

EDUCATIONAL ARTICLE

Open Access



# Art driven by visual representations of chemical space

Daniela Gaytán-Hernández<sup>1</sup>, Ana L. Chávez-Hernández<sup>1</sup>, Edgar López-López<sup>1,2</sup>, Jazmín Miranda-Salas<sup>1</sup>, Fernanda I. Saldívar-González<sup>1</sup> and José L. Medina-Franco<sup>1\*</sup>

## Abstract

Science and art have been connected for centuries. With the development of new computational methods, new scientific disciplines have emerged, such as computational chemistry, and related fields, such as cheminformatics. Cheminformatics is grounded on the chemical space concept: a multi-descriptor space in which chemical structures are described. In several practical applications, visual representations of the chemical space of compound datasets are low-dimensional plots helpful in identifying patterns. However, the authors propose that the plots can also be used as artistic expressions. This manuscript introduces an approach to merging art with cheminformatics through visual and artistic representations of chemical space. As case studies, we portray the chemical space of food chemicals and other compounds to generate visually appealing graphs with twofold benefits: sharing chemical knowledge and developing pieces of art driven by cheminformatics. The art driven by chemical space visualization will help increase the application of chemistry and art and contribute to general education and dissemination of cheminformatics and chemistry through artistic expressions. All the code and data sets to reproduce the visual representation of the chemical space presented in the manuscript are freely available at <https://github.com/DIFAC-QUIM/Art-Driven-by-Visual-Representations-of-Chemical-Space>. *Scientific contribution:* Chemical space as a concept to create digital art and as a tool to train and introduce students to cheminformatics.

**Keywords** Artwork, Chemical space, Chemical multiverse, Cheminformatics, Data visualization, Education, Food chemistry, Foodinformatics, Molecular representation, Open science

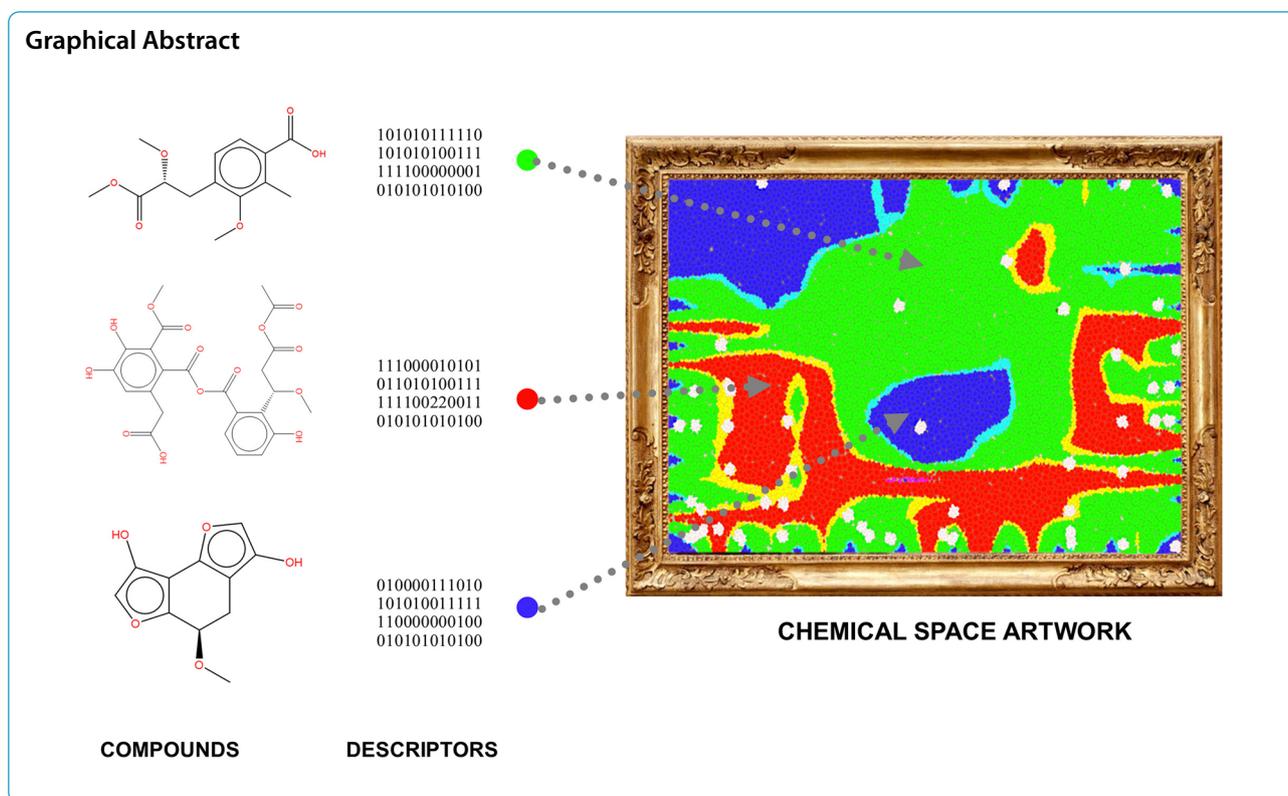
\*Correspondence:

José L. Medina-Franco  
medinajl@unam.mx

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



## Introduction

Art can be considered as the set of activities and products of human beings with aesthetic, ethical, and communication objectives that impact individuals or societies [1]. Its impact may seek to transmit ideas, emotions, needs, concerns, or values [2]. Science can be considered an art tool that makes the materialization of ideas possible and delimits the ideas of artists. What is important about science is not only that it has served to enable the work to be executed. What is fundamental is that it has allowed it to be imagined. Furthermore, scientific knowledge allows for a more profound interpretation of art.

Historically, the relationship between science and art has existed since humans created art. One example is chemistry, a scientific discipline that historically has had a symbiotic relationship with art and has determined its respective evolutions. Among the many interactions of chemistry in art are the development of pigments and spectroscopic techniques, materials for conservation and restoration, to name just a few [3, 4].

The advent of computers gave rise first to computational chemistry and then chemoinformatics. Chemoinformatics, also frequently referred to in the literature as cheminformatics [5] aims to manage and organize information, visualize chemical space, perform

data mining, and establish mathematical relationships between chemical structures and properties. While bioinformatics focuses on biologically relevant macromolecules, chemoinformatics is focused on small compounds [6]. As an independent theoretical discipline, chemoinformatics relies on the chemical space concept [7–10]. Understanding the concept of chemical space within and outside cheminformatics can be complicated. Generally, this concept has been accompanied by various images that seek to represent characteristics that chemists have assigned according to the inherent purposes of their research, leaving aside the aesthetic composition that, in turn, can contribute to deepening and communicating beyond the common sense, which associates thinking to an operation that excludes its connections with the affections, sensitivity, and creation. In Cheminformatics, chemical space has been defined as a chemical descriptor vector space (cf. Fig. 1A) set by the numerical vector  $X$  encoding property or molecular structure aspects as elements of the descriptor vector  $X$  [11]. As such, cheminformatics methods strongly depend on molecular representation and numerical descriptors [12]. There are many descriptors whose selection will depend on the type of molecules studied, for example, organic, inorganic, small molecules, peptides (whose

size can differ significantly), natural products, and food chemicals, to name a few. For small molecules (e.g., molecular weight < 1000 Da), it is common to use as descriptors molecular fingerprints [13, 14], whole molecule properties (e.g., properties of pharmaceutical relevance [15, 16]), and sub-structures such as molecular scaffolds [17]. Figure 1A shows a schematic representation of the concept of chemical space, e.g., a chemical space table as a matrix where compounds are the rows and the numerical descriptors are the columns. Graphical and reduction dimension techniques are used to map the usually large multi-dimensional spaces into two or three dimensions that can be plotted and easily visualized.

Since the chemical space of a set of compounds is not unique and will depend on the set of descriptors chosen to describe it, multiple chemical spaces are theoretically possible for the same data set. Continuing this line of thinking, a chemical multiverse was proposed recently and defined as “the group of numerical vectors that describe differently the same set of molecules.” An alternative definition of the chemical multiverse is a “group of multiple chemical spaces, each defined by a given set of descriptors—a group of “descriptor universes” [7]. The chemical multiverse concept is represented in Fig. 1B.

Chemical spaces and chemical multiverses are, like many other types of analysis, frequently analyzed through data visualization techniques (Fig. 1). Indeed, data visualization is widely used in science and other areas to effectively summarize and communicate data to produce information and, ultimately, knowledge. Extensive reviews have been published concerning the visualization of chemical spaces [9, 10]. As reviewed, there are multiple methods of visualization, such as principal component analysis (PCA) [18], t-distributed stochastic neighbor embedding (t-SNE) [19], Tree MAP (TMAP) [20], self-organizing map (SOM) [21–23], and the generative topographic mapping (GTM) [24]. Each one will have advantages and disadvantages. As emphasized above, the visualization of a given data set will depend on the type of descriptors used.

The visual representation of chemical spaces can lead to visually appealing figures, particularly if appropriate color schemes are used. The visually attractive settings are used to emphasize patterns in the chemistry data to facilitate visual information extraction. For instance, to

highlight grouping or clustering in the chemistry data or to rapidly identify patterns in the structure–property landscapes. At the same time, the visually attractive graphs can be for the chemistry expert and non-expert, a visually appealing graph, or a digital “painting” or work of art. In other words, the graph or digital painting is driven by chemical structures and descriptors. Therefore, the person generating the chemical space representation could be considered a chemical space artist who can communicate not only chemical data and information but even emotions if the chemical structures are associated with a personal, emotional, or another type of feeling the “artist” / author want to communicate through the visualization, e.g., an artistic expression.

In this sense, the concept of chemical space also opens up the possibility of searching for new representations that have to do with the need to configure another image of thought, and think in a novel fashion; it is a creative task and is similar to art.

This manuscript proposes the general notion of generating visual representations of chemical space and chemical multiverses as a means of chemical communication that produces new experiences and, in parallel, artistic expressions. To illustrate the proposal, we generated chemical space visualizations of four flavor categories from an extensive public database of food chemicals, FooDB [25], using different descriptors and molecular fingerprints. We considered four flavor categories, as detailed in the Methods section. The concept would further promote art driven by cheminformatics and can be expanded to other information-related disciplines, such as bioinformatics. Using different descriptors and visualization methods, we show examples of chemical multiverse visualizations of four flavor categories from FooDB and other chemical compounds.

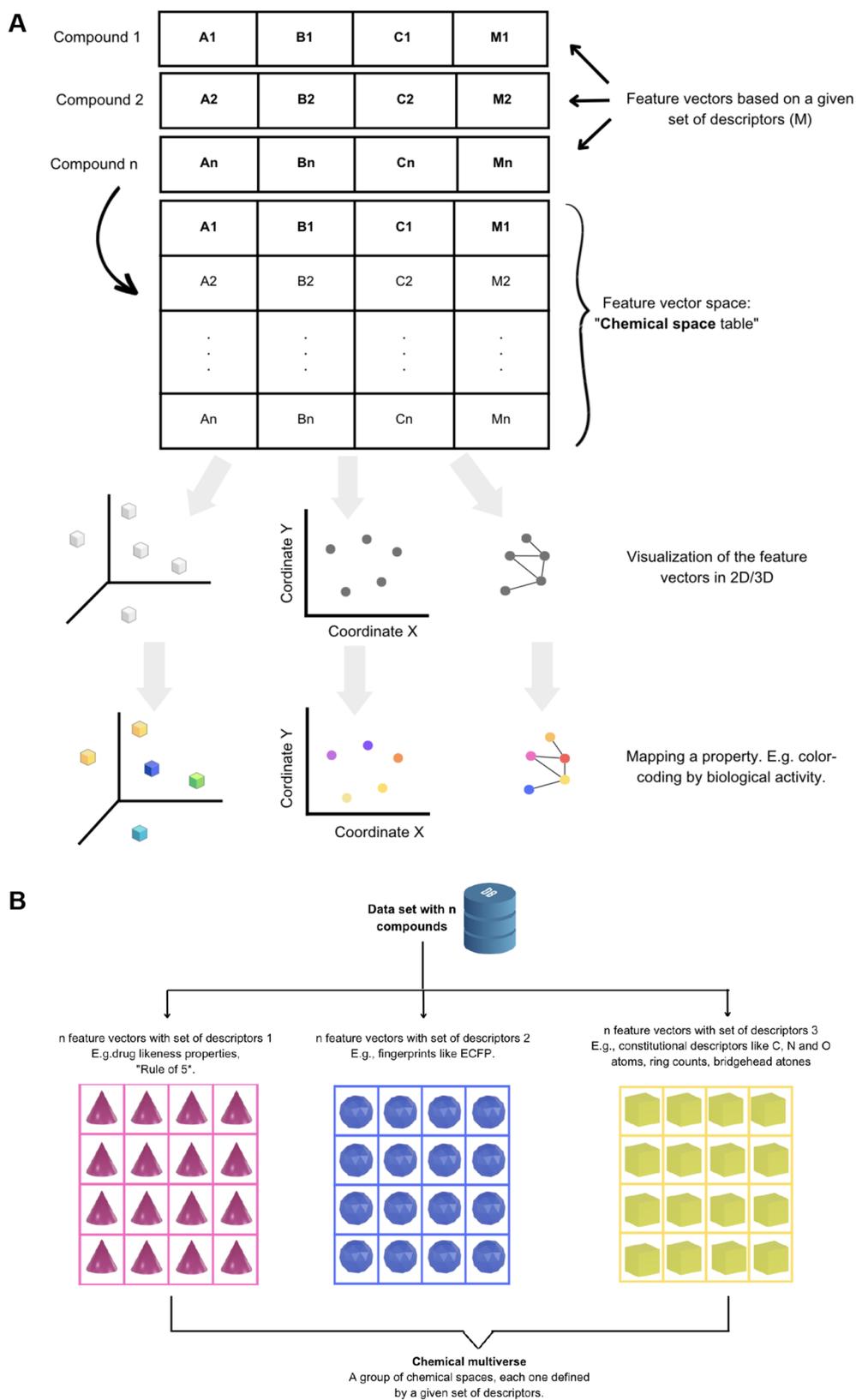
## Methods

### Data sets

Herein, we used food chemicals to generate visual representations of the chemical space as artworks. Food and its flavors, colors, textures, and aromas are generally associated with the great pleasures of life; for this reason, they have been a source of inspiration in art world. However, an approximation at the structural level of the molecules has yet to be addressed. Specifically, we used chemical structures from the public database

(See figure on next page.)

**Fig. 1** Schematic concept of **A** chemical space and its visual representation in low-dimensions. **B** Schematic representation of a chemical multiverse for a hypothetical data set of  $n$  compounds: descriptors of different design (continuous properties, molecular fingerprints, constitutional descriptors, etc.) can lead to alternative chemical spaces for the same data set



**Fig. 1** (See legend on previous page.)

FoodDB [25]. The current version of FoodDB contains 70,477 compounds, and after data set standardization (described in detail in Sect. "Data set standardization") has 52,856 molecules. FoodDB has information about macronutrients, micronutrients, and food chemicals that give food flavor, color, taste, texture, and aroma to foods. Each chemical item in FoodDB contains more than 100 separate data fields providing detailed compositional, biochemical, and physiological information [25]. From FoodDB, 4964 natural flavorings derived from food compounds were identified across twenty flavor categories. Figure 2 summarizes the frequency of the seven most populated categories.

From the twenty-seven flavor categories, we defined four new flavor categories: (1) ground flavors, (2) wine-tasting, (3) contrast between fatty and spicy, and (4) natural remedies. Additional file 1: Table S1 shows the number of compounds in each of the four categories considered in this work. Flavors of the ground/flavor similar to herbaceous are earthy, herbaceous, and green flavors. Wine tasting is composed of fruity and floral flavors. The contrast between fatty and spicy is composed of fatty and spicy flavors. Medicinal comprises balsamic, chemical, and medicinal, which are characteristic flavors found in ointments, alcohol, and syrups. Additional file 1: Fig. S1 shows the overlapping compounds between the selected flavor categories.

#### Data set standardization

Compounds in FoodDB, encoded as SMILES strings [12], were standardized using the open-source cheminformatics toolkit RDKit [26] and Standardizer, LargestFragmentChoser, Uncharger, Reionizer y TautomerCanonicalizer functions implemented in MolVS [27]. Compounds with valence errors or any chemical element other than H, B, C, N, O, F, Si, P, S, Cl,

Se, Br, and I were removed. Stereochemistry information, when available, was retained. Compounds with multiple components were split, and the largest component was retained. The remaining compounds were neutralized and reionized to generate the corresponding canonical tautomer.

#### Molecular descriptors

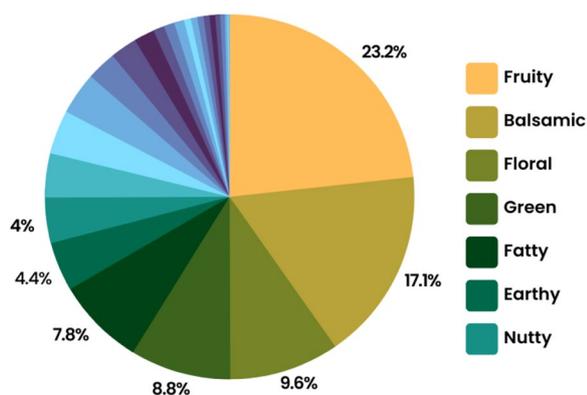
For each molecule, physicochemical properties and molecular fingerprints were calculated as descriptors using Python language and RDKit. The whole molecule descriptors computed were hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), topological polar surface area (TPSA), number of rotatable bonds (RB), molecular weight (MW), and partition coefficient octanol/water (LogP). Molecular fingerprints computed were Molecular Access System (MACCS) Keys (166-bits) [13], extended connectivity fingerprint (ECFP) [14] of 1024-bits with diameter 4 (ECFP4). Of note, virtually any other descriptors can be used, as further commented in the Sect. "Discussion".

#### Visualization methods

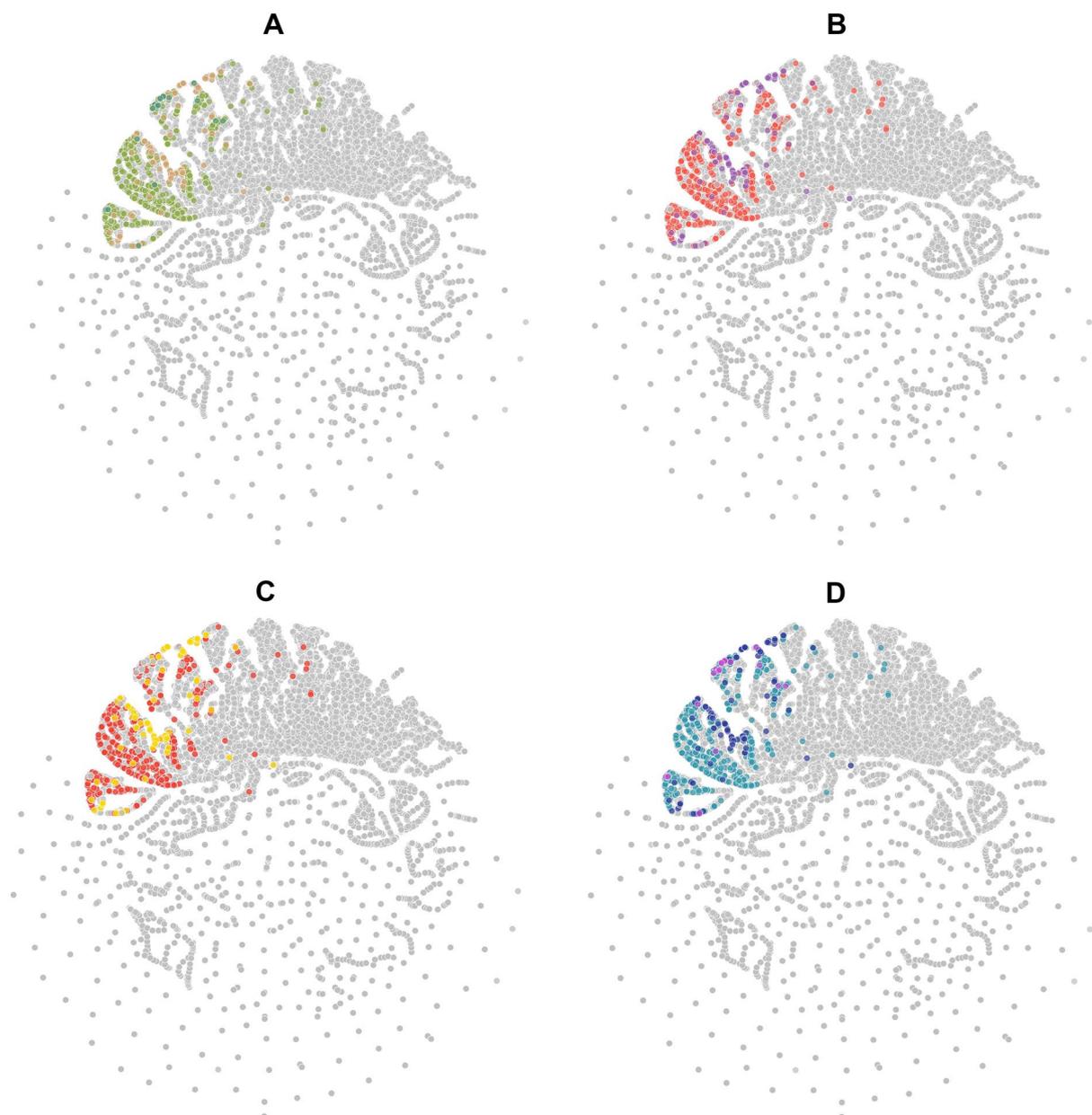
In this study, we used three well-known dimensionality reduction methods: t-SNE, PCA, and TMAPs, although additional visualization methods can be used. Briefly, t-SNE generates plots that organize compounds. Similar compounds form clusters and dissimilar compounds are distant from each other. PCA is a linear dimensionality reduction technique that transforms data with many dimensions (i.e., descriptors) into a lower dimensional space and keeps the different relationships between the data points as much as possible [18]. PCA was generated from six whole molecule descriptors (MW, HB, HBA, SlogP, TPSA, and RB). TMAPs allow visualization of many chemical compounds through the distance between clusters and the detailed structure of these through branches and sub-branches. Local sensitive hashing allows each compound to be grouped hierarchically according to common substructures using molecular fingerprints. In this work, we use MACCS keys (166-bits) [13] fingerprints. Then, each chemical compound was encoded using the MinHash algorithm. The number of nearest neighbors,  $k=50$ , and the factor used by the augmented query algorithm,  $kc=10$ , were used to generate the TMAPs [20].

#### Results

Figures 3, 4, 5, 6 show examples of so-called "Art Galleries" composed by visualization of the chemical space of different food chemical categories. The visual representations of chemical space were generated with t-SNE (Figs. 3 and 4), PCA (Fig. 5), and TMAPs



**Fig. 2** The seven most frequent flavor categories identified in FoodDB

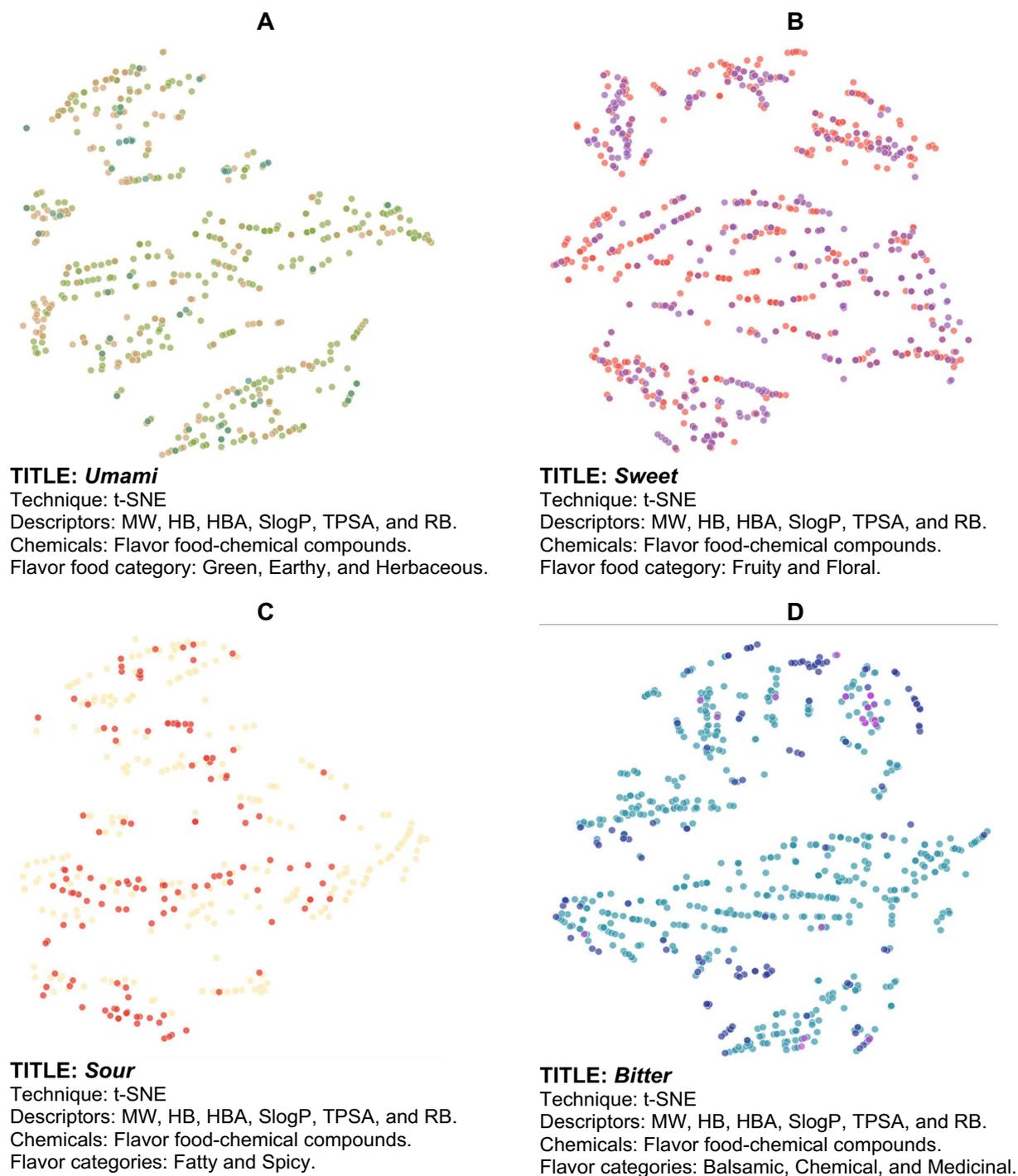


**Fig. 3** Four flavor categories and full FooDB. The flavor categories are **A** Ground flavors (655 compounds), **B** Wine-tasting (1024 compounds), **C** Contrast between fatty and spicy (430 compounds), and **D** Natural remedies (762 compounds)

(Fig. 6). Below each image (i.e., “digital paintings”) is presented basic information of the “technique” (visualization method, allusive to the techniques used in paintings), descriptors, and chemicals (that would be meaningful information for a chemistry-oriented person to understand the data presented). Each visual representation of the chemical space or Artwork includes a “Title” that is reminiscent of the name of the piece of art or digital painting.

## Discussion

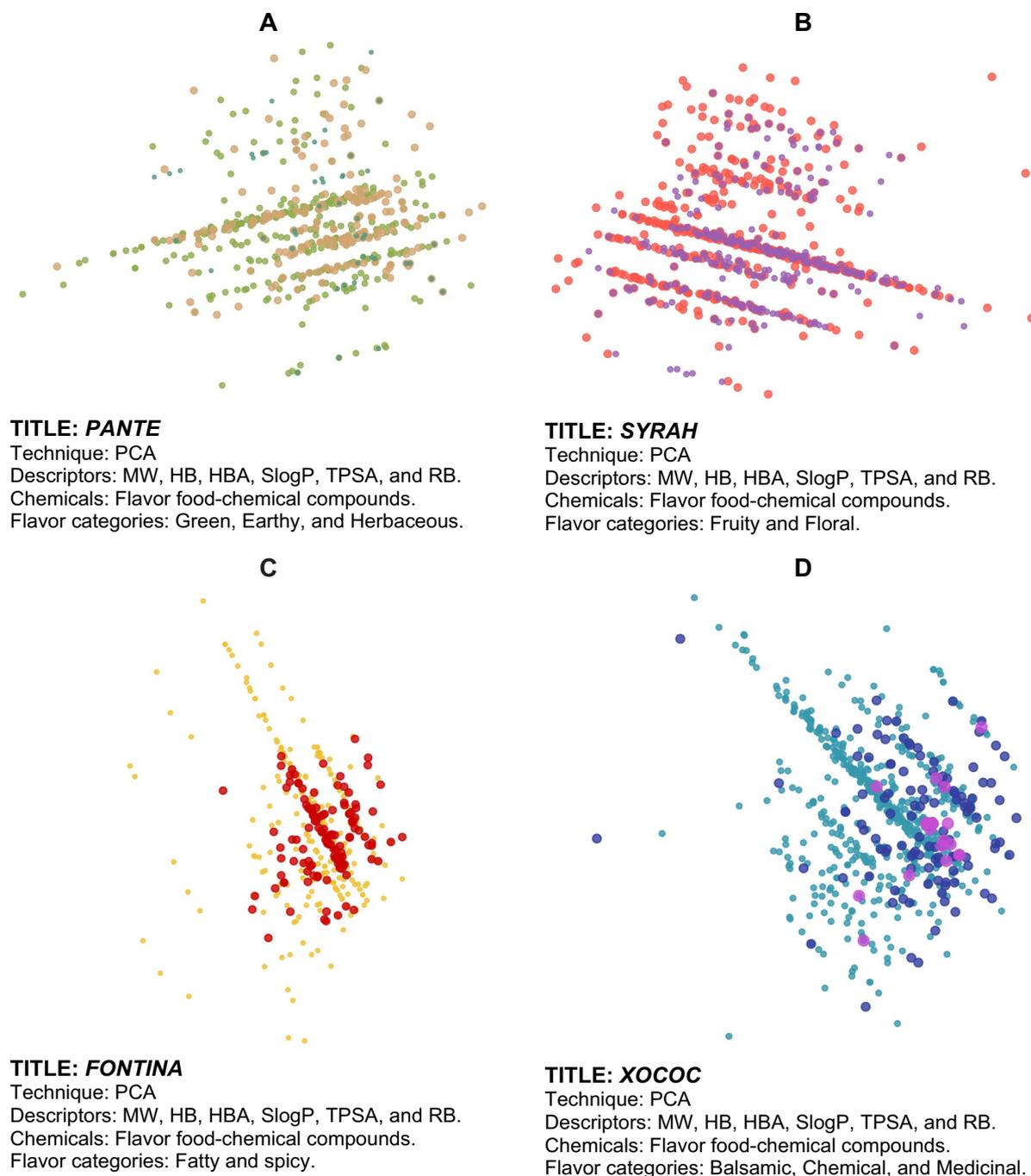
Cheminformatics has been broadly used in drug discovery. Still, it has many more applications in chemistry, with increasing applications in food chemistry, as evidenced by the emergence of the research areas of food chemical informatics or food informatics [28, 29]. There are others, such as natural products [30, 31], polymers, and materials, to name a few [6]. Herein, we propose expanding the realm of cheminformatics’ applications through the visual



**Fig. 4** Four flavor categories: **A** Ground flavors (655 compounds), **B** Wine-tasting (1024 compounds), **C** Contrast between fatty and spicy (430 compounds), and **D** Natural remedies (762 compounds)

representation of the chemical space of compound data sets—herein illustrated with food chemicals—to yield exemplary “art pieces.” The connection or synergy between cheminformatics and art has a strong

potential to bring together at least two sectors of the population that might be otherwise disconnected. From an educational point of view, which is a central need in

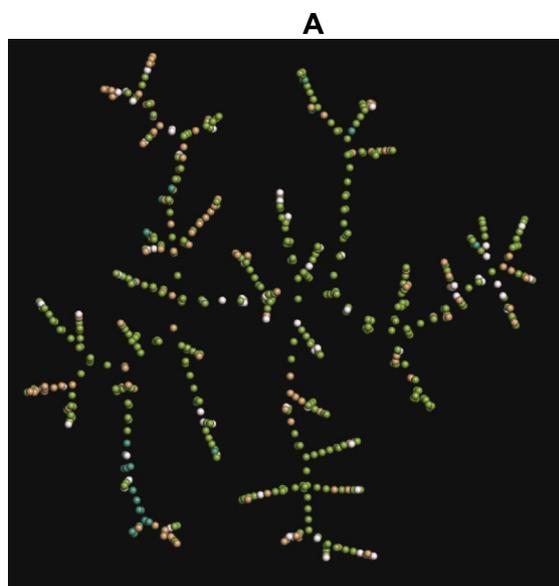


**Fig. 5** Four flavor categories: **A** Ground flavors (655 compounds), **B** Wine-tasting (1024 compounds), **C** Contrast between fatty and spicy (430 compounds), and **D** Natural remedies (762 compounds)

chemoinformatics—the synergy might attract young students and kids to chemistry through art.

The subdiscipline of food informatics was proposed in 2014 as a specific application of chemoinformatics to food chemistry [28]. Since then, numerous applications of chemoinformatics to different aspects of food

chemistry have been published, including analysis of the chemical space of food chemicals to characterize the structural diversity [32]. In Sect. "Results" we showed examples of visual representations of the chemical space of food chemicals as an artistic expression and scientific dissemination through art. There are many



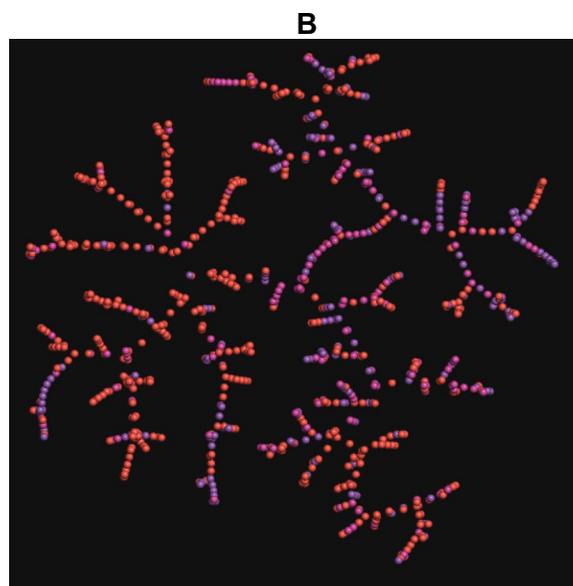
**TITLE: *Ébano (Ebony)***

Technique: TMAP

Descriptors: MACCS keys fingerprints

Chemicals: Flavor food-chemical compounds.

Flavor categories: Green, Earthy, and Herbaceous.



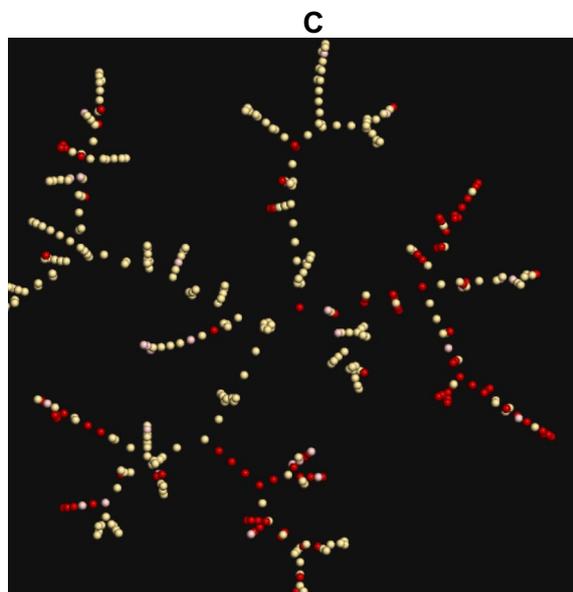
**TITLE: *Flor de corazón (Heart flower)***

Technique: TMAP

Descriptors: MACCS keys fingerprints

Chemicals: Flavor food-chemical compounds.

Flavor categories: Fruity and Floral



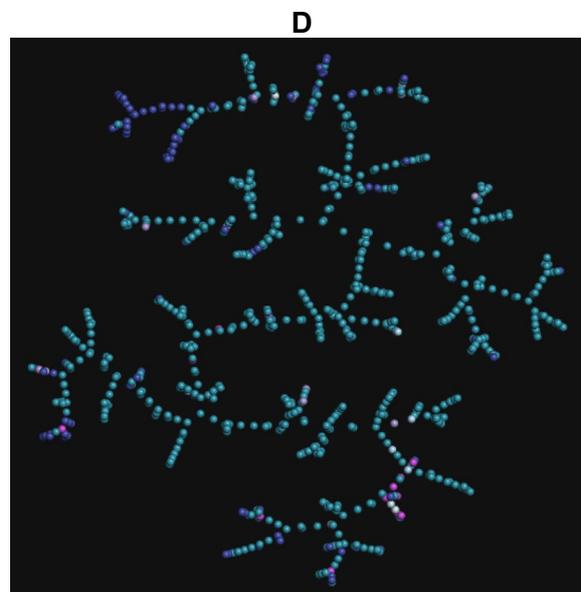
**TITLE: *Amaranto (Amaranth)***

Technique: TMAP

Descriptors: MACCS keys fingerprints

Chemicals: Flavor food-chemical compounds.

Flavor categories: Fatty and Spicy



**TITLE: *Huele a miel (Honey's smell)***

Technique: TMAP

Descriptors: MACCS keys fingerprint

Flavor food-chemical compounds.

Flavor categories: Balsamic, Chemical, and Medicinal

**Fig. 6** Four flavor categories: **A** Ground flavors (655 compounds), **B** wine-tasting (1024 compounds), **C** Contrast between fatty and spicy (430 compounds), and **D** Natural remedies (762 compounds)

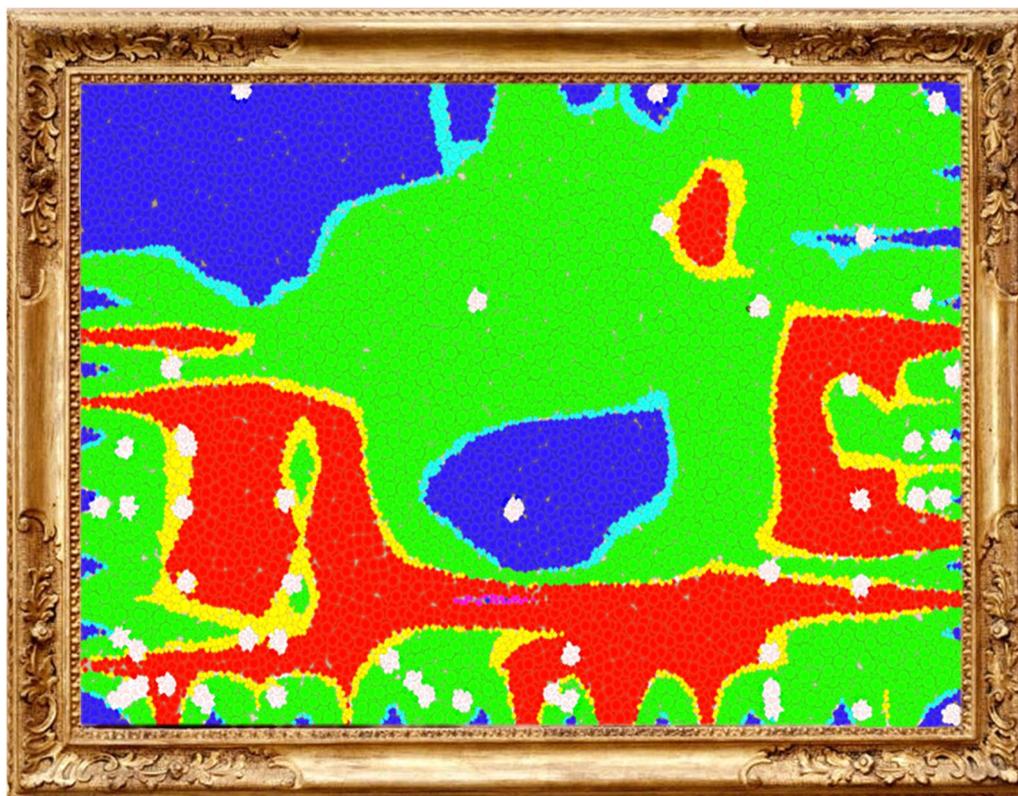
**Table 1** Exemplary potential paintings based on the visualization of the chemical space of compound data sets

Data set	Artistic meaning	Artwork name
Random compounds	Aleatory molecules represent the vastness of our universe and daily life. We are in contact with many chemicals every time, but we don't look at their complexity and intrinsic disorder in our universe and daily life	"Chaos"
Diverse data set	The diversity offers many colors, flavors, tastes, and experiences. In nature, diversity (in all senses) is a constant feature	"Diversity"
Marine natural products	We don't understand the sea; It has life, death, color, and darkness. It's constantly changing	"The Ocean" "Immensity"
Drugs approved for the treatment of HIV	Everything happens in a positive HIV human; Fear, memories, happiness, and normality. The drugs help... but are not a complete answer	"Living with AIDS"
Hormones—neurotransmitters	Love = hormones + neurotransmitters + special persons	"The chemistry of love"
Chemicals associated with depression	Depression = hormones + neurotransmitters—purpose	"Darkness"
Food chemicals	The great pleasures of life are often accompanied by flavors, colors, textures, and aromas	"Bellyful" "Flavor trip"
ZINC database vs. drug-like compounds	We know a lot about our nature and composition, but we don't know much more. Our knowledge is a mere stain on an entire canvas that we do not yet understand	"Our knowledge"

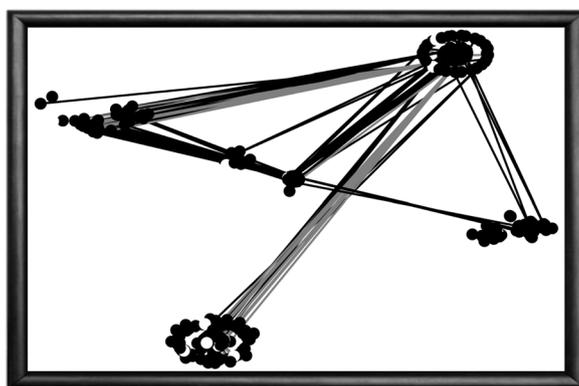
possibilities to expand the genesis of the proposed "art-cheminformatics," as further elaborated in Sect. "Conclusions and outlook".

#### Exemplary art-related chemical spaces and multiverses

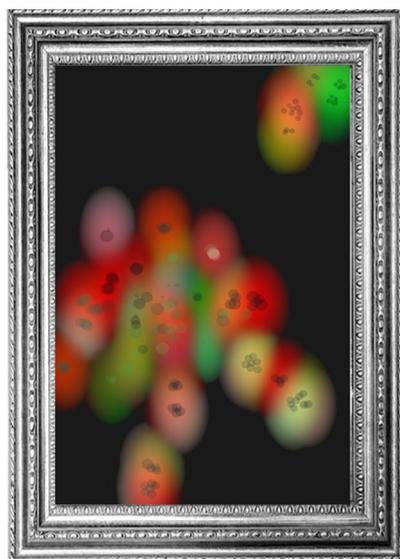
The examples of visual representation of chemical space as artistic representations presented in Sect. "Results"



**Fig. 7** Chemical space art example. Title: "Wise nature"; Autor: Edgar López-López; Technique: SOM—using DataWarrior software [33, 34]; Dataset: Random natural products (1000 compounds); Descriptors: predicted mutagenic, tumorigenic, Reproductive effective, and Irritant; Technical description: Each white point is a natural product, the regions colored in red represent the chemical space with a high predicted probability of containing compounds with side effects, the opposite for the blue color; Artistic interpretation: The "nature" is not always healthy, in nature, there has always been a duality between what fills us with life and what takes it away

**TITLE: Chemical umbrella**

Technique: PCA + Data fusion (chemical multiverse approach).  
Descriptors: Cell-based and enzymatic inhibition data. Dots are connected based on their inhibitory activity against different types of cytochromes (proteins related to hepatic protection).  
Hepatotoxic compounds.

**TITLE: Broked cancer**

Technique: Constellation plots.  
Descriptors: Anticancer cell inhibition data.  
Anticancer drugs.

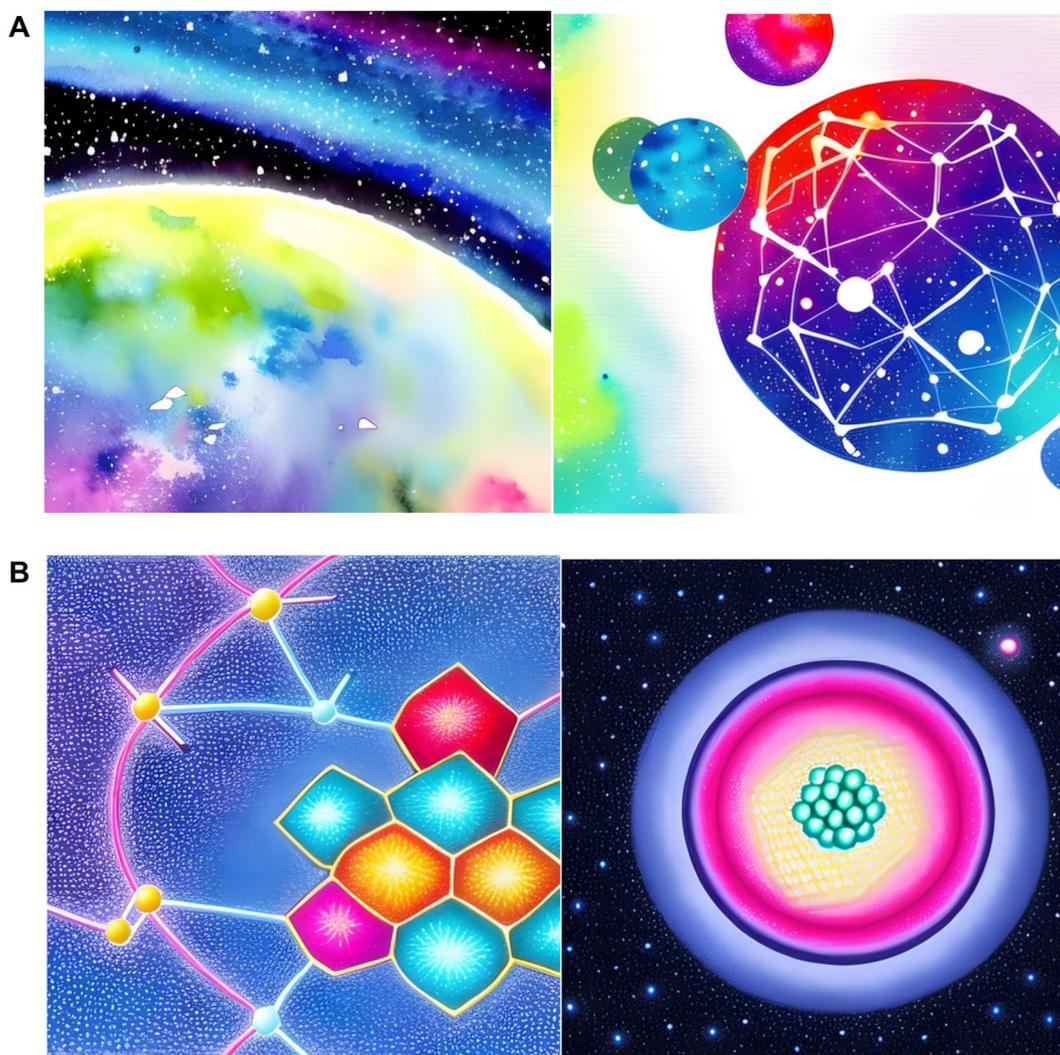
**Fig. 8** Chemical space art examples. Chemical artworks were generated with public data [35–37]

are focused on food chemicals and molecular descriptors suitable to represent such chemical compounds. Also, examples of visualization methods used in the previous section are t-SNE, PCA, and TMAPs. However, as commented in the Introduction, the number of established visualization techniques, molecular descriptors, and, perhaps most importantly, the number of chemical structures are immense. Therefore, there are thousands or millions of ways to generate chemical space-driven works of art. To glimpse the artistic possibilities, Table 1 summarizes examples of the cheminformatics-driven visualization of chemical space and multiverses. The table summarizes examples of compound data sets with chemicals of different types that could be used to represent their vastness, complexity, diversity, and chaotic intrinsic features from an artistic perspective. Many more compound data sets and multiple combinations of descriptors and visualization techniques

could be used. However, as with any other artistic vehicle, the real importance of any type of art is its capacity to tell histories or convey a message that sometimes is hidden.

To illustrate further the potential of generating artistic representations through visualization of chemical space, Fig. 7 shows an example of chemical space artwork from a random natural products dataset, decoding by their side effects descriptors (e.g., mutagenesis, tumorigenesis, and negative reproductive effects, etc.). Their color palette, from red to blue, represents the probability of each natural product generating side effects. The “canvas” was “painted” with a dotted technique, reflecting another possible set of textures that can be developed with this technique. Like in Fig. 7, we intrinsically know that “nature” is not always healthy and that within us, there is a delicate balance that is very easy to break.

Figure 8 shows additional examples of chemical space artwork that combine different reduction data



**Fig. 9** Example(s) of artificial intelligence-driven art with the free application Canva (<https://www.canva.com/>) using the keyword chemical space and **A** Watercolor and **B** color pencil

**Table 2** Representative developments of combining art with chemoinformatics through artistic visualizations of chemical space

Development	Putative outcome or application
Continue developing a digital collection focused on the artistic representation of the chemical space	<i>The Chemical Space Art Museum</i>
Generate automated workflows using open software or informatic tools to improve the accessibility of this kind of art to people with different academic/artistic backgrounds	<i>ChemArt Generators</i>
Establish a free, open-access, and permanent repository of art pieces. This encourages open science and open art. The scientific and artistic community could support the repository	<i>ChemART Gallery</i> . An example is at <a href="https://www.difacquim.com/chemical-art-gallery/">https://www.difacquim.com/chemical-art-gallery/</a>
Set up a sustained educational or cultural program as a continued open and permanent exposition	<i>Art Driven by Chemical Space Visualization</i> program

methods and descriptors to generate an artistic visual representation of the chemical data. We encourage the readers to reflect and find other artistic interpretations that these figures could have. The examples of chemical

space visualization as work art have been included in a Chemical Space Art Gallery freely available at <https://www.difacquim.com/chemical-art-gallery/>

### Artificial intelligence and digital art

Artificial intelligence (AI) is used to generate artistic representations [38, 39]. Although it is not the central point of this manuscript, Fig. 9 illustrates images generated with free resources using keywords associated with “chemical space.” Specifically, the figure shows an example of a chemical multiverse/chemical space driven by an AI-web server training on words. Although the images are attractive, a striking difference with the chemical space artworks presented in previous sections (Figs. 3, 4, 5, 6, 7, 8) is that the images in Fig. 9 are based on keyword training. The former are derived directly from chemical structures encoded with molecular descriptors. Another important aspect is a greater understanding and human intervention in the former representations, something questionable in AI-guided pictures.

### Conclusions and outlook

Science and art have long been intimately related. A typical example is summarized by the phrase, “Drug discovery is as much an art as it is a science.” Certainly, chemistry is substantially used in art, such as in art restoration and preservation. However, an emerging trend exists to apply chemistry and its concepts to generate artwork. Herein, we discuss an approach to combining art with chemoinformatics through the visual representations of chemical space. We presented a few examples of chemical space artworks that can be “digital paintings.” The author of the low-dimensional graphs can use the plots with dual general purposes: communicate data and generate chemical information (as generally done with the visualizations of chemical space) and convey an emotional or personal meaning to the graph (driven by chemistry and informatics principles).

We also conclude that chemical space-driven works of art can be tools to promote science in general and chemistry in particular for the broad audience. Thus, chemistry informatic-driven artistic expressions can be an approach to disseminating science. Such an approach aligns with the graphical abstracts frequently used in peer-reviewed journals. The “chemical art” could be useful to represent complex data by using an artistic and attractive perspective. The person generating the chemical space representation could be considered a “chemical space artist.”

We envision several further developments and areas of opportunity for art driven by visual representations of chemical space. Table 2 summarizes ongoing chemical arts projects, from the generation of “easy to use” tools, the first chemical art gallery, and the implementation of this artistic mode to introduce the new generation of chemoinformaticians to the chemical

space concept. In parallel, AI methods will continue expanding and exploring the chemical space, offering new types of molecules and descriptors that could be used to increase the possibilities of representing chemical space from an artist’s perspective.

### Abbreviations

AI	Artificial intelligence
ECPF	Extended connectivity fingerprint
HBD	Hydrogen bond donors
HBA	Hydrogen bond acceptors
GTM	Generative topographic mapping
LogP	Partition coefficient octanol/water
MACCS	Molecular ACCes System
MW	Molecular weight
PCA	Principal component analysis
TMAP	TreeMap
t-SNE	T-Distributed stochastic neighbor embedding
TPSA	Topological polar surface area
SOM	Self-organizing map

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13321-023-00770-4>.

**Additional file 1: Table S1.** The number of flavor compounds, flavor notes, and flavor categories. **Figure S1.** Unique and overlapping structures of four flavor categories from FooDB. All the code and data sets to reproduce the visual representation of the chemical space presented in the manuscript are freely available at <https://github.com/DIFACQUIM/Art-Driven-by-Visual-Representations-of-Chemical-Space>.

### Acknowledgements

We are grateful for the rich and useful discussions with Dra. Karina Martínez-Mayorga. D.G.-H. thanks the support from CONAHCYT. A.L.-C.-H., E.L.-L., and F.I.S.-G thank CONAHCYT, Mexico, for the Ph.D. scholarships number 847870, 894234, and 848061, respectively.

### Author contributions

D. G-H and E. L-L. generated visual representations of the chemical space and interpreted the data. A.L. C-H supervised the work of D. G-H, participated in the conceptualization of the study, and helped to draft the first version of the draft. F.I. S-G and J. M-S contributed to the discussion of the concepts. J.L. M-F conceptualized the project, supervised the project and drafted the first version of the manuscript. All authors contributed to edit and prepare the final version of the manuscript.

### Funding

No funding was received to perform this research.

### Availability of data and materials

All data related to this manuscript can be accessed in the Supplementary material.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Competing interests

Authors declare that have no competing interests.

#### Author details

<sup>1</sup>DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Avenida Universidad 3000,

04510 Mexico City, Mexico. <sup>2</sup>Department of Chemistry and Graduate Program in Pharmacology, Center for Research and Advanced Studies of the National Polytechnic Institute, 07000 Mexico City, Mexico.

Received: 13 July 2023 Accepted: 13 October 2023

Published online: 21 October 2023

## References

1. La Galván-Madrid JL (2011) Química y el Arte: ¿Cómo mantener el vínculo? *Educ Quim* 22(3):207–211. [https://doi.org/10.1016/S0187-893X\(18\)30136-8](https://doi.org/10.1016/S0187-893X(18)30136-8)
2. Bello DG (2023) *La química de lo bello*, 2nd edn. Ediciones Paidós, Barcelona
3. Orna MV (2001) Chemistry, color, and art. *J Chem Educ* 78(10):1305. <https://doi.org/10.1021/ed078p1305>
4. Kafetzopoulos C, Spyrellis N, Lymperopoulou-Karaliota A (2006) The chemistry of art and the art of chemistry. *J Chem Educ* 83(10):1484. <https://doi.org/10.1021/ed083p1484>
5. Miranda-Salas J, Peña-Varas C, Valenzuela Martínez I, Olmedo DA, Zamora WJ, Chávez-Fumagalli MA, Azevedo DQ, Castilho RO, Maltarollo VG, Ramírez D, Medina-Franco JL (2023) Trends and challenges in chemoinformatics research in Latin America. *Artif Intell Life Sci* 3(1):100077. <https://doi.org/10.1016/j.jalisci.2023.100077>
6. López-López E, Bajorath J, Medina-Franco JL (2020) Informatics for chemistry, biology, and biomedical sciences. *J Chem Inf Model* 61(1):26–35. <https://doi.org/10.1021/acs.jcim.0c01301>
7. Medina-Franco JL, Chávez-Hernández AL, López-López E, Saldívar-González FI (2022) Chemical multiverse: an expanded view of chemical space. *Mol Inf* 41(11):2200116. <https://doi.org/10.1002/minf.202200116>
8. Medina-Franco JL, Sánchez-Cruz N, López-López E, Díaz-Eufracio BI (2022) Progress on open chemoinformatic tools for expanding and exploring the chemical space. *J Comput Aided Mol Des* 36(5):341–354. <https://doi.org/10.1007/s10822-021-00399-1>
9. Osolodkin DI, Radchenko EV, Orlov AA, Voronkov AE, Palyulin VA, Zefirov NS (2015) Progress in visual representations of chemical space. *Expert Opin Drug Discov* 10(9):959–973. <https://doi.org/10.1517/17460441.2015.1060216>
10. Medina-Franco J, Martínez-Mayorga K, Giulianotti M, Houghten R, Pinilla C (2008) Visualization of the chemical space in drug discovery. *Curr Comput Aided Drug Des* 4(4):322–333. <https://doi.org/10.2174/157340908786786010>
11. Saldívar-González FI, Medina-Franco JL (2022) Approaches for enhancing the analysis of chemical space for drug discovery. *Expert Opin Drug Discov* 17(7):789–798. <https://doi.org/10.1080/17460441.2022.2084608>
12. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 28(1):31–36. <https://doi.org/10.1021/ci00057a005>
13. Durant JL, Leland BA, Henry DR, Nourse JG (2002) Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* 42(6):1273–1280. <https://doi.org/10.1021/ci010132r>
14. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. *J Chem Inf Model* 50(5):742–754. <https://doi.org/10.1021/ci100050t>
15. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46(1–3):3–26. [https://doi.org/10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0)
16. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 45(12):2615–2623. <https://doi.org/10.1021/jm020017n>
17. Singh N, Guha R, Giulianotti MA, Pinilla C, Houghten RA, Medina-Franco JL (2009) Chemoinformatic analysis of combinatorial libraries, drugs, natural products, and molecular libraries small molecule repository. *J Chem Inf Model* 49(4):1010–1024. <https://doi.org/10.1021/ci800426u>
18. Greener JG, Kandathil SM, Moffat L, Jones DT (2021) A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* 23(1):40–55. <https://doi.org/10.1038/s41580-021-00407-0>
19. van der Maaten L, Hinton G (2023) Visualizing data using t-SNE. <https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbclid=IwAR1186/s13321-020-0416-x>. Accessed 1 Jun 2023
20. Probst D, Reymond J-L (2020) Visualization of very large high-dimensional data sets as minimum spanning trees. *J Cheminf* 12(1):1–13. <https://doi.org/10.1186/s13321-020-0416-x>
21. Kohonen T (2001) *Self-organizing maps*. Springer, Berlin Heidelberg, pp 105–176
22. Schneider P, Tanrikulu Y, Schneider G (2009) Self-organizing maps in drug discovery: compound library design, Scaffold-Hopping. *Repurpos Curr Med Chem* 16(3):258–266. <https://doi.org/10.2174/092986709787002655>
23. Virshup AM, Contreras-García J, Wipf P, Yang W, Beratan DN (2013) Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds. *J Am Chem Soc* 135(19):7296–7303. <https://doi.org/10.1021/ja401184g>
24. Bishop CM, Svensén M, Williams CKI (1998) Developments of the generative topographic mapping. *Neurocomputing* 21(1):203–224. [https://doi.org/10.1016/S0925-2312\(98\)00043-5](https://doi.org/10.1016/S0925-2312(98)00043-5)
25. FooDB <https://foodb.ca/>. Accessed 20 Apr 2023
26. RDKit <https://www.rdkit.org>. Accessed 8 Jan 2022
27. MolVS <https://molvs.readthedocs.io/en/latest/>. Accessed 8 Jan 2022
28. Martínez-Mayorga K, Medina-Franco JL, Eds (2014) *Foodinformatics: applications of chemical information to food chemistry*. Springer International Publishing: Cham
29. Peña-Castillo A, Méndez-Lucio O, Owen JR, Martínez-Mayorga K, Medina-Franco JL (2018) Chemoinformatics in food science. In *Applied chemoinformatics*, Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, pp 501–525
30. Kirchmair J (2020) Molecular informatics in natural products research. *Mol Inf* 39(11):2000206. <https://doi.org/10.1002/minf.202000206>
31. Medina-Franco JL, Saldívar-González FI (2020) Cheminformatics to characterize pharmacologically active natural products. *Biomolecules* 10(11):1566. <https://doi.org/10.3390/biom10111566>
32. Naveja JJ, Rico-Hidalgo MP, Medina-Franco JL (2018) Analysis of a large food chemical database: chemical space, diversity, and complexity. *F1000Res*. <https://doi.org/10.12688/f1000research.15440.2>
33. Sander T, Freyss J, von Korff M, Rufener C (2015) DataWarrior: an open-source program for chemistry aware data visualization and analysis. *J Chem Inf Model* 55(2):460–473. <https://doi.org/10.1021/ci500588j>
34. López-López E, Naveja JJ, Medina-Franco JL (2019) DataWarrior: an evaluation of the open-source drug discovery tool. *Expert Opin Drug Discov* 14(4):335–341. <https://doi.org/10.1080/17460441.2019.1581170>
35. López-López E, Medina-Franco JL (2023) Towards decoding hepatotoxicity of approved drugs through navigation of multiverse and consensus chemical spaces. *Biomolecules* 13(1):176. <https://doi.org/10.3390/biom13010176>
36. Medina-Franco JL, Naveja JJ, López-López E (2019) Reaching for the bright StARs in chemical space. *Drug Discov Today* 24(11):2162–2169. <https://doi.org/10.1016/j.drudis.2019.09.013>
37. López-López E, Cerda-García-Rojas CM, Medina-Franco JL (2021) Tubulin inhibitors: a chemoinformatic analysis using cell-based data. *Molecules* 26(9):2483. <https://doi.org/10.3390/molecules26092483>
38. DALL-E 2 <https://openai.com/dall-e-2/>. Accessed 20 Jun 2023
39. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M (2022) Hierarchical text-conditional image generation with CLIP latents. [arXiv:2204.06125](https://arxiv.org/abs/2204.06125)

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.