The topology of drug–target interaction networks: implicit dependence on drug properties and target families†‡

Jordi Mestres,*[a] Elisabet Gregori-Puigjané,[a] Sergi Valverde[bc] and Ricard V. Solé[bd]

Received 23rd March 2009, Accepted 26th May 2009
First published as an Advance Article on the web 8th July 2009
DOI: 10.1039/b905821b

The availability of interaction data between small molecule drugs and protein targets has increased substantially in recent years. Using seven different databases, we were able to assemble a total of 4767 unique interactions between 802 drugs and 480 targets, which means that on average every drug is currently acknowledged to interact with 6 targets. The application of network theory to the analysis of these data reveals an unexpectedly complex picture of drug–target interactions. The results confirm that the topology of drug–target networks depends implicitly on data completeness, drug properties, and target families. The implications for drug discovery are discussed.

Introduction

The traditional view of drugs interacting selectively with a specific protein target has been recently challenged by growing evidence that they possess instead an inherently rich polypharmacology.[1] For example, celecoxib (Celebrex) is still being referred to as a selective cyclooxygenase-2 non-steroidal anti-inflammatory drug, even though relevant affinities for at least two additional targets (namely, carbonic anhydrase II and 5-lipoxygenase) have been lately identified,[2,3] and pergolide (Permax) was not long ago still regarded as a member of a class of drugs known as dopamine agonists, despite its currently accepted promiscuity over multiple G protein-coupled receptors (GPCRs).[4] As more data on the interaction between drugs and targets are being generated and made publicly accessible,[5] it is becoming evident that selective drugs acting exclusively on single targets seem to be the exception rather than the norm.

The core of this historical misconception of drug selectivity lies in the fact that, mainly due to limited time and resources, drugs are usually not screened systematically through a large panel of protein targets for the sake of acquiring knowledge about their complete pharmacological profile. Instead, drugs are only tested against a limited number of off-target proteins assumed to be relevant for a particular drug discovery project and selected mainly on the basis of safety concerns and phylogenetic relationships to the primary target.[6,7] In addition, our biased perception of drug selectivity is aggravated by the fact that, from all data generated, only a portion is ultimately published, and even then it is found scattered over numerous bibliographic sources often using different names for the same drug and target entities.[8] In this respect, recent initiatives to collect and store drug–target interaction data from the literature have contributed highly to the modern appreciation of the polypharmacology of drugs.[5,9]

As a consequence, even though currently available interaction data may still be largely incomplete, nonhomogeneous and biased toward certain areas of interest, an unexpectedly complex picture of drug–target interactions has started to emerge.[10] Analysing this increasingly complex ensemble of interactions requires specialised visualisation tools. The simplest approach is to depict directly the entire drug–target interaction matrix in which each cell of the matrix is coloured according to the affinity between a particular drug–target pair.[11] But in recent years, the application of network theory to visualise and analyse drug–target interaction data has become very popular for its ability to capture complexity in a compact and illustrative manner.[10–14] Accordingly, the aim of this contribution is to perform a more in-depth exploration of the topology of networks constructed from drug–target interaction data and, in particular, analyse the effect that drug properties and target families might have on the overall structure of drug–target networks.

Results

Drug–target interaction data were extracted from seven annotated chemical libraries, collecting globally a total of 6284 interactions between 802 drugs and 480 targets. Table 1 summarises the number of drugs, targets, drug–target interactions, and target interactions per drug provided by each particular database. Additional drug–target interaction data were predicted computationally by means of a ligand-based approach to target profiling.[11] Then, drug–target networks were constructed by linking all drug–target pairs for which an interaction was either known experimentally from annotated

[a] *Chemogenomics Laboratory, Research Unit on Biomedical Informatics, Institut Municipal d'Investigació Mèdica, Parc de Recerca Biomèdica, Doctor Aiguader 88, Catalonia, 08003 Barcelona, Spain. E-mail: jmestres@imim.es; Fax: +34-93-3160550; Tel: +34-93-3160540*
[b] *Complex Systems Laboratory, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain*
[c] *CNRS-UMR, Université Paul Sabatier, Toulouse, France*
[d] *Santa Fe Institute, NM, USA*

† This article is part of the 2009 *Molecular BioSystems* 'Emerging Investigators' issue: highlighting the work of outstanding young scientists at the chemical- and systems-biology interfaces.

‡ Electronic supplementary information (ESI) available: One table and two figures that complement Fig. 2 and Fig. 4. See DOI: 10.1039/b905821b

Table 1 List of databases with drug–target interaction data used in this work

Database[a]	Drugs	Targets	Interactions	Interactions per drug
DrugBank 2008.07	774	289	1404	1.8
Wombat 2007.2	336	257	1178	3.5
DB + WB	785	413	2288	2.9
PDSP 2008	296	187	2590	8.8
BindingDB 2008.09	70	41	109	1.6
CYPdb	153	34	418	2.7
GPCRdb	145	59	573	3.9
NRdb	12	7	12	1.7
DB + WB + OD	802	480	4767	5.9
DB + WB + OD + IS	802	675	10343	12.9

[a] DB: DrugBank, WB: Wombat, OD: other databases, IS: *in silico* predictions.

chemical libraries or predicted computationally by *in silico* target profiling and then visualised using a common layout algorithm that disperses the nodes in the plane and minimises their overlap.[15] Apart from the apparent visual differences, the topology of drug–target interaction networks was assessed quantitatively in terms of the fraction of nodes belonging to the largest connected component (nLCC) in the corresponding drug network, in which nodes are drugs connected if they share at least one target, and target network, in which nodes are targets connected if they share at least one drug, and these values were then compared to the corresponding values obtained from one hundred randomised networks keeping the total number of nodes and interactions unchanged. Given the relatively large degree of connectivity found in all drug and target networks derived from the drug–target networks presented here, standard deviations for the corresponding nLCC values were in most cases lower than the number of decimal digits provided and thus they have been omitted for the sake of simplicity. The following sections analyse the effects that data completeness, drug properties, and target families may have on the topology of drug–target networks.

Topology of drug–target interaction networks

A recent study highlighted the dramatic effect that data completeness had on the topology of drug–target interaction networks when systematically completing the interaction data present in DrugBank[5] with additional literature-based experimental data available in Wombat[16] and subsequently with estimated data obtained from an *in silico* target profiling method.[10] A revised analysis is presented here, now including experimental drug–target interaction data added in the recent updates of DrugBank and Wombat, and complemented with new data available from five additional databases, namely, BindingDB,[17] PDSP,[18] CYPdb, GPCRdb, and NRdb[19] (Table 1).

The first drug–target network (Fig. 1a) was based on the 1404 interactions available in DrugBank connecting 774 drugs to 289 targets, resulting in an average number of interactions per drug of 1.8. As can be observed, the topology of this drug–target network reveals a well-organised modular structure, with many proteins clustering around phylogenetic families. The nLCC values for the corresponding drug and target networks are 0.66 and 0.47, respectively. Both these numbers are significantly smaller than the values of 0.93 and 0.77 obtained from the corresponding randomised networks. The two nLCC values are also found to be very much in agreement with those reported recently from networks derived using the same data source.[14]

A second drug–target network (Fig. 1b) was then derived by supplementing the original data present in DrugBank with additional interaction data available in Wombat, which resulted in 2288 drug–target interactions between 785 drugs and 413 targets. This means that Wombat enriched the original DrugBank network with 11 drugs, 124 targets, and 884 drug–target interactions. The result is that the topology of the drug–target network is visibly affected, becoming more complex and interconnected, with an average number of interactions per drug of 2.9. Quantitatively, this is reflected by nLCC values of 0.87 and 0.78 for the corresponding drug and target networks, respectively, which become now closer to the values of 0.95 and 0.91 derived from randomised networks.

A third drug–target network (Fig. 1c) was constructed with all drug–target interaction data from the seven annotated databases considered in this work (Table 1). In total, they accumulate 4767 interactions connecting 802 drugs and 480 targets, that is 17 drugs, 67 targets, and 2479 interactions more than the network constructed from DrugBank and Wombat only (Fig. 1b). This translates into a significant increase in the number of interactions per drug, doubling the value from 2.9 to 5.9, which in turn results in a drug–target network with a highly-connected structure. The nLCC values for the corresponding drug and target networks are 0.92 and 0.87, respectively, almost reaching the corresponding values of 0.97 and 0.94 from the randomised networks. Interestingly, the value of 5.9 interactions per drug obtained from all experimental drug–target interaction data currently available is now approaching the projected value of 6.3 reported previously by *in silico* target profiling.[10]

Finally, a fourth drug–target network (Fig. 1d) was derived by complementing all experimental drug–target interaction data with annotations assigned using a ligand-based approach to *in silico* target profiling.[11] Note that, compared to previous works,[10,11] the ligand-based target models used here were significantly augmented with the latest drug–target interaction data included in the most recent updates of all annotated chemical libraries (Table 1). The resulting network contains 10 343 interactions linking 802 drugs with 675 targets, which now gives a projected average number of interactions per drug of 12.9. The topology of the network reveals a high

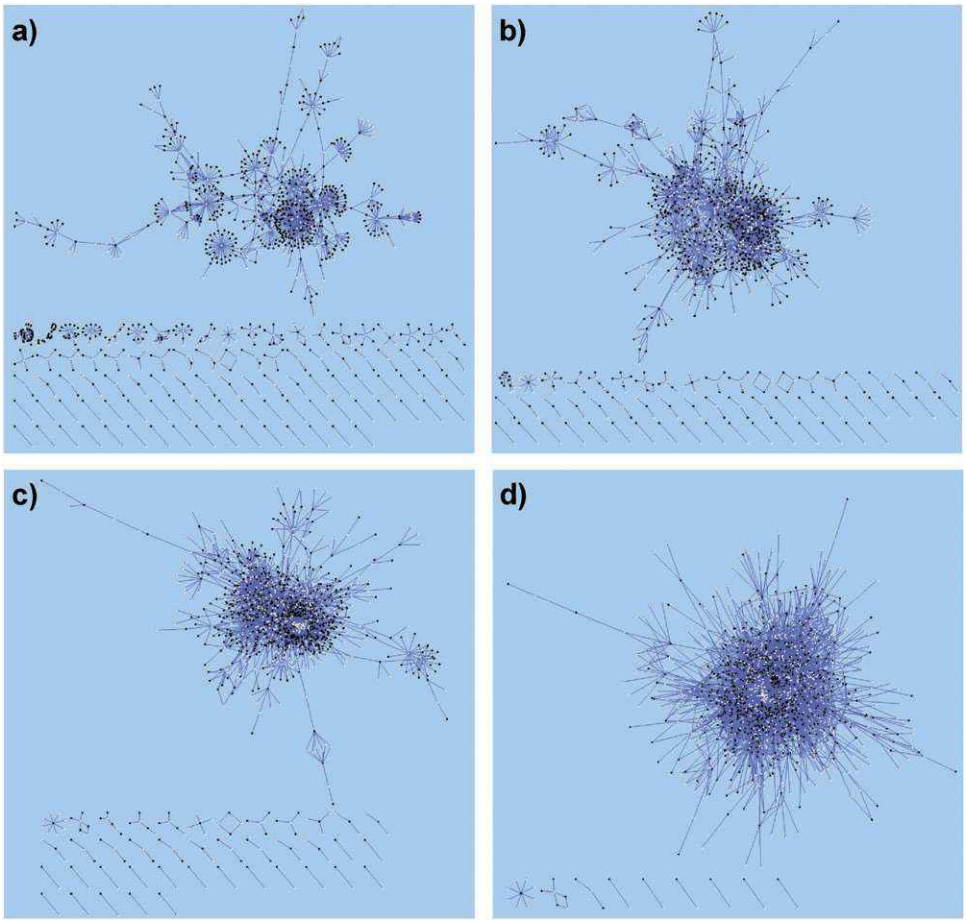

Fig. 1 Drug–target interaction networks derived cumulatively from various data sources: (a) DrugBank, (b) DrugBank and Wombat, (c) DrugBank, Wombat, BindingDB, PDSP, CYPdb, GPCRdb and NRdb, and (d) DrugBank, Wombat, BindingDB, PDSP, CYPdb, GPCRdb, NRdb and *in silico* annotations. Drugs and targets are indicated as black and white circles, respectively.

interconnected structure between nodes, with nLCC values for the corresponding drug and target networks of 0.98 and 0.97, respectively, very close to the values of 0.99 and 0.99 obtained from the analogous randomised networks.

Overall, the exercise of systematically adding currently known experimental drug–target interaction data contained in different databases confirms the dramatic effect that data completeness has on the topology of drug–target networks and consequently on any conclusions that can be derived from them.[10] In this respect, the well-organised modular structure of the drug–target network constructed with interaction data from DrugBank (Fig. 1a) was gradually transformed into highly-connected topologies as additional experimental data were considered. Likewise, the average number of 2 targets per drug derived from DrugBank and 3 targets per drug, when supplemented with Wombat, increased to 6 targets per drug when all experimental drug–target interaction data available at present were used. This value may go up to 13 targets per drug, according to the projection obtained by means of an *in silico* target profiling method.

Dependence on drug properties

Using a simple model of ligand–receptor interactions, one can deduce that as the number of potentially-interacting features in a molecule increases, the chance of observing a useful interaction for a randomly chosen ligand falls dramatically.[20] The direct consequences of this theoretical hypothesis are that large complex molecules should in principle be more selective than small simple molecules and thus the topology of drug–target interaction networks could depend implicitly on the properties of drugs.

In order to investigate this aspect, two properties were calculated for all drugs, namely, molecular weight (MW, in Dalton), which correlates well with molecular size, and hydrophobicity (clog*P*, calculated logarithm of *n*-octanol–water partition coefficient), taken as a rough estimate of pharmacophoric complexity. Then, the set of 802 drugs was partitioned into several ranges for each property and the projected mean promiscuity (number of interactions per drug) for all drugs within each property range calculated from the values used to derive the drug–target network completed with computationally-predicted interactions (Fig. 1d). The distribution of the mean promiscuities of drugs upon variation of MW and clog*P* is presented in Fig. 2. For the sake of completeness, the exact number of drugs, the mean number of interactions and the degree of the drug–target network within each property block is provided in Table S1 (ESI‡) and the corresponding distribution obtained from the drug–target

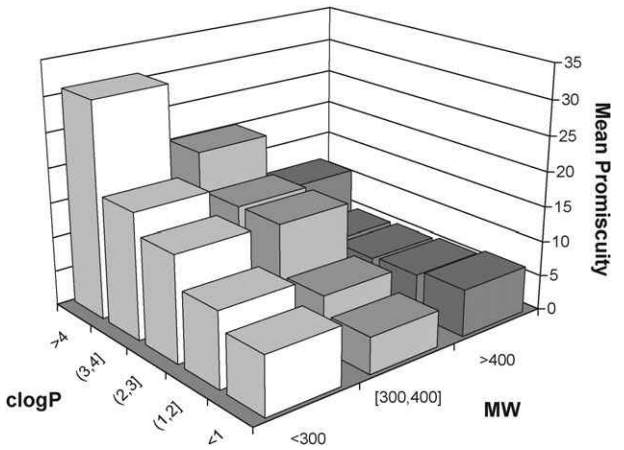

Fig. 2 Variation of the mean promiscuity of drugs contained within value ranges of molecular weight (MW) and hydrophobicity (clog*P*).

network derived from all currently available experimental interaction data (Fig. 1c) is also given in Fig. S1 of the ESI.‡

The general trends observed are that, within a given range of MW values, promiscuity tends to increase with hydrophobicity but also that, within a given range of clog*P* values, promiscuity has a tendency to decrease with size. The effect is clearly visible for the smallest (MW $<$ 300) and the more hydrophobic (clog*P* $>$ 4) drugs, but it becomes less evident for drugs in the other ranges of property values. For example, the mean promiscuity obtained for the least hydrophobic drugs (clog*P* $<$ 1) remains almost invariable with increased MW. The same holds true for the largest drugs (MW $>$ 400) upon increased hydrophobicity, although in this case a change is markedly observed at high hydrophobicity values (clog*P* $>$ 4). Likewise, for drugs in the medium range of hydrophobicity values (2 $<$ clog*P* $<$ 4), the mean promiscuity experiences minor changes with increased molecular size until it drops abruptly at high values (MW $>$ 400). Qualitatively, similar trends are already obtained from the analysis of experimental interaction data (Fig. S1, ESI‡), the differences being mainly associated with the issue of data completeness. Broadly, these results agree well with the inverse correlation between mean MW and promiscuity, and the trend that promiscuous compounds tend to have clog*P* values above 2.5–3.0, observed recently from an analysis of a different source of drug–target interaction data.[21]

Following our previous observations (Fig. 1), one should expect that low mean promiscuities will be potentially associated with highly disconnected drug–target networks with well-structured topologies for the largest connected component, whereas high mean promiscuity values may result in more dense and connected network topologies. In order to highlight the topological differences among the networks constructed from drugs included in the various property blocks of Fig. 2, focus was given to the sets of drugs contained in the two property poles. On one hand, the set of 27 drugs with MW $>$ 400 and clog*P* $<$ 1 have, on average, 2.4 experimentally known interactions per drug, with a projected

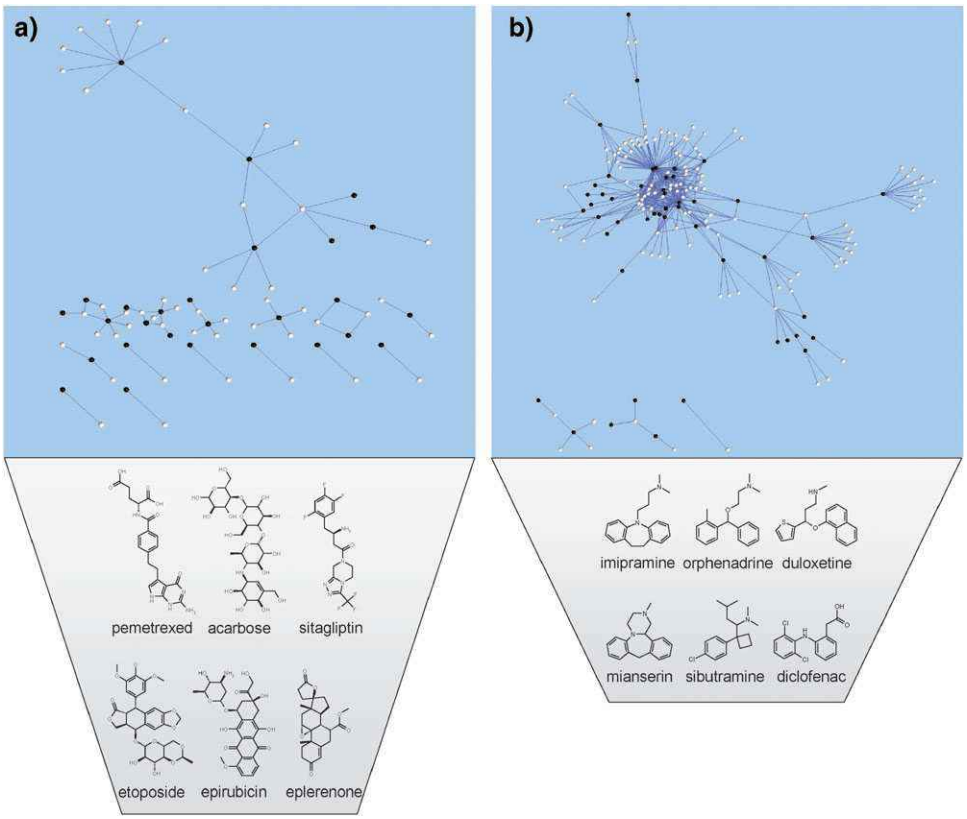

Fig. 3 Drug–target interaction networks and representative structures of drugs contained in the two blocks of property ranges: (a) MW $>$ 400 and clog*P* $<$ 1, and (b) MW $<$ 300 and clog*P* $>$ 4. Drugs and targets are indicated as black and white circles, respectively.

value of 6.7. Accordingly, the drug–target network associated with this property block displays a relatively small and modular giant component alongside many isolated interactions (Fig. 3a). On the other hand, the set of 49 drugs with MW $<$ 300 and clog$P$ $>$ 4 have, on average, 9.2 experimentally known interactions per drug, with a projected value of 30.9. Consequently, the drug–target network derived for this property block shows comparably a much larger and connected giant component with only a few isolated interactions (Fig. 3b). These two examples illustrate clearly the implicit dependence of the topology of drug–target networks on drug properties.

A close inspection of the composition and structural features of drugs contained within the two most distant property blocks may provide some clues for the trends observed. Therefore, representative structures of the two sets of drugs were also included in Fig. 3. As can be observed, the structures of drugs with MW $>$ 400 and clog$P$ $<$ 1 are of highly complex nature, complexity here is understood to be the presence and combination of multiple potentially-interacting features (Fig. 3a). In contrast, the structures of drugs with MW $<$ 300 and clog$P$ $>$ 4 all share a rather simple arrangement of hydrophobic features combined with the presence of a few potentially-interacting features (Fig. 3b), visually quite distinct from the complex composition of the previous drug set. Interestingly, while many of the large hydrophilic drugs are enzyme inhibitors, the majority of the small hydrophobic drugs interact with GPCRs, an indication that the degree of drug–target interactions may vary among target families.

Dependence on target families

A recent analysis highlighted the dependence of molecular properties on target families for a set of 642 marketed oral drugs.[22] For example, it was found that the mean MW for drugs acting on ion channels and proteases is 305.5 and 430.6, respectively, and the mean clog$P$ for drugs acting on aminergic GPCRs and nuclear receptors is 2.8 and 4.1, respectively. Therefore, on one hand, it has been reported already that the properties of drugs vary significantly among target families and, on the other hand, it was shown in the previous section that the topology of drug–target networks depends implicitly on the properties of drugs. Consequently, one should expect also an implicit dependence of the topology of drug–target networks on the target family. To investigate this aspect, the list of 480 targets experimentally known to interact with any of the 802 drugs was organised in four main target families of therapeutic relevance, namely, enzymes, GPCRs, ion channels/transporters, and nuclear receptors. Then, drugs were assigned to each target family if interaction data between

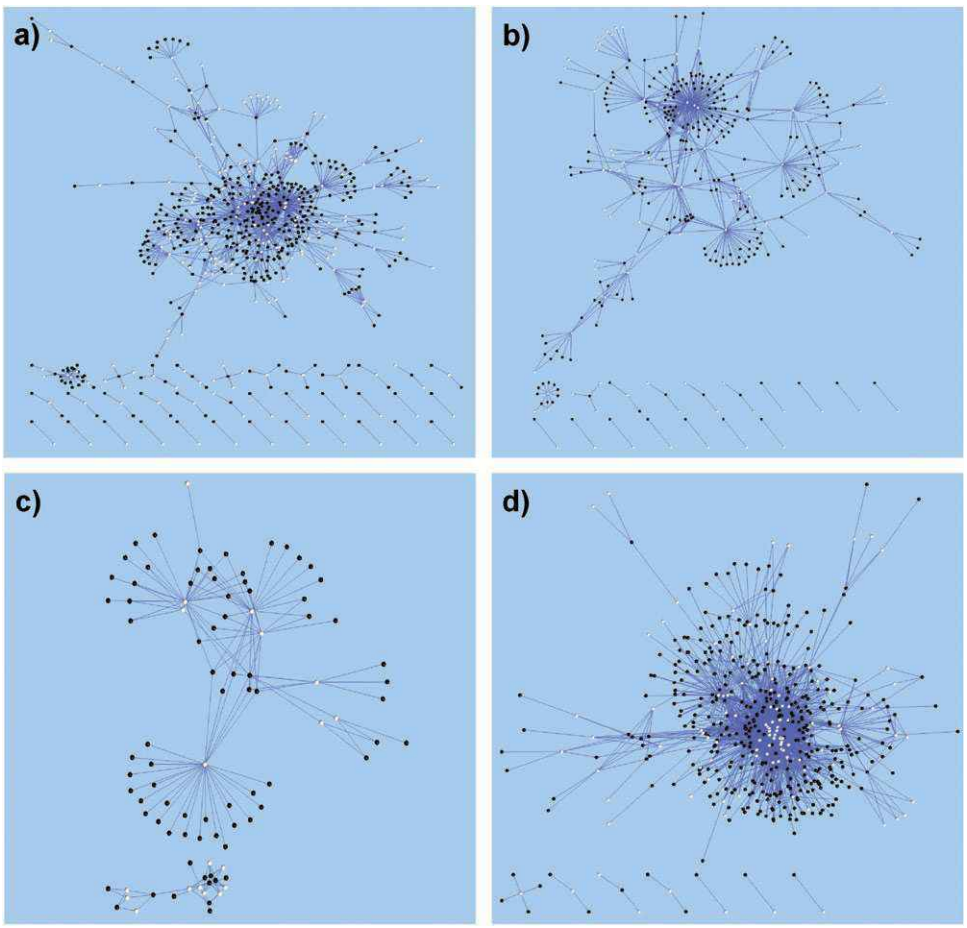

Fig. 4 Drug–target interaction networks derived for the different target families: (a) enzymes, (b) ion channels/transporters, (c) nuclear receptors, and (d) G protein-coupled receptors. Drugs and targets are indicated as black and white circles, respectively.

the drug and a member of that family existed. The resulting drug–target networks are presented in Fig. 4, which represents a deconvolution into target families of Fig. 1c. For the sake of completeness, the corresponding networks derived from drug–target interaction data completed by computational means are provided in Fig. S2 of the ESI.‡

The drug–target network derived for enzymes contains 1112 interactions connecting 431 drugs with 191 targets, resulted in an average number of experimentally known interactions per drug of 2.6, with a projected value of 4.8 (Fig. 4a). Visually, the topology of the network reveals a well-organised structure. The nLCC values for the corresponding drug and target networks are 0.84 and 0.74, respectively, both numbers being smaller than the values of 0.95 and 0.86 obtained from the corresponding randomised networks.

A similar topological structure is observed for the drug–target network derived for ion channels/transporters, containing 623 interactions between 268 drugs and 93 targets, with an average number of 2.3 experimentally known interactions per drug and a projected value of 3.0 (Fig. 4b). In this case, the nLCC values for the drug and target networks derived are 0.90 and 0.74, respectively, comparably smaller than the values of 0.94 and 0.87 obtained from the randomised networks and all similar to the corresponding values obtained above for the enzyme network.

The drug–target network constructed for the 77 drugs connected to 19 nuclear receptors through 155 interactions reveals a much simpler topology due to the relatively small size of the network, with an average number of experimentally known interactions per drug of 2.0 and a projected value of 2.3 (Fig. 4c). The nLCC values of the associated drug and target networks are 0.84 and 0.47, respectively. In fact, the network is composed of two subnetworks, dividing nuclear receptors into two main cross-pharmacology sets. The largest connected component contains 7 members of the estrogen-like class and the 2 thyroid hormone receptors, whereas the smallest component includes the 3 retinoic acid receptors, the 3 retinoic X receptors, the 3 peroxisome proliferator activated receptors, and the vitamin D3 receptor.

In contrast, the drug–target network for GPCRs contains 2646 interactions between 396 drugs and 106 targets, resulted in an average number of interactions per drug of 6.7, with a projected value of 10.0 (Fig. 4d). Compared to the drug–target networks obtained for the other three target families, the topology of this network is visually more dense and interconnected, with many targets collapsed in the centre of the network. From a quantitative viewpoint, the nLCC values for the corresponding drug and target networks of 0.97 and 0.93, respectively, are now much closer to the values of 0.99 and 0.94 obtained from the analogous randomised networks. In fact, only 8 targets are left outside the largest connected component derived from all experimental interaction data available, those being the two endothelin type A and B receptors, the gonadotropin-releasing, follicle-stimulating, and luteinizing hormone receptors, and the purinoreceptor,[11] in addition to a calcium-sensing and a taste receptor. Overall, these analyses illustrate clearly the implicit dependence of the topology of drug–target networks on target families which, beyond the effect that the properties of bioactive molecules might have, are a reflection of how the intrinsic phylogenetic relationships among the targets of a given family translate into cross-pharmacologies between them.

Conclusions

It is a fact that our conventional notion of target selectivity should be reconsidered and recognised as being as truthful as our limited degree of knowledge on the complete pharmacological profile of drugs permits. In this respect, data completeness is a critical issue to our understanding of drug–target interactions and more efforts should be devoted to systematically derive complete interaction matrices. In the meantime, caution is advised on any conclusion derived from analyses performed uniquely on currently available drug–target interaction data.

On the basis of current interaction data, but largely complemented also with the projections performed by means of *in silico* target profiling methods, a complex picture of drug–target interactions is starting to emerge. Within this context, network theory offers both an illustrative and quantitative framework with which to analyse this complexity. In this work, network approaches were used to analyse the implicit dependence of the topology of drug–target interaction networks on drug properties and target families. A remarkable convergence of the results was obtained. From the analysis of drug properties, it was observed that small hydrophobic drugs appear to be significantly more promiscuous than large hydrophilic drugs, whereas from the analysis of target families, the drug–target interaction network composed by drugs targeting GPCRs was found to be significantly more connected than any of the networks derived for the other target families. Both findings converge when realising that small hydrophobic drugs are known to interact primarily with GPCRs, whereas the majority of the large hydrophilic drugs appear to be enzyme inhibitors.

These results may have potential implications for drug discovery. On one hand, the properties of the chemical space surrounding drugs may be linked implicitly to some degree of target promiscuity. On the other hand, drug properties seem to determine also to a certain extent the scope of the biological space that could be potentially targeted by those drugs. Thus, ultimately, target spaces may also be implicitly linked to some basal degree of target promiscuity. Understanding the complexity associated with the implicit dependencies of promiscuity on drug properties and target families may facilitate our quest to generate safer, more efficient drugs.

Databases and methods

Annotated chemical libraries

Drug–target interaction data available experimentally were extracted from a representative sample of annotated chemical libraries (Table 1). It included DrugBank, the major public resource of drug–target interaction data;[5] Wombat, a commercial collection of small molecules with known biological activity from medicinal chemistry literature;[16] BindingDB, a publicly accessible database of experimentally determined binding affinities of protein–ligand complexes;[17]

and PDSP K_i database, a repository of experimental K_i data on receptors available from the Psychoactive Drug Screening Program.[18] These four databases were supplemented with three additional annotated libraries assembled internally and directed to cytochromes (CYPdb), G protein-coupled receptors (GPCRdb) and nuclear receptors (NRdb).[19] In total, 6284 drug–target interactions between 802 drugs and 480 targets were collected (Table 1). Over 40% of those interactions are reported in two or more databases, which results in a total number of 4767 unique drug–target interactions. Of those, 3749 interactions are found exclusively in one database only, with PDSP (39.0%), DrugBank (24.2%), and Wombat (21.6%) gathering almost 85% of them.

In silico target profiling

Several approaches to *in silico* pharmacology have appeared in recent years, some exploiting the vast amount of information on bioactive ligands (ligand-based methods), others making use of protein structures (structure-based methods).[23] Additional drug–target interaction data generated computationally were obtained here using a ligand-based target profiling method developed recently in our laboratory.[11] Our approach relies on the assumption that the set of bioactive ligands collected for a given target provides in fact a complementary description of the target from a ligand perspective. In order to be able to process this information efficiently, molecular structures need to be encoded using some sort of mathematical descriptors. In this respect, a novel set of low-dimension molecular descriptors called SHED were used.[24] SHED are derived from distributions of atom-centered feature pairs extracted directly from the topology of molecules. The collection of SHED values reflecting the overall distribution of pharmacophoric features in a molecule constitutes its SHED profile. The ensemble of SHED profiles representing all bioactive molecules annotated to a particular protein target constitutes a mathematical description of the target from a chemical perspective. Ligand-based descriptor models were derived for 1491 targets using a total of 156 749 unique bioactive molecules stored in the annotated chemical libraries described above. The probability of a drug interacting with a particular target is assumed to be related to the degree of similarity relative to the set of known bioactive ligands for that target. Accordingly, Euclidean distances between the SHED profile of a drug and all SHED profiles associated with a target are first calculated and a final target scoring is given by the minimum value of all Euclidean distances.[25] Following previous validation analyses,[11,25] if that minimum Euclidean distance is below 0.6, a new drug–target interaction is assigned.

Acknowledgements

This research was funded by the Spanish Ministerio de Ciencia y Educación (BIO2008-02329) and the Instituto de Salud Carlos III.

References

1 B. L. Roth, D. J. Sheffler and W. K. Kroeze, *Nat. Rev. Drug Discovery*, 2004, **3**, 353–359.
2 A. Weber, A. Casini, A. Heine, D. Kuhn, C. T. Supuran, A. Scozzafava and G. Klebe, *J. Med. Chem.*, 2004, **47**, 550–557.
3 G. F. Sud'ina, M. A. Pushkareva, P. Shephard and T. Klein, *Prostaglandins, Leukotrienes Essent. Fatty Acids*, 2008, **78**, 99–108.
4 T. Kvernmo, J. Houben and I. Sylte, *Curr. Top. Med. Chem.*, 2008, **8**, 1049–1067.
5 D. S. Wishart, C. Knox, A. C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam and M. Hassanali, *Nucleic Acids Res.*, 2008, **36**, D901–D906.
6 M. Cases and J. Mestres, *Drug Discovery Today*, 2009, **14**, 479–485.
7 S. Whitebread, J. Hamon, D. Bojanic and L. Urban, *Drug Discovery Today*, 2005, **10**, 1421–1432.
8 L. J. Jensen, J. Saric and P. Bork, *Nat. Rev. Genetics*, 2006, **7**, 119–129.
9 N. P. Savchuk, K. V. Balakin and S. E. Tkachenko, *Curr. Opin. Chem. Biol.*, 2004, **8**, 412–417.
10 J. Mestres, E. Gregori-Puigjané, S. Valverde and R. V. Solé, *Nat. Biotechnol.*, 2008, **26**, 983–984.
11 E. Gregori-Puigjané and J. Mestres, *Comb. Chem. High Throughput Screening*, 2008, **11**, 669–676.
12 G. V. Paolini, R. H. B. Shapland, W. P. van Hoorn, J. S. Mason and A. L. Hopkins, *Nat. Biotechnol.*, 2006, **24**, 805–815.
13 M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin and B. K. Shoichet, *Nat. Biotechnol.*, 2007, **25**, 197–206.
14 M. A. Yıldırım, K.-I. Goh, M. E. Cusick, A.-L. Barabási and M. Vidal, *Nat. Biotechnol.*, 2007, **25**, 1119–1126.
15 V. Batagelj and A. Mrvar, *Connections*, 1998, **21**, 47–57.
16 M. Olah, M. Mracec, L. Ostopovici, R. Rad, A. Bora, N. Hadaruga, I. Olah, M. Banda, Z. Simon and T. I. Oprea, in *Chemoinformatics in Drug Discovery*, ed. T. I. Oprea, Wiley-VCH, New York, 2004, pp. 223–239.
17 T. Liu, Y. Lin, X. Wen, R. N. Jorissen and M. K. Gilson, *Nucleic Acids Res.*, 2007, **35**, D198–D201.
18 B. L. Roth, E. Lopez, S. Beischel, R. B. Westkaemper and J. M. Evans, *Pharmacol. Ther.*, 2004, **102**, 99–110.
19 M. Cases, R. García-Serna, K. Hettne, M. Weeber, J. van der Lei, S. Boyer and J. Mestres, *Curr. Top. Med. Chem.*, 2005, **5**, 763–772.
20 M. M. Hann, A. R. Leach and G. Harper, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 856–864.
21 A. L. Hopkins, J. S. Mason and J. P. Overington, *Curr. Opin. Struct. Biol.*, 2006, **16**, 127–136.
22 M. Vieth and J. J. Sutherland, *J. Med. Chem.*, 2006, **49**, 3451–3453.
23 S. Ekins, J. Mestres and B. Testa, *Br. J. Pharmacol.*, 2007, **152**, 9–20.
24 E. Gregori-Puigjané and J. Mestres, *J. Chem. Inf. Model.*, 2006, **46**, 1615–1622.
25 J. Mestres, L. Martin-Couce, E. Gregori-Puigjané, M. Cases and S. Boyer, *J. Chem. Inf. Model.*, 2006, **46**, 2725–2736.