

Cheminformatic Resources for Organometallic Drug Discovery

José L. Medina-Franco* , Yesenia Cruz-Lemus, Yazmin Percastre-Cruz

Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Mexico City, Mexico

Email: *jose.medina.franco@gmail.com, medinajl@unam.mx

How to cite this paper: Medina-Franco, J.L., Cruz-Lemus, Y. and Percastre-Cruz, Y. (2020) Cheminformatic Resources for Organometallic Drug Discovery. *Computational Molecular Bioscience*, 10, 1-11.
<https://doi.org/10.4236/cmb.2020.101001>

Received: January 13, 2020

Accepted: February 10, 2020

Published: February 13, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Medicinal Organometallic Chemistry keeps contributing to drug discovery efforts including the development of diagnostic compounds. Despite the limiting issues of metal-based molecules, e.g., such as toxicity, there are drugs approved for clinical use and several others are under clinical and pre-clinical development. Indeed, several research groups continue working on organometallic compounds with potential therapeutic applications. For arguably historical reasons, cheminformatic methods in drug discovery have been applied thus far mostly to organic compounds. Typically, metal-based molecules are excluded from compound data sets for analysis. Indeed, most software and algorithms for drug discovery applications are focused and parametrized for organic molecules. However, considering the emerging field of material informatics, the objective of this Commentary we emphasize the need to develop cheminformatic applications to further develop metallodrugs. For instance, one of the starting points would be developing a compound database of organometallic molecules annotated with biological activity. It is concluded that cheminformatic methods can boost the research area of Medicinal Organometallic Chemistry.

Keywords

Cheminformatics, Chemical Space, D-InnoDB, Drug Discovery, Medicinal Organometallic Chemistry, Medicinal Inorganic Chemistry, Metal-Based Compounds, Structure-Activity Relationships

1. Introduction

Organometallic and inorganic compounds attract a large interest because of their broad range of biological activities [1]. Metal-containing compounds in clinical use and under clinical development as well as relevant in diagnostic applica-

tions have been discussed extensively in the literature. Indeed, metallodrugs and some inorganic molecules offer large benefits as diagnostic tools [2] and are associated with the field of metalloimaging. Metallodrugs have diverse mechanisms of action [3] and therapeutic applications of which one of the more explored are anti-cancer followed by antibacterial activity. Another important point is the therapeutic application that covers not only human but also veterinary consumption which expands the field of application and the opportunities for success.

Organometallic compounds also are attractive because they can explore novel molecular targets not addressed by the currently available chemical space (defined mostly by organic compounds). Similarly, emerging molecular targets such as epigenetic could be conveniently addressed by novel compounds located outside the traditional drug-like space. In addition, metallodrugs offer distinct features that could be useful for complex diseases best addressed by multi-target approaches [4].

Despite the fact organometallic compounds have general concerns such as toxicity and cost (particularly considering a “large” production), organometallic drugs have attractive and distinct structural features with the ability to augment the relevant medicinal chemical space, ideally balancing novelty with relevance [5]. One of the distinct structural features of organometallic drugs is molecular complexity that is well-known to have a major impact on drug discovery [6].

While molecular docking of metal-complexes and other modeling approaches are commonly conducted to explore compound-target interactions [7] [8], chemoinformatic studies addressing aspects such as chemical diversity, visual representation of the chemical space, and similarity searching, to name a few, have been done on a more limited basis. However, they represent major areas of opportunity. Chemoinformatics arises from the combination of scientific and technological tools with the 3D understanding and manipulation of the chemistry applied to therapeutic research. Although it has been applied to a larger extent to organic compounds, it is proposed that using these tools in organometallic chemistry can provide good outcomes and generate significant advances in therapeutics.

The main goal of this Commentary is to highlight major areas where, in the author’s opinion, chemoinformatic methods used to explore organic-based molecules can be extended to address the needs of the organometallic-based drug discovery. After this short Introduction, several areas of opportunity or application are discussed. They are not arranged in strict order of priority.

2. Areas of Application

Chemoinformatic methods are broadly used across several stages of the drug discovery process [9]. For historical reasons these approaches, boosted by the needs of pharmaceutical companies, have been developed and applied to organic compounds. Unlike organic compounds, organometallic compounds have had

much less therapeutic applications, which is even more noticeable when it comes to the application of chemoinformatics for its development. In fact, in several methods and chemoinformatic protocols metal-based compounds are excluded from the analysis. The practice of filtering out compounds is typically done when working with medium-to-large chemical databases for analysis of chemical diversity or to develop predictive models, to name a few examples. However, as mentioned earlier, these types of practices have had little or no application in organometallic molecules. One reason to remove metal-based compounds is its overall low frequency in most major chemical databases used in drug discovery currently available. A second and perhaps more strong reason to exclude metal-containing molecules is the lack of appropriate parameters to address the presence of metal atoms in the chemical structures.

In this section of the Commentary we outline several areas where chemoinformatic methods commonly used in current drug discovery can be applied to study organometallic compounds.

2.1. Database of Organometallic Compounds for Drug Discovery

Perhaps one of the most relevant and straightforward applications of cheminformatics to study organometallic compounds is related to compound databases. Indeed, compound databases play a significant role in drug discovery [10]. Either public, in-house (mostly private), virtual, and on-demand [11], are key repositories to store, organize, and mine chemical and biological information. Major compound databases used in drug discovery have been reviewed extensively elsewhere [12]. Despite the fact these large compound databases include some organometallic molecules, the vast majority are small organic molecules and the few organometallic molecules that are included do not have information that is useful for the development of other new molecules, but they are limited to common parameters. In fact, as commented in the Introduction, while characterizing the chemical diversity of such databases, a common practice is filtering out metal-containing molecules.

To the best of our knowledge, there are no large compound databases that store and organize the information of organometallic molecules annotated with biological activity. Therefore, building, curating and maintaining a database of this kind is a major area of opportunity to integrate informatics methods to organometallic drug discovery. To address this need, a proof-of-principle database is D-InnoDB [13]. This database with still a limited number of compounds so far, contains information of molecules approved for clinical use and under clinical development.

Compounds databases with organometallic compounds can be further developed in a web-based application (*vide infra*). Similar to other databases of organic compounds, a database of organometallic compounds annotated with biological activity can facilitate a large number of analysis such as structure-property (activity) relationships—QSP(A)R—including activity landscape modeling [14], data mining, and virtual screening, to name a few. The compound database can be

made publicly accessible and be like ChEMBL and or PubChem [15]. With such a database, accessibility to organometallic compounds would increase, promoting progress in research and proposing new studies.

2.2. Molecular Representation

A cornerstone in chemoinformatics is molecular representation [16]. The appropriate description of the molecules is the most important first step towards virtual any qualitative or quantitative analysis. This can be clearly seen in studies such as QSP(A)R where the selection of the descriptors is key to obtain a predictive model. In some instances, “simple” 2D descriptors can be enough to obtain a useful and predictive model. In other instances, more accurate descriptors are required to capture the molecular shape and 3D information for explaining and/or predicting biological activity or assessing metal-binding sites in proteins [17].

The accuracy and speed of the calculation to compute 1D, 2D or 3D descriptors are one of the most sensitive points considering that as the overall accuracy of the descriptors increases, the calculation speed decreases. This is particularly relevant for metal-based molecules while selecting the descriptors to be computed. Therefore, it is essential to keep in mind the application of the description to optimize resources. While there are several methods based on quantum mechanics to describe accurately metal-based compounds, such methods are still not suited to manage efficiently large amounts of structures. In addition, the existing (including calculated) information and descriptors on these molecules is not uniform and is not available for use.

Molecular Fingerprints

In cheminformatics, molecular fingerprints are common representations of organic compounds and several different types have been developed. In general, such fingerprints are computed very rapidly and are appropriate to analyze even thousands or millions of compounds efficiently. In turn, such representations are the basis to perform several analyses such as similarity searching, diversity, and clustering analysis (including qualitative, quantitative and visual analysis).

A general approach to generate appropriate molecular fingerprints for organometallic compounds is developing a typical (dictionary or topological) fingerprint for the organic portion of the molecule and then adding a fingerprint developed for the metal portion. A bottleneck of this approach is the speed of the calculations to compute the metal portion. A workaround to address this issue can be to generate large compound databases with the values pre-calculated for different metals.

2.3. Diversity Analysis

In drug discovery, common practice and useful chemoinformatic analysis is the quantification of the molecular diversity of compound databases [18]. For instance, to identify novel hits it is generally desirable to screen compound data-

bases with large structural diversity. The working hypothesis is that testing several different scaffolds with varied side chains will increase the probability to identify one or more promising compounds. That is why it is necessary to design libraries that have a chemical space with relevance and with tools that help to maximize the possibility of identifying leads molecules, responding to the detection demands of structurally diverse compounds.

To address the need for generating diverse libraries, organic chemists have developed “diversity-oriented-libraries” [19]. Another approach is the “libraries-from-libraries” [20]. In lead optimization, in contrast, a general approach is screening compound data sets with lower molecular diversity e.g., high structural similarity to the active, lead molecule. In other words, in lead optimization it is more common to explore focused regions in chemical space. Examples of data sets aimed to address this need are the “focused” and “targeted” libraries. For all these cases, *i.e.*, to select diverse or focused and less diverse libraries, experimental chemists (medicinal, organic or inorganic) can readily identify and select compounds that meet the desired criteria of diversity. However, when dealing with medium-to-large compound databases it becomes more difficult to assess molecular diversity in an accurate manner. This is clear when purchasing data sets available from third parties (commercial vendors, for instance). Therefore, diversity analysis is standard practice when analyzing organic small molecules. The methods available for this type of molecules can be readily extended to analyze the diversity of organometallic compounds. To this end, the development of molecular fingerprints appropriate for organometallic molecules (*vide supra*) can be the basis to measure the diversity. Such molecular fingerprint representations or other appropriate representation based on continuous values can be used as the basis to apply diversity metrics such as the Tanimoto coefficient, Euclidean distance or other diversity metrics available [16]. It would remain to assess the most suitable fingerprint representations and diversity metrics tuned for organometallic compounds.

2.4. Chemical Space

The concept of chemical space [21] is also quite relevant in drug discovery for several purposes. Although there is not a single, correct definition, one concept is a multi-dimensional space for set (ideally all chemical possible) compounds [22]. The concept of chemical space is the basis to perform studies that include but are not limited to QSP(A)R studies (e.g., it is used as a matrix that contains the descriptors and biological activity); diversity analysis; clustering and visual assessment of diversity; comparative studies assessing the similarity or differences among compound data sets. A suitable chemical space can be used as a standard for profiling most structural sets of interest [23].

Since the chemical space depends on chemical representation, there are no “unique” or “invariant” chemical spaces. Despite the large dependence of the chemical space with structural representation, quantitative and qualitative anal-

ysis of the chemical space of organic molecules is now relatively straightforward to study. Indeed, it is fairly common to find visual representation of the chemical space of compound data sets from different sources such as synthetic molecules (e.g., from diverse designs), natural products (e.g., from different geographical sources or natural origin) [24] [25]. This is largely in part due to there are “standard” representations and descriptors available for organic compounds. However, there is no visual representation of the chemical space comparing bioactive organic vs. organometallic compounds. This is due to largely in part, the lack of appropriate molecular descriptors suited to represent a large number of organometallic molecules in an efficient manner. As commented above, fingerprint representations of organometallic molecules will boost the qualitative and quantitative analysis of its chemical space.

2.5. QSP(A)R, Machine Learning, and AI

QSP(A)R and more recently machine learning and other artificial intelligence (AI) methods are being used heavily in drug discovery [26] [27]. In retrospective studies, one of the main interests of QSP(A)R models is to explain the biological activity at the molecular level or provide a rational basis of activity through the association of quantitative descriptors with a biological endpoint. The QSP (A) R studies are necessary to handle the information obtained from the detection methods already used. In machine learning and artificial intelligence, retrospective studies aim to learn as much as possible of the data to then predict the activity of new molecules. The prediction of biological activity is also one of the major goals of QSP(A)R studies. While QSPR methods are conducted regularly for organometallic and inorganic compounds to explain or predict properties such as chemical reactivity and other properties of interest in material science [28], the methods have not been largely used to predict biological activity. This is another major area of opportunity for chemoinformatics. A starting point to develop such methods can be the assembly, curation and maintenance of compound databases of organometallic molecules such as D-InnoDB discussed above. A database like that containing information of organometallic molecules annotated with biological activity can be the starting point to do SAR or structure-multiple activity relationships (SmART) [29].

2.6. Virtual Screening

In silico also called virtual screening of compound databases has been a very useful approach to identify hit compounds [30]. Compound databases from different sources such as synthetic libraries, natural product data sets [31], and even virtual libraries (where compounds are cherry-picked for synthesis and experimental testing) are screened regularly. To this end, two general approaches structure-based and ligand-based methods are employed. As discussed in detail elsewhere, the method of choice will depend on the experimental information available for the system e.g. if the 3D structure of the molecular target is known

or not. Examples of virtual screening techniques include but are not limited to docking, pharmacophore-based screening, similarity searching, and combinations of the above. In the latter cases, also named cascade or sequential approaches, fast (but less accurate) methods are applied first to rapidly filter large amounts of compounds followed by more accurate but slower methods to select molecules for experimental testing. At the end of the process factors such as availability of the physical samples are considered (e.g., commercial availability and cost, for instance, if the compounds are commercially available).

Similarity Searching

One of the ligand-based techniques frequently used in virtual screening is similarity searching. The rationale of these approaches is that similar compounds have similar activity (if there are no activity cliffs [27], that is, molecules with similar chemical structure but very different and unexpected large activity difference). In this case, the chemical structures of all the molecules in a compound library are compared systematically with the chemical structure of one or several active molecules that are used as reference or queries. Two key components to perform similarity searching are the molecular representation and a similarity measure [16]. For molecular representation is common to employ a molecular fingerprint because they are quite fast to compute (*vide supra*). Thus, applying chemoinformatics tools as a whole such as virtual screening, the use of fingerprints and the search for molecular similarity, we can expect to find a result that, although not definitive, gives an overview or a guide to what is being sought.

Thus far, similarity searching has not been reported for organometallic compounds, but it can be easily performed once an efficient molecular representation is developed. To this end, molecular fingerprints suited for organometallic molecules can boost the application of this technique to identify novel hit compounds. In addition to molecular fingerprints, other molecular representations can be employed.

2.7. ADME/Tox Profile

Absorption, distribution, metabolism, excretion, and toxicity (ADMETox) are key properties in drug development that characterize the evolution of a drug candidate within the body. Several lead compounds and clinical candidates fail due to inappropriate ADMETox characteristics. The reliable prediction *in silico* of these properties for small organic compounds remains an active area of research [32]. This is particularly true for organometallic compounds that are usually excluded from the studies due to their both challenging behavior and not so attractive applications (*vide supra*). However, there is a significant amount of ADMETox data reported in the literature that could be used to develop predictive models in the same or similar manner they are generated and optimized for small organic molecules. Thus, the predictive models can be further used to predict the properties of compounds in external data sets considering, of course, that they fall in the applicability domain of the structures used to generate the

models. Based on these models, programs or servers can be generated that allow predictions of the toxicity of new compounds. Predictive and validated models to predict the ADMETox profile of organometallic compounds can be stored and used through webservers, similar to the tools currently available to predict the properties of organic small molecules (*vide infra*).

2.8. Webservers

A broad number of chemoinformatic resources and methods are now available to the scientific community through webservers. A considerable number of such servers are publicly accessible [33] [34]. Chemoinformatic servers focused on organic small molecules include, but are not limited to, the generation and analysis of molecular descriptors, visual representation of the chemical space, diversity analysis, and servers to predict the ADMETox profile of compound data sets. Other servers are dedicated to hosting compounds and predictive models. As discussed above, the servers are focused on organic molecules such that a common preparation or curation step to analyze compound data sets is to remove compounds containing metals. Therefore, we consider that a significant area of opportunity to advance Medicinal Organometallic Chemistry is to develop, maintain and update web-severs able to manage and deal with organometallic molecules. Such servers can be used for hosting maintaining and mining compound databases, calculation of descriptors including molecular fingerprints, and the prediction of properties—including ADMETox—using validated QSP(A)R models, to name a few.

3. Conclusion

Organometallic- and inorganic-based compounds are promising resources to address novel and emerging molecular targets. Similarity, metal-based medicinal agents can represent new alternatives to tackle difficult targets poorly addressed by the current traditional chemical space typically defined by small organic molecules. In addition, organometallic-based compounds can be part of multi-target approaches used in combination with other biologics or organic small-molecules. While molecular modeling and docking of metal-complexes compounds are performed regularly to explore compound-target interactions, chemoinformatic approaches aimed to organize and manage the information of organometallic compound databases annotated with biological activity are still limited. Similarly, chemoinformatic approaches to study systematically the chemical diversity, visual representation of the chemical space, similarity searching, and SAR involving organometallic compounds are not fully developed. One of the key starting points to extend the cheminformatics to organometallic compounds is developing appropriate and efficient molecular representations to describe, as accurately as possible, the structure of the compounds. Such representations will largely depend on the intended purpose of the analysis, for instance, data mining, exploration of the chemical space, diversity analysis, molecular interactions drug/compound-molecular

targets, and SAR, to name a few. To this end, developing appropriate fingerprints can be a key component to treat organometallic compounds with chemoinformatic methods. We believe that the tools and practices of chemoinformatics can board the challenges presented by current trends in drug discovery and design, however, there is a need to expand the traditional search field to obtain new and better results that with the application of other disciplines, a common objective is achieved, such as the development and discovery of drugs. We expect that this Commentary contributes to further develop the fields of Medicinal Organometallic and Inorganic Chemistry.

Acknowledgements

Fruitful discussions with Emma Belem Andrade Hernández are acknowledged. We thank the School of Chemistry of the National Autonomous University of Mexico (UNAM) for funding through the *Programa de Apoyo a la Investigación y al Posgrado* (PAIP). We also thank that program NUATEI (*Nuevas Alternativas para el Tratamiento de Enfermedades Infecciosas*), IIB-UNAM, for funding.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Mjos, K.D. and Orvig, C. (2014) Metallodrugs in Medicinal Inorganic Chemistry. *Chemical Reviews*, **114**, 4540-4563. <https://doi.org/10.1021/cr400460s>
- [2] Markwalter, C.F., Kantor, A.G., Moore, C.P., Richardson, K.A. and Wright, D.W. (2019) Inorganic Complexes and Metal-Based Nanomaterials for Infectious Disease Diagnostics. *Chemical Reviews*, **119**, 1456-1518. <https://doi.org/10.1021/acs.chemrev.8b00136>
- [3] Wang, Y., Li, H. and Sun, H. (2019) Metalloproteomics for Unveiling the Mechanism of Action of Metallodrugs. *Inorganic Chemistry*, **58**, 13673-13685. <https://doi.org/10.1021/acs.inorgchem.9b01199>
- [4] Kenny, R.G. and Marmion, C.J. (2019) Toward Multi-Targeted Platinum and Ruthenium Drugs—A New Paradigm in Cancer Drug Treatment Regimens? *Chemical Reviews*, **119**, 1058-1137. <https://doi.org/10.1021/acs.chemrev.8b00271>
- [5] Medina-Franco, J.L., Martinez-Mayorga, K. and Meurice, N. (2014) Balancing Novelty with Confined Chemical Space in Modern Drug Discovery. *Expert Opinion on Drug Discovery*, **9**, 151-165. <https://doi.org/10.1517/17460441.2014.872624>
- [6] Méndez-Lucio, O. and Medina-Franco, J.L. (2017) The Many Roles of Molecular Complexity in Drug Discovery. *Drug Discovery Today*, **22**, 120-126. <https://doi.org/10.1016/j.drudis.2016.08.009>
- [7] Kircheva, N. and Dudev, T. (2019) Novel Insights into Gallium's Mechanism of Therapeutic Action: A DFT/PCM Study of the Interaction between Ga³⁺ and Ribonucleotide Reductase Substrates. *The Journal of Physical Chemistry B*, **123**, 5444-5451. <https://doi.org/10.1021/acs.jpcc.9b03145>
- [8] Sciortino, G., Garribba, E. and Maréchal, J.-D. (2019) Validation and Applications

- of Protein-Ligand Docking Approaches Improved for Metalloligands with Multiple Vacant Sites. *Inorganic Chemistry*, **58**, 294-306. <https://doi.org/10.1021/acs.inorgchem.8b02374>
- [9] Duffy, B.C., Zhu, L., Decornez, H. and Kitchen, D.B. (2012) Early Phase Drug Discovery: Cheminformatics and Computational Techniques in Identifying Lead Series. *Bioorganic & Medicinal Chemistry*, **20**, 5324-5342. <https://doi.org/10.1016/j.bmc.2012.04.062>
- [10] Miller, M.A. (2002) Chemical Database Techniques in Drug Discovery. *Nature Reviews Drug Discovery*, **1**, 220-227. <https://doi.org/10.1038/nrd745>
- [11] Walters, W.P. (2019) Virtual Chemical Libraries. *Journal of Medicinal Chemistry*, **62**, 1116-1124. <https://doi.org/10.1021/acs.jmedchem.8b01048>
- [12] Yang, J.F., Wang, D., Jia, C.Y., Wang, M.Y., Hao, G.F. and Yang, G.F. (2019) Freely Accessible Chemical Database Resources of Compounds for *in Silico* Drug Discovery. *Current Medicinal Chemistry*, **26**, 7581-7597. <https://doi.org/10.2174/0929867325666180508100436>
- [13] Percastre-Cruz, Y. and Medina-Franco, J.L. (2019) Towards a Compound Database of Organometallic Drugs for Chemoinformatic Simulations. *3rd International Conference in Bioinformatics, Simulations and Modeling (iCBSM)*, Talca, Chile, 4-8 November 2019.
- [14] Yoshimori, A., Tanoue, T. and Bajorath, J. (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. <https://doi.org/10.1021/acsomega.9b00595>
- [15] Bender, A. (2010) Databases Compound Bioactivities Go Public. *Nature Chemical Biology*, **6**, 309-309. <https://doi.org/10.1038/nchembio.354>
- [16] Maggiora, G., Vogt, M., Stumpfe, D. and Bajorath, J. (2014) Molecular Similarity in Medicinal Chemistry. *Journal of Medicinal Chemistry*, **57**, 3186-3204. <https://doi.org/10.1021/jm401411z>
- [17] Sciortino, G., Garribba, E., Pedregal, J.R.-G. and Maréchal, J.-D. (2019) Simple Coordination Geometry Descriptors Allow to Accurately Predict Metal-Binding Sites in Proteins. *ACS Omega*, **4**, 3726-3731. <https://doi.org/10.1021/acsomega.8b03457>
- [18] Medina-Franco, J.L. (2012) Interrogating Novel Areas of Chemical Space for Drug Discovery Using Chemoinformatics. *Drug Development Research*, **73**, 430-438. <https://doi.org/10.1002/ddr.21034>
- [19] Lenci, E., Menchi, G., Saldívar-Gonzalez, F.I., Medina-Franco, J.L. and Trabocchi, A. (2019) Bicyclic Acetals: Biological Relevance, Scaffold Analysis, and Applications in Diversity-Oriented Synthesis. *Organic & Biomolecular Chemistry*, **17**, 1037-1052. <https://doi.org/10.1039/C8OB02808G>
- [20] López-Vallejo, F., Nefzi, A., Bender, A., Owen, J.R., Nabney, I.T., *et al.* (2011) Increased Diversity of Libraries from Libraries: Chemoinformatic Analysis of Bis-Diazacyclic Libraries. *Chemical Biology & Drug Design*, **77**, 328-342. <https://doi.org/10.1111/j.1747-0285.2011.01100.x>
- [21] Medina-Franco, J.L., Naveja, J.J. and López-López, E. (2019) Reaching for the Bright Stars in Chemical Space. *Drug Discovery Today*, **24**, 2162-2169. <https://doi.org/10.1016/j.drudis.2019.09.013>
- [22] Virshup, A.M., Contreras-García, J., Wipf, P., Yang, W. and Beratan, D.N. (2013) Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds. *Journal of the American Chemical Society*, **135**, 7296-7303. <https://doi.org/10.1021/ja401184g>

- [23] Naveja, J. and Medina-Franco, J. (2017) Chemmaps: Towards an Approach for Visualizing the Chemical Space Based on Adaptive Satellite Compounds. *Fl000Research*, **6**, 1134. <https://doi.org/10.12688/fl000research.12095.1>
- [24] Saldívar-González, F.I., Valli, M., Andricopulo, A.D., da Silva Bolzani, V. and Medina-Franco, J.L. (2019) Chemical Space and Diversity of the Nubbe Database: A Chemoinformatic Characterization. *Journal of Chemical Information and Modeling*, **59**, 74-85. <https://doi.org/10.1021/acs.jcim.8b00619>
- [25] Olmedo, D.A., González-Medina, M., Gupta, M.P. and Medina-Franco, J.L. (2017) Cheminformatic Characterization of Natural Products from Panama. *Molecular Diversity*, **21**, 779-789. <https://doi.org/10.1007/s11030-017-9781-4>
- [26] Lavecchia, A. (2019) Deep Learning in Drug Discovery: Opportunities, Challenges and Future Prospects. *Drug Discovery Today*, **24**, 2017-2032. <https://doi.org/10.1016/j.drudis.2019.07.006>
- [27] Maggiora, G.M. (2006) On Outliers and Activity Cliffs—Why QSAR Often Disappoints. *Journal of Chemical Information and Modeling*, **46**, 1535. <https://doi.org/10.1021/ci060117s>
- [28] Ramprasad, R., Batra, R., Pailania, G., Mannodi-Kanakkithodi, A. and Kim, C. (2017) Machine Learning in Materials Informatics: Recent Applications and Prospects. *NPJ Computational Materials*, **3**, 54. <https://doi.org/10.1038/s41524-017-0056-5>
- [29] Saldívar-González, F.I., Naveja, J.J., Palomino-Hernández, O. and Medina-Franco, J.L. (2017) Getting Smart in Drug Discovery: Chemoinformatics Approaches for Mining Structure-Multiple Activity Relationships. *RSC Advances*, **7**, 632-641. <https://doi.org/10.1039/C6RA26230A>
- [30] Lavecchia, A. and Di Giovanni, C. (2013) Virtual Screening Strategies in Drug Discovery: A Critical Review. *Current Medicinal Chemistry*, **20**, 2839-2860. <https://doi.org/10.2174/09298673113209990001>
- [31] Prieto-Martinez, F.D. and Medina-Franco, J.L. (2018) Flavonoids as Putative Epi-Modulators: Insight into Their Binding Mode with Brd4 Bromodomains Using Molecular Docking and Dynamics. *Biomolecules*, **8**, 61. <https://doi.org/10.3390/biom8030061>
- [32] Wang, Y., Liu, H., Fan, Y., Chen, X., Yang, Y., *et al.* (2019) *In Silico* Prediction of Human Intravenous Pharmacokinetic Parameters with Improved Accuracy. *Journal of Chemical Information and Modeling*, **59**, 3968-3980. <https://doi.org/10.1021/acs.jcim.9b00300>
- [33] Villoutreix, B.O., Lagorce, D., Labbé, C.M., Sperandio, O. and Miteva, M.A. (2013) One Hundred Thousand Mouse Clicks Down the Road: Selected Online Resources Supporting Drug Discovery Collected over a Decade. *Drug Discovery Today*, **18**, 1081-1089. <https://doi.org/10.1016/j.drudis.2013.06.013>
- [34] Gonzalez-Medina, M., Naveja, J.J., Sanchez-Cruz, N. and Medina-Franco, J.L. (2017) Open Chemoinformatic Resources to Explore the Structure, Properties and Chemical Space of Molecules. *RSC Advances*, **7**, 54153-54163. <https://doi.org/10.1039/C7RA11831G>