MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **85** (2021) 465-498

ISSN 0340 - 6253

An Intuitionistic Fuzzy Set Analysis of Drug-Target Interactions

Gerry Maggiora[1], Eulalia Szmidt[2,3]

[1]*BIO5 Institute, University of Arizona*
Tucson, AZ 85718 USA
maggiora@email.arizona.edu

[2]*Systems Research Institute Polish Academy of Sciences,*
Newelska 6, 01-447 Warsaw, Poland, and
[3]*Warsaw School of Information Technology, ul. Newelska 6, 01-447 Warsaw, Poland*
szmidt@ibspan.waw.pl

(Received December 31, 2019)

Abstract

The use of drugs in modern medicine is widespread and no doubt has played a significant role in improving outcomes in the treatment of many medical conditions. Hence, the discovery of drugs for new therapeutic indications as well drugs of enhanced efficacy for existing indications are of paramount importance. Information on drug-target interactions plays an important role in this task—its importance is reflected by the fact that searching the Internet based on 'drug-target interaction' yields 280,000,000 hits. The interaction of a given drug with respect to multiple drug targets—also known as *polypharmacology*—is becoming increasingly important in modern drug research as it bears on the role of the biological pathways underlying a drug's mechanism(s) of action, the existence of side-effects, and the repurposing of existing drugs for new therapeutic indications. The ability of multiple, structurally dissimilar drugs to interact with a given target, also known as *polyspecificity*, is the complement of polypharmacology, and while not as well-known has nevertheless played a role in drug discovery research under the rubric of multiple lead series. Both of these concepts are considered in this work. Although sets of *n* drugs and *m* targets give rise to $n \times m$ virtual drug-target pairs, experimental and computational interaction data typically exist for only a small subset of these pairs. Thus, polypharmacology and polyspecificity values will, in all likelihood, be underestimated. As is shown in this work, taking account of drug-target pairs of unknown interaction (*i.e.* 'null' pairs) yields an upper bound to these values. However, in order to include such information in their analysis it is desirable that a methodology be able to handle null pairs in a reasonably straightforward manner. This is not the case with classical sets where the interaction of drug-target pairs is typically represented by set-theoretic

relations, where a '1' indicates the presence of an interactive drug-target pair—with respect to a given interaction threshold—and a '0' indicates the presence of a non-interactive pair. A new set-theoretic formalism based on *extreme-value intuitionistic fuzzy sets* (IFS) is utilized in this work since it *explicitly* accounts for the uncertainty inherent in null drug-target pairs. As will be demonstrated in future work, it can be generalized to include drug-target pairs with real-valued interactions. A simple example dataset is provided illustrating the application of IFSs to the analysis of drug-target datasets.

1 Introduction

1.1 Overview of data sources

The use of drugs in modern medicine is widespread and no doubt has played a significant role in improving outcomes in the treatment of many medical conditions. Hence, the discovery of drugs for new therapeutic indications as well as drugs of enhanced efficacy for existing indications are of paramount importance. In this regard, the field of *chemogenomics*, which is based largely on the study of drug-target interactions, plays a central role. The growing importance of such interactions is clearly of great significance in drug research as shown by a recent Internet search of 'drug-target interaction' that yielded 280,000,000 hits. Heretofore, most of the data on drug-target interactions resided in the corporate databases of pharmaceutical companies, but over the last decade there has been a significant increase in the amount of drug-target information that now can be found in a number of publicly available databases [1–9].[1]

Even though a wide variety of drugs and their targets have been studied, the interactions of only a small percentage of drug-target pairs have been determined experimentally, as exemplified by the four databases given in Table 1. In fact, the known interactions of drug-target pairs in many databases represent less than five percent of the possible interactions, though there are exceptions. Mestres, *et al.* have characterized this condition as a lack of *data completeness* [10,11]. Addressing this issue experimentally is an enormous task. And although obtaining large, diverse sets of compounds for screening campaigns is fairly straightforward [12-14], the same cannot be said for drug targets, even with the availability of the modern tools afforded by molecular biology. Thus, the lack of experimental drug-target data can largely be attributed to a lack of target availability, although the resources needed are also considerable, further frustrating efforts within the public domain to create large drug-target datasets (*Cf.* [9]).

[1] The terminology 'drug' rather than the more general 'ligand' is employed in this work because of its extensive usage in drug research. It should, however, be noted that the interaction of a wide variety of ligands, of which drugs form a subset, are typically contained in such databases. In addition, the term 'interaction' rather than the more restrictive 'activity' is used here as many of the values that describe the state of drug-target pairs, such as IC_{50}'s and scoring function values, are not directly associated with activities or binding free energies.

Table 1. Examples of drug-target databases [2-4]. The number of active and inactive drug-target pairs are given for two different activity thresholds, 1.0 μM and 0.1 μM.

Data-bases	Number of Drugs	Number of Targets	1.0 μM		0.1 μM		Total Interactions	Total Possible Interactions	Percent Data Complete[b]
			Active	Inactive	Active	Inactive			
Binding DB[a]	10,460	681	53,446	24,964	35,816	42,594	89,815	7,123,260	1.0126
PDSP	1,042	147	2,739	1,951	1,831	2,860	4,690	153,174	3.062
ChEMBL Kinase	28,032	259	34,222	18,011	20,022	32,209	52,233	7,260,288	0.719
ChEMBL GPCR	36,659	136	60,913	20,500	40,280	41,133	81,413	4,985,624	1.633

[a] Note that there are 11,405 entities where the activity values are not given explicitly but rather are given as less than or greater than with respect to the activity thresholds.

[b] The percent data completeness is given by the ratio of the number of drug-target pairs that are equal to or exceed the activity threshold divided by the total number of possible pairs.

Table 1 also provides an indication of how sensitive the number of interactive pairs is to the value of the interaction threshold. Since interaction values that fall below $\sim 20\ \mu M$ are rarely of interest in the search for bioactive compounds, such weakly interactive pairs are rarely if ever included in drug-target databases. This results in a severe undercounting of these pairs, further exacerbating the already significant lack of data completeness present in all public drug-target databases (*vide supra*). As discussed in Section 5, neglecting weakly-interactive pairs can significantly influence the magnitude of computed upper bounds for a number of drug-target parameters.

Not surprisingly, data completeness is significantly biased by therapeutic areas of interest, such as kinases, ion channels, G-protein coupled receptors, and proteases, that drive much of drug research. In addition to the lack of completeness, data in these databases tend to be sporadic and of uneven quality due to the fact they are obtained from multiple datasets acquired from many sources, using a number of experimental protocols, under a variety of experimental conditions. Although a number of rela-tively small, effectively complete datasets have been assembled, they are generally confined to industrial pharmaceutical laboratories. In any case, publicly available drug-target databases fall far short of attaining the professed, but highly unrealistic, goal of chemogenomics, namely, the description of all possible drugs with respect to all possible druggable targets [15-18].

1.2 Computational approaches to drug-target interactions

Numerous attempts have been made to overcome the lack of data completeness by applying computational methods. Because of the sheer size of the problem, an early trend appeared to favor a more qualitative approach aimed at assessing whether a given drug does or does not interact in some fashion with a specific target. This is consistent with Rognan's view that "*changing paradigms in rational drug design—making a binary yes-or-no answer instead [of] quantitative predictions—are pushing again structure-based algorithms under the spotlight*" [19]. Koeppen and Bieler have provided an excellent overview of many of the computational aspects of chemogenomics [20], as has the book edited by Jacoby [21]. Other relevant reviews include [22-25]. Table 1 of the publication by Nogueira and Koch [26] provides a reasonably up to date summary of the available web-based servers for target prediction.

Recently, a host of powerful computational algorithms have been developed, some of which are exemplified by the following references [26-37]. As is the case for mathematical models in general, these methods rely on existing experimental drug-target interaction data, and hence are plagued by its general sparsity and inhomogeneity. Because of this they are unable to deal entirely with the problem of data completeness, although they hold some promise for limited sets of drugs and targets.

Structure-based 'docking' methods offer an alternative to mathematical modeling-based approaches since they do not depend on experimental drug-target interaction data, although they do depend on detailed three-dimensional target-structure data. There are two general types of approaches that fall under this rubric—scoring-function methods [26,38-43] and free-energy methods [44-47]. The former tends to treat molecular interactions phenomenologically while the latter tends to adopt a more fundamental (*a priori*) approach based on the use of molecular potential-energy functions and molecular dynamics or Monte Carlo simulations. Both approaches require structural information on the target (usually a protein), which can be obtained from the PDB (Protein Data Bank and its affiliates: www.rcsb.org) and from the BMRB (Biological Magnetic Resonance Bank: www.bmrb,wisc.edu). Although there are nearly 150,000 macromolecular structures in these repositories, the three-dimensional structures of many biologically relevant targets, especially the membrane-bound receptors that provide a significant number of drug targets, are still unavailable. An added complication of both methods is the need to account for the structural flexibility of both drugs and targets. Although partial drug and target flexibility can be handled rather efficiently, if fully accounted for it adds considerably to the computational burden of these methods.

Since the values obtained in many instances from scoring functions are phenomenological, thresholds used to separate interactive from non-interactive drug-target pairs are by their nature somewhat artificial. Moreover, since there is no measure common to all scoring functions, values determined by different computational methods cannot be easily compared to one another. This is not the case for drug-target interactions estimated by free-energy methods, which have a fundamental thermodynamic basis. Thus, computed free energies can, in principle, be compared directly, although their accuracy may differ depending on a variety of factors associated with their determination. Hence, all of these approaches must be used with caution.

Scoring function-based methods have recently been refined to an amazing degree and now can deal with *computational screening experiments* that involve the docking of millions to even billions of compounds to a given target [48]. Moreover, these methods have identified active molecules with novel chemotypes [49,50]. Such methods are, however, extremely computationally intensive, and while they have identified a number of interesting chemotypes they have no doubt also missed some. In any case, they represent a significant advance in computational docking methodology, and as such hold considerable promise as a means for identifying interesting, new structural classes of drugs. The difficulty with applying these methods to the current work is the need to carry out docking experiments not only on a large number of compounds but also on large a number of targets as well, a significant computational burden to say the least. Although free-energy methods afford a more fundamental approach to the determination of drug-target interactions, they are considerably more computationally intensive. Hence, they do not at present provide a practical alternative to scoring-function

methods. In any case, it does not appear that the issue of data completeness for large drug-target databases can be fully addressed experimentally or computationally any time soon.

As is the case with experimental methods, a number of issues also arise with respect to the results obtained by computational methods, as a given method can usually be implemented in a variety of ways, *e.g.* through the use of more accurate and complete datasets in mathematical modeling and improved scoring and potential-energy functions in structure-based methods. Hence, use of the same method by two different laboratories does not guarantee identical results—some implementations of a given method will produce better results than others. Although these are important issues, dealing with them in more detail in this work would take us too far afield.

1.3 A set-theoretical view of drug-target interactions

Drug-target pairs are taken to be the elements of sets or relations, the latter also being sets. Pairs with interaction values obtained experimentally or computationally that are equal to or greater than a given threshold value are typically designated as interactive pairs. Pairs whose interaction fall below the threshold are designated as non-interactive—use of such thresholds is common practice in the analysis of drug-target databases. A major, often neglected, problem involves drug-target pairs whose activities or interactions have neither been experimentally determined nor computationally estimated. Such pairs are typically treated *de facto* as non-interactive pairs or are neglected entirely. This is routinely done in drug-target networks [51-53], which have become a popular way to represent the drug-target information. Although rarely done, it is possible to include information on non-interactive and null pairs using 'edge-colored' networks or the matrix representations [54].

Because of the differences posed by experimental and computational assessments of drug-target activities or interactions, it is desirable to adopt a consistent language for the remainder of this paper. Thus, the following terminology is used to describe the three types of drug-target pairs. Active or interacting pairs will be designated as ***interactive pairs***, inactive or non-interacting pairs will be designated as ***non-interactive pairs***, and pairs of unknown activity or interaction will be designated as ***null pairs***. As shown in Table 1, null drug-target pairs constitute a substantial fraction of the total number of possible pairs in most databases. Hence, assuming that such pairs are non-interactive or neglecting them entirely can significantly bias any data analysis. Thus, a way is needed to take some account of null drug-target pairs.

As will be clear in the following, *intuitionistic fuzzy sets* (IFSs) provide a suitable mathematical framework [55-61] for doing so. Nevertheless, the question arises as to whether the issues treated in this work could not be handled by more traditional methods. This is, in fact, the case in the current work, which can also be handled using classical set-theoretic ternary relations. However, they cannot be easily extended to treat drug-target pairs with explicit,

continuous interaction values, a subject of on-going research in our laboratories. Moreover, working within a single, consistent mathematical framework such as IFSs ensures that the parameters and inferences derived therefrom provide an internally consistent framework for treating a range of related problems.

1.4 Polypharmacology and polyspecificity

Polypharmacology is the ability of drugs to interact with multiple targets that generally belong to different activity classes [62-64]. Paolini *et al.* [65] have provided an interesting view of pharmacological space from an integrated structure-activity and polypharmacolgy point of view. Although until relatively recently the term was unknown, its influence has long been recognized in the form of the side effects and toxicities observed for virtually all drugs. Recently, it has also been the basis for the repurposing of a number of drugs for new therapeutic indications [66].

By combining experimental data and computational estimates, Mestres *et al.* [10] have suggested that on average nearly seven protein targets interact with each drug. Hu and Bajorath [67] have presented an excellent overall summary of polypharmacology data. Their work, which is based on the stringent requirement that drug-target interactions be determined from equilibrium measurements (*e.g.* K_{eq}'s) and not assay-dependent measurements (*e.g.* IC_{50}'s), states that polypharmacology values of two to three are most probable, although greater values are possible but less probable. Their data is roughly consistent with that reported by Jalencas and Mestres [68].

By contrast, polyspecificity, which has been shown to be the complement of polypharmacology [69], arises from the fact that multiple, structurally dissimilar drugs can interact with the same target. Although one might assume that such behavior would be unlikely due to the widely held belief that drug-target interactions are fairly specific, polyspecificity is nevertheless quite prevalent [70]. It is, in fact, the basis for the well-known practice of using 'multiple lead series' in lead optimization within the pharmaceutical industry [69,71]. While we are unaware of any published general, systematic study of polyspecificities, such studies could be of interest in drug discovery research. Because of the large amount of missing data due to the presence of drug-target pairs of unknown interaction, both polypharmacologies and polyspecificities are likely to be significantly underestimated.

Joint analogs of polypharmacologies and polyspecificities also exist, although they have been less well characterized than their singular counterparts. Joint polypharmacologies arise due to the number of targets that are interactive with respect to a pair of drugs, while joint polyspecificities arise due to the number of drugs that are interactive with respect to a given pair of targets. As shown in Section 6, the values of both of these entities, not surprisingly, cannot exceed that of their corresponding singular values.

Sections 5 and 6 show that missing data can to some extent be compensated for by determining error bounds on the polypharmacology and polyspecificity values. However, the error bounds are only useful when the number of null drug-target pairs is much less than the corresponding number of interactive pairs, otherwise the bounds become too large to be of practical value—this is not usually the case as seen by the data in Table 1. Vogt, *et al.* [72] have developed an automated approach for removing null pairs that holds considerable promise as a means for addressing this issue, although a side-effect of their method is the concurrent loss of some interactive pairs that may also be of interest in a given study.

A largely unappreciated issue that can also significantly influence the magnitude of the bounds on polypharmacologies and polyspecificities is the reluctance of many scientists, for a variety of reasons, to publish data on non-interactive drug-target pairs. Because of this, such pairs are counted in the class of null drug-target pairs, artificially increasing its size as well as the size of its corresponding error bound. Thus, it is important to include drug-target pairs with known, but small interaction values in drug-target databases and datasets. All of the above factors can influence the determination of bounds to polypharmacologies and polyspecificities. Nevertheless, such bounds may serve in some cases as suitable measures for these values. Such methods are described in Sections 5 and 6.

1.5 Overview

The material in the remainder of the paper is as follows: Section 2 provides a brief introduction to intuitionistic fuzzy sets (IFSs) and intuitionistic fuzzy relations (IFRs); Section 3 discusses classical set-theoretic drug-target relations that are the basis of many analyses including those based on bipartite drug-target networks; Section 4 describes the development of the extreme-value intuitionistic fuzzy sets ($\text{IFS}_{\mathbf{EV}}$'s) and relations ($\text{IFR}_{\mathbf{EV}}$'s) needed to handle typical threshold-based drug-target datasets; Section 5 presents an intuitionistic fuzzy set analysis of polypharmacologies and polyspecificities; Section 6 extends the analysis to joint polypharmacologies and polyspecificities and provides several examples of all four of these drug-target parameters; Section 7 presents a summary of the work and several conclusions derived from it.

2 Intuitionistic fuzzy sets (IFSs) and relations (IFRs)

2.1 General description

Intuitionistic fuzzy sets [55-61] are generalizations of the fuzzy sets developed by Lofti Zadeh [73]. As is the case for all types of finite sets, the reference or universal set (*a.k.a.* the Universe of Discourse), X, is given by

$$X = \left\{x_1, x_2, ..., x_n\right\}, \tag{2.1}$$

which is a classical, crisp set. In contrast to both classical, crisp sets and fuzzy sets, IFSs are represented by the set of ordered triples

$$\tilde{A} = \left\{ \left(\mu_{\tilde{A}}(x), \nu_{\tilde{A}}(x), \pi_{\tilde{A}}(x) \right) | \forall \ x \in X \right\}, \tag{2.2}$$

where $\mu_{\tilde{A}}(x) \in [0,1]$ is a function that indicates membership of the *x*-th element of *X* in $\tilde{A}$; $\nu_{\tilde{A}}(x) \in [0,1]$ is a new function that explicitly indicates non-membership of *x* in $\tilde{A}$; and $\pi_{\tilde{A}}(x) \in [0,1]$ is also a new function that provides a measure of the uncertainty due to a lack of knowledge of whether *x* does or does not belong to $\tilde{A}$. Although it has been suggested that IFSs are equivalent to interval-value fuzzy sets, it was shown that they are in fact unique and, moreover, are of practical value in many, varied applications [59].

The π-parameter, which is generally called an intuitionistic fuzzy index or hesitation margin, is treated here as a measure of 'knowledge-based uncertainty' such that $\pi_{\tilde{A}}(x) = 1$ if there is complete uncertainly regarding our knowledge of whether *x* does or does not belong to $\tilde{A}$ and $\pi_{\tilde{A}}(x) = 0$ if there is no uncertainty in our knowledge of whether *x* does or does not belong to $\tilde{A}$. The importance of considering the information contained in all three parameters in order to fully characterize IFSs has been addressed in two relatively recent publications [60,61]. Note that knowledge-based uncertainty is *not* the same as measurement-based uncertainty, which pervades many areas of science and engineering. The tilde ('~') over the set symbol indicates that it is an IFS and not a classical, crisp set. This notation will be used throughout the paper.

The values of the three characteristic functions satisfy

$$\mu_{\tilde{A}}(x) + \nu_{\tilde{A}}(x) + \pi_{\tilde{A}}(x) = 1 \tag{2.3}$$

for all $x \in X$. Thus, only two of the entities in Eq. (2.3) are independent. As will be seen in the sequel**,** the uncertainty function plays a key role in the analysis of drug-target databases since it deals with null drug-target pairs. If $\nu_{\tilde{A}}(x)$ is set equal to $1 - \mu_{\tilde{A}}(x)$, as is the case for fuzzy sets, $\pi_{\tilde{A}}(x) = 0$ for all $x \in X$ and $\tilde{A}$ becomes equivalent to a fuzzy set. The complement of an IFS is defined as

$$\tilde{A}^c = \left\{ \left(\nu_{\tilde{A}}(x), \mu_{\tilde{A}}(x), \pi_{\tilde{A}}(x) \right) | \forall \ x \in X \right\}, \tag{2.4}$$

where the first two terms in the ordered triple have been transposed.

2.2 Intuitionistic fuzzy set operations

The intersection of a pair of IFSs is given by

$$\tilde{A} \cap \tilde{B} = \left\{ \left(\mu_{\tilde{A} \cap \tilde{B}}(x), \nu_{\tilde{A} \cap \tilde{B}}(x), \pi_{\tilde{A} \cap \tilde{B}}(x) \right) | \forall \ x \in X \right\} \tag{2.5}$$

where

$$\left. \begin{array}{l} \mu_{\tilde{A} \cap \tilde{B}}(x) = \min \left[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x) \right] \\ \nu_{\tilde{A} \cap \tilde{B}}(x) = \max \left[\nu_{\tilde{A}}(x), \nu_{\tilde{B}}(x) \right] \\ \pi_{\tilde{A} \cap \tilde{B}}(x) = 1 - \mu_{\tilde{A} \cap \tilde{B}}(x) - \nu_{\tilde{A} \cap \tilde{B}}(x) \end{array} \right\} \tag{2.6}$$

and the expression for $\pi_{\tilde{A} \cap \tilde{B}}(x)$ in Eq. (2.6) follows from Eq. (2.3) since $\tilde{A} \cap \tilde{B}$ is also a set. The union of a pair of IFSs is given by

$$\tilde{A} \cup \tilde{B} = \left\{ \left(\mu_{\tilde{A} \cup \tilde{B}}(x), \nu_{\tilde{A} \cup \tilde{B}}(x), \pi_{\tilde{A} \cup \tilde{B}}(x) \right) | \forall \ x \in X \right\} \ , \tag{2.7}$$

where

$$\left. \begin{array}{l} \mu_{\tilde{A} \cup \tilde{B}}(x) = \max \left[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x) \right] \\ \nu_{\tilde{A} \cup \tilde{B}}(x) = \min \left[\nu_{\tilde{A}}(x), \nu_{\tilde{B}}(x) \right] \\ \pi_{\tilde{A} \cup \tilde{B}}(x) = 1 - \mu_{\tilde{A} \cup \tilde{B}}(x) - \nu_{\tilde{A} \cup \tilde{B}}(x) \end{array} \right\} \ , \tag{2.8}$$

and the expression for $\pi_{\tilde{A} \cup \tilde{B}}(x)$ in Eq. (2.8) again follows from Eq. (2.3). Note the reversal in the order of the max and min functions with respect to intersections and unions.

2.3 Cardinalities of intuitionistic fuzzy sets

Cardinalities of IFSs are based on the notion of 'sigma-counts' given by the function $\sum Count(\cdot)$, which was initially described by Lofti Zadeh for fuzzy sets [73]; sigma-counts play an important role in this work**.** Because IFSs are more complex than fuzzy sets determining their cardinalities is more complex as well [55,56]. In contrast to the case of fuzzy sets, the overall cardinality of an IFS is made up of several terms. The minimum value of the cardinality of $\tilde{A}$ is equal to its sigma-count, which is given by

$$\min \sum Count(\tilde{A}) = \sum_{x \in X} \mu_{\tilde{A}}(x) \ , \tag{2.9}$$

while the largest value of the sigma-count is given by

$$\max\sum Count(\tilde{A}) = \sum_{x\in X}\left[\mu_{\tilde{A}}(x) + \pi_{\tilde{A}}(x)\right] \ , \tag{2.10}$$

and is due to the fact that if the unknown data associated with the elements of $\pi_{\tilde{A}}(x)$ were actually available they would, in the most optimistic case, positively affect the values of individual membership functions, $\mu_{\tilde{A}}(x)$, and thus increase the sigma-count. Hence, the value of the cardinality of $\tilde{A}$ lies in an interval rather than at a single point, i.e.

$$\sum Count(\tilde{A}) \in \left[\min\sum Count(\tilde{A}), \max\sum Count(\tilde{A})\right] . \tag{2.11}$$

Now consider the sigma-count of the complement $\tilde{A}^c$. Because, as shown in Eq. (2.4), the first two terms in the ordered triple associated with $\tilde{A}^c$ are transposed from that of $\tilde{A}$ given in Eq. (2.2), the minimum sigma-count of the complement is defined by

$$\min\sum Count(\tilde{A}^c) = \sum_{x\in X}\nu_{\tilde{A}}(x) \ , \tag{2.12}$$

and the corresponding maximum sigma-count becomes

$$\max\sum Count(\tilde{A}^c) = \sum_{x\in X}\left[\nu_{\tilde{A}}(x) + \pi_{\tilde{A}}(x)\right] . \tag{2.13}$$

Thus, the sigma-count of $\tilde{A}^c$ also lies in an interval similar to that given in Eq. (2.11) for $\tilde{A}$, *i.e.*

$$\sum Count(\tilde{A}^c) \in \left[\min\sum Count(\tilde{A}^c), \max\sum Count(\tilde{A}^c)\right]. \tag{2.14}$$

Lastly, it can be shown that

$$\begin{aligned} \Delta\tilde{A} &= \max\sum Count(\tilde{A}) - \min\sum Count(\tilde{A}) \\ &= \max\sum Count(\tilde{A}^c) - \min\sum Count(\tilde{A}^c) \ . \\ &= \sum_{x\in X}\pi_{\tilde{A}}(x) \end{aligned} \tag{2.15}$$

□

Hence, the difference between the max and min counts of $\tilde{A}$ and $\tilde{A}^c$ are identical and are equal to the total knowledge-based uncertainty.

2.4 Intuitionistic fuzzy relations (IFRs)

Relations can be defined for IFSs in an analogous manner to that for classical, crisp sets given in Section 3. In the case of IFRs the reference or universal sets are *X*, given in Eq. (2.1), and *Y*, given in Eq. (2.16) below,

$$Y = \left\{ y_1, y_2, ..., y_m \right\} . \tag{2.16}$$

and the binary Cartesian product $X \times Y$ represents the universal set of the relation—technically these are binary relations as $X \times Y$ is a binary Cartesian product. Higher-order relations can be defined but are not addressed here. Thus, for simplicity the modifier 'binary' is omitted, as all relations employed in this work are binary relations. As illustrated by Eq. (2.17),

$$\tilde{R}_{\tilde{A},\tilde{B}} = \left\{ \left(\mu_{\tilde{R}}(x,y), \nu_{\tilde{R}}(x,y), \pi_{\tilde{R}}(x,y) \right) | \forall \ x \in X \text{ and } y \in Y \right\} . \tag{2.17}$$

Hence, IFRs involve two sets of ordered entities: (1) the usual ordered-pairs that are made up from the elements of $\tilde{A}$ and $\tilde{B}$, such that $(x,y) \in \tilde{R}_{\tilde{A},\tilde{B}}$ for all $x \in X$ and $y \in Y$,and (2) the ordered-triples associated with the respective membership, non-membership, and knowledge-based uncertainty functions, $\mu_{\tilde{A},\tilde{B}}(x,y)$, $\nu_{\tilde{A},\tilde{B}}(x,y)$, and $\pi_{\tilde{A},\tilde{B}}(x,y)$, all of whose values lie on the unit interval of the real line. As was the case for the IFSs shown in Eq. (2.3),

$$\mu_{\tilde{A},\tilde{B}}(x,y) + \nu_{\tilde{A},\tilde{B}}(x,y) + \pi_{\tilde{A},\tilde{B}}(x,y) = 1, \ \forall \ x,y \in X \times Y. \tag{2.18}$$

The relation $\tilde{R}_{\tilde{A},\tilde{B}}$ can also be represented as an $n \times m$-dimensional matrix of ordered triples,

$$\tilde{\mathbf{R}}_{\tilde{A},\tilde{B}} = \begin{pmatrix} \left(\mu_{\tilde{A},\tilde{B}}(x_1,y_1), \nu_{\tilde{A},\tilde{B}}(x_1,y_1), \pi_{\tilde{A},\tilde{B}}(x_1,y_1) \right) & \cdots & \left(\mu_{\tilde{A},\tilde{B}}(x_1,y_m), \nu_{\tilde{A},\tilde{B}}(x_1,y_m), \pi_{\tilde{A},\tilde{B}}(x_1,y_m) \right) \\ \vdots & \ddots & \vdots \\ \left(\mu_{\tilde{A},\tilde{B}}(x_n,y_1), \nu_{\tilde{A},\tilde{B}}(x_n,y_1), \pi_{\tilde{A},\tilde{B}}(x_n,y_1) \right) & \cdots & \left(\mu_{\tilde{A},\tilde{B}}(x_n,y_m), \nu_{\tilde{A},\tilde{B}}(x_n,y_m), \pi_{\tilde{A},\tilde{B}}(x_n,y_m) \right) \end{pmatrix}. \tag{2.19}$$

3 Classical, crisp relations describing drug-target interactions

The set of drug-target pairs is composed of the set of *n* drugs,

$$D = \left\{ d_1, d_2, ..., d_i, ..., d_n \right\} \tag{3.1}$$

and *m* targets

$$T = \left\{ t_1, t_2, ..., t_j, ..., t_m \right\}, \tag{3.2}$$

which are typically represented by classical, set-theoretic relations, where $\left(d_i, t_j\right) \in R_{DT} \subseteq D \times T$, (d_i, t_j) are ordered-pairs of the elements of D and T, and $D \times T$ is their Cartesian product. Set-theoretic relations are typically represented as $n \times m$-dimensional relational matrices (*Cf.* Eq. (2.19))

$$\mathbf{R}_{D,T} = \begin{pmatrix} \chi_{D,T}(d_1, t_1) & \chi_{D,T}(d_1, t_2) & \cdots & \chi_{D,T}(d_1, t_m) \\ \chi_{D,T}(d_2, t_1) & \chi_{D,T}(d_2, t_2) & \cdots & \chi_{D,T}(d_2, t_m) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_{D,T}(d_n, t_1) & \chi_{D,T}(d_n, t_2) & \cdots & \chi_{D,T}(d_n, t_m) \end{pmatrix}, \tag{3.3}$$

each of whose elements are given by the characteristic function $\chi_{D,T}(d_i, t_j)$, which satisfies

$$\chi_{D,T}(d_i, t_j) = \begin{cases} 1 & \text{if the } i\text{-th the drug interacts with the } j\text{-th target} \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

for $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$. Thus, $\mathbf{R}_{D,T}$ is a binary-valued matrix. Summing the elements of the *i*-th row will yield the *polypharmacology* of the i-th drug, d_i, while summing the elements of the *j*-th column will yield the *polyspecificity* of the j-th target, t_j [40,41,46]. Joint analogs of polypharmacologies and polyspecificities, which involve pairs of drugs and targets also exist and are treated in Section 6.

If $\mathbf{R}_{D,T}$ provides an accurate representation of drug-target interactions, each row represents the interaction of the drug associated with that row with respect to each target in the set of targets—what can nominally be called a 'target-based drug interaction profile' or '***drug profile***' for short. A given column represents the activity of the target associated with that column with respect to each drug in the set of drugs—what can nominally be called a 'drug-based target interaction profile' or '***target profile***' for short.

While this is fairly straightforward, it neglects an important aspect of the problem, namely, that many elements of $\mathbf{R}_{D,T}$ correspond to null drug-target pairs whose activities have neither been experimentally determined nor computationally estimated. Although it is inappropriate to assume that a given null pair is inactive, it is implicitly done in many instances. Assuming that to be the case can significantly bias any analysis based on drug-target activities (*vide supra*). Since, as shown in Eq. (3.4), classical, crisp relations only allow the possibility that an element is either a member or not a member of a given relation, they do not have the capability to faithfully represent drug-target interactions as they are unable to represent null pairs (Note, as discussed in Section 1.3, that ternary and higher-order classical, set-theoretic

relations can extend this concept). Table 1 shows that such pairs generally constitute by far the largest number of elements in drug-target relations. Hence, drug-target relations are typically sparsely populated by interactive and non-interactive pairs, which constitute only about one to three percent of the number of possible pairs. Although for some databases the percent of interactive and non-interactive pairs may reach 20 percent, this represents an upper bound in essentially all cases. Thus, a new, more general formalism is needed that can address the crucial issue of null pairs. As we shall see, the IFSs described in the previous section provide a suitable mathematical framework for accomplishing this task, but an adaptation is needed to handle drug-target datasets, as explained in the following section.

4 Extreme-value intuitionistic fuzzy sets and relations

Because drug-target pairs are either 'in', 'not in', or of 'indeterminant' membership their corresponding membership, non-membership, and knowledge-based uncertainty functions are restricted to the set of binary values, $\{0,1\}$. The fact that these values represent extremes is indicated by the subscript 'EV' so that IFSs and IFRs are now designated by IFS$_{\mathbf{EV}}$ and IFR$_{\mathbf{EV}}$, respectively. In order to clearly distinguish normal intuitionistic fuzzy sets and relations from their extreme-value counterparts, the latter are labeled with a bar '–' so that $\tilde{A}$ and $\tilde{B}$ become $\overline{A}$ and $\overline{B}$ and $\tilde{\mathbf{R}}_{\tilde{A},\tilde{B}}$ becomes $\overline{\mathbf{R}}_{\overline{A},\overline{B}}$. As we are explicitly dealing with drug-target datasets in this work, $\overline{A}$ and $\overline{B}$ will be replaced by $\overline{D}$ and $\overline{T}$, respectively, and $\overline{\mathbf{R}}_{\overline{A},\overline{B}}$ by $\overline{\mathbf{R}}_{\overline{D},\overline{T}}$.

4.1 Decomposing IFR$_{\mathbf{EV}}$'s to IFS$_{\mathbf{EV}}$'s

As we are interested in drug-target interactions in this work, Eq. (2.27) becomes

$$\overline{\mathbf{R}}_{\overline{D},\overline{T}} = \begin{pmatrix} \left(\mu_{\overline{D},\overline{T}}(d_1,t_1),\nu_{\overline{D},\overline{T}}(d_1,t_1),\pi_{\overline{D},\overline{T}}(d_1,t_1)\right) & \cdots & \left(\mu_{\overline{D},\overline{T}}(d_1,t_m),\nu_{\overline{D},\overline{T}}(d_1,t_m),\pi_{\overline{D},\overline{T}}(d_1,t_m)\right) \\ \vdots & \ddots & \vdots \\ \left(\mu_{\overline{D},\overline{T}}(d_n,t_1),\nu_{\overline{D},\overline{T}}(d_n,t_1),\pi_{\overline{D},\overline{T}}(d_n,t_1)\right) & \cdots & \left(\mu_{\overline{D},\overline{T}}(d_n,t_m),\nu_{\overline{D},\overline{T}}(d_n,t_m),\pi_{\overline{D},\overline{T}}(d_n,t_m)\right) \end{pmatrix}. \tag{4.1}$$

Each row of $\overline{\mathbf{R}}_{\overline{D},\overline{T}}$ contains the set of interactions of a given drug with respect to each of the targets in the reference set T, and hence represent drug profiles, which are given by

$$\overline{D}_i = \left\{\left(\mu_{\overline{D},\overline{T}}(d_i,t_1),\nu_{\overline{D},\overline{T}}(d_i,t_1),\pi_{\overline{D},\overline{T}}(d_i,t_1)\right)\cdots\left(\mu_{\overline{D},\overline{T}}(d_i,t_m),\nu_{\overline{D},\overline{T}}(d_i,t_m),\pi_{\overline{D},\overline{T}}(d_i,t_m)\right)\right\} \tag{4.2}$$

for $i = 1,2,...,n$. This is exemplified by the rows of drug-target data in Table 2. The terms within the parentheses in the right most column of the table correspond to $\min Count(\overline{D}_i)$, $\min Count(\overline{D}_i^c)$, and $\Delta\overline{D}_i$, given respectively by Eqs. (2.9), (2.12), and (2.15). Their sum is

equal to $m = \left| T \right|$. Correspondingly, each column of $\overline{\mathbf{R}}_{\bar{D},\bar{T}}$ contains the set of activities of a given target with respect to each of the drugs in the reference set D, and hence represent target profiles, which are given by

$$\overline{T}_j = \left\{ \left(\mu_{\bar{D},\bar{T}}(d_1,t_j), \nu_{\bar{D},\bar{T}}(d_1,t_j), \pi_{\bar{D},\bar{T}}(d_1,t_j) \right) \cdots \left(\mu_{\bar{D},\bar{T}}(d_n,t_j), \nu_{\bar{D},\bar{T}}(d_n,t_j), \pi_{\bar{D},\bar{T}}(d_n,t_j) \right) \right\} \quad (4.3)$$

for $j = 1,2,...,m$. This is exemplified by the columns of drug-target data in Table 2. The elements in parentheses in the bottom row of the table correspond to $\min Count(\overline{T}_j)$, $\min Count(\overline{T}_j^c)$, and $\Delta \overline{T}_j$, given respectively by Eqs. (2.9), (2.12), and (2.15). Their sum is equal to $n = \left| D \right|$.

Because the values of the membership, non-membership, and knowledge-based uncertainty functions for each of the elements of the IFS$_{\mathbf{EV}}$'s and IFR$_{\mathbf{EV}}$'s all lie in $\{0,1\}$, and because they satisfy Eqs. (2.3) and (2.19), if one of the values for a given ordered-triple is equal to unity the other two must perforce each be equal to zero, as is clear from the ordered-triples in Table 2. This relationship can be written in logical form as

$$\begin{gathered} \mu_{\bar{D}_i}(d_i,t_k) \veebar \nu_{\bar{D}_i}(d_i,t_k) \veebar \pi_{\bar{D}_i}(d_i,t_k) = 1 \\ \forall \ d_i \in D \text{ and } t_k \in T \end{gathered}, \quad (4.4)$$

where $\veebar$ is the symbol for 'exclusive-or'. It provides a means for partitioning drug and target profiles into three non-intersecting subsets such that for drug profiles

$$\begin{aligned} &\overline{D}_i^{\mu} \cap \overline{D}_i^{\nu} = \overline{D}_i^{\mu} \cap \overline{D}_i^{\pi} = \overline{D}_i^{\nu} \cap \overline{D}_i^{\pi} = \varnothing \\ &\overline{D}_i = \overline{D}_i^{\mu} \cup \overline{D}_i^{\nu} \cup \overline{D}_i^{\pi} \end{aligned} \quad (4.5)$$

for all $D_i \subseteq D$ and for target profiles

$$\begin{aligned} &\overline{T}_j^{\mu} \cap \overline{T}_j^{\nu} = \overline{T}_j^{\mu} \cap \overline{T}_j^{\pi} = \overline{T}_j^{\nu} \cap \overline{T}_j^{\pi} = \varnothing \\ &\overline{T}_j^{\mu} \cup \overline{T}_j^{\nu} \cup \overline{T}_j^{\pi} = \overline{T}_j \end{aligned} \quad (4.6)$$

For all $T_j \subseteq T$.

As is clear from Table 2, the elements in each row, which constitute a drug profile, are partitioned. For example, consider the fifth row where

$$\begin{aligned}\overline{D}_5^{\mu} &= \left\{(d_5,t_2),(d_5,t_3),(d_5,t_5)\right\}\\ \overline{D}_5^{\nu} &= \left\{(d_5,t_1),(d_5,t_6),(d_5,t_8)\right\}\\ \overline{D}_5^{\pi} &= \left\{(d_5,t_4),(d_5,t_7)\right\}\end{aligned} . \tag{4.7}$$

Clearly the elements of $\overline{D}_5 = \left\{\overline{D}_5^{\mu}, \overline{D}_5^{\nu}, \overline{D}_5^{\pi}\right\}$ form a partition since $\overline{D}_5^{\mu} \cap \overline{D}_5^{\nu} = \overline{D}_5^{\mu} \cap \overline{D}_5^{\pi} = \overline{D}_5^{\nu} \cap \overline{D}_5^{\pi} = \varnothing$ and $\overline{D}_5 = \overline{D}_5^{\mu} \cup \overline{D}_5^{\nu} \cup \overline{D}_5^{\pi}$.

The above discussion shows that $\mathbf{R}_{\overline{D},\overline{T}}$ can be decomposed into a family of drug profiles

$$\mathbf{R}_{\overline{D},\overline{T}} = \bigcup_{i=1}^{n} \overline{D}_i \ . \tag{4.7}$$

Similar arguments show that $\mathbf{R}_{\overline{D},\overline{T}}$ can also be decomposed into a family of target profiles

$$\mathbf{R}_{\overline{D},\overline{T}} = \bigcup_{k=1}^{m} \overline{T}_k \ . \tag{4.8}$$

Table 2. Extreme-value intuitionistic fuzzy relations. The rows correspond to drug profiles and the columns to target profiles. The terms in parentheses in the right most column are described in Section 4 and are given by Eqs. (2.9), (2.12), and (2.15), respectively. The corresponding terms in the bottom row are described in Section 4 and are also given by the same equations.

	$\overline{T}_1$	$\overline{T}_2$	$\overline{T}_3$	$\overline{T}_4$	$\overline{T}_5$	$\overline{T}_6$	$\overline{T}_7$	$\overline{T}_8$	*Counts*
$\overline{D}_1$	(0,0,1)	(1,0,0)	(0,1,0)	(1,0,0)	(0,0,1)	(1,0,0)	(0,0,1)	(1,0,0)	{4,1,3}
$\overline{D}_2$	(1,0,0)	(1,0,0)	(0,1,0)	(0,0,1)	(0,1,0)	(1,0,0)	(0,1,0)	(1,0,0)	{4,3,1}
$\overline{D}_3$	(0,1,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(0,0,1)	(1,0,0)	(0,1,0)	{5,2,1}
$\overline{D}_4$	(0,0,1)	(1,0,0)	(0,1,0)	(1,0,0)	(0,0,1)	(0,0,1)	(0,1,0)	(1,0,0)	{3,2,3}
$\overline{D}_5$	(0,1,0)	(1,0,0)	(1,0,0)	(0,0,1)	(1,0,0)	(0,1,0)	(0,0,1)	(0,1,0)	{3,3,2}
$\overline{D}_6$	(1,0,0)	(1,0,0)	(0,1,0)	(0,0,1)	(0,1,0)	(1,0,0)	(0,0,1)	(0,1,0)	{3,3,2}
$\overline{D}_7$	(0,1,0)	(0,0,1)	(0,0,1)	(0,0,1)	(1,0,0)	(0,1,0)	(1,0,0)	(0,0,1)	{2,2,4}
$\overline{D}_8$	(0,0,1)	(0,0,1)	(0,1,0)	(1,0,0)	(0,0,1)	(0,0,1)	(0,1,0)	(1,0,0)	{2,2,4}
$\overline{D}_9$	(1,0,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,1,0)	(1,0,0)	(0,1,0)	(1,0,0)	{3,2,3}
$\overline{D}_{10}$	(1,0,0)	(0,0,1)	(0,1,0)	(0,0,1)	(1,0,0)	(1,0,0)	(0,1,0)	(0,1,0)	{3,3,2}
$\overline{D}_{11}$	(0,0,1)	(1,0,0)	(0,0,1)	(1,0,0)	(0,1,0)	(0,1,0)	(1,0,0)	(0,0,1)	{3,2,3}
$\overline{D}_{12}$	(1,0,0)	(1,0,0)	(0,1,0)	(0,0,1)	(0,0,1)	(0,1,0)	(0,0,1)	(1,0,0)	{3,2,3}
Counts	{5,3,4}	{8,0,4}	{2,7,3}	{5,0,7}	{4,4,4}	{6,4,2}	{4,5,3}	{7,2,3}	—

5 Polypharmacologies and polyspecificities

5.1 Polypharmacologies

As noted in the introduction, polypharmacology [62-64] has taken on an increased importance in drug research, due to the fact that it is largely responsible for side effects and provides a basis for the repositioning of existing drugs for new therapeutic indications [66]. Although they are less well known, polyspecificities have appeared in a number of phases of discovery research, albeit under different guises, *e.g.* as 'multiple lead series' in lead optimization studies [69,71]. Hence, determining the values of both of these parameters may be of importance in a number of areas of drug research.

As shown by Maggiora and Gokhale [54,69], polypharmacologies and polyspecificities are mathematically related by the elements of the drug-target interaction matrix given by Eq. (3.3). And, as we shall see, they are also related through the corresponding elements of the extreme-value drug-target matrix, Eq. (4.1). If, however, drug-target pairs are missing from the set of interactive pairs, that is if the set of interactive pairs is incomplete, then both the polypharmacology and polyspecificity values will generally, but not in all cases, be less than their true values. The degree that they are in error will, however, not necessarily be the same for both entities. For example, a given polypharmacology value could be 33 percent in error, while its corresponding polyspecificity value may only be 17 percent in error, it all depends on what drug-target pairs are present in the set of interactive pairs. In addition, the presence of substantial numbers of null pairs can impose additional difficulties in the use of classical set-theoretic methods (*vide supra*). IFR$_{\mathbf{EV}}$'s can, under certain conditions, provide a means for overcoming many of these difficulties.

The model data in Table 2 affords a simple example of an IFR$_{\mathbf{EV}}$, where each cell of the table is represented by a triple of elements. Because we are dealing with IFR$_{\mathbf{EV}}$'s the value of each element in a given triple lies in the binary set $\{0,1\}$ with the constraint that all three elements must sum to unity as required by Eq. (2.18). As noted earlier, when one of the terms has a value of unity the other two must perforce be of value zero. The advantage of this formalism is that it not only provides a means for dealing with interactive and non-interactive pairs, but, more importantly, for also dealing with null pairs. The latter can present significant problems in the analysis of many real drug-target databases as illustrated by the following example.

The minimum value of the polypharmacology of the *i*-th drug is given by the min-count of the elements of the *i*-th drug profile, $\bar{D}_i$, associated with the IFR$_{\mathbf{EV}}$ in Table 2. Hence,

$$\mathcal{P}(d_i)_{\min} = \min Count\left(\bar{D}_i\right) = \sum_{j=1}^{m} \mu_{\bar{D}_i}(d_i, t_j), \quad \forall\ d_i \in D\ , \tag{5.1}$$

which is given by the first term in curly brackets in the far-right column of Table 2. Since the values of the elements of the extreme-value IFS_{EV}'s and IFR_{EV}'s are binary integers it is no longer necessary to use sigma-counts. Thus, they are replaced by the simpler *Count* function, where the '$\sum$' is removed so that $\sum Count(\cdot)$ becomes $Count(\cdot)$ for Eqs. (2.9) through (2.14).

Consider, for example, the first drug, d_1, in the list of drugs in Table 2. Its polypharmacology is obtained by substituting the values in the leftmost term in each of the parentheses in the first row of Table 2 into Eq. (5.1). This gives a value of '4' as indicated by the corresponding term within the curly brackets found in the right most cell in the first row of the table. The comparable term in each row of the table gives the polypharmacology value for each of the fictitious drugs, d_i, where $i = 1,2,...,12$. The third term within the curly brackets located in the right most column of Table 2 corresponds to the count of null pairs with respect to the drug associated with each row of the table. As an example, again consider the first row of Table 2, which gives a value of '3' for the number of drug-target pairs of unknown interaction, a value that can be obtained by summing the third term in each of the parentheses in the first row of the table.

As shown by Szmidt and Baldwin [74] (*Cf.* [55,56]), an upper bound to $\mathcal{P}(d_i)_{\min}$ can be obtained by computing

$$\mathcal{P}(d_i)_{\max} = \max Count(\bar{D}_i) = \sum_{j=1}^{m} \left[\mu_{\bar{D}_i}(d_i, t_j) + \pi_{\bar{D}_i}(d_i, t_j) \right], \ \forall d_i \in D, \tag{5.2}$$

where it is tacitly assumed that null drug-target pairs are considered to be interactive, such that $\pi_{\bar{D},\bar{T}}\left(d_i, t_j\right) = 1 \rightarrow \mu_{\bar{D},\bar{T}}\left(d_i, t_j\right) = 1$. Since, only a fraction of the pairs of unknown interaction are apt to actually interact, the upper bound is likely to be a considerable overestimate. In any case, the value of the polypharmacology lies in an interval, which may also include its endpoints, rather than at a single value, *i.e.*

$$\mathcal{P}(d_i) \in \left[\mathcal{P}(d_i)_{\min}, \mathcal{P}(d_i)_{\max} \right], \ \forall \ d_i \in D \ , \tag{5.3}$$

where the size of the interval is equal to the knowledge-based uncertainty in the polypharmacology value,

$$\Delta \mathcal{P}(d_i) = \sum_{j=1}^{m} \pi_{\bar{D}_i}(d_i, t_j), \ \forall \ d_i \in D \ , \tag{5.4}$$

(*Cf.* Eq. (2.15)).

A largely unappreciated issue that can also influence the magnitude of the bounds on polypharmacology values is the reluctance of many scientists, for a variety of reasons, to publish data on non-interactive drug-target pairs. Because of this, such pairs are necessarily counted in the class of null drug-target pairs. However, Eq. (5.5), which can be obtained by rearranging Eq. (2.18) and substituting it into Eq. (5.4),

$$\begin{aligned}\Delta\mathcal{P}(d_i) &= \sum_{j=1}^{m}\left[1-\mu_{\bar{D}_i}(d_i,t_j)-\nu_{\bar{D}_i}(d_i,t_j)\right] \\ &= m-\mathcal{P}(d_i)-\sum_{j=1}^{m}\nu_{\bar{D}_i}(d_i,t_j) \quad , \\ &\forall\ d_i \in D\end{aligned} \tag{5.5}$$

shows that the presence of non-interactive pairs, $\nu_{\bar{D}_i}\left(d_i,t_j\right)$, can significantly reduce the interval of knowledge-based uncertainty. This follows since m and $\mathcal{P}(d_i)$ are constant, thus $\Delta\mathcal{P}(d_i)$ is determined by the number of non-interactive drug-target pairs associated with the *i*-th drug, $\sum_{j=1}^{m}\nu_{\bar{D}_i}\left(d_i,t_j\right)$. Not reporting non-interactive drug-target pairs increases the interval of uncertainty in the value of the polypharmacology, as should indeed be expected. This emphasizes the importance of reporting all drug-target interaction data, even those that correspond to non or weakly interactive pairs. Lastly, in the rare case that there are no null pairs for d_i, $\pi_{\bar{D},\bar{T}}\left(d_i,t_j\right)=0$ for $j=1,2,...,m$, and the interval collapses so that $\mathcal{P}(d_i)=\mathcal{P}(d_i)_{\min}=\mathcal{P}(d_i)_{\max}$ for $i=1,2,...,n$.

5.2 Polyspecificities

Comparable expressions for polyspecificities, which involve the columns rather than the rows of Table 2, are given in Eqs. (5.6) to (5.8):

$$\mathcal{S}(t_j)_{\min}=\min Count(\bar{T}_j)=\sum_{i=1}^{n}\mu_{\bar{T}_j}(d_i,t_j),\ \forall\ t_j\in T, \tag{5.6}$$

$$\mathcal{S}(t_j)_{\max}=\max Count(\bar{T}_j)=\sum_{i=1}^{n}\left[\mu_{\bar{T}_j}(d_i,t_j)+\pi_{\bar{T}_j}(d_i,t_j)\right],\ \forall\ t_j\in T\ , \tag{5.7}$$

and

$$\mathcal{S}(t_j)\in\left[\mathcal{S}(t_k)_{\min},\,\mathcal{S}(t_k)_{\max}\right],\ \forall\ t_j\in T \tag{5.8}$$

In the case polyspecificities, the size of the interval is equal to the knowledge-based uncertainty in the polyspecificity,

$$\Delta S(t_j) = \sum_{i=1}^{n} \pi_{\bar{T}_j}(d_i, t_j)\,, \;\; \forall \; t_j \in T\,. \tag{5.9}$$

(*Cf.* Eq. (2.15)).

As was described for polypharmacology values in the previous section, the presence of non-interactive drug-target pairs reduces the interval of knowledge-based uncertainty for polyspecificity values as shown by Eq. (5.9), which is identical in form to that given in Eq. (5.5) for polypharmacology values. Hence, for all $t_j \in T$

$$\begin{aligned} \Delta S(t_j) &= \sum_{i=1}^{n} \left[1 - \mu_{\bar{T}_j}(d_i, t_j) - \nu_{\bar{T}_j}(d_i, t_j) \right] \\ &= n - S(t_j) - \sum_{i=1}^{n} \nu_{\bar{T}_j}(d_i, t_j) \end{aligned} \tag{5.10}$$

The same argument applies in this case as well, and the size of the interval, $\Delta S(t_j)$, is determined by the number of inactive drug-target pairs associated with the *j*-th target, $\sum_{i=1}^{n} \nu_{\bar{T}_j}\left(d_i, t_j\right)$.

Again, in the rare case that there are no null pairs in the column associated with a specific t_j, $\pi_{\bar{D},\bar{T}}\left(d_i, t_j\right) = 0$ for $i = 1, 2, \ldots, n$, the interval collapses so that $S(t_j) = S(t_j)_{\min} = S(t_j)_{\max}$ for $j = 1, 2, \ldots, m$.

5.3 Examples of polypharmacologies and polyspecificities

The maxima associated with polypharmacologies, given in Eq. (5.2), and polyspecificities, in Eq. (5.6), represent extremes since they assume that all of the null drug-target pairs are interactive, an assumption, as noted earlier, that leads to overly optimistic estimates of the of these quantities. Table 3 gives the max and min values of the polypharmacologies and polyspecificities obtained from the data in Table 2.

Although the naive examples described here illustrate a number of features of our approach, they do not adequately capture the difficulties brought about by the distribution of interactive and non-interactive pairs in real drug-target databases since in many cases the activity or lack thereof of drug-target pairs has only been experimentally measured or computationally estimated for less than 20 percent of the possible pairs in any dataset and in some cases less than five percent (*Cf.* Table 1). In any case, the percent of null pairs generally far exceeds the percent of interactive and non-interactive pairs. Although this may appear to be extreme, it is not. As noted earlier, most drug-target databases are, in fact, plagued by a

significant lack of data completeness [9], which gives rise to upper bounds of polypharmacologies and polyspecificities that are much too large to be of practical use. For these bounds to be useful the percentage of known interactive drug-target pairs should be significantly greater than the number of null pairs, a condition that is rarely met in practice.

Vogt, *et al*. [72] have developed a method for obtaining drug-target matrices with a minimal number of null pairs, but their method may also remove some pairs of interest. By combining their method with the one described here it may be possible to retain a significant portion of the desired pairs along with a reduced number of null pairs so that improved estimates of upper bounds to polypharmacologies and polyspecificities can be obtained. In such instances, the estimated bounds may be tight enough so that the values of these two entities will be suitable for the further analysis of drug-target data.

Table 3. Interval values and knowledge-based uncertainties of polypharmacologies and polyspecificities are obtained from the data in Table 2. Polypharmacology values lie in the interval defined by corresponding their min and max values defined by Eq. (5.3), and the knowledge-based uncertainty is the difference between these two values as given by Eq. (5.4). Polyspecificity values lie in the interval defined by corresponding their min and max values defined by Eq. (5.7), and the knowledge-based uncertainty is the difference between these two values as given by Eq. (5.8).

Polypharmacologies		
Drug	$[\mathcal{P}_{\min}, \mathcal{P}_{\max}]$	$\Delta\mathcal{P}$
1	[4,7]	3
2	[4,5]	1
3	[5,6]	1
4	[3,6]	3
5	[3,5]	2
6	[3,5]	2
7	[2,6]	4
8	[2,6]	4
9	[3,6]	3
10	[3,5]	2
11	[3,6]	3
12	[3,6]	3

Polyspecificities		
Target	$[S_{\min}, S_{\max}]$	ΔS
1	[5,9]	4
2	[8,12]	4
3	[2,5]	3
4	[5,12]	7
5	[4,8]	4
6	[6,8]	2
7	[4,7]	3
8	[7,10]	3

6 Joint polypharmacologies and polyspecificities

6.1 Joint polypharmacologies

Joint analogs of polypharmacologies and polyspecificities can also be defined. The minimum value of joint polypharmacologies is given by the number of common targets that are known to interact with the (i,j)-th pair of drugs, *i.e.*

$$\begin{aligned}\mathcal{P}(d_i,d_j)_{\min} &= \min Count\left(\bar{D}_i \cap \bar{D}_j\right) \\ &= \sum_{k=1}^{m} \min\left[\mu_{\bar{D}_i}(d_i,t_k),\, \mu_{\bar{D}_j}(d_j,t_k)\right]. \qquad (6.1)\\ \text{for } i &= 1,2,...,n \text{ and } j = i+1,...,n\end{aligned}$$

As set intersections are symmetric, only terms in the upper triangle of the joint polypharmacology and joint polyspecificity matrices are considered as those in the lower triangle of these matrices are equivalent. Diagonal terms are also omitted because they are equivalent to the polypharmacologies and polyspecificities discussed in Section 5. Hence, there are $n(n-1)/2$ unique joint polypharmacologies. A related definition of joint polypharmacology was given by Paolini, et al. [43]. Because

$$\begin{aligned}\sum_{k=1}^{m} \min\left[\mu_{\bar{D}_i}(d_i,t_k),\, \mu_{\bar{D}_j}(d_j,t_k)\right] &\le \begin{cases}\sum_{k=1}^{m}\mu_{\bar{D}_i}(d_i,t_k) \\ \sum_{k=1}^{m}\mu_{\bar{D}_i}(d_j,t_k)\end{cases} \qquad (6.2)\\ \text{for } i &= 1,2,...,n \text{ and } j = i+1,...,n\end{aligned}$$

it follows that

$$\begin{aligned}\mathcal{P}(d_i,d_j)_{\min} &\le \begin{cases}\mathcal{P}(d_i)_{\min} \\ \mathcal{P}(d_j)_{\min}\end{cases}. \qquad (6.3)\\ \text{for } i &= 1,2,...,n \text{ and } j = i+1,...,n\end{aligned}$$

The *maximum* values of the joint polypharmacologies are given by the number of targets that are known to interact with the (i,j)-th pair of drugs plus the number of targets whose interaction with these drugs is unknown, *i.e.*

$$\begin{aligned}\mathcal{P}(d_i,d_j)_{\max} &= \max Count\left(\bar{D}_i \cap \bar{D}_j\right) \\ &= \sum_{k=1}^{m}\left[\mu_{\bar{D}_i}(d_i,t_k)+\pi_{\bar{D}_i}(d_i,t_k)\right]\wedge\left[\mu_{\bar{D}_j}(d_j,t_k)+\pi_{\bar{D}_i}(d_j,t_k)\right]. \\ &\text{for } i=1,2,...,n \text{ and } j=i+1,...,n\end{aligned} \tag{6.4}$$

In a similar fashion to the case of the polypharmacology of single drugs, this *assumes* that all of the drug-target pairs of unknown interaction are now considered to be interactive, and hence all of the pairs of unknown interaction no longer exist. In analogy to Eqs. (6.2) and (6.3), it follows that

$$\begin{aligned}\sum_{k=1}^{m}\left[\mu_{\bar{D}_i}(d_i,t_k)+\pi_{\bar{D}_i}(d_i,t_k)\right]\wedge\left[\mu_{\bar{D}_j}(d_j,t_k)+\pi_{\bar{D}_i}(d_j,t_k)\right] \le \begin{cases}\sum_{k=1}^{m}\left[\mu_{\bar{D}_i}(d_i,t_k)+\pi_{\bar{D}_i}(d_i,t_k)\right] \\ \sum_{k=1}^{m}\left[\mu_{\bar{D}_j}(d_j,t_k)+\pi_{\bar{D}_i}(d_j,t_k)\right]\end{cases} \\ \text{for } i=1,2,...,n \text{ and } j=i+1,...,n\end{aligned} \tag{6.5}$$

so that

$$\begin{aligned}\mathcal{P}(d_i,d_j)_{\max} \le \begin{cases}\mathcal{P}(d_i)_{\max} \\ \mathcal{P}(d_j)_{\max}\end{cases} . \\ \text{for } i=1,2,...,n \text{ and } j=i+1,...,n\end{aligned} \tag{6.6}$$

Also, since

$$\begin{aligned}\sum_{k=1}^{m}\left[\mu_{\bar{D}_i}(d_i,t_k)\wedge\mu_{\bar{D}_j}(d_j,t_k)\right] \le \sum_{k=1}^{m}\left[\mu_{\bar{D}_i}(d_i,t_k)+\pi_{\bar{D}_i}(d_i,t_k)\right]\wedge\left[\mu_{\bar{D}_j}(d_j,t_k)+\pi_{\bar{D}_j}(d_j,t_k)\right] \\ \text{for } i=1,2,...,n \text{ and } j=i,i+1,...,n\end{aligned} \tag{6.7}$$

it follows that, as expected,

$$\begin{aligned}\mathcal{P}\left(d_i,d_j\right)_{\min} \le \mathcal{P}\left(d_i,d_j\right)_{\max} . \\ \text{for } i=1,2,...,n \text{ and } j=i,i+1,...,n\end{aligned} \tag{6.8}$$

Again, as was the case for polypharmacologies in Eq. (5.3), the values of joint polypharmacologies are also given by an interval,

$$\mathcal{P}(d_i,d_j)\in\left[\mathcal{P}(d_i,d_j)_{\min},\mathcal{P}(d_i,d_j)_{\max}\right] \quad \text{for } i=1,2,\ldots,n \text{ and } j=i+1,i+2,\ldots,n \tag{6.9}$$

6.2 Examples of joint polypharmacologies

These relationships are exemplified by the examples in Table 4. $\bar{D}_1$, $\bar{D}_2$, $\bar{D}_3$, $\bar{D}_5$, $\bar{D}_9$, and $\bar{D}_{10}$ all correspond to single drug profiles, while $\bar{D}_1\cap\bar{D}_2$, $\bar{D}_3\cap\bar{D}_5$, and $\bar{D}_9\cap\bar{D}_{10}$ correspond to joint drug profiles. The first term in square brackets in the right most column of the table gives the minimum values of the polypharmacologies or joint polypharmacologies, and the second term corresponds to their maximum values—their 'true' values lie in these intervals, which could include their endpoints. An examination of the table shows that the results are consistent with the relationships described in Eqs. (6.3) and (6.6).

Table 4. Example of IFSEV's taken from Table 2. The rows correspond to drug profiles and the columns to individual targets. The far-right column corresponds interval defined by the min and max polypharmacology values.

Drug Profiles	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	$\left[\mathcal{P}_{\min},\mathcal{P}_{\max}\right]$
$\bar{D}_1$	(0,0,1)	(1,0,0)	(0,1,0)	(1,0,0)	(0,0,1)	(1,0,0)	(0,0,1)	(1,0,0)	[4,7]
$\bar{D}_2$	(1,0,0)	(1,0,0)	(0,1,0)	(0,0,1)	(0,1,0)	(1,0,0)	(0,1,0)	(1,0,0)	[4,5]
$\bar{D}_1\cap\bar{D}_2$	(0,0,1)	(1,0,0)	(0,1,0)	(0,0,1)	(0,1,0)	(1,0,0)	(0,1,0)	(1,0,0)	[3,5]
$\bar{D}_3$	(0,1,0)	(1,0,0)	(1,0,0)	(1,0,0)	(1,0,0)	(0,0,1)	(1,0,0)	(0,1,0)	[5,6]
$\bar{D}_5$	(0,1,0)	(1,0,0)	(1,0,0)	(0,0,1)	(1,0,0)	(0,1,0)	(0,0,1)	(0,1,0)	[3,5]
$\bar{D}_3\cap\bar{D}_5$	(0,1,0)	(1,0,0)	(1,0,0)	(0,0,1)	(1,0,0)	(0,1,0)	(0,1,0)	(0,1,0)	[3,4]
$\bar{D}_9$	(1,0,0)	(0,0,1)	(0,0,1)	(0,0,1)	(0,1,0)	(1,0,0)	(0,1,0)	(1,0,0)	[3,6]
$\bar{D}_{10}$	(1,0,0)	(0,0,1)	(0,1,0)	(0,0,1)	(1,0,0)	(1,0,0)	(0,1,0)	(0,1,0)	[3,5]
$\bar{D}_9\cap\bar{D}_{10}$	(1,0,0)	(0,0,1)	(0,1,0)	(0,0,1)	(0,1,0)	(1,0,0)	(0,1,0)	(0,1,0)	[2,4]

6.3 Joint polyspecificities

Similar expressions arise for joint polyspecificities, which are defined as the number of drugs that bind to two targets, t_k and t_l, that are listed below for completeness. Their derivation the meaning of these expressions are similar to those shown above for joint polypharmacologies and will not be repeated here.

$$\begin{aligned} \mathcal{S}(t_k,t_l)_{\min} &= \min Count\left(\overline{T}_k \cap \overline{T}_l\right) \\ &= \sum_{i=1}^{n}\left[\mu_{\overline{T}_k}(d_i,t_k) \wedge \mu_{\overline{T}}(d_i,t_l)\right] \\ &\text{for } k=1,2,\ldots,m \text{ and } l=k+1,k+2,\ldots,m \end{aligned} \quad , \tag{6.10}$$

$$\begin{aligned} \mathcal{S}(t_k,t_l)_{\max} &= \max Count\left(\overline{T}_k \cap \overline{T}_l\right) \\ &= \sum_{i=1}^{n}\left[\mu_{\overline{T}_k}(d_i,t_k) + \pi_{\overline{T}_k}(d_i,t_k)\right] \wedge \left[\mu_{\overline{T}_l}(d_i,t_l) + \pi_{\overline{T}_l}(d_i,t_l)\right], \\ &\text{for } k=1,2,\ldots,m \text{ and } l=k+1,k+2,\ldots,m \end{aligned} \tag{6.11}$$

and

$$\begin{aligned} &\mathcal{S}(t_k,t_l) \in \left[\mathcal{S}(t_k,t_l)_{\min}, \mathcal{S}(t_k,t_l)_{\max}\right] \\ &\text{for } k=1,2,\ldots,m \text{ and } l=k+1,k+2,\ldots,m \end{aligned} \quad . \tag{6.12}$$

Lastly, the polyspecificity inequalities comparable to those given in Eqs. (6.3), (6.6), and (6.8) for polypharmacologies, are given by

$$\begin{aligned} \mathcal{S}(t_k,t_l)_{\min} &\le \begin{cases} \mathcal{S}(t_k)_{\min} \\ \mathcal{S}(t_l)_{\min} \end{cases} \\ \mathcal{S}(t_k,t_l)_{\max} &\le \begin{cases} \mathcal{S}(t_k)_{\max} \\ \mathcal{S}(t_l)_{\max} \end{cases} \\ &k=1,2,\ldots,m \text{ and } l=k+1,k+2,\ldots,m \end{aligned} \quad . \tag{6.13}$$

Because of the similarities of the mathematics of polyspecificities to those of polypharmacologies and their joint counterparts, examples such as those given above for polypharmacologies in Table 4, are not given here for polyspecificities.

7. Summary and Conclusions

Despite rapid growth in the size and coverage of publicly available drug-target databases, interactive and non-interactive drug-target pairs typically make up less than five percent of the total. The extremely low percentage suggests that polypharmacolgy and polyspecificity values are likely to be significantly underestimated. Even if only a small percentage of the null pairs where found to be interactive—a very mild assumption—they would nevertheless contribute to the polypharmacology and polyspecificity values, in some cases significantly. Increasing the number of targets and/or the number of drugs, given the above assumption, could also add to these values.

As an example, consider the Binding DB database given in Table 1. There are essentially 7,123,260 – 89,815 = 7,033,445 null pairs. If only one percent of them interact with the 681

targets it will yield about 70,000 additional interactive pairs. As there are about 10,000 drugs in the database, the polypharmacology of each drug would *on average* be augmented by about seven interactive pairs. This is admittedly a very crude calculation, but it illustrates the potential magnitude by which polypharmacologies may be underestimated—since seven is a mean value, the polypharmacology for some of the drugs may be increase by less, while others may increase by more than seven. Moreover, since the number of druggable targets is about 2,000, there are approximately 1,300 more targets whose interaction with the set of ~10,000 drugs considered above could be investigated. Even if only one percent of the new targets interact with any of these drugs, it would result in an average of 13 new interactive drug-target pairs for each drug. Although these are obviously very crude estimates, they do provide additional, albeit tentative, evidence which indicates that current polypharmacology values may be significantly underestimated. Hence, completely neglecting null pairs, or assuming that they correspond to inactive pairs, could seriously effect polypharmacology and polyspecificity values. Totally addressing this lack experimentally in a way that maintains public accessibility of the data is, however, beyond the capability of essentially all non-industrial pharmaceutical laboratories, the PubChem database supported by the United States National Institutes of Health being one possible exception **[**9**]**.

Because of the significant lack of data completeness in most drug-target databases (*Cf.* Mestres, *et al*. [10]), computational methods are necessary if substantial progress is to be made in alleviating this problem. Imputing interaction data for such a large number of null drug-target pairs from a relatively sparse set of known interactive and non-interactive pairs is problematic at best, especially given the highly inhomogeneous distribution of these pairs in drug-target space. While applying computational chemogenomic methods [17-25] could help, most such methods have not been tested on diverse enough sets of drugs and targets to provide meaningful results in all cases. In addition, there may be a lack of the data needed to implement some of the methods for certain classes of drug-target pairs. Hence, their reliability or applicability is likely to be somewhat questionable when applied broadly. Although this precludes a complete assessment of the interaction of all drug-target pairs in a typical publicly accessible drug-target database, this could in some instances be achievable for smaller more directed sets of drug-target pairs using experimental methods, supplemented where necessary with computational methods.

Using the methodology described in this work it is possible to determine bounds for polypharmacologies and polyspecificities. However, as discussed in Section 5, the size of the bounds will generally be much too large to be of any practical value. They would be more useful if the percent of null pairs could be reduced to something significantly less that the percent of interactive pairs. An approach for accomplishing such a reduction was recently proposed by Vogt, *et al*. [72]. Their approach, which was discussed in Sections 1 and 5, is based on a method they developed for removing null pairs, thus creating datasets more fully

populated with pairs whose interactivities have been determined experimentally or computationally, but smaller in size than the original database. However, it has the unfortunate side effect of also removing some interacting or non-interacting pairs that may be of interest. As noted earlier, by combining their method with that described in this work it may be possible to retain many of the desired drug-target pairs and also to estimate useful bounds for the resulting polypharmacology and polyspecificity values. But this is a subject for future research.

A number of other issues have been largely overlooked in many studies of drug-target databases. One issue involves weakly or non-interacting drug-target pairs, which are assumed to be of little value and thus are largely ignored due to a preoccupation of many researchers with the identification of strongly interacting pairs. This is exemplified by a number of studies based on drug-target networks, which typically ignore the information on weakly or non-interacting pairs [51-53] (*Cf.* [54]), since edges are only drawn between drug and target nodes if the pair strongly interacts. This is tantamount to placing weakly and non-interacting pairs in the same category as null pairs, which should not be the case as these pairs contain structure-activity information while the latter do not. In addition, since the presence of non-interactive pairs reduces the number of null pairs, the magnitude of the knowledge-based uncertainties for drugs or targets, which depend on the size of these intervals, are correspondingly reduced, leading to a general tightening of the bounds on the polypharmacology and polyspecificity values, as discussed in Section 5.

Currently, there is interest in determining the similarity of drugs and targets based on their drug and target profiles [75-79]. In a following paper, we will extend the current approach to include these similarities, which can be described as functions of their single and joint polypharmacologies or polyspecificities, and we will show that rigorous bounds for these quantities can be obtained that are related to, but slightly more complex than, the bounds described in the current work. In addition, we will describe a generalization of the method presented here, which will remove the restriction of activity thresholds in the case of experimental data and allow for continuous activity values.

Acknowledgements: GM would like to thank Dr. Vijay Gokhale for helpful discussion on subjects related to this work and David Bishop for help with preparing the manuscript.

References

[1] D. Szklarczyk, A. Santos, C. von Mering, L. J. Jensen, P. Bork, M. Kuhn, STITCH 5: augmenting protein-chemical interaction networks with tissue affinity data. *Nucleic Acids Res.* **44** (2016) D380-D384.

[2] T. Liu, Y. Lin, X. Wen, R. N. Jorrisen, M. K. Gilson, BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **35** (2007) D198-D201.

[3] PDSP database website: pdspdb.unc.edu/pdspWeb/, accessed March 25, 2019.

[4] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hershey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40** (2011) D1100-D1107.

[5] S. Günther, M. Kuhn, M. Dunkel, M. Campillos, C. Senger, E. Petsalaki, J. Ahmed, E. G. Urdiales, A. Gewiess, L. J. Jensen, R. Schneider, R. Skoblo, R. B. Russell, P. E. Bourne, P. Bork, R. Preisser, SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.* **36** (2008) D919-D923.

[6] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynllaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46** (2018) D1074-D1082.

[7] J. von Eichborn, M. S. Murgueitio, M. Dunkel, S. Koerner, P. E. Bourne, R. Preissner, PROMISCUOUS: a database for network-based drug-repositioning. Nucleic Acids Res. **39** (2010) D1060-D1066.

[8] O. Taboureau, T. I. Oprea, ChemProt: A disease chemical biology database, in: E. Jacoby (Ed.), *Computational Chemogenomics*, Pan Stanford Publishing, Singapore, 2014, pp. 193-213.

[9] S. Kim, P. A. Thiessen, E. E. Bolton, J. Chien, G. Fu, A. Gindulyte, L. Han, J. He, B.A. Shoemaker, J. Wang, B. Yu, J. Zhang, S. H. Bryant, PubChem substance and compound databases. *Nucleic Acids Res.* **44** (2016) D1202-D121.

[10] J. Mestres, E. Gregori-Puigjané, S. Valverde, R. V. Solé, Data completeness—Achilles heel of drug-target networks. *Nature Biotech.* **26** (2008) 983-984.

[11] J. Mestres, E. Gregori-Puigjané, S. Valverde, R. V. Solé, The topology of drug–target interaction networks: implicit dependence on drug properties and target families. *Mol. Biosys.* **5** (2009) 1051-1057.

[12] A. J. Williams, Public chemical compound databases. *Curr. Opin. Drug Disc. Dev.* **11** (2008) 393-404.

[13] T. Sterling, J. Irwin, ZINC15 – Ligand discovery for everyone. *J. Chem. Inf. Model.* **55** (2015) 2324-2337.

[14] Real Space: www.chem-space.com—*viewed* May 11, 2020.

[15] P. R. Caron, M. D. Mullican, R. D. Mashal, K. P. Wilson, M. S. Su, M. Murko, Chemogenomic approaches to drug discovery, *Curr. Opin. Chem. Biol.* **5** (2001) 464-470.

[16] A. L. Hopkins, C. R. Groom, The druggable genome. *Nature Rev. Drug Disc.* **1** (2002) 727-730.

[17] T. A. Halgren, Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **49** (2009) 377-389.

[18] J. P. Overington, B. Al-Lazikani, A. L. Hopkins, How many drug targets are there? *Nature Rev. Drug Disc.* **5** (2006) 993-996.

[19] D. Rognan, Structured-based approaches to target fishing and ligand profiling. *Mol. Inf.* **29** (2010) 176-187.

[20] H. Koeppen, M. Bieler, Chemogenomic approaches for the quantitative comparisons of biological targets, in: E. Jacoby (Ed.), *Computational Chemogenomics*, Pan Stanford Pub., Singapore, 2014, pp. 1-38.

[21] E. Jacoby (Ed.), *Computational Chemogenomics*, Pan Stanford Pub., Singapore, 2014.

[22] D. Rognan, Computational approaches to target fishing and ligand profiling, in: E. Clementi, J. M. André, J. A. McCammon (Eds.), *Theory and Applications in Computational Chemistry: The First Decade of the Second Millennium – AIP Conference Proceedings*, 1456 (2012) 157-164, doi: 10.1063/1.4730655.

[23] D. Rognan, Chemogenomic approaches to rational drug design. *Brit. J. Pharm.* **152** (2007) 38-52.

[24] J. Bajorath, Computational analysis of ligand relationships within target families. *Curr. Opin. Chem. Biol.* **12** (2008) 352-358.

[25] H. Strömbergsson, G. J. Kleywegt, A chemogenomics view of protein-ligand spaces, *BMC Bioinformatics* **10** (2009) #S13.

[26] M. S. Nogueira, O. Koch, The development of target-specific machine learning models as scoring function for docking-based target prediction methods, *J. Chem. Inf. Model.* **59** (2019) 1238-1252.

[27] F. Cheng, C. Liu, J. Jiang, W. Lu, W. Li, G. Liu, W. Zhou, J. Huang, Y. Tang, Prediction of drug-target interactions and drug repositioning via network inference, *PLoS Comput. Biol.* **8** (2012) #e1002503.

[28] X. Chen, C. Yan, X. Zhang, X. Zhang, F. Dai, J. Yin, Y. Zhang, Drug–target interaction prediction: databases, web servers and computational models, *Brief. Bioinf.* **17** (2016) 696-712.

[29] W. Zhou, S. Anusuya, M. Kesherwani, K. V. Priya, A. Vimala, V. Shanmugam, D. Velmurugan, M. M. Gromiha, Drug-target interactions – Prediction methods and applications, *Curr. Prot. Pept. Sci.* **19** (2018) 537-561.

[30] Z. Wu, W. Li, G. Liu, Y. Tang, Network-based methods for prediction of drug-target interactions, *Front. Pharmacol.* **9** (2018) #1134.

[31] K. Sachdev, M. K. Gupta, A comprehensive review of feature-based methods for drug-target interaction prediction, *J. Biomed. Inf.* **93** (2019) doi:10.1016/j.jbi.2019.103159.

[32] R. Chen, X. Liu, S. Jin, J. Lin and J. Liu, Machine learning for drug-target interaction prediction. *Molecules* **23** (2018) #103159.

[33] Z.R. Tanoli, Z. Alam, M. Vähä-Koskela, B. Ravikumar, A. Malyutina, A. Jaiswal, J.g Tang, K. Wennerberg, T. Aittokallio, Drug Target Commons 2.0: a community platform for systematic analysis of drug–target interaction profiles, *Database* **2018** (2018) #bay083.

[34] L. Y. Xia, Z. Y. Yang, H. Zhang, Y. Liang, Improved prediction of drug-target interactions using self-paced learning with collaborative matrix factorization, *J. Chem. Inf. Model.* **59** (2019) 3340-3351.

[35] W. Torng, R. B. Altman, Graph convolution neural networks for predicting drug-target interactions, *J. Chem. Inf. Model.* **59** (2019) 4131-4149.

[36] Y. Li, Y. A. Huang, Z. H. You, L. P. Li, Z. W. Au, Drug-target interaction prediction-based drug fingerprint information and protein sequence, *Molecules* **24** (2019) #2999.

[37] T. M. Whitehead, B. W. J. Irwin, P. Hunt, M. D. Segall, G. J. Conduit, imputation of assay bioactivity data using deep learning, *J. Chem. Inf. Model.* **59** (2019) 1197-1204.

[38] H. Gohlke, M. Hendlich, G. Klebe. Knowledge-based scoring function to predict protein-ligand interactions, *J. Mol. Biol.* **295** (2000) 337-356.

[39] E. Perola, W. P. Walters, P. S. Charifson, A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance, *Proteins* **56** (2004) 235-249.

[40] T. Schulz-Gasch, M. Stahl, Scoring functions for protein-ligand interactions: a critical perspective, *Drug Disc. Today: Technologies* **1** (2004) 231-239.

[41] O. Korb, T. Stüand, T. E. Exner, Empirical scoring functions for advance protein-ligand docking with PLANTS, *J. Chem. Inf. Model.* **49** (2009) 84-96.

[42] J. C. Wang, J. H. Lin, Scoring functions for prediction of protein-ligand interactions, *Curr. Pharm. Des.* **1** (2013) 2174-2182.

[43] Y. Li, M. Su, Z. Liu, J. Li, J. Liu, L. Han, R. Wang, Assessing protein-ligand interaction scoring functions with the CASF-2013 benchmark, *Nature Protocols* **13** (2018) 666-680.

[44] N. Singh, A. Warshel, Absolute binding free energy calculations: on the accuracy of computational scoring of protein-ligand interactions, *Proteins: Struct. Funct. Bioinf.* **78** (2010) 1705-1723.

[45] K. M. Merz, Limits of free energy computation for protein-ligand interactions, *J. Chem. Theory Comput.* **6** (2010) 1769-1776.

[46] L. Duan, X. Liu, J. Z. H. Zhang, Interaction entropy: a new paradigm for highly efficient and reliable computation of protein-ligand binding free energy, *J. Am. Chem. Soc.* **138** (2016) 5722-5728.

[47] C. D. Parks, Z. Gaieb, M. Chiu, H. Yang, S. Chenghua, W. P. Walters, J. M. Jantzen, G. McGaughey, R. A. Lewis, S. D. Bembenek, M. K. Ameriks, T. Mirzadegan, S. K. Burley, R. E. Amaro, M. K. Gilson, D3R Grand Challenge 4: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies, *J. Comput. Aided Mol. Design* **34** (2020) 99-119.

[48] C. Gorgulla, A. Boeszoermenyi, Z. F. Wang, P. D. Fischer, P. W. Coote, K. M. Padmanabha Das, Y. S. Malets, D. S. Radchenko, Y. S. Moroz, D. A. Scott, K. Fackeldey, M. Hoffman, I. Iavniuk, G. Wagner, H. Arthanari, An open-source drug-discovery platform enables ultra-large virtual screens, *Nature* **580** (2020) 663-668.

[49] J. Lyu, S. Wang, T. E. Balius, I. Singh, A. Levit, Y. S. Moroz, M. J. O'Meara, T. Che, E. Algaa, K. Tolmachova, A. A. Tolmachev, B. K. Shoichet, B. L. Roth, J. J. Irwin, Ultra-large library docking for discovering new chemotypes, *Nature* **566** (2019) 224-229.

[50] R. M. Stein, H. J. Kang, J. D. McCorvy, G. C. Glatfelter, A. J. Jones, T. Che, S. Slocum, X. P. Huang, O. Savych, Y. S. Moroz, B. Stauch, L. C. Johansson, V. Cherezov, T. Kenakin, J. J. Irwin, B. K. Shoichet, B. L. Roth, M. L. Dubocovich, Virtual discovery of melatonin receptor ligands to modulate circadian rhythms, *Nature* **579** (2020) 609-614.

[51] M. A. Yildrim, K. I. Goh, M. E. Cusick, A. L. Barábasi, M. Vidal, Drug-target network, *Nature Biotech.* **25** (2007) 1119-1126.

[52] A. L. Hopkins, Network pharmacology, *Nature Biotech.* **25** (2007) 1110-1111.

[53] I. Vogt, J. Mestres, Drug-target networks, *Mol. Inf.* **29** (2010) 1-14.

[54] G. M. Maggiora, V. Gokhale, A simple mathematical approach to the analysis of polypharmacology and polyspecificity data, *F1000Research* **6** (2017) #788.

[55] K. T. Atanassov, *On Intuitionistic Fuzzy Sets Theory*, Springer, Berlin, 2012.

[56] E. Szmidt, *Distances and Similarities in Intuitionistic Fuzzy Sets*, Springer, Berlin, 2014.

[57] E. Szmidt, J. Kacprzyk, Distances between intuitionistic fuzzy sets, *Fuzzy Sets Sys.* **114** (2000) 505-518.

[58] E. Szmidt, J. Kacprzyk, Entropy for intuitionistic fuzzy sets, *Fuzzy Sets Sys.* **118** (2001) 467-477.

[59] E. Szmidt, J. Kacprzyk, A perspective on differences between Atanassov's intuitionistic fuzzy sets and interval-valued fuzzy sets, in: V. Torra, A. Dahlbom, Y. Narukawa, (Eds.), *Fuzzy Sets, Rough Sets, Multisets and Clustering*, Springer, Heidelberg , 217, pp. 221-237.

[60] E. Szmidt, J. Kacprzyk, Intuitionistic fuzzy sets - two and three term representations in the context of a Hausdorff distance, *Acta Universitatis Matthiae Belii, Series Mathematics* **19** (2011) 53-62.

[61] E. Szmidt, J. Kacprzyk, P. Bujnowski, How to measure the amount of knowledge conveyed by Atanassov's intuitionistic fuzzy sets, *Inf. Sci.* **257** (2014) 276-285.

[62] J. U. Peters (Ed.), *Polypharmacology in Drug Discovery*, Wiley, New York, 2012.

[63] A. Anighoro, J. Bajorath, G. Rastelli, Polypharmacology: challenges and opportunities in drug discovery, *J. Med. Chem.* **57** (2014) 7874-7887.

[64] W. Zhang, Y. Bai, Y. Wang, W. Xiao, Polypharmacology in drug discovery: A review from [a] systems pharmacology perspective, *Curr. Pharm. Design* **22** (2016) 3171-3181.

[65] G. V. Paolini, R. H. B. Shapland, W. P. van Hoorn, J. S. Mason, A. L. Hopkins, Global mapping of pharmacological space, *Nature Biotech.* **24** (2006) 805-815.

[66] M. J. Barratt, D. E. Frail (Eds.), *Drug Repositioning – Bringing New Life to Shelved Assets and Existing Drugs*, Wiley, New York, 2012.

[67] Y. Hu, J. Bajorath, High-resolution view of compound promiscuity, *F1000Research* **2** (2013) #144.

[68] X. Jalencas, J. Mestres, On the origins of drug polypharmacology, *Med. Chem. Comm.* **4** (2013) 80-87.

[69] G. M. Maggiora, V. Gohkale, Non-specificity of drug-target interactions – Consequences for drug discovery, in: R. Bienstock, V. Shanmugasundaram, J. Bajorath (Eds.), *Frontiers in Molecular Design and Chemical Information Science – Skolnik Award Symposium 2015*, *Am. Chem. Soc.*, Oxford Univ. Press, 2016, pp. 91-142.

[70] P. Iyer, D. Stumpfe, M. Vogt, J. Bajorath, G. M. Maggiora, Activity landscapes, information theory, and structure-activity relationships, *Mol. Inf.* **32** (2013) 421-430.

[71] G.M. Maggiora, M. Vogt, D. Stumpfe, J. Bajorath, Molecular similarity in medicinal chemistry, *J. Med. Chem.* **57** (2014) 3186-3204.

[72] M. Vogt, S. Jasial, J. Bajorath, Extracting compound profiling matrices from screening data, *ACS Omega* **3** (2018) 4706-4712.

[73] L. A. Zadeh, Fuzzy sets, *Inform. Contr.* **8** (1965) 338-353.

[74] E. Szmidt, J. F. Baldwin, A “family” of similarity measures for intuitionistic fuzzy set theory and mass assignment theory, *NIFS* **12** (2006) 1-23.

[75] A. Schuffenhauer, P. Floersheim, P. Acklin, E. Jacoby, Similarity metrics for ligands reflecting the similarity of the target proteins, *J. Chem. Inf. Comput. Sci.* **43** (2003) 391-405.

[76] A. Steffen, T. Kogej, C. Tyrchan, O. Engkvist, Comparison of molecular fingerprint methods on the basis of biological profile data, *J. Chem. Inf. Model.* **49** (2009) 338-347.

[77] T. Cheng, Q. Li, Y. Wang, S. F. Bryant, Identifying compound-target associations by combining bioactivity profile similarity search and public databases mining, *J. Chem. Inf. Model.* **51** (2011) 2440-2448.

[78] M. Bieler, R. Heilker, H. Köppen, G. Schneider, Assay related target similarity (ARTS) – Chemogenomics approach for quantitative comparison of biological targets, *J. Chem. Inf. Model.* **51** (2011) 1897-1905.

[79] P. M. Petrone, B. Simms, F. Nigsch, E. Lounkine, P. Kutchukian, A. Cornett, Z. Deng, J. W. Davies, J. L. Jenkins, M. Glick, Rethinking molecular similarity: Comparing compounds on the basis of biological activity, *ACS Chem. Biol.* **7** (2012) 1399-1409.