Reaching for the bright StARs in chemical space

José L. Medina-Franco[1], J. Jesús Naveja[1,2] and Edgar López-López[1]

[1] Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico
[2] PECEM, School of Medicine, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico

Visualization of activity data in chemical space is common in drug discovery. Navigating the space in a systematic manner is not trivial, given its size and huge coverage. To this end, methods for data visualization have been developed charting biological activity into chemical space. Herein, we review the progress in different visualization approaches to explore the chemical space aiming at reaching insightful structure–activity relationships (SARs) in the chemical space. We discuss recent methods including consensus diversity plots, ChemMaps, and constellation plots. Several of the methods we review can be extended to analyze other properties of interest in medicinal chemistry, such as structure–toxicity relationships, and can be adapted to postprocess results of virtual screening (VS) of large compound libraries.

Introduction

Identifying structure–property relationships (SPRs) is a common practice in many areas of chemistry. To this end, data visualization, although it might not reveal all patterns in the data in an automatic manner, is a helpful approach to uncover such patterns. Drug discovery teams constantly conduct visualization of SARs to augment efficiency. This applies to the individual level, where the scientist uses visualization methods to help convert data into knowledge and, eventually, knowledge into wisdom [1]. Similarly, data visualization is key in research teams of either a handful of scientists or across large teams in different sites. Moreover, data visualization is also a key component in teaching and research publications, where it can assist the reader in interpreting the data more easily and identifying underlying patterns in them.

Visualization of SARs/SPRs is challenging considering the increasing amount of data that is being generated by, for example, high-throughput screening (HTS) of large collections, the size and features of which have evolved over the years [2]. In addition, virtual collections are screened first with computational methods followed by experimental approaches [3]. There are also large public repositories available of compound databases annotated with biological activity, such as ChEMBL 25 [4] and PubChem [5], that could benefit from methods to visualize high-dimensional data.

Over the past 20 years, several strategies for SAR visualization have been developed, some of which are customized to analyze the output of HTS [6]. Overall, these methods have two major parts: (i) clustering or organization of information of chemical structures; and (ii) visualization approaches to project activity data into the cluttered information [7]. The first is frequently associated with the organization or visualization of the chemical space covered by the clustered compounds, whereas the second part of mapping activity data into the chemical space (whether purely numerically and/or visually as well) is associated with the concept of activity landscape modeling [8].

Several useful reviews collect different approaches for the visualization of SAR developed since the early methods presented by Gedeck and Willett [9]. For instance, in 2009 Villar and Hansen reviewed chemoinformatic approaches for data mining in drug discovery [10]. A year later, Wawer *et al.* published a comprehensive survey on data structures and *in silico* methods to uncover SAR insights from large compound collections [11]. In 2013, Bajorath reviewed methods for large-scale SAR analysis, including

Corresponding author: Medina-Franco, J.L. (jose.medina.franco@gmail.com), (medinajl@unam.mx)

https://doi.org/10.1016/j.drudis.2019.09.013

numerical and visualization approaches [12]. In 2016, Stumpfe and Bajorath reviewed developments for visualizing SAR [13].

Visualization of the chemical space *per se* (i.e., not necessarily mapping activity data into such space) is an active area of research [14–18]. There are several other review papers covering recent developments to visualize the chemical space [19–21]. Here, we review approaches developed to visualize SAR data with a focus on broadly used and recent methods to visualize large chemical spaces and multidimensional activity data. To emphasize the connection between SAR analysis in the context of visualization of chemical space we refer in the title to 'StARs' (**st**ructure-**a**ctivity **r**elationship**s**). We discuss general tools for SAR visualization with an emphasis on the most recent approaches. Whenever possible we briefly comment on the sector that developed the software (e.g., industry, academic), availability (e.g., commercial, free, standalone, web-based) and application examples. We then review representative commercial and free software, including open source codes, and then survey web-based tools with a focus on public servers. Throughout, we discuss illustrative examples of applications of the different tools, software, and servers to explore the SAR of screening data, and the outcome from VS (e.g., post processing VS results, and multi-target prediction). Some tools have been used to analyse structure–multiple activity relationships or SmART [22] and cell data. The methods can also be easily adapted to visualize and postprocess results of large VSs.

Visualization methods

Several chemoinformatics tools, software, and applications use an approach to cluster compounds and then another method(s) to visualize the data that are clustered. Clustering and other data analysis methods can be considered direct visualizations of the data themselves.

Perhaps one of the most straightforward ways to visualize the SAR of compound series is through R-group tables. Indeed, for its ease of interpretation, such R-group tables continue to be used and are implemented in commercial software used by medicinal chemists. However, although these tables are a great resource for practicing medicinal chemists, they are more difficult to adapt for visualizing heterogenous and large data, for instance, generated through HTS [12]. However, different methods, based on either 2D or 3D depictions, have been proposed by academic or industry groups and are used for internal purposes without the need to release a software or web-based application. Examples of these approaches, with an emphasis on representative common methods and recent ones, are summarized in Table 1 and discussed here.

Molecular Property eXplorer

Molecular Property eXplorer (MPX) is an example of an in-house software developed by a pharmaceutical company. It visualizes clustered data using tree-maps and heat maps [7]. For clustering, MPX uses agglomerative hierarchical clustering to group data based on their similarity between 2D chemical structures or biological assay values. MPX has implemented tools for the user to query data sets by activity or chemical structures.

SAR maps and enhanced SAR maps

SAR maps [23] and enhanced SAR maps [24] are also in-house programs developed by a pharmaceutical company. Both are based on R-group analysis of series of compounds. R-group decomposition is one of the most common and interpretable representations used by medicinal chemists. The visual representation of the SAR is a heat map (rectangular matrix of cells) where the columns and rows represent substituents of two selected modification sites. Each cell in the heat maps represents one compound that is associated with the attaching substituents. The cells can be color coded in different manners, such as by chemical property or biological activity, thus allowing for complex but insightful visualizations.

Visual representation of activity landscapes

Activity landscape modeling has been defined as any representation (numerical or graphical) that compares in a systematic manner, structure similarity with potency relationships of pairs of compounds that share a specific activity (e.g., quantified with one or more biological endpoints) [25]. A recent review of modeling activity landscapes for drug discovery is available [26], with a focus on mathematical, statistical, and machine-learning approaches for modeling global landscapes. Bajorath, his group and other research teams have worked extensively on developing numerical SAR functions and visual representations of activity landscapes. Exemplary approaches are structure–activity similarity (SAS) maps and their variants, structure–activity landscape index (SALI) graph, network-like similarity graphs (NSGs), and related bipartite matching molecular series graphs. These and other approaches for multitarget activity landscapes, including ligand–target differentiation maps, have been reviewed extensively [12]. A recent example was

TABLE 1

Examples of methods for the visualization of SARs

Method	Brief description	Refs
SAR matrices (SARMs)	Systematic extraction of structurally related compounds, as well as visualization of SAR patterns and generation of virtual analogs	[55]
SAR maps and enhanced SAR maps	Intuitive method for study and visualization of substituents (R groups); allows identification of missing analogs in the set, in addition to the influence of these on activity (or other property of interest)	[23,24]
SAR Analyser	Interactive tool that combines techniques, such as Map SAR and Tree SAR	[56]
Consensus Diversity Plots	Integral study of chemical space; combines different criteria to measure molecular similarity (e.g., fingerprints, physicochemical and pharmaceutical properties, and scaffold similarity)	[29]
ChemMaps	Uses range of reference molecules (satellites) to systematically explore chemical space (ChemGPS drugspace)	[15]
Shinyheatmap	Ultrafast low memory heatmap web interface for big data -omic approach	[57]

presented in 3D activity landscape views adding a biological activity surface on different 2D projections of the chemical space [27].

Drug Discovery Maps

Drug Discovery Map is a useful machine-learning model for mapping and clustering chemical and biological information [28]. These maps are based on the t-distributed stochastic neighbor embedding (t-SNE) algorithm to generate a visualization of molecular and biological similarity. In the original and a recent report, authors of that work applied the 'drug discovery maps' to a group of kinases. Janssen *et al*. analyzed a heterogeneous series of 2272 compounds approved as inhibitors of kinases, obtaining clustering (31 chemotypes) and biological differentiation among these (200 kinases). In the same way, by aligning the amino acids of the catalytic site of these kinases and analyzing their physicochemical properties, they were able to cluster kinases and reproduce some previous phylogenetic studies of these proteins. This illustrates that methods such as t-SNE are able to capture the chemical space in an intuitive way without losing high resolution (and predictive power). However, its potential for the classification and study of peptide and peptidomimetic compounds is yet to be explored.

Consensus diversity plots

Consensus diversity (CD) plots are 2D graphs that represent, for a given data set, fingerprint similarity versus scaffold diversity (along the *X*- and *Y*-axes, respectively) [29]. Each data set is represented by a data point in the graph. The diversity based on whole properties (e.g., physicochemical properties of pharmaceutical relevance) is represented in a continuous color scale. The relative size of the database is captured in the size of the data point. Data points can be color coded by any other average/representative measure of the activity of the compounds or any other activity related metric. Thus, CD plots can be interpreted as an approach to visualize the overall diversity of compound data sets. Alternatively, CD plots can be interpreted as a chemical space defined by a set of diversity metrics that are used as reference (diversity space).

ChemMaps

ChemMaps is an approach for the visual representation of chemical space based on the similarity matrix of compound databases generated with the similarity of the molecular fingerprints [15]. The approach uses a 'satellites' approach [30,31] where satellites are, in principle, molecules the similarity of which to the rest of the molecules in the database provides sufficient information for generating a visualization of the chemical space. ChemMaps, as with any other method to represent the chemical space, can be used as basis to visualize activity data. For instance, ChemMaps were recently used to explore the SAR of 52 epigenetic target data sets. There, the fraction of activity cliffs was represented with a continuous color scale leading the rapid identification of clusters of epigenetic data sets with an overall continuous SAR (such as inhibitors of histone deacetylase, HDAC). In addition, data sets with a rough landscape and a large proportion of activity cliffs were identified, such as inhibitors of disruptor of telomeric silencing 1, DOT1L [32].

Constellation plots

Constellation plots were developed as a graphical representation of SAR that integrates coordinate-based chemical space representation with analog series [33]. Constellation plots enable the identification of whole regions in chemical space rich in SAR annotations, either active regions ('bright' SAR in analogy with cosmic space) or 'dark regions', where few or none active molecules have been found for that target. The 'constellations' are given by groupings of analog series in the coordinate-based chemical space.

Constellation plots attempt to reduce the number of points depicted in chemical space representations, while increasing the quality and volume of data legibly represented in a 2D chemical space plot. Their aim is to group analogs that clearly share a common core, which can be then compared with other cores and their corresponding associated molecules.

Fig. 1 provides an example of a constellation plot generated with the same data but different dimensionality reduction analyses (i.e., t-SNE, PCA, and GTM). The data analyzed to produce this plot was a subset of 2789 molecules from a virtual library focused on DNMT inhibitors generated by Chemspace (https://chem-space.com/). From the original data set, 472 molecules could be mapped to 152 analog series and summarized in 188 cores as described in [33]. Sometimes, cores in the same analog series also share compounds that could be mapped to more than a single core; this is represented by lines linking the cores. The color represents the average docking score against DNMT1 [Protein Data Bank (PDB) ID: 3PTA] obtained in a molecular operating environment (MOE) after standard preprocessing of the protein and ligands. The size of the circle indicates the relative number of compounds sharing the core. The plot readily reveals 'highly' populated core structures with overall favorable docking scores. Constellation plots have also been used to explore the SAR of AKT1 inhibitors and the SmART of small molecules tested as inhibitors of three different DNMTs [33].

As can happen for other visualization tools of experimental SAR or SmARTs, constellation plots can be adapted to post processing the results of VS of large data sets, such as in the example presented in Fig. 1. The only practical difference with an experimentally derived activity value is that the 'property' to map/represent with a color can be a calculated score/activity. Of course, the mapped property could be the result of almost any VS approach, such as similarity searching, docking scores, or a combination-consensus score.

Standalone software for SAR visualization

Some SAR visualization methods have been developed as either free or commercial software and web-based platforms. These are summarized in Table 2. Over the past few years, several software packages or applications have been developed, such as LeadScope, LeadPharmer, and SARNavigator, which have been reviewed elsewhere [7,9]. Herein, we focus on the most recently developed software applications. Free software is highlighted.

Open source

DataWarrior

DataWarrior is a free and interactive software used mainly among the drug discovery community to support data-driven approaches. It focuses on extracting knowledge from chemical information

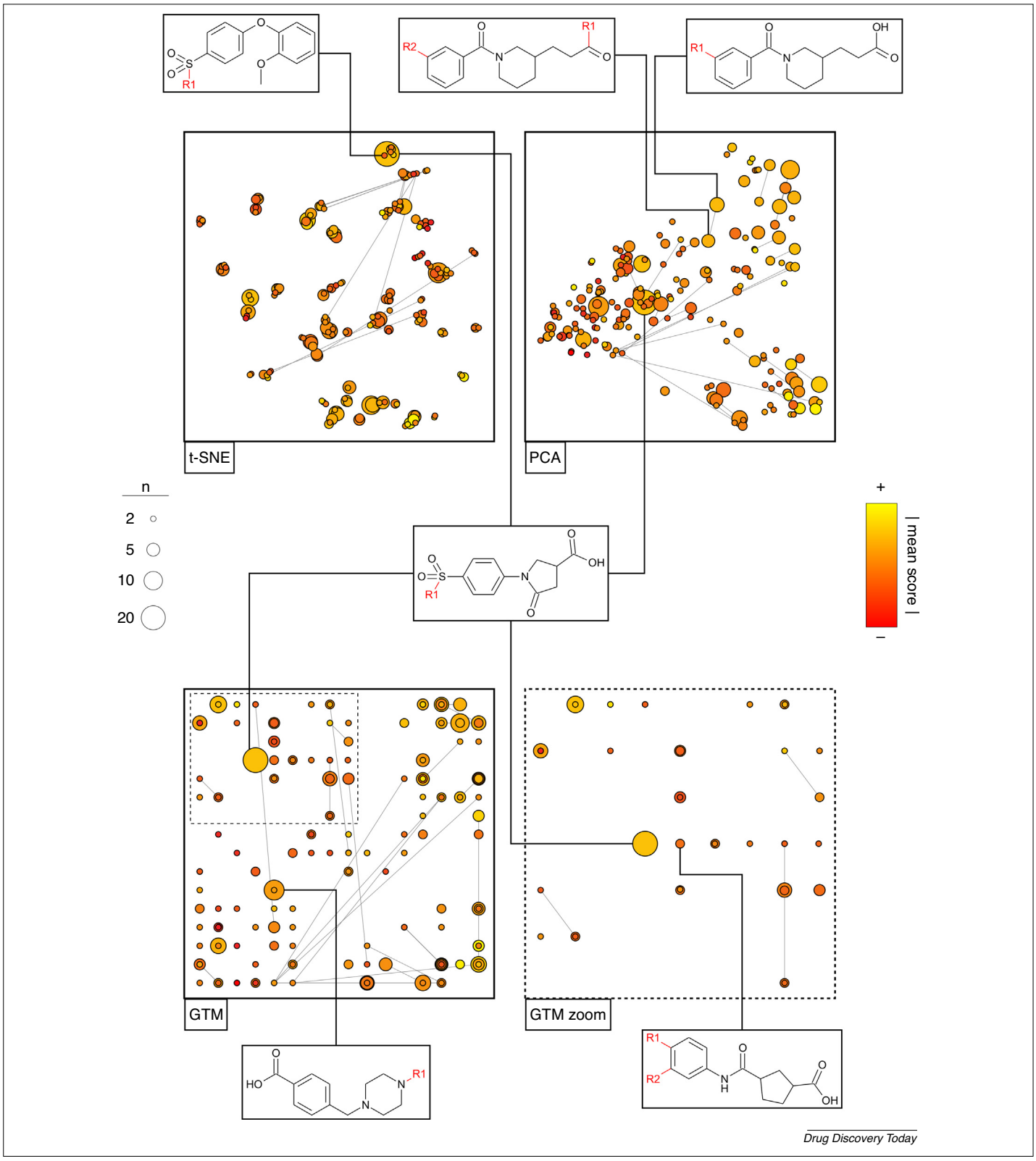

FIGURE 1

Constellation plots showing the 188 cores found for a data set of 472 molecules in 153 analog series. Linking lines represent shared molecules between two cores. The color represents the average of the docking score for all molecules in a core (red: less active; yellow: more active). All three panels have the same information, but the chemical space representation is different: t-SNE; PCA; and GTM. A zoom-in into a delimited area in GTM chemical space is also shown.

TABLE 2

Examples of free and commercial applications available for visualization of SAR

Software	Brief description	Refs
Standalone		
DataWarrior	Open, multipurpose software for data analysis and visualization broadly used in drug discovery projects	[37]
Tableau	Tools for preparation and interactive data analysis (www.tableau.com)	[36,58]
SARANEA	Free tool that supports SAR and selectivity analysis; based on network graph visualization and oriented to explore SAR and activity landscape of a full set of molecules	[38]
Scaffold Hunter	Free tool used to find associations between structural features and different activity levels; visualization based on a radial tree representation showing a hierarchical scaffold structure and activity level of scaffolds	[59]
Spotfire	Commercial software for visualization, data analysis, and pattern recognition; has predictive, classification, and grouping methods in a dynamic graphics environment (www.tibco.com/products/tibco-spotfire)	[34,60]
Molecular Property eXplorer (MPX)	Commercial software developed and used within a pharmaceutical company; visualizes clustered data using tree-maps and heatmaps	[7]
LeadScope	Software that allows structural exploration of large chemical libraries	[61]
Optibrium: StarDrop	Commercial software useful in study of pharmaceutical descriptors, construction, visualization, and interactive screening of chemical space as well as construction of quantitative SAR models	[46]
Miner3D	Software program for data visualization and interactive data analysis. (https://secure.miner3d.com)	[35]
Molecular Operating Environment – SAR Report	Allows structural study of substituents and scaffolds of compounds, as well as creation of SAR diagrams	[50]
Web based		
VisualiSAR	Web-based application for clustering, structure browsing, and SAR study	[49]
MOESaic	Browser-based commercial application for analyzing series of chemical structures and related property data; allows search of substructures and analysis of paired molecular pairs (MMP), analysis of the R group with defined scaffolds, statistical and visual analysis of properties, and virtual design of structures	[50]
Activity Landscape Plotter	Tools for activity landscape analysis using SAS maps and dual-activity difference maps	[52]

and identifying associations between chemical structures and alphanumeric information. Similar to Spotfire [34] and Miner 3D [35] (which are both commercial), DataWarrior is a visualization software program useful for its ability to interactively represent multidimensional activity data in 2D or 3D. It enables further dimensionality by adding a physical meaning to te color, shape, and size of the data points. DataWarrior was developed by a pharmaceutical company that released a public version in 2014 [36]. Currently, it is used by researchers from the academic, government, and industry sectors. Its applications have spanned basic and applied research to teaching [36]. In addition to Spotfire (see later) and Miner 3D, other related software are Dotmatics and Tableau, recently reviewed and compared directly elsewhere [37].

SARANEA

SARANEA is a free graphical user interface to NSGs and NSG-based data structures and accepts customized molecular fingerprint representations and potency data as input. The main focus of SARANEA is the simultaneous interactive exploration of multiple NSGs. Compound selection in one (potency or selectivity) NSG representation can be automatically synchronized with all other NSG views, making it possible to compare multiple SARs and SSRs for different targets. The program also integrates user-defined descriptors and depicts molecules, providing immediate interactive access to the molecular structures represented by nodes in the graphs. SAR and SSR trees, pathways, and chemical neighborhood graphs are integrated [38]. The source code is also freely available.

Scaffold Hunter

Scaffold Hunter facilitates the exploration of the chemical space of a compound data set by integrating different graphical data visualizations in complementary views with appropriate analysis methods such as clustering. A key component of Scaffold Hunter are the scaffold trees, which enable sets of molecules with a shared scaffold to be grouped and reveal the relationships between those scaffolds based on the structural inclusion relation [39,40]. The scaffold tree concept is reminiscent of a hierarchical-based classification of scaffolds and their assessment using enrichment factors [41]. Scaffold Hunter was recently used for the hierarchical classification and visualization of the SAR nucleoside and non-nucleoside reverse transcriptase inhibitors of HIV-1 [42].

Commercial

Herein, we mention briefly three examples of commercial software tailored for visualization of SARs. This section is not intended to be comprehensive, providing a representation of the importance of data visualization integrated in software suites. These and other examples are presented in Table 2.

Spotfire

Spotfire has been around for >20 years (reviewed in [37]). The software is used by research groups in industry, academic, and government. It has also been used as basis to develop applications within pharmaceutical companies for data mining and visualization. An example is SARConnect, developed and used in a pharmaceutical company to integrate biological and chemistry data. SARConect represents an integration between chemoinformatics and bioinformatics [43].

MOE: SAR Report

The software MOE, developed by te Chemical Computing Group, implements a structure–activity report that is reminiscent of the

SAR maps discussed earlier. The application includes an algorithm for scaffold detection and fragmentation [44]. The output is a detailed and interactive graphical representation of the SAR and the user can map one or more properties of interest for a set of compounds. Recently, the SAR report of MOE was used to explore the SmART of a data set of synthetic molecules with reported activity against DNMTs and G9a, two major epigenetic targets [45].

StarDrop

StarDrop is a comprehensive suite of integrated and software developed by Optibrium. It includes a module called 'SAR Analysis' that enables the user to explore the SAR using R-group analysis, compound clustering, and activity landscape analysis [46].

Web-based applications

Web-based applications have a large impact on the chemoinformatics community and drug discovery teams. Villoutreix *et al.* compiled an extensive list of web-based applications for different purposes [47]. Additional servers have also been reviewed [48]. Overall, several molecular modeling tools include tools for data visualization, for example, for rendering chemical structures in 3D. Likewise, a large number of chemoinformatic tools include graphs to display clustered compounds or generate visual representations of calculated or experimental properties. Here, we focus on web-based applications that primarily focus on SAR visualization or have a strong component of features for such purposes. We also discuss large public compound data sets that have implemented tools for visualization. Representative applications are summarized in Table 2.

VisualiSAR

VisualiSAR was one of the first web-based visualization methods released. It is a proprietary application that was designed to visualize chemical structures, and to identify and emphasize similarities between them. VisualiSAR combines cluster analyses of chemical structures and modal analyses of molecular fingerprints. The program is tuned to analyze large data sets [49].

MOESaic: SAR explorer

Chemical Computing Group recently implemented a web-based application called MOESaic [50] that was designed to enhance common medicinal chemistry workflows to explore SAR/SPRs and contribute to the efficient visualization, analysis, and profiling of compounds. Methodologically, MOESaic is based on an interactive matched molecular pairs (MMP) analysis and R-group profiling. The web-server is available to those that hold a MOE license.

Activity Landscape Plotter

Based on the concept of activity landscape modeling, a web-based application to generate SAS and Dual-Activity Difference (DAD) maps [51] was recently developed. The latter are convenient to explore SmARTs. The application, Activity Landscape Plotter is intended to generate SAS and DAD maps from user-supplied data sets in a text file that contains structure (e.g., SMILES) and activity data [52]. The server automatically computes pairwise SARs using different available fingerprint representations. The user can further map and/or project different information in the SAS/DAD maps, such as SALI values and maximum activity on the compound pair. The free webserver has been used to identify activity cliffs [45,53]. Another approach to visualize SmARTs is radial coordinate-based multitarget visualization [54].

To further illustrate selected methods discussed herein, Fig. 2a shows a chemical space representation of 2789 molecules from a virtual library focused on DNMT inhibitors (as used in Fig. 1) with the ChemMaps methodology. The docking score is represented by

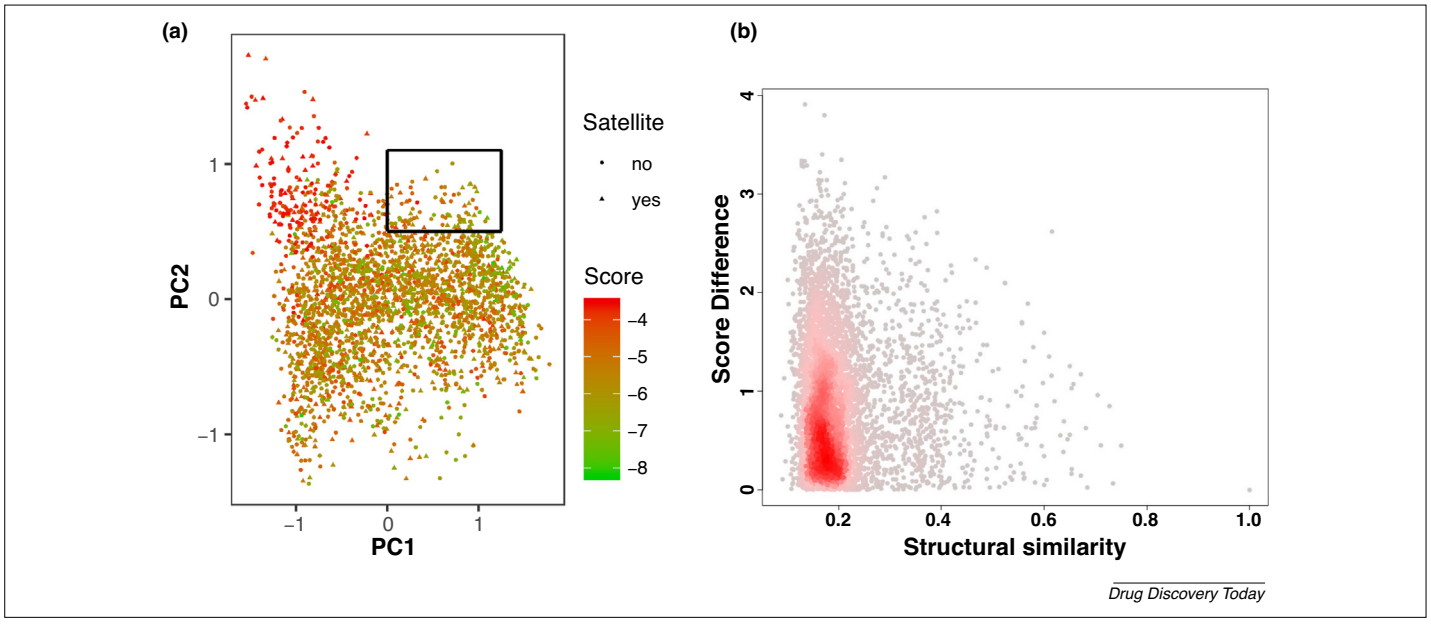

FIGURE 2

Visualization tools to visualize chemical space and post-process results of virtual screening. (a) Chemical space representation generated with the ChemMaps methodology. Axes represent the first two principal components. The same data set of 2789 molecules as in Fig. 1 was used. An illustrated subset of 836 molecules as satellites reached convergence in ChemMaps. (b) An adapted density structure–activity similarity (SAS) map is presented using a subset of molecules selected as shown in (a). The SAS map was adapted to represent differences in docking scores (instead of values of biological activity).

a continuous color scale from green (most favorable score) to red (least favorable score). The molecules used as satellites (836) are distinguished from all the molecules using a different shape. Fig. 2b illustrates a density SAS map of a subset of molecules selected from ChemMaps, marked with a rectangle. The density SAS map was generated with Activity Landscape Plotter.

Concluding remarks and outlook

In drug discovery, visualization methods have a major role for data mining and information extraction from constantly increasing data sets. Indeed, visualization methods can be crucial tools for interpreting experimental data and bridge experimental and *in silico* information. Although data visualization alone does not solve problems *per se* and is not currently tuned to automatically suggest to the medicinal chemist the compound to synthesize next, it is important in decision-making by multidisciplinary research teams and also teaching. To address the needs of data visualization, in-house, public, and commercial tools have been developed by academic and private research groups. Despite several advances in the field, new or improved methods continue to emerge and evolve. Such developments have been implemented in a range of formats, including concepts, easy-to-adapt graphs, stand-alone or web-services (public or commercial). The pharmaceutical industry also continues to develop in-house tools, some of which have been released to the public. Conceptually, SAR and SmART visualization is at the interface of chemical space, data mining, and activity landscape modeling. Similar to many cheminformatics methods, data visualization and its interpretation depends on the molecular representation.

It is anticipated that large public repositories of screening data will incorporate and improve interactive visualization methods so that the user can uncover patterns and help transforming data into knowledge and knowledge into wisdom. The visualization methods discussed herein can also be adapted to visualize *in silico* profiling data, for instance to post process results of VS of large compound libraries and multitarget profiling campaigns. Moreover, these visualization approaches can be extended to analyze other properties of interest in medicinal chemistry, such as structure–toxicity relationships.

Acknowledgments

We thank Consejo Nacional de Ciencia y Tecnología (CONACyT) for grant 282785. J.J.N. is thankful to CONACyT for the granted scholarship number 622969. We thank Rajarshi Guha for his comments and proofreading the manuscript.

References

1 Frické, M. (2019) The knowledge pyramid: the DIKW hierarchy. *Knowl. Organ.* 46, 33–46
2 Volochnyuk, D.M. *et al.* (2019) Evolution of commercially available compounds for HTS. *Drug Discov. Today* 24, 390–402
3 Walters, W.P. (2019) Virtual chemical libraries. *J. Med. Chem.* 62, 1116–1124
4 Gaulton, A. *et al.* (2017) The ChEMBL database in 2017. *Nucleic Acids Res.* 45, D945–D954
5 Kim, S. *et al.* (2019) PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 47, D1102–D1109
6 Strobelt, H. *et al.* (2012) HiTSEE KNIME: a visualization tool for hit selection and analysis in high-throughput screening experiments for the KNIME platform. *BMC Bioinformatics* 13 (Suppl. 8), S4
7 Kibbey, C. and Calvet, A. (2005) Molecular Property eXplorer: a novel approach to visualizing SAR using tree-maps and heatmaps. *J. Chem. Inf. Model.* 45, 523–532
8 Bajorath, J. (2012) Modeling of activity landscapes for drug discovery. *Expert Opin. Drug Discov.* 7, 463–473
9 Gedeck, P. and Willett, P. (2001) Visual and computational analysis of structure–activity relationships in high-throughput screening data. *Curr. Opin. Chem. Biol.* 5, 389–395
10 Villar, H.O. and Hansen, M.R. (2009) Mining and visualizing the chemical content of large databases. *Curr. Opin. Drug Discov. Dev.* 12, 367–375
11 Wawer, M. *et al.* (2010) Data structures and computational tools for the extraction of SAR information from large compound sets. *Drug Discov. Today* 15, 630–639
12 Bajorath, J. (2013) Large-scale SAR analysis. *Drug Discov. Today Technol.* 10, e419–e426
13 Stumpfe, D. and Bajorath, J. (2016) Recent developments in SAR visualization. *Med. Chem. Commun.* 7, 1045–1055
14 de la Vega de León, A. and Bajorath, J. (2016) Chemical space visualization: transforming multidimensional chemical spaces into similarity-based molecular networks. *Future Med. Chem.* 8, 1769–1778
15 Naveja, J.J. and Medina-Franco, J.L. (2017) ChemMaps: towards an approach for visualizing the chemical space based on adaptive satellite compounds. *F1000Res* 6, 1134
16 Probst, D. and Reymond, J.-L. (2018) Exploring Drugbank in virtual reality chemical space. *J. Chem. Inf. Model.* 58, 1731–1735
17 Karlov, D.S. *et al.* (2019) Chemical space exploration guided by deep neural networks. *RSC Adv.* 9, 5151–5157
18 Capecchi, A. *et al.* (2019) PubChem and ChEMBL beyond Lipinski. *Mol. Inform.* 38, e1900016
19 Medina-Franco, J.L. and Aguayo-Ortiz, R. (2013) Progress in the visualization and mining of chemical and target spaces. *Mol. Inform.* 32, 942–953
20 Reymond, J.-L. (2015) The chemical space project. *Acc. Chem. Res.* 48, 722–730
21 Osolodkin, D.I. *et al.* (2015) Progress in visual representations of chemical space. *Expert Opin. Drug Discov.* 10, 959–973
22 Saldívar-González, F.I. *et al.* (2017) Getting SMARt in drug discovery: chemoinformatics approaches for mining structure–multiple activity relationships. *RSC Adv.* 7, 632–641
23 Agrafiotis, D.K. *et al.* (2007) SAR maps: a new SAR visualization technique for medicinal chemists. *J. Med. Chem.* 50, 5926–5937
24 Kolpak, J. *et al.* (2009) Enhanced SAR maps: expanding the data rendering capabilities of a popular medicinal chemistry tool. *J. Chem. Inf. Model.* 49, 2221–2230
25 Wassermann, A.M. *et al.* (2010) Activity landscape representations for structure-activity relationship analysis. *J. Med. Chem.* 53, 8209–8223
26 Vogt, M. (2018) Progress with modeling activity landscapes in drug discovery. *Expert Opin. Drug Discov.* 13, 605–615
27 Miyao, T. *et al.* (2019) Three-dimensional activity landscape models of different design and their application to compound mapping and potency prediction. *J. Chem. Inf. Model.* 59, 993–1004
28 Janssen, A.P.A. *et al.* (2019) Drug Discovery Maps, a machine learning model that visualizes and predicts kinome-inhibitor interaction landscapes. *J. Chem. Inf. Model.* 59, 1221–1229
29 González-Medina, M. *et al.* (2016) Consensus Diversity Plots: a global diversity analysis of chemical libraries. *J. Cheminform.* 8, 63
30 Larsson, J. *et al.* (2005) Expanding the ChemGPS chemical space with natural products. *J. Nat. Prod.* 68, 985–991
31 Oprea, T.I. and Gottfries, J. (2001) Chemography: the art of navigating in chemical space. *J. Comb. Chem.* 3, 157–166
32 Oviedo-Osornio, C.I. *et al.* (2018) Computational methods for epigenetic drug discovery: a focus on activity landscape modeling. *Adv. Protein Chem. Str.* 113, 65–83
33 Naveja, J.J. and Medina-Franco, J.L. (2019) Finding constellations in chemical space through core analysis. *Front. Chem.* 7, 510
34 TIBCO Software Inc (1996) *TIBCO Spotfire.* 1996 [accessed 17 September 2019]www.tibco.com/products/tibco-spotfire
35 Addinsoft (2018) *Miner3D.* 2018 [accessed 17 September 2019] https://secure.miner3d.com/
36 Sander, T. *et al.* (2015) DataWarrior: an open-source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.* 55, 460–473

37 López-López, E. *et al.* (2019) DataWarrior: an evaluation of the open-source drug discovery tool. *Expert Opin. Drug Discov.* 14, 335–341
38 Lounkine, E. *et al.* (2010) SARANEA: a freely available program to mine structure-activity and structure-selectivity relationship information in compound data sets. *J. Chem. Inf. Model.* 50, 68–78
39 Klein, K. *et al.* (2013) Visual analysis of biological activity data with Scaffold Hunter. *Mol. Inform.* 32, 964–995
40 Anon. Scaffold Hunter; http://scaffoldhunter.sourceforge.net/ [Accessed 17 September 2019].
41 Medina-Franco, J.L. *et al.* (2006) Hierarchical strategy for identifying active chemotype classes in compound databases. *Chem. Biol. Drug Des.* 67, 395–408
42 Viira, B. *et al.* (2017) Chemical structure and correlation analysis of HIV-1 NNRT and NRT inhibitors and database-curated, published inhibition constants with chemical structure in diverse datasets. *J. Mol. Graph. Model.* 76, 205–223
43 Eriksson, M. *et al.* (2012) Sarconnect: a tool to interrogate the connectivity between proteins, chemical structures and activity data. *Mol. Inform.* 31, 555–568
44 Clark, A.M. and Labute, P. (2009) Detection and assignment of common scaffolds in project databases of lead molecules. *J. Med. Chem.* 52, 469–483
45 López-López, E. *et al.* (2018) Activity landscape and molecular modeling to explore the SAR of dual epigenetic inhibitors: a focus on g9a and DNMT1. *Molecules* 23
46 Optibrium (2019) *StarDrop*. 2019 [accessed 17 September 2019]www.optibrium.com/
47 Villoutreix, B.O. *et al.* (2013) One hundred thousand mouse clicks down the road: selected online resources supporting drug discovery collected over a decade. *Drug Discov. Today* 18, 1081–1089
48 González-Medina, M. *et al.* (2017) Open chemoinformatic resources to explore the structure, properties and chemical space of molecules. *RSC Adv.* 7, 54153–54163
49 Wild, D.J. and Blankley, C.J. (1999) VisualiSAR: a web-based application for clustering, structure browsing, and structure-activity relationship study. *J. Mol. Graph. Model.* 17, 85–89
50 Fortin, G. (2019) MOEsaic: the application of matched molecular pair analysis to SAR exploration. *Abstr. Pap. Am. Chem. Soc.* 257, 42
51 Pérez-Villanueva, J. *et al.* (2011) Structure–activity relationships of benzimidazole derivatives as antiparasitic agents: dual activity-difference (DAD) maps. *Med. Chem. Commun.* 2, 44–49
52 González-Medina, M. *et al.* (2017) Activity Landscape Plotter: a web-based application for the analysis of structure-activity relationships. *J. Chem. Inf. Model.* 57, 397–402
53 Chávez-Villarreal, K.G. *et al.* (2018) Synthesis, antimycobacterial evaluation, and QSAR analysis of meso-dihydroguaiaretic acid derivatives. *Med. Chem. Res.* 27, 1026–1042
54 de la Vega de León, A. and Bajorath, J. (2012) Design of a three-dimensional multitarget activity landscape. *J. Chem. Inf. Model.* 52, 2876–2883
55 Yoshimori, A. *et al.* (2019) Integrating the structure–activity relationship matrix method with molecular grid maps and activity landscape models for medicinal chemistry applications. *ACS Omega* 4, 7061–7069
56 Peltason, L. and Stoffler, D.S. (2012) SAR Analyser: a tool for interactive SAR data visualization and navigation. *J. Cheminform.* 4 (Suppl 1), P31
57 Khomtchouk, B.B. *et al.* (2017) shinyheatmap: ultra fast low memory heatmap web interface for big data genomics. *PLoS One* 12, e0176334
58 Tableau Software (2018) *Tableau*. 2018 [accessed 17 September 2019]www.tableau.com
59 Schäfer, T. *et al.* (2017) Scaffold Hunter: a comprehensive visual analytics framework for drug discovery. *J. Cheminform.* 9, 28
60 Dotmatics (2005) *Dotmatics Informatics Suite*. 2005 [accessed 17 September 2019] In: https://www.dotmatics.com/
61 Roberts, G. *et al.* (2000) Leadscope: software for exploring large sets of screening data. *J. Chem. Inf. Comput. Sci.* 40, 1302–1314