The many roles of molecular complexity in drug discovery

Oscar Méndez-Lucio and José L. Medina-Franco

Facultad de Química, Departamento de Farmacia, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico

Molecular complexity is becoming a crucial concept in drug discovery. It has been associated with target selectivity, success in progressing into clinical development and compound safety, among other factors. Multiple metrics have been developed to quantify molecular complexity and explore complexity–property relationships. However, there is no general agreement regarding how to measure this molecular feature. Herein, we have surveyed the many roles of molecular complexity in drug discovery discussing in a critical manner different quantification methods. Through the analysis of various reference compound databases, common pitfalls and workarounds of the quantification of molecular complexity are discussed.

Introduction

Molecular complexity, like molecular similarity and chemical space, is an intuitive but subjective concept that is not easy to quantify in a unique and 'best' manner. Furthermore, molecular complexity has been associated with major aspects in the drug development process such as success in progressing into clinical development [1], target selectivity [2,3] and compound safety [4]. To establish rigorous complexity–property relationships is necessary to measure the molecular complexity of chemical structures. Although the Oxford dictionary defines 'complex' as 'consisting of many different and connected parts' it seems that this definition is not always followed when talking about molecular complexity. To date, there is no global definition for molecular complexity; in fact, it is not unusual that chemists do not agree with each other regarding the complexity of a molecule [5,6]. Usually, molecular complexity is defined based on heuristic parameters influenced by each person's background, past experience or subdiscipline. Furthermore, molecular complexity is frequently associated with synthetic accessibility but these concepts have major differences. The first one should be regarded as a molecular property that only depends on the molecular structure and hence is not directly influenced by external factors. By contrast, synthetic challenges can be variable because they depend on the skills of the chemist, solvents, reagents and reactions available at the moment of the evaluation.

The lack of a unique definition of molecular complexity brings about the challenge of finding a universal metric to quantify it. This is reflected by the numerous approaches developed so far to measure complexity [1,7–13]. Among these attempts there is a broad range of possibilities from simple structural parameters, such as molecular weight, number of chiral centers or fraction of sp^3 (Fsp^3) atoms, to more elaborated formulations such as the one recently proposed by Böttcher [9]. Herein we have surveyed a broad range of implications of quantifying molecular complexity in drug discovery and development. The manuscript is organized in six main sections. After this introduction, different approaches to quantify molecular complexity are discussed. The next section presents applications of this concept in drug discovery. After that, the dependence of molecular complexity with different metrics through a survey of the molecular complexity of benchmark databases is discussed, followed by a brief presentation on the implications of molecular complexity beyond drug discovery. The review finishes by covering general recommendations based on the authors' experience before the concluding remarks.

Quantification of molecular complexity

Several different metrics have been proposed to date to assess molecular complexity. Some of them are more elaborate than

Corresponding authors: Méndez-Lucio, O. (oscarmen@comunidad.unam.mx), Medina-Franco, J.L. (medinajl@unam.mx)

http://dx.doi.org/10.1016/j.drudis.2016.08.009

others but all metrics capture molecular complexity to a certain degree. It should be mentioned that a simple complexity metric is not a synonym of poor performance.

Topological and physicochemical descriptors

A recent study showed that simple topological or physicochemical descriptors (e.g., number of chiral centers, number of unique topological torsions, a Wiener index, 'compactness' descriptors, etc.) are able to predict the perception of molecular complexity of trained chemists [5]. To some extent, this is in line with the famous quote: ''simplicity is the ultimate sophistication'' (attributed to several authors including Leonardo Da Vinci). In addition, many physicochemical or topological properties have been used independently to assess molecular complexity. For example, Fsp^3 is a common descriptor of complexity because saturation enables the preparation of more-complex molecules with larger three-dimensionality [1,2]. Similarly, the fraction of chiral centers (FCC) is also associated with molecular complexity and not because of synthetic difficulty but because their presence increases the number of unique molecules with the same formula and molecular weight [2,3]. Another common example is molecular weight, which is intuitively associated with larger and possibly more-complex structures.

Substructure-based approaches

Substructure methods assess complexity as a measure of feature richness by counting specific chemical features and combining them into a single score [6]. Some examples are the Whitlock [14] and Barone-Chanon [11] indices which are calculated from the number of rings, unsaturations, heteroatoms and chiral centers. Other formulations can include different features, for example atomic electronegativities and bond parameters as proposed by Allu and Oprea [13]. It is important to mention that these methods are broadly used despite the fact that they are usually based on empirical knowledge and optimization.

Graph-theoretical methods

An alternative to substructure-based approaches are those known as graph-theoretical approaches [6]. These methods usually take several topological parameters into account using graph theory to quantify molecular complexity, including: size, branching, cyclicity, symmetry, among others [15]. For example, Bertz proposed the use of subgraphs of a molecular graph to quantify molecular complexity [8]. Other approaches include the one proposed by Randić based on augmented vertex-degree [16], the one of Bonchev who proposed two indices (TC and TC1) based on connectivity of subgraphs [17] and the additive index recently proposed by Böttcher [9] just to name a few from a large list of methods available [7,18–20] including the indices based on quantum mechanics calculations developed by Luzanov and Babich [21].

Applications of molecular complexity in drug discovery

Molecular complexity has been associated with a number of properties relevant to drug discovery and development. The most representative are discussed in the next sections.

Lead optimization and drug development

In recent years a number of molecular complexity metrics have been associated with drug-likeness or with clinical success rate [22,23]. Lovering *et al.* showed that compounds were more complex (measured by sp^3 and number of chiral centers) as they advanced through different stages of clinical trials [1]. This observation could be explained at least in part because compounds with higher Fsp^3 carbons have more suitable physicochemical properties such as higher solubility or improved log*P* [23]. It could also be related to an increased compound potency because more-complex and less-flat structures are more likely to have a better drug–target complementarity. The later hypothesis was evaluated independently by Selzer *et al.* and Schuffenhauer *et al.* who analyzed the relationship between activity and molecular complexity of a large historical dataset from Novartis [6,24]. Both studies showed that, considering the datasets analyzed, highly active compounds ($IC_{50} < 1$ nM) are more complex than medium active ($1\ \mu M < IC_{50} > 1$ nM) or inactive compounds. In fact, putting these results together with the conclusions obtained by Hann *et al.* [12] we have a picture of the classic drug discovery pipeline: weak hits or leads typically present low molecular complexity, which is increased during lead optimization as part of the efforts to increase potency or improve ADME properties.

Two subsequent logical questions are: which is the minimum molecular complexity for a lead and which is the maximum complexity value for a drug? The first question was addressed by Hann *et al.* using Einstein's words: ''as simple as possible, but not simpler''. In other words, leads should be as simple as possible but keeping crucial molecular features to be active. Addressing the second question, a drug could be as complex as synthetically feasible without affecting ADMET properties [6]. For instance, natural products, which represent a large source of bioactive molecules, usually present high molecular complexity [25] as recently emphasized by González-Medina *et al.* using a dataset of fungal compounds [4].

Selectivity and promiscuity association

Based on previous observations that link activity to molecular complexity, it is not surprising that it can also be linked to compound selectivity or promiscuity. The rationale is that, at least in principle, a more-complex ligand will present better complementarity for a specific target, but not for off-targets. In fact, the theoretical study of Hann *et al.* supports this hypothesis because their results suggest that a higher molecular complexity reduces the chance of observing interactions between a random ligand and a protein target [12]. Nevertheless this hypothesis was not proved with real data until Lovering showed that promiscuity decreases as a function of molecular complexity using Fsp^3 and number of chiral centers as descriptors [2]. In a more recent work, experiments by Clemons *et al.* supported Lovering's observations after testing >15 000 molecules against 100 diverse protein targets [3]. Clemons *et al.* found that compounds from different sources (i.e., commercial, academic or natural products) presented different selectivity patterns. More importantly, these binding patterns were highly correlated to the same complexity metrics previously used by Lovering [1–3]. The Clemons study is relevant not only because it supports Lovering's and Hann's hypotheses but also because it showed that the complexity–selectivity relationship is not biased by lead optimization. In fact, Fsp^3 has already been used to guide the design of potent and selective inhibitors [26]. It must be mentioned that the fraction of chiral centers and Fsp^3 are not

the only complexity measures that could be associated with promiscuity, but they might be the most used so far because they are easy to calculate and straightforward to interpret. However, other measures can be employed as indicators of selectivity, for example the size of the molecular framework. Notably, Yang *et al.* showed that the fraction size of the molecular framework compared to the whole molecule (f_{MF}) is also associated with promiscuity and/or selectivity [27]. Molecules with large values of f_{MF} (i.e., large framework and few side-chains) are more promiscuous. A direct application in drug discovery would be using molecular complexity to develop polypharmacological drugs with the minimum number of off-targets possible [28,29].

Library design and compound selection

Given the increased awareness of the relationship between molecular complexity and drug-likeness, it is not surprising that this property is being taken into account to design chemical libraries and select compounds. Indeed, different research groups working on drug discovery projects have noted the need for including more-complex molecules in screening libraries to achieve higher success rates [30]. In fact, an analysis published in 2010 showed that commercial vendor libraries had the lowest molecular complexity compared with an academic collection or natural products [30]. Nevertheless, this has been changing; currently more and more chemical vendors are making available screening libraries particularly designed to have a broad range of complexity. An example is the Beyond the Flatland Library (http://www.chemdiv.com/products/beyond-flatland-library/) which is enriched with compounds with high Fsp^3 content as suggested by Lovering et al. [1,2]. Other examples include the Fsp^3-Enriched Fragment Library (http://lifechemicals.midnighters.eu/screening-libraries/fragment-libraries/fsp3-enriched-fragment-library), the PPI Focused BBs Collection and the 3D BBs Collection (http://www.princetonbio.com/products/targeted_bbl).

Dependence of metrics in complexity–property relationships: a case study

Despite the fact that a number of metrics have been developed to quantify molecular complexity, to the best of our knowledge there are no benchmark studies that compare the metrics directly with each other. Similar to other representations such as molecular fingerprints [31,32], molecular complexity methods provide different but complementary information. To illustrate the dependence of the molecular complexity values with different metrics, we surveyed the molecular complexity of publicly available datasets with different measures of easy access to the drug discovery community (i.e., straightforward to compute, implemented in open-source software or available online). As reference datasets the compounds published by Clemons *et al.* (natural products, commercial and academic compounds) [3] were analyzed. In addition, the molecular complexity of three additional datasets that were used as references was evaluated: (i) a subset of the ToxCast data collection (US Environmental Protection Agency, US Tox21 Program, PubChem Bioassay database, Source=Tox21; http://www.ncbi.nlm.nih.gov/sites/entrez?db=pcassay&term=Tox21) that includes 5593 compounds tested on 33 different assays; (ii) 1814 approved drugs from DrugBank [33]; and (iii) the NIH Molecular Libraries Small Molecule Repository (MLSMR) which is a large set of >400 000 small molecules typically evaluated using HTS (http://www.ncbi.nlm.nih.gov/sites/entrez?db=pcsubstance&term=MLSMR). In this survey, the metrics compared were the fraction of chiral atoms, Fsp^3, PubChem complexity (using the Bertz/Hendrickson/Ihlenfeldt formula [10]) and a complexity index calculated with DataWarrior [34]. Of course, beyond this short review, the analysis can be extended comparing more complexity metrics discussed in the previous section and analyzing many more compound datasets. As stated above, herein we selected a number of complexity metrics of easy access for the drug discovery community.

Compound selectivity

As reviewed above, one of the major interests to quantify molecular complexity in drug development campaigns is to find relationships between this molecular feature with compound selectivity. In this subsection we explored such associations with two independent sets of compounds annotated with biological information for a number of biological endpoints: Clemons' and ToxCast datasets. In contrast to Clemons' study [3], we survey the effect of using different complexity metrics. By analogy with the work of Clemons *et al.* (see above), compounds in both datasets were classified in three categories based on the relative distribution of the complexity values: (i) simple: if complexity $\leq$ mean − 1 standard deviation (SD); (ii) intermediate: if mean − 1 SD < complexity < mean + 1 SD; and (iii) complex: if complexity $\geq$ mean + 1 SD. Similarly, compounds were also classified into four categories based on bioactivity data: (i) compounds inactive in all assays; (ii) active in only one assay; (iii) active in 2–5 assays; and (iv) active in six or more assays. Fig. 1 summarizes the results of the molecular complexity measured for the Clemons' and ToxCast sets using four metrics outlined above: PubChem Complexity, DataWarrior Complexity, Fsp^3 atoms and fraction of chiral atoms – the pie charts summarize the distribution of compounds considering activity and selectivity.

The results in Fig. 1 are in agreement with the trends observed in the Clemons publication, in that simple compounds (based on fraction of chiral atoms) tend to be more promiscuous (active in six or more assays) compared with complex molecules. Similar results were obtained for the ToxCast set, where 450 simple compounds were active in more than six assays; by contrast only 22 complex compounds presented a similar degree of promiscuity. A similar trend (less complex compounds are more promiscuous than simple molecules) can be observed for both datasets when fraction of sp^3 carbons was used as the complexity metric. However, results obtained with the complexity values computed in PubChem and DataWarrior slightly differ from the ones obtained with fraction of chiral and sp^3 atoms. In particular, PubChem and DataWarrior complexity metrics suggest that complex compounds are less promiscuous for Clemons' data (108 and 110 compounds, respectively), but more-promiscuous for ToxCast (249 and 285 compounds, respectively) compared with simple compounds. As discussed below, these results are a consequence of the metric complexity and how compounds were classified as simple or complex.

Compound classification: the importance of using a reference dataset

In general, assessing whether a compound is 'complex' or 'simple' is a difficult task because complexity is a subjective concept and it

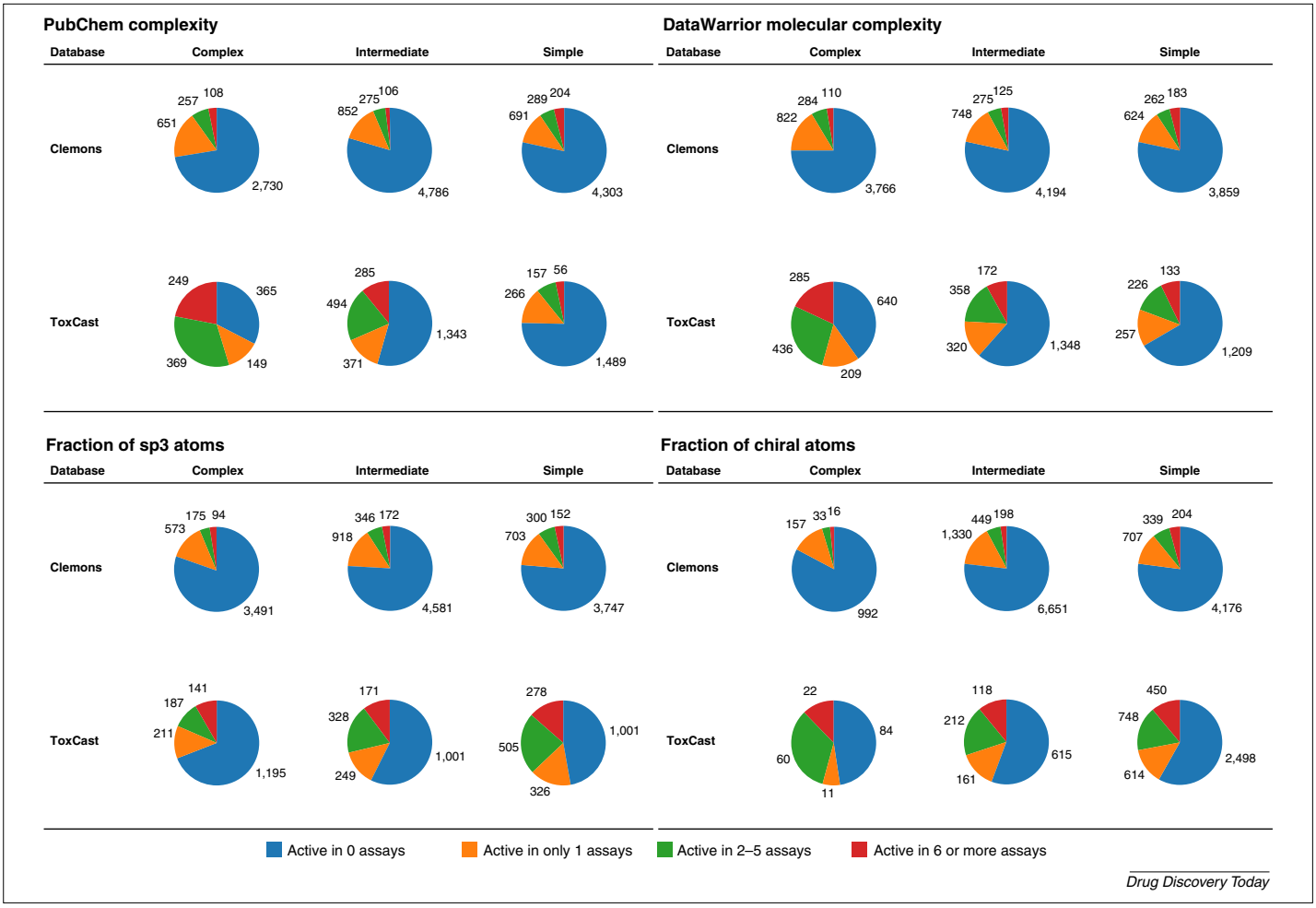

FIGURE 1

Molecular complexity measured for the Clemons and ToxCast sets using four metrics.

is usually based on comparisons. Many studies using complexity metrics classify compounds into complex or simple by using arbitrary cutoffs or by comparing the complexity of each compound to other molecules in the dataset (as in Clemons' study and the survey discussed in the previous section). In general, based on the findings of our survey above, the use of complexity cutoff values is not advisable unless they are well supported by previous knowledge or evidence. For example, the distribution of complexity values of a large compound database can be used as reference. Nevertheless, this potential solution should be implemented with caution addressing the questions: is the dataset large enough to make a fair judgment of complexity and how can the results obtained with one dataset be compared with a different set of molecules? For instance, Fig. 2 depicts the complexity distributions of four reference datasets (Clemons, ToxCast subset, MLSMR and DrugBank) using four complexity metrics. It can be seen that, in general, the four datasets have different distributions for each complexity metric except for the fraction of chiral centers (FCC) where all datasets presented complexity values close to zero. This is not surprising because most compounds have few stereo-centers to facilitate the chemical synthesis. It is noteworthy that the ToxCast subset and the Clemons dataset presented different distributions of PubChem and DataWarrior complexity metrics. More specifically, the Clemons dataset contains more complex compounds than ToxCast. This fact directly affects compound classification and cross-comparison between datasets. For instance, if compounds are classified based on the absolute distribution of complexity values in each dataset, the most complex compounds of the ToxCast subset have complexity values similar to the ones presented by the regarded simple compounds of the Clemons dataset. Of note, the conclusions obtained from the study will be valid only for the specific dataset and can be generally applicable if the dataset is large enough. Therefore, as a workaround of issues associated with discrepancies in classification (e.g., high, medium, low complexity), it is proposed to classify compounds based on the distribution of complexity values of a large and comprehensive dataset as reference. The selection of the reference dataset will depend on the goals of the study and the type of molecules being analyzed. For example, in a drug discovery project that involves typical drug-like compounds, MLSMR and/or a collection of approved drugs such as DrugBank are suitable reference databases because they are large and diverse collections available in the public domain.

Relationship between complexity metrics

As discussed above, it is remarkable the broad range of metrics used to assess molecular complexity. This raises the question: are different complexity metrics giving the same information? As a

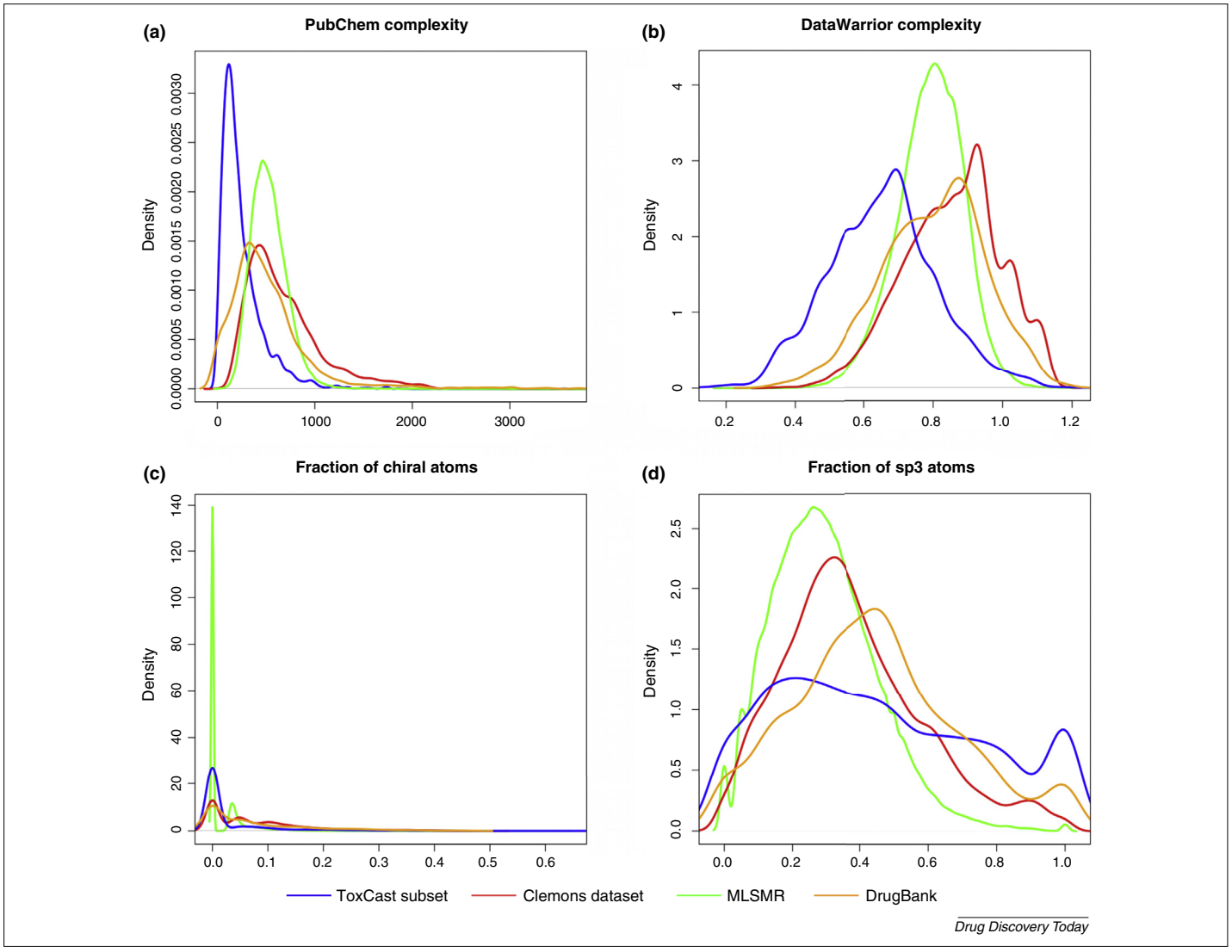

FIGURE 2

Complexity distributions of four reference datasets using four complexity metrics.

contribution to address this question in this short survey we compared the complexity values of four different metrics (see above) computed for a large and diverse set. In this survey, MLSMR was used as a case study. As mentioned above, MLSMR is a large (>400 000 small molecules), diverse and publicly available dataset. Fig. 3 shows density plots of all six pair-wise comparisons of the four different metrics of complexity. The density of data points is represented using a continuous color scale from yellow (most populated region) to gray (least populated). At a glance, Fig. 3 indicates that the metrics have no correlation. In particular, metrics specially designed to capture the whole complexity of a molecule (e.g., PubChem and DataWarrior complexity) did not correlate with metrics typically used to measure complexity, such as FCC and Fsp^3. For example, compounds with high FCC values are associated with low PubChem complexity values, whereas the same molecules have high DataWarrior complexity (see Figure S1 in the supplementary material online). This suggests that each of these complexity metrics are capturing different information: a compound classified as complex with one metric can be regarded as simple with the other. In other words, there is a dependence of molecular complexity with the metric used to quantify it. These results suggest that molecular complexity cannot be defined by a single or individual property, but instead it should be associated to a group of properties. This problem is related to molecular representation using structural fingerprints: a single fingerprint cannot capture all the information of a molecule [35]. In analogy with molecular fingerprints, it is suggested to measure molecular complexity as a combination of complexity metrics [36–40].

Beyond drug discovery

The relevance and possible application of molecular complexity is not restricted to drug discovery and has been noted by researchers in different fields. A novel and interesting application is the association of complexity with sensorial responses such as olfactory notes [41]. An original paper by Kermen *et al.* describes the quantitative structure–odor relationship using molecular complexity. Authors of that work showed that more-complex structures (as measured using PubChem complexity) evoke more numerous olfactory notes. Although molecular complexity cannot distinguish between functional groups or different smells, it was

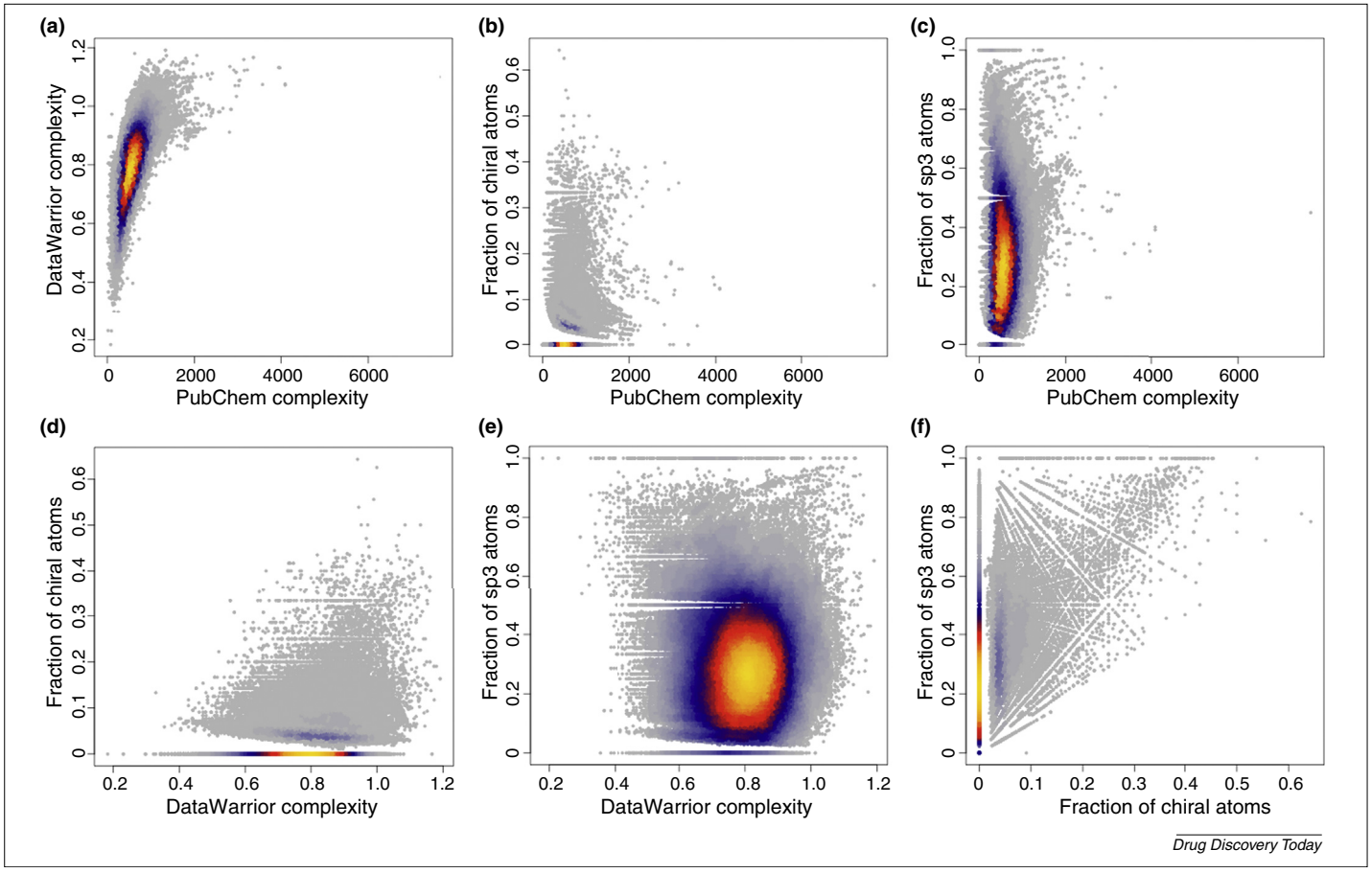

FIGURE 3

Density plots of all six pair-wise comparisons of the four different metrics of complexity computed for the Molecular Libraries Small-Molecule Repository (MLSMR). The density of data points is represented with a continuous scale from more-dense (orange) to less dense (gray).

correlated with the number of notes and odor pleasantness. It is worth noting that the neural mechanism by which molecular complexity is associated with olfactory notes remains unknown given the intricacy of the olfactory system. One of the hypotheses is that high complexity odorants activate more types of olfactory receptor than low complexity odorants [41], which contrasts with different complexity–selectivity studies (see above) [2,3]. An alternative explanation is the known relationship between complexity metrics and physicochemical properties such as logP or solubility [1,23], because they have been associated with structure–odor relationships and odor intensity [42–44].

Concluding remarks

Throughout this review we highlighted the importance of molecular complexity for drug design and development. Although there is not an absolute definition of complexity different metrics have been developed to apply the complexity concept in numerous drug discovery situations. However, the broad range of approaches to quantify molecular complexity can lead to different pitfalls. To avoid these pitfalls it is advised to: (i) use a large and diverse reference dataset when labeling compounds as complex instead of using arbitrary cutoffs; and (ii) apply more than one complexity metric to reduce the dependency of the conclusions on how complexity is measured.

Quantification of molecular complexity and its use in drug discovery is not a fully solved problem and it is in continued development. For instance, a comprehensive research study analyzing multiple reference collections with multiple complexity metrics is yet to be conducted (including many more than the four used in this short review). It also remains to investigate different ways to combine the complexity metrics to develop robust consensus measures of complexity. In our opinion, most of the questions that remain unanswered are associated with the 'best' way to measure molecular complexity and how to interpret the results for specific purposes. For example, find a concrete complexity predictor to guide the design of new molecules, identify the ideal range of complexity values for drug development, determine whether the conclusions obtained from complexity indices are universal or if they are target or case-dependent, and develop a global molecular complexity metric. As discussed in this review, there has been significant progress to advance the field of molecular complexity including uses in different areas such as library design and target selectivity, among others. However, there is still a long way to go.

Conflicts of interest

The authors declare that they do not have any conflicts of interest related to this manuscript.

Acknowledgments

We thank the funding from the Programa de Apoyo a la Investigación y el Posgrado (PAIP) 5000-9163, Facultad de Química, UNAM.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.drudis.2016.08.009.

References

1 Lovering, F. *et al.* (2009) Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* 52, 6752–6756
2 Lovering, F. (2013) Escape from flatland 2: complexity and promiscuity. *MedChemComm* 4, 515–519
3 Clemons, P.A. *et al.* (2010) Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles. *PNAS* 107, 18787–18792
4 González-Medina, M. *et al.* (2016) Chemoinformatic expedition of the chemical space of fungal products. *Future Med. Chem.* 8 (12), 1399–1412
5 Sheridan, R.P. *et al.* (2014) Modeling a crowdsourced definition of molecular complexity. *J. Chem. Inf. Model.* 54, 1604–1616
6 Selzer, P. *et al.* (2005) Complex molecules: do they add value? *Curr. Opin. Chem. Biol.* 9, 310–316
7 von Korff, M. and Sander, T. (2013) About complexity and self-similarity of chemical structures in drug discovery. In *Chaos and Complex Systems: Proceedings of the 4th International Interdisciplinary Chaos Symposium* (Stavrinides, G.S. *et al.* eds), pp. 301–306, Springer, Berlin Heidelberg
8 Bertz, S.H. (1981) The first general index of molecular complexity. *J. Am. Chem. Soc.* 103, 3599–3601
9 Böttcher, T. (2016) An additive definition of molecular complexity. *J. Chem. Inf. Model.* 56, 462–470
10 Hendrickson, J.B. *et al.* (1987) Molecular complexity: a simplified formula adapted to individual atoms. *J. Chem. Inf. Comput. Sci.* 27, 63–67
11 Barone, R. and Chanon, M. (2001) A new and simple approach to chemical complexity. Application to the synthesis of natural products. *J. Chem. Inf. Comput. Sci.* 41, 269–272
12 Hann, M.M. *et al.* (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* 41, 856–864
13 Allu, T.K. and Oprea, T.I. (2005) Rapid evaluation of synthetic and molecular complexity for *in silico* chemistry. *J. Chem. Inf. Model.* 45, 1237–1243
14 Whitlock, H.W. (1998) On the structure of total synthesis of complex natural products. *J. Org. Chem.* 63, 7982–7989
15 Nikolić, S. *et al.* (2000) Complexity of molecules. *J. Chem. Inf. Comput. Sci.* 40, 920–926
16 Randić, M. and Plavšić, D. (2002) On the concept of molecular complexity. *Croat. Chem. Acta* 75, 107–116
17 Bonchev, D. (2000) Overall connectivities/topological complexities: a new powerful tool for QSPR/QSAR. *J. Chem. Inf. Comput. Sci.* 40, 934–941
18 Randić, M. and Plavšić, D. (2003) Characterization of molecular complexity. *Int. J. Quantum Chem.* 91, 20–31
19 Randić, M. *et al.* (1979) Search for all self-avoiding paths for molecular graphs. *Comput. Chem.* 3, 5–13
20 Rücker, G. and Rücker, C. (2000) Walk counts, labyrinthicity, and complexity of acyclic and cyclic graphs and molecules. *J. Chem. Inf. Comput. Sci.* 40, 99–106
21 Luzanov, A.V. and Babich, E.N. (1995) Quantum-chemical quantification of molecular complexity and chirality. *J. Mol. Struct. Theochem.* 333, 279–290
22 Waring, M.J. *et al.* (2015) An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* 14, 475–486
23 Meanwell, N.A. (2016) Improving drug design: an update on recent applications of efficiency metrics, strategies for replacing problematic elements, and compounds in nontraditional drug space. *Chem. Res. Toxicol.* 29, 564–616
24 Schuffenhauer, A. *et al.* (2006) Relationships between molecular complexity, biological activity, and structural diversity. *J. Chem. Inf. Model.* 46, 525–535
25 López-Vallejo, F. *et al.* (2012) Expanding the medicinally relevant chemical space with compound libraries. *Drug Discov. Today* 17, 718–726
26 Hirata, K. *et al.* (2016) SAR exploration guided by LE and Fsp3: discovery of a selective and orally efficacious rorγ inhibitor. *ACS Med. Chem. Lett.* 7, 23–27
27 Yang, Y. *et al.* (2010) Investigation of the relationship between topology and selectivity for druglike molecules. *J. Med. Chem.* 53, 7709–7714
28 Hu, Y. and Bajorath, J. (2013) How promiscuous are pharmaceutically relevant compounds?. A data-driven assessment. *AAPS J.* 15, 104–111
29 Reddy, A.S. and Zhang, S. (2013) Polypharmacology: drug discovery for the future. *Expert Rev. Clin. Pharmacol.* 6, 41–47
30 Dandapani, S. and Marcaurelle, L.A. (2010) Grand challenge commentary: accessing new chemical space for 'undruggable' targets. *Nat. Chem. Biol.* 6, 861–863
31 Koutsoukas, A. *et al.* (2014) How diverse are diversity assessment methods?. A comparative analysis and benchmarking of molecular descriptor space. *J. Chem. Inf. Model.* 54, 230–242
32 Duan, J. *et al.* (2010) Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. *J. Mol. Graphics Modell.* 29, 157–170
33 Law, V. *et al.* (2014) Drugbank 4.0: shedding new light on drug metabolism. *Nucl. Acids Res.* 42, D1091–D1097
34 Sander, T. *et al.* (2015) Datawarrior: an open-source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.* 55, 460–473
35 Sheridan, R.P. and Kearsley, S.K. (2002) Why do we need so many chemical similarity search methods? *Drug Discov. Today* 7, 903–911
36 Medina-Franco, J.L. *et al.* (2009) Characterization of activity landscapes using 2D and 3D similarity methods: consensus activity cliffs. *J. Chem. Inf. Model.* 49, 477–491
37 Dimova, D. *et al.* (2014) Method for the evaluation of structure–activity relationship information associated with coordinated activity cliffs. *J. Med. Chem.* 57, 6553–6563
38 Maggiora, G. *et al.* (2014) Molecular similarity in medicinal chemistry. *J. Med. Chem.* 57, 3186–3204
39 Stumpfe, D. *et al.* (2014) Recent progress in understanding activity cliffs and their utility in medicinal chemistry. *J. Med. Chem.* 57, 18–28
40 Hu, Y. *et al.* (2016) Computational exploration of molecular scaffolds in medicinal chemistry. *J. Med. Chem.* 59, 4062–4076
41 Kermen, F. *et al.* (2011) Molecular complexity determines the number of olfactory notes and the pleasantness of smells. *Sci. Rep.* 1, 206
42 Schiffman, S.S. (1974) Physicochemical correlates of olfactory quality. *Science* 185, 112–117
43 Greenberg, M.J. (1979) Dependence of odor intensity on the hydrophobic properties of molecules. A quantitative structure odor intensity relationship. *J. Agric. Food Chem.* 27, 347–352
44 Turin, L. and Yoshii, F. (2003) Structure–odor relations: a modern perspective. In *Handbook of Olfaction and Gustation* (Doty, R., ed.), Marcel Dekker