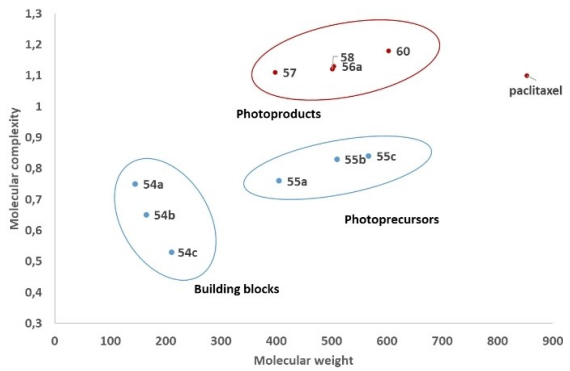


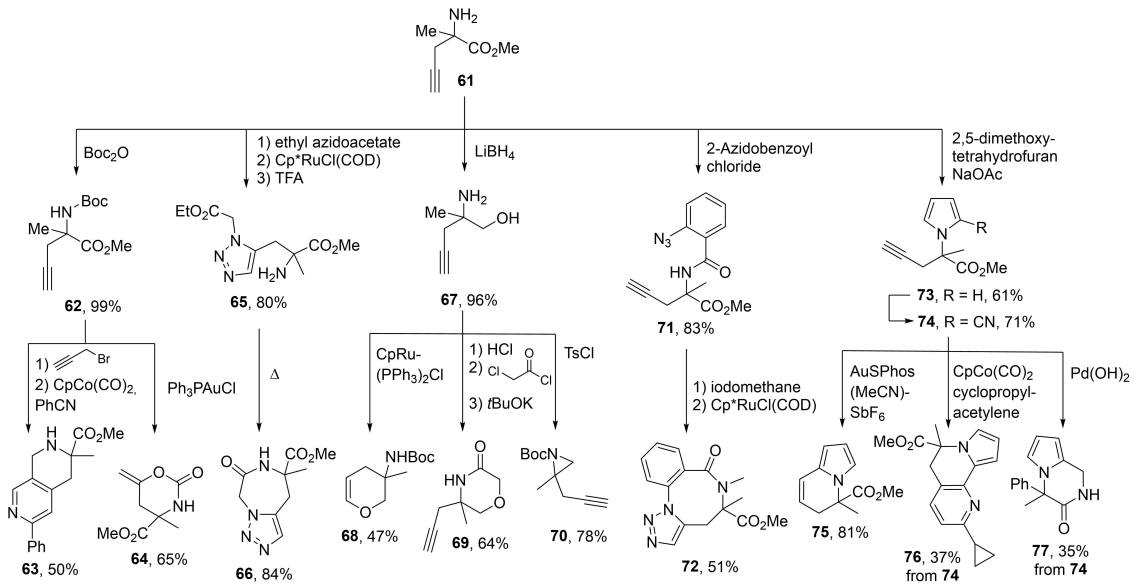
Figure 5. Molecular complexity graph reporting complexity metrics versus molecular weight of compounds of the photoassisted DOS: building blocks **54**, photoprecursors **55** and products **56**–**60**.[65]

The differences in molecular complexity of the molecular scaffolds obtained with the Photoassisted DOS was quantified using a Molecular Complexity graph in which the complexity index are shown plotted versus the molecular weight.[66] Several different metrics have been developed to quantify molecular complexity,[67,68] such as: 1) the ratio of sp^3 carbon atoms (Fsp^3), 2) the number of heavy atoms, 3) the number of covalent bonds between carbon atoms, 4) the number of heavy atomic numbers, or other metrics that combined more than one of these descriptors. In the graph reported by Kutateladze and coworkers (Figure 5), the complexity index was calculated using DataWarrior. Photoprecursors **55** have complexity indices around 0.7 similar to those of anilines building blocks **54**, whereas the products of the photoassisted cascade (**56**–**60**), still possessing similar molecular weights, have a much higher complexity index (more than 1), comparable with those of paclitaxel. The increase of molecular complexity per synthetic step is an important metric, particularly useful to evaluate how a methodology is a „complexity-generating transformation".

Cycloadditions have been exploited also by Spring and coworkers to report a high quality chemical library of more than forty different sp^3-rich compounds starting from α,α-disubstituted propargyl amino esters **61**.[69] As shown in Scheme 9, the propargylation and Co-induced [2 + 2 + 2] cyclotrimerisation of intermediate **62** gave tetrahydro-2,6-naphthyridine **63**, whereas the gold-catalysed hydroamination led to oxo-1,3-oxazinane **64**. Then, reaction of **61** with ethyl 2-azidoacetate and subsequent Ru-mediated 1,5-click chemistry was used to obtain intermediate **65**, which gave the rigid scaffold **66** by lactamisation. Reduction of the carboxylic moiety of **61** into the corresponding alcohol provided intermediate **67**, which was further diversified into dihydropyran **68** by Ru-mediated cyclisation of the alcohol with the terminal alkyne, into morpholinone **69** by acylation with chloroacetyl chloride followed by base-mediated cyclisation and into aziridine **70** by intramolecular attack of the amino group on the alcohol. The introduction of 2-azido benzoyl moiety into **61** gave intermediate **70** which was used as a starting point for the synthesis of triazolo-diazocine **72**, obtained by intramolecular Ru-mediated click chemistry.

Finally, pyrrole **73** was prepared via Paal-Knorr reaction and this new scaffold was found to be useful for the preparation of a panel of different compounds, including the dihydroindolizine **75**, obtained by gold catalysed-intramolecular cyclisation between the terminal alkyne and the pyrrole moiety, and scaffolds **76** and **77** obtained from 2-cyano derivative **74** through Co-catalysed [2 + 2 + 2]-cyclotrimerization and Pd-catalysed hydrogenation, respectively.

As for the chemoinformatics analysis, the authors reported a Multi-Dimensional Scaling (MDS) Plot.[70] Similarly to PCA, this plot gives a visual representation of the chemical space of a library. However, while PCA uses physicochemical descriptors to distribute molecules in chemical space areas, the MDS plot uses



Scheme 9. DOS of sp^3-rich polyheterocyclic scaffolds from α,α-disubstituted propargyl amino esters **61**.

molecular fingerprints. In the graph obtained by Spring and coworkers,[69] Morgan fingerprints are calculated for each compound to obtain a 2D-similarity matrix based on Euclidean distance (Figure 6).

Then, the similarity matrix was transformed into a 2D MDS plot, in which Dimension 1 and Dimension 2 contain the pairwise similarity observed in an underlying distribution of Euclidean distances. As for PCA, the authors used different reference sets to compare the distribution of the DOS library, in particular they used two combinatorial chemistry libraries,[71,72] and seven different sets of compounds obtained by CHEMBL, each one containing compounds with biological activity versus a defined protein target. These sets of compounds allows to identify precise area of the bioactive space, thus suggesting how the DOS library of sp^3-rich polyheterocyclic scaffolds may contain compounds acting as potential kinase inhibitors or membrane receptor ligands, as the area covered by these compounds overlaps considerably with these bioactive space.

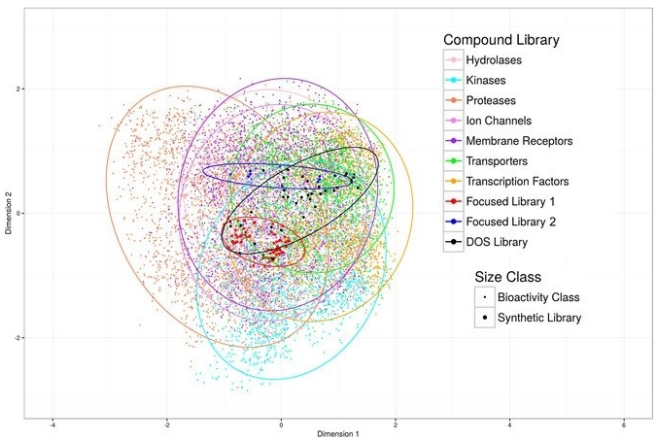
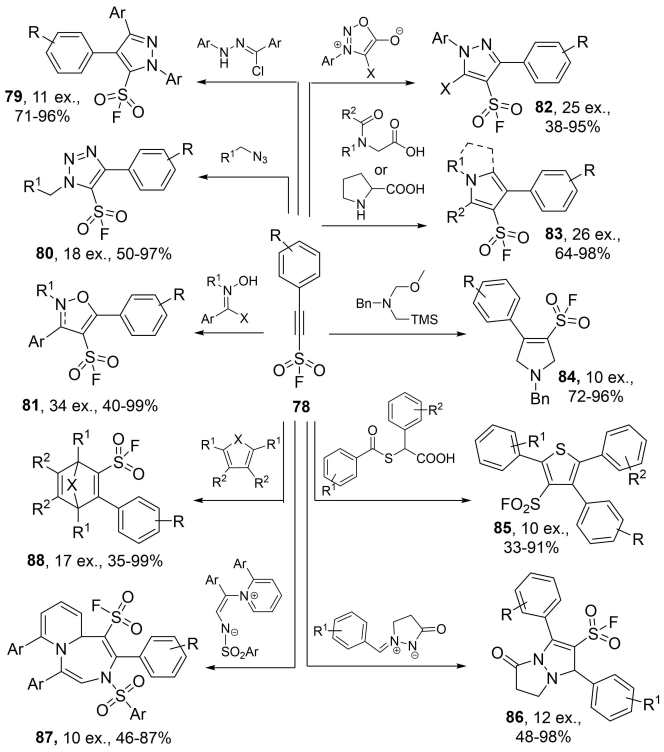


Figure 6. Multidimensional scaling plot based on Morgan fingerprints of the DOS library, compared to two focused libraries and seven different bioactivity classes of common protein targets. Reproduced from Ref. [69], Copyright 2018, Wiley-VCH.

In the context of exploiting cycloadditions as complexity-generating reactions, a powerful approach, recently proposed by Sharpless, Moses and coworkers is the Diversity Oriented Clicking (DOC).[73] Under this name, they defined a novel strategy for the divergent synthesis of libraries through the application of click-chemistry transformations. They reported the click reactions of 2-substituted-alkynyl-1-sulfonyl fluoride **78** with a variety of 1,3- and 1,5-dipoles and cyclic dienes, providing a library of 173 compounds characterized by 10 discrete heterocyclic scaffolds. As shown in Scheme 10, the 1,3-dipolar cycloadditions of **78** with nitrile imines, azides, nitrile oxides, Sydnones, glycine or proline, azomethine ylides and thio-Munchnones gave respectively pyrazoles **79**, triazoles **80**, isoxazoles **81**, pyrazoles **82**, pyrroles **83**, dihydropyrrole **84** and thiophene **85**. Then, *N,N*-bicyclic pyrazolidinone products **86**, bicyclic-1,4-diazepines **87** were obtained respectively by [3 + 2] and [5 + 2] cycloaddition with azomethine imines, and bicyclic scaffolds **88** by [4 + 2] Diels-Alder cycloadditions with cyclic dienes. The authors used the LLAMA (Lead-Likeness and Molecular Analysis) tool, developed by Nelson and coworkers,[74] to assign the lead-likeness penalty (LLP) value to each product, depending on how far molecular properties are from Congreve's rules, and to distribute compounds in the lead-like and drug-like chemical space.

As shown in Figure 7, most of the compounds are inside the Lipinski space, even though with quite large lead-likeness penalty. Despite that, the application of this DOC library into a phenotypic assays against methicillin-resistant *Staphylococcus aureus* strain allowed for the identification of 16 compounds that were able to inhibit cell growth, with improved activity



Scheme 10. Diversity-Oriented Clicking (DOC) from 2-substituted-alkynyl-1-sulfonyl fluoride intermediates **77.**

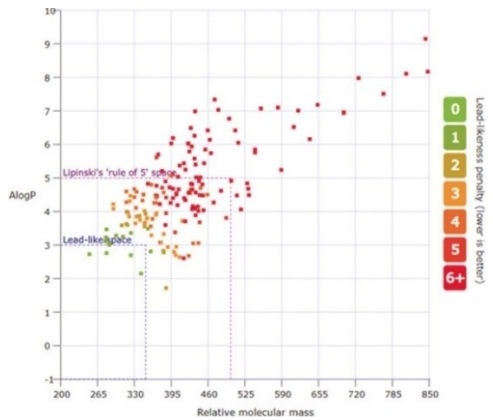


Figure 7. LLAMA analysis of the Diversity-Oriented Clicking (DOC) library, reporting molecular weight against AlogP value of each compounds, coloured according to lead-likeness penalty. Reproduced from Ref. [73], Copyright 2020, Wiley-VCH.

over methicillin, and most active compounds were actually found to be those outside the Lipinski drug like space, thus revealing that chemoinformatics analysis should be always considered in a critical way.

5. Our Contribution in the DOS and Chemoinformatics of Lactam Libraries

Our efforts in the field of Diversity-Oriented Synthesis consists in the development of Build/Couple/Pair strategies, starting from sugar- and amino-acid derived building blocks with the aim of developing skeletally complex scaffolds containing glyco- and/or peptidomimetic moieties.[75–79] In this context, we became interested in the chemoinformatics characterization of the chemical space occupied by compounds possessing precise chemotypes, such as bicyclic acetals[80] and lactams.[81] Thus, in collaboration with Prof. Medina Franco at the National Autonomous University of Mexico, we systematically investigated databases of natural products, drugs and ChEMBL, searching for all the possible topologically different ring combinations that can be found for isolated, fused and spiro lactams, finding scaffolds, such as β-spiro or ε-bridged, that while exhibiting good biological activity are underrepresented in the literature.[81] Thus, with the aim of producing a library of novel skeletally complex lactams, we developed an algorithm to generate lactam compounds that do not differ only for the appendages, as it is commonly found in virtual libraries,[82–84] but also for their scaffolds. We implemented in the algorithm the principles of DOS by applying a Build/Couple/Pair (B/C/P) strategy, starting from commercially available building blocks that contain amine or carboxylic acid functional groups (Figure 8).[85] Briefly, in the first block (A), building blocks from a selected database were retrieved and translated into strings of chemical type. Then, in the second part (B), building blocks were selected in order to comply with the Congreve's 'rule of three' and subdivided into three classes, depending on the presence of a primary amine, a carboxylic acid or a secondary amine. By using the same node, building blocks containing only one functional group were discarded, as they could not undergo any intramolecular reaction in the pairing phase, and compounds containing more than two functional groups were removed to prevent that intramolecular reactions could occur between functional groups within the same building block. In the third part (C), the amide bond formation between carboxylic acids and amines was formed and then different intramolecular cyclization reactions were applied for the pairing phase (block D), such as lactamization; lactonization, ether formation by acid-catalyzed alcohol condensation, Williamson ether synthesis, Buchwald-Hartwig cross coupling reaction, olefin metathesis, reductive amination and CuAAC. Finally, the lactams obtained from the DOS B/C/P workflow were separated into macrocycles (more than 7-membered rings) and non-macrocycles (3- to 7-membered rings). In general, esterification, lactamization, and cross coupling reactions generated the largest amount of compounds. Reactions such as the enyne metathesis and the 'click' reaction generated mainly macrocyclic compounds, due to the steric constraints imposed by the formed heterocycle.

Using Enamine database as building blocks collection, a library of 90845 macrocylic and 20558 non-macrocylic lactams

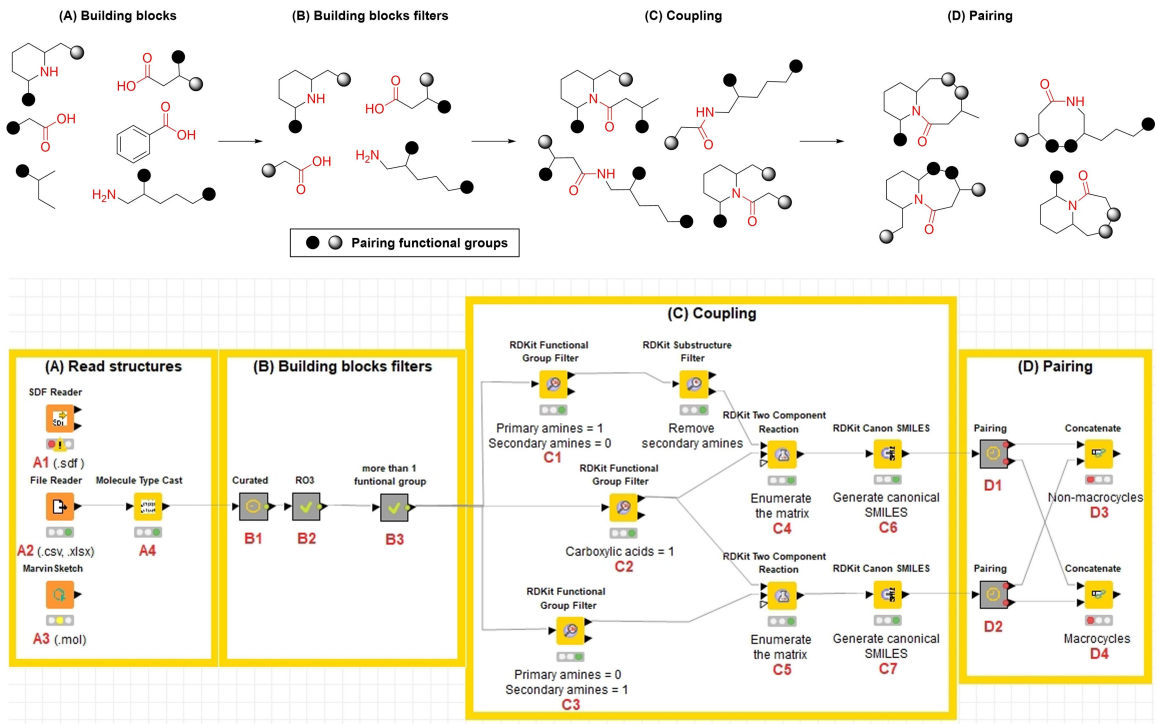


Figure 8. Workflow for the design of DOS library of lactams based on the Build/Couple/Pair approach. Reproduced with permission from Ref. [85], Copyright 2020, Elsevier.

was obtained, some examples of the latter ones are reported in Figure 9.

Non-macrocyclic lactams were classified depending on their scaffold in β-, γ-, δ-, and isolated, fused and spiro-lactams, as reported in the previous work,[81] thus revealing a great variety of spiro- and fused- δ- and ε-lactams, that were not so much populated in CHEMBL. Also, the chemical space of non-macrocyclic lactams was evaluated using PCA and PMI analysis, revealing that the DOS library has a broader physicochemical space than approved drugs, and a shape complexity comparable with the lactams of the ChEMBL database. Also, Consensus Diversity Plot (CDP) was performed to compare the global structural diversity of the DOS libraries in comparison with lactams contained in natural products (UNPD), in bioactive compounds (CHEMBL) and in approved drugs (Figure 10). In this plot, the size of the data points represents the relative size of each data set and the color of each data point represents the diversity of six physicochemical properties of pharmaceutical relevance.[86] To measure the structural diversity (x axis), molecular fingerprint was used applying the ECFP4 Tanimoto coefficient, considering not only the core scaffold but also the appendages. On the other hand, scaffold diversity was measured using Cyclic System Retrieval (CSR) curves (y axis), using the area under the curve (AUC) as a quantitative measurement, that reports the fraction of cyclic systems against the cumulative fraction of the database. AUC value ranges from 0.5 (maximum diversity, when each compound in the library has a different cyclic system) to 1.0 (minimum diversity, when a single cyclic system encompasses all the compounds) (Figure 10).[87]

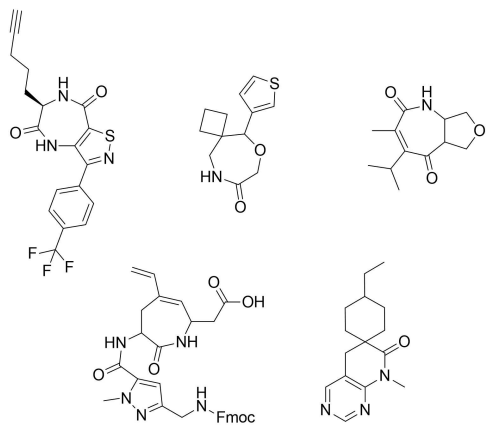


Figure 9. Selected examples of the skeletally different lactams obtained applying this workflow using Enamine as database of building blocks.

As shown by Figure 10, the DOS library is diverse both in structural terms (fingerprints) and in terms of scaffolds (AUC metric). Finally, we curated the databases of non macrocyclic lactams by removing compounds possessing already reported scaffolds, those with potential PAINS liability, and those that do not follow Lipinski's rules. In this way, a database of 2581 compounds with drug-like properties was obtained, ready to be used for virtual screening studies.

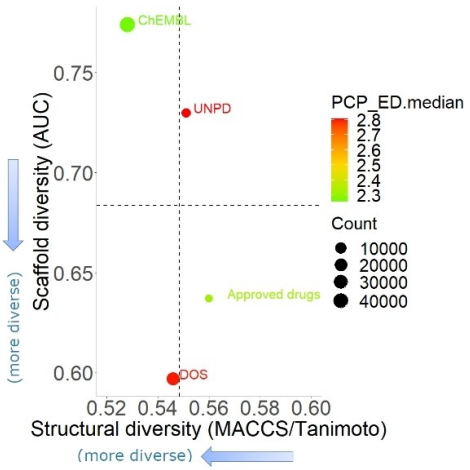


Figure 10. Consensus Diversity Plot comparing the diversity of DOS library with the reference data sets. The median MACCS/Tanimoto of the data set is plotted on the X axis and the AUC of the Cyclic System Retrieval (CSR) curves on the Y axis. Data points are colored by the diversity of six properties of pharmaceutical relevance from red (more diverse), to green (less diverse). The relative size of the data set is represented by the size of the point. Reproduced with permission from Ref. [85], Copyright 2020, Elsevier.

5. Conclusions

In these two decades, Diversity-Oriented Synthesis (DOS) has changed significantly: from the historical focus on structural diversity, new approaches have been developed in order to increase the biological outcome of the molecules within the library. Although the synthesis of complex natural products requires the application of more challenging chemistry, the investments in the chemical synthesis is perfectly balanced by the smaller size of the library needed and also by the highest rate of chance of finding hit compounds. In this context, chemoinformatics offer the possibility of the identification of area of the chemical space enriched by bioactive compounds, or those that have not been explored yet. Also, computational tools useful to analyse and graphically visualize the structural diversity and complexity of the compounds produced in a DOS library are nowadays required to better communicate with medicinal chemists and chemical biologists involved in the biological screening, as seen by the increasing number of DOS papers that include a chemoinformatics graph. Each method has some advantages and some limitations and choosing the right chemoinformatics tool can be trivial, as the results may be subjected to multiple interpretations. However, additional advances are expected from the chemoinformatics community, for example to predict the bioactivity of a newly synthetized molecule, to facilitate drug repurposing, or to suggest more efficien synthetic strategies, that reduce the need of protecting groups and maximize the structural and skeletal diversity of the final compounds.

Acknowledgements

MUR – Italy (PRIN 2020, project 2020833Y75) and Fondazione Cassa di Risparmio di Pistoia e Pescia (Bando Giovani@Ricerca scientifica 2018) are acknowledged for financial support. Open Access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: Chemical libraries · Drug discovery · Scaffolds · Small molecules · Synthetic methods

[1] S. L. Schreiber, *Science* **2000**, *287*, 1964–1969.
[2] "*Diversity-Oriented Synthesis: Basics and Applications in Organic Synthesis, Drug Discovery, and Chemical Biology*", A. Trabocchi, Ed.; Wiley and Sons, **2013**.
[3] W. R. J. D. Galloway, A. Isidro-Llobet, D. R. Spring, *Nat. Commun.* **2011**, *1*, 80.
[4] T. E. Nielsen, S. L. Schreiber, *Angew. Chem. Int. Ed.* **2008**, *47*, 48–56; *Angew. Chem.* **2008**, *120*, 52–61.
[5] N. T. Patil, V. S. Shinde, B. Sridhar, *Angew. Chem. Int. Ed.* **2013**, *52*, 2251–2255; *Angew. Chem.* **2013**, *125*, 2307–2311.
[6] I. Pavlinov, E. M. Gerlach, L. N. Aldrich, *Org. Biomol. Chem.* **2019**, *17*, 1608–1623.
[7] R. S. Bon, H. Waldmann, *Acc. Chem. Res.* **2010**, *43*, 1103–1114.
[8] C. J. Gerry, S. L. Schreiber, *Nat. Rev. Drug Discovery* **2018**, *17*, 333–352.
[9] E. Lenci, A. Guarna, A. Trabocchi, *Molecules* **2014**, *19*, 16506–16528.
[10] B. M. Ibbeson, L. Laraia, E. Alza, C. J. O'Connor, Y. S. Tan, H. M. L. Davies, G. McKenzie, A. R. Venkitaraman, D. R. Spring, *Nat. Commun.* **2014**, *5*, 3155–3163.
[11] M. J. Wawer, K. Li, S. M. Gustafsdottir, V. Ljosa, N. E. Bodycombe, M. A. Marton, K. L. Sokolnicki, M.-A. Bray, M. M. Kemp, E. Winchester, B. Taylor, G. B. Grant, C. S. Hon, J. R. Duvall, J. A. Wilson, J. A. Bittker, V. Dančík, R. Narayan, A. Subramanian, W. Winckler, T. R. Golub, A. E. Carpenter, A. F. Shamji, S. L. Schreiber, P. A. Clemons, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10911–10916.
[12] P. M. Petrone, A. M. Wassermann, E. Lounkine, P. Kutchukian, B. Simms, J. Jenkins, P. Selzer, M. Glick, *Drug Discovery Today* **2013**, *18*, 674–680.
[13] K. Liu, C. Zhu, J. Min, S. Peng, G. Xu, J. Sun, *Angew. Chem. Int. Ed.* **2015**, *54*, 12962–12967; *Angew. Chem.* **2015**, *127*, 13154–13159.
[14] G. Karageorgis, M. Dow, A. Aimon, S. Warriner, A. Nelson, *Angew. Chem. Int. Ed.* **2015**, *54*, 13538–13544; *Angew. Chem.* **2015**, *127*, 13742–13748.
[15] F. K. Brown, *Annu. Rep. Med. Chem.* **1998**, *33*, 375–384.
[16] D. K. Agrafiotis, M. K. Holloway, S. A. Johnson, C. H. Reynolds, T. R. Stouch, A. Tropsha, C. L. Waller, *J. Comput.-Aided Mol. Des.* **2018**, *32*, 723–729.
[17] E. Lenci, A. Trabocchi, *ChemBioChem* **2019**, *20*, 1115–1123.
[18] M. A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casaulta, A. Odermatt, P. Ertl, H. Waldmann, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17272–17277.
[19] S. Wetzel, R. S. Bon, K. Kumar, H. Waldmann, *Angew. Chem. Int. Ed.* **2011**, *50*, 10800–10826; *Angew. Chem.* **2011**, *123*, 10990–11018.
[20] H. van Hattum, H. Waldmann, *J. Am. Chem. Soc.* **2014**, *136*, 11853–11859.
[21] Y. Wang, T. Suzek, J. Zhang, J. Wang, S. He, T. Cheng, B. A. Shoemaker, A. Gindulyte, S. H. Bryant, *Nucleic Acids Res.* **2014**, *42* 1075–1082.
[22] A. Varnek, I. I. Baskin, *Mol. Inf.* **2011**, *30*, 20–32.
[23] J. L. Medina-Franco, G. M. Maggiora, *Molecular Similarity Analysis* In Chemoinformatics for Drug Discovery, J. Bajorath (Ed.), pp. 343–399, Wiley. Hoboken, NJ, USA, **2013**.
[24] For a more comprehensive explanation of all the different chemo-informatics tools reported so far, see "*Chemoinformatics approaches to assess chemical diversity and complexity of small molecules*": F. I. Saldivar-Gonzalez, J. L. Medina-Franco, in *Small Molecule Drug Discovery:*

Methods, Molecules and Applications (Eds.: A. Trabocchi, E. Lenci), Elsevier, Amsterdam, **2019**, pp. 83–102.
[25] H. Kim, T. T. Tung, S. Bum Park, *Org. Lett.* **2013**, *15*, 5814–5817.
[26] T. J. Pawar, H. Jiang, M. A. Vázquez, C. Villegas Gómez, D. Cruz Cruz, *Eur. J. Org. Chem.* **2018**, 1835–1851.
[27] R. Surakanti, S. Sanivarapu, C. Thulluri, P. S. Iyer, R. Tangirala, R. Gundla, U. Addepally, Y. L. Murthy, L. Velide, S. Sen, *Chem. Asian J.* **2013**, *8*, 1168–1176.
[28] H. Zhu, Y. Cai, S. Ma, Y. Futamura, J. Li, W. Zhong, X. Zhang, H. Osada, H. Zou, *ChemSusChem.* **2021**, *14*, 5320–5327.
[29] Y. Choi, H. Kim, S. Bum Park, *Chem. Sci.* **2019**, *10*, 569–575.
[30] W.-H. Shin, X. Zhu, M. G. Bures, D. Kihara, *Molecules* **2015**, *20*, 12841–12862.
[31] T. A. D. Holth, M. A. Walters, O. E. Hutt, G. I. Georg, *ACS Comb. Sci.* **2020**, *22*, 150–155.
[32] O. E. Hutt, T. L. Doan, G. I. Georg, *Org. Lett.* **2013**, *15*, 1602–1605.
[33] N. Nikolova, J. Jaworska, *QSAR Comb. Sci.* **2003**, *22*, 1006–1026.
[34] H. Kikuchi, T. Nishimura, E. Kwon, J. Kawai, Y. Oshima, *Chem. Eur. J.* **2016**, *22*, 15819–15825.
[35] F. Kopp, C. F. Stratton, L. B. Akella, D. S. Tan, *Nat. Chem. Biol.* **2012**, *8*, 358–365.
[36] W. H. B. Sauer, M. K. Schwarz, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 987–1003.
[37] http://lifescience.opensource.epam.com/indigo/.
[38] N. Yang, J. S. Kays, T. R. Skillman, L. Burris, T. W. Seng, C. Hammond, *J. Pharmacol. Exp. Ther.* **2005**, *312*, 127–133.
[39] X. He, X. Chen, S. Lin, X. Mo, P. Zhou, Z. Zhang, Y. Lu, Y. Yang, H. Gu, Z. Shang, Y. Lou, J. Wu, *ChemistryOpen* **2017**, *6*, 102–111.
[40] Z. Li, G. Ngojeh, P. DeWitt, Z. Zheng, M. Chen, B. Lainhart, V. Li, P. Felpo, *Tetrahedron Lett.* **2008**, *49*, 7243–7245.
[41] *Principal Component Analysis*, I. T. Jolliffe, Ed.; Springer-Verlag: New Jork, **2002**.
[42] A. Bender, J. L. Jenkins, J. Scheiber, S. C. K. Sukuru, M. Glick, J. W. Davies, *J. Chem. Inf. Model.* **2009**, *49*, 108–119.
[43] J. Dong, D. S. Cao, H.-Y. Miao, S. Liu, B.-C. Deng, Y.-H. Yun, N.-N. Wang, A.-P. Lu, W.-B. Zeng, A. F. Chen, *J. Cheminf.* **2015**, *7*, 60.
[44] M. E. Maffei, G. Arimura, A. Mithofer, *Nat. Prod. Rep.* **2012**, *29*, 1288–1303.
[45] S. Yun, N. Kang, S. Park, H. Ha, Y. K. Kim, J. Lee, Y. Chang, *Acc. Chem. Res.* **2014**, *47*, 1277–1286.
[46] N. S. Finney, *Curr. Opin. Chem. Biol.* **2006**, *10*, 238–245.
[47] X. Luo, L. Qian, Y. Xiao, Y. Tang, Y. Zhao, X. Wang, L. Gu, Z. Lei, J. Bao, J. Wu, T. He, F. Hu, J. Zheng, H. Li, W. Zhu, L. Shao, X. Dong, D. Chen, X. Qian, Y. Yang, *Nat. Commun.* **2019**, *10*, 258.
[48] T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal, S. W. Krska, *Chem. Soc. Rev.* **2016**, *45*, 546–576.
[49] M. Moir, J. J. Danon, T. A. Reekie, M. Kassiou, *Expert Opin. Drug Discovery* **2019**, *14*, 1137–1149.
[50] B. Hong, T. Luo, X. Lei, *ACS Cent. Sci.* **2020**, *27*, 622–635.
[51] E. Driggers, S. Hale, J. Lee, N. K. Terrett, *Nat. Rev. Drug Discovery* **2008**, *7*, 608–624.
[52] H. S. G. Beckmann, F. Nie, C. E. Hagerman, H. Johansson, Y. S. Tan, D. Wilcke, D. R. Spring, *Nat. Chem.* **2013**, *5*, 861–867.
[53] A. Grossmann, S. Bartlett, M. Janecek, J. T. Hodgkinson, D. R. Spring, *Angew. Chem. Int. Ed.* **2014**, *53*, 13093–13097; *Angew. Chem.* **2014**, *126*, 13309–13313.
[54] F. Nie, D. L. Kunciw, D. Wilcke, J. E. Stokes, W. R. J. D. Galloway, S. Bartlett, H. F. Sore, D. R. Spring, *Angew. Chem. Int. Ed.* **2016**, *55*, 11139–11143; *Angew. Chem.* **2016**, *128*, 11305–11309.
[55] A. Guarnieri-Ibáñez, F. Medina, C. Besnard, S. L. Kidd, D. R. Spring, J. Lacour, *Chem. Sci.* **2017**, *8*, 5713–5720.
[56] L. Li, Z.-L. Li, F.-L. Wang, Z. Guo, Y.-F. Cheng, N. Wang, X.-W. Dong, C. Fang, J. Liu, C. Hou, B. Tan, X.-Y. Liu, *Nat. Commun.* **2016**, *7*, 13852.
[57] S. L. Kidd, T. J. Osberger, N. Mateu, H. F. Sore, D. R. Spring, *Front. Chem.* **2018**, *6*, 460.
[58] T. Flagstad, G. Min, K. Bonnet, R. Morgentin, D. Roche, M. H. Clausen, T. E. Nielsen, *Org. Biomol. Chem.* **2016**, *14*, 4943–4946.
[59] D. J. Foley, A. Nelson, S. P. Marsden, *Angew. Chem. Int. Ed.* **2016**, *55*, 13650–13657; *Angew. Chem.* **2016**, *128*, 13850–13857.
[60] S. E. Motika, P. J. Hergenrother, *Nat. Prod. Rep.* **2020**, *37*, 1395–1403.
[61] T.-G. Chen, L. M. Barton, Y. Lin, J. Tsien, D. Kossler, I. Bastida, S. Asai, C. Bi, J. S. Chen, M. Shan, H. Fang, F. G. Fang, H.-W. Choi, L. Hawkins, T. Qin, P. S. Baran, *Nature* **2018**, *560*, 350–354.
[62] L. Chen, L. Yu, Y. Deng, Z.-J. Zheng, Z. Xu, J. Cao, L.-W. Xu, *Adv. Synth. Catal.* **2016**, *358*, 480–485.
[63] Y.-H. Sun, T.-Y. Sun, Y.-D. Wu, X. Zhang, Y. Rao, *Chem. Sci.* **2016**, *7*, 2229–2238.
[64] Y. Xia, A. Studer, *Angew. Chem. Int. Ed.* **2019**, *58*, 9836–9840; *Angew. Chem.* **2019**, *131*, 9941–9945.
[65] D. M. Kuznetsov, A. G. Kutateladze, *J. Am. Chem. Soc.* **2017**, *139*, 16584–16590.
[66] M. von Korff, T. Sander, *Sci. Rep.* **2019**, *9*, 967.
[67] S. H. Nilar, N. L. Ma, T. H. Keller, *J. Comput.-Aided Mol. Des.* **2013**, *27*, 783–792.
[68] T. Böttcher, *J. Chem. Inf. Comput. Sci.* **2016**, *56*, 462–470.
[69] N. Mateu, S. L. Kidd, L. Kalash, H. F. Sore, A. Madin, A. Bender, D. R. Spring, *Chem. Eur. J.* **2018**, *24*, 13681–13687.
[70] I. Borg, P. J. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. Springer Science & Business Media, **2005**.
[71] V. Nesterenko, K. S. Putt, P. J. Hergenrother, *J. Am. Chem. Soc.* **2003**, *125*, 14672–14673.
[72] M. C. Schuster, D. A. Mann, T. J. Buchholz, K. M. Johnson, W. D. Thomas, L. L. Kiessling, *Org. Lett.* **2003**, *5*, 1407–1410.
[73] C. J. Smedley, G. Li, A. S. Barrow, T. L. Gialelis, M.-C. Giel, A. Ottonello, Y. Cheng, S. Kitamura, D. W. Wolan, K. B. Sharpless, J. E. Moses, *Angew. Chem. Int. Ed.* **2020**, *59*, 12460–12469; *Angew. Chem.* **2020**, *132*, 12560–12569.
[74] I. Colomer, C. Jempson, P. Craven, Z. Owen, R. G. Doveston, I. Churcher, S. P. Marsden, A. Nelson, *Chem. Commun.* **2016**, *52*, 7209–7212.
[75] E. Lenci, G. Menchi, A. Guarna, A. Trabocchi, *J. Org. Chem.* **2015**, *80*, 2182–2191.
[76] E. Lenci, R. Innocenti, G. Menchi, C. Faggi, A. Trabocchi, *Org. Biomol. Chem.* **2015**, *13*, 7013–7019.
[77] E. Lenci, R. Innocenti, G. Menchi, A. Trabocchi, *Front. Chem.* **2018**, *6*, 522.
[78] R. Innocenti, E. Lenci, L. Baldini, C. Faggi, G. Menchi, A. Trabocchi, *Eur. J. Org. Chem.* **2019**, *36*, 6203–6210.
[79] R. Innocenti, E. Lenci, G. Menchi, A. Trabocchi, *Beilstein J. Org. Chem.* **2020**, *16*, 200–211.
[80] E. Lenci, G. Menchi, F. I. Saldívar-Gonzalez, J. L. Medina-Franco, A. Trabocchi, *Org. Biomol. Chem.* **2019**, *17*, 1037–1052.
[81] F. I. Saldívar-González, E. Lenci, A. Trabocchi, J. L. Medina-Franco, *RSC Adv.* **2019**, *9*, 27105–27116.
[82] T. Hoffmann, M. Gastreic, *Drug Discovery Today* **2019**, *24*, 1148–1156.
[83] W. P. Walters, *J. Med. Chem.* **2019**, *62*, 1116–1124.
[84] L. Humbeck, S. Weigang, T. Schfer, P. Mutzel, O. Koch, *ChemMedChem* **2018**, *13*, 532–539.
[85] F. I. Saldívar-González, E. Lenci, L. Calugi, J. L. Medina-Franco, A. Trabocchi, *Bioorg. Med. Chem.* **2020**, *28*, 115539.
[86] M. González-Medina, F. D. Prieto-Martínez, J. R. Owen, J. L. Medina-Franco, *J. Cheminf.* **2016**, *8*, 63.
[87] M. González-Medina, J. R. Mariana Owen, T. El-Elimat, C. J. Pearce, N. H. Oberlies, M. Figueroa, J. L. Medina-Franco, *Front. Pharmacol.* **2017**, *8*, 180.

Manuscript received: May 16, 2022
Revised manuscript received: June 13, 2022
Accepted manuscript online: June 17, 2022