ELSEVIER

Drugs in other drugs: a new look at drugs as fragments

Miles G. Siegel and Michal Vieth*

Discovery Chemistry Research, Lilly Research Laboratories, Lilly Corporate Center, DC 1513 Indianapolis, IN 46285, USA

The authors of this review have examined the complete set of marketed drugs, with regards to looking for structural similarities between drugs. By comparing the structures of all drugs, it has been established how many times one marketed drug occurred as a substructure within another marketed drug. A total of 209 from 1386 marketed drugs sized between 100 and 1500 Da (i.e. 15% of the 1386 total) are contained within other drugs, differing by one or more continuous chemical fragment, and as many as 418 drugs from the total of 1386 (i.e. 30%) contain other drugs as substructure fragments. Many smaller drugs occur in multiple larger drugs. Most of the small changes tend to retain primary indicated pharmacology, whereas larger changes more often lead to different primary pharmacology. We identify a subset of drugs that can be used in fragment-based drug discovery strategies. In addition, the analysis enhances understanding of marketed drug space from the chemical building-block perspective.

Introduction

Many recent analyses of marketed oral drugs provide clues to the medicinal chemist for the development of new pharmaceuticals [1–5]. A lot of this research is focused on defining and using druglike property space, distinguishing drugs from non-drugs, and the evolution of drug properties over time [2,3,6]. The focus on use of the information present in drug databases began with the breakthrough work of Lipinski and co-workers [7], who concluded the distribution of molecular properties in compounds with favorable permeation properties could be used to better guide medicinal chemists in creating new, chemical druglike space. The Rule of Five [7], and related follow-up work [3,5,7–12], has helped to define the boundaries of druglike property space. Subsequent work has attempted to define a corresponding leadlike property space for molecules that serve as starting points in drug discovery efforts [13,14]. The druglike and leadlike property spaces, although different, cover overlapping ranges [12]. There have also been some attempts to organize known drugs from the scaffold and building block perspectives, and to link these to observed property trends [3].

Fragment-based drug discovery efforts have occurred in parallel with the analyses of druglike and leadlike property space [3,15,16]. Fragment-based approaches use growth or linking paradigms of small molecules to create larger ones [15,17], with some parallel attempts to use building blocks that are present in existing marketed drugs [18]. The authors of this article recently presented the analysis of marketed drugs from the perspective of properties as a function of route of administration [3], proteomic family [5] and structural fragments [19]. In many cases, molecular fragments of drugs contain biological activity, and it is not unusual for fragments of marketed drugs to be marketed drugs in their own right. In this review, we attempt to shed light on drug interrelationships by comparing and contrasting small- and large-fragment differences between existing drugs. We do this by examining the subset of marketed drugs that wholly contain other marketed drugs as continuous substructure fragments (i.e. drugs containing other drugs; DCODs) and the corresponding drugs that are wholly contained in other drugs (i.e. drugs in other drugs; DIODs), and analyzing the DIOD–DCOD pairs in terms of the most common differences between the DIOD and DCOD, including structure, properties, pharmacology and indication.

Initial observations

The drug dataset used for this study was based on the one described in previous publications [3,5], with some modifications. Drugs <100 Da and >1500 Da were excluded from the analysis, as were

Corresponding author: Vieth, M. (m.vieth@lilly.com)
* The authors contributed equally to this work.

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2006.11.011

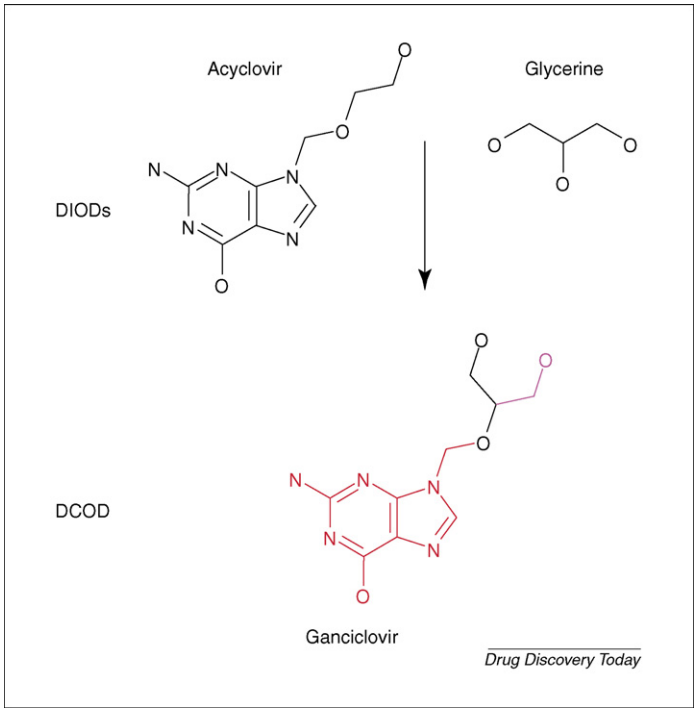


FIGURE 1

Example of two DIOD–DCOD pairs. Two small drugs (acyclovir and glycerine) are DIODs, and are contained in the same DCOD (ganciclovir). Color-coding in ganciclovir exemplifies the differences between ganciclovir and acyclovir (magenta), and gancilcovir and glycerine (red). The two pairs also exemplify cases where DIODs are introduced before the DCODs.

3% of drugs listed as investigational that, although available to physicians [20], were not marketed. Finally, to avoid the trivial relationship of drugs and prodrugs, all prodrugs and drugs containing carboxylic- and phosphonic-esters were removed from the dataset, leaving a total set of 1386 drugs. This also resulted in the elimination of interesting covalently linked drug dimers such as sultamicillin, which hydrolyzes into its constituent elements *in vivo*, a prodrug that resulted from the combination of the antibiotic ampicillin and the β-lactamase inhibitor sulbactam. We then performed substructure searches for each of the 1386 selected marketed drugs – against all other marketed drugs. Note that our searches do not identify cases where a single atom or group of atoms in one drug is mutated to a different atom, thus excluding some drugs that might be similar but do not share smaller drug substructure. From the searches, we identified 965 DIOD–DCOD pairs of drugs with one or more matches (Figure 1). We were able to assign structural differences between DIODs and DCODs to 642 of the DIOD–DCOD pairs automatically. The analysis of these 642 DIOD–DCOD pairs reveals that 209, or ~15%, of drugs in the set of 1386 examined are DIODs and 418, or 30%, of drugs in the set examined are DCODs; 565 (i.e. 41%) of drugs are either DIODs or DCODs and 62 (i.e. 4.5%) of drugs are mutually DIODS and DCODs.

We first examined the magnitude of the molecular weight (MW) change in an average DIOD–DCOD pair. The mean DIOD MW is 293 Da (the median MW is 272 Da) and the mean DCOD MW is 428 Da (the median MW is 358 Da). The mean MW difference within DIOD–DCOD pairs is 214 Da (i.e. higher than the simple 135 Da difference between the DIOD and DCOD because many

TABLE 1

Detailed statistics of (a) DIODs and (b) DCODs as a function of drug size for drugs with molecular weights (MWs) >100 Da

(a) DIODs

MW bin of DIODS	Number of DIODS	Number of DIOD–DCOD pairs	Mean number of pairs per DIOD	Maximum number of DCODs for one DIOD	Median MW change per pair	Median cLogP change per pair	Total number of drugs	% of DIODS
100–200	55	347	6	62	209	0.5	158	35
200–300	75	144	2	11	51	0.5	416	18
300–400	52	115	2	11	36	0	453	11
400–500	16	23	1.5	3	34	−0.2	198	8
500–1500	11	13	1	2	30	−0.2	161	7

(b) DCODs

MW bin of DCODs	Number of DCODs	Number of DIOD–DCOD pairs	Mean number of pairs per DCOD	Maximum DIODS for this DCOD	Median MW change per DCOD	Median clogP change per DCOD	Total number of drugs	% of DCODs
100–200	26	35	1	3	−30	0	158	16
200–300	90	129	1	7	−72	−0.5	416	22
300–400	142	207	1	6	−73	−0.3	453	31
400–500	77	111	1	4	−125	−0.5	198	39
500–1500	83	160	2	5	−654	0.25	161	52

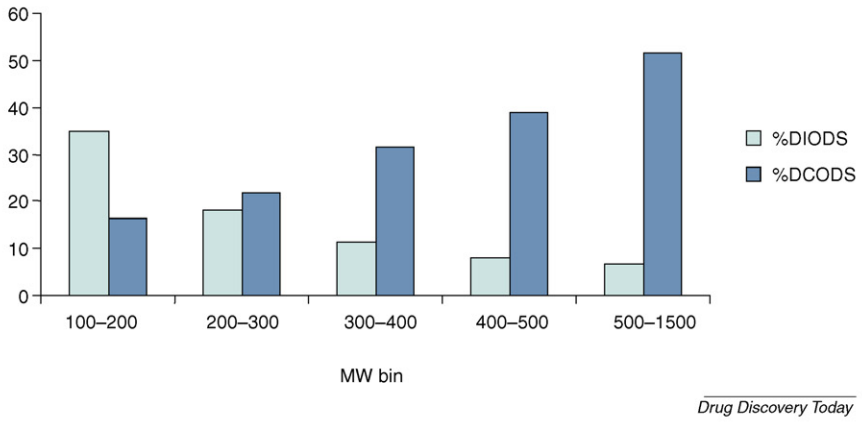


FIGURE 2

Graphical representation of the data from Table 1. The percentage of DIODs and DCODs in all drugs is shown for each molecular weight (MW) bin.

smaller DIODs are parts of multiple DCODs). On average, this suggests large structural changes between a DCOD and its contained DIOD.

To look at the size distribution of the DIOD and DCOD drugs we divided the 209 identified DIODs and 418 identified DCODs into 100 Da MW bins, starting at 100 Da (Table 1; Figure 2). Unsurprisingly, the percentage of DIODs increases as the MW bin decreases; 35% of drugs with MW <200 Da (i.e. 55 from 158 drugs) occur in other drugs. This group also shows by far the largest mean increase in MW and calculated logP (clogP) [21] between pairs, with MW more than doubling in most cases and clogP increasing by 0.5 units in the DCOD. By comparison, in most cases, the larger DIODs require a 30 Da increase in MW and a minimal, or no, clogP change to create larger drugs. Table 2a shows the 18 most frequently occurring DIODs. As can be seen in Table 2a, >67% of these DIODs (i.e. 13 from 18) are between 100 and 200 Da in MW.

Also as expected, the percentage of DCODs increases as MW increases (Table 1b; Figure 2). In particular, >57% (i.e. 132 from 233) of the drugs >500 Da contain other drugs. Interestingly, the mean increase in MW per DCOD drug is substantial for this category, with only a small clogP change. As can be seen from Table 2b, drugs containing the largest number of other drugs are moderately sized (i.e. 200–350 Da), which confirms the overall trend from Table 1b for the DCOD that contains the most DIODs.

DIODs versus DCODS: most-common drug-pair differences

The statistics for all matching pairs, as a function of the number of differing continuous fragments, is shown in Table 3. Of the 642 DIOD–DCOD drug pairs, 274 (i.e. 43%) differ by a single continuous fragment, 154 differ by two fragments and 101 differ by three fragments. We focused this part of our analysis on the 274 pairs that differ by only one continuous fragment. The high-level analysis of these pairs shows that there are 167 drugs that, upon addition of a single continuous chemical fragment, create another 243 DCOD drugs. A total of 132 unique fragments account for drug differences in these 274 pairs. Figure 3 presents the statistics of the most frequently occurring fragments for all pairs in this group. A single atom change [e.g. carbon, 42 pairs; oxygen (hydroxyls or carbonyl), 31 pairs; chlorine, 16 pairs; or fluorine, 11 pairs], accounts for ~40% of all single-fragment changes that differentiate one drug from another.

DIODs and DCODs as a function of MW difference: pharmacological implications

To characterize the profile of DIOD–DCOD pairs with respect to MW differences, we divided DIOD–DCOD pairs into groups with small MW differences (i.e. a change in MW <150 Da) versus pairs with greater MW differences (i.e. MW >150 Da). Our hypothesis was that many DIOD–DCOD pairs were caused by minor changes in one drug, and we anticipated that if this was the case the DIOD–DCOD pairs with MW changes <150 Da would substantially outnumber the pairs with larger MW changes. Interestingly, 166 DCODs (i.e. 40% from a total of 418) seem to result from larger differences from DIODs – trivial and nontrivial additions are equally probable. This observation has interesting implications in a fragment-based approach to drug discovery, where a smaller fragment (e.g. in this case the DIOD drug) serves as the starting point for building a larger, pharmacologically relevant molecule.

We then analyzed indication differences between DIOD–DCOD pairs as a function of MW difference. We anticipated that, as the MW difference increased, the likelihood that the DIOD and the DCOD would display different pharmacology would increase. To understand these differences, we chose to look at the primary indicated pharmacology [5]; we used the online databases Micromedex (http://www.micromedex.com), Drugbank (http://redpoll.pharmacy.ualberta.ca/drugbank/) or PharmGKB (http://www.pharmgkb.org/). In the analysis, the small-MW-change group (i.e. <150 Da) changed indication <18% of the time, whereas the large-MW-change group (i.e. >150 Da) changed indication in almost 53% of cases. This large difference is further-magnified on closer examination of the two groups. In the small-MW-change group the apparent difference in pharmacology often results not from an actual change in receptor-binding profile of the DIOD as it is grown into the DCOD but from a DIOD or DCOD that displays poorly defined, unknown or polypharmacology [22]. For

example, amphetamine, a stimulant, is a DIOD for six DCODs with different listed pharmacologies. Amphetamine itself, however, is known to exhibit a variety of effects, including norepinephrine- and dopamine-transporter inhibition, as well as adrenergic activation. Amphetamine DCODs such as dexfenfluramine and phentermine (i.e. anorectics) or bupropion (i.e. an antidepressant) probably exert their effects for these indications via these norepinephrine and dopamine transporters (Table 4a).

TABLE 2

Most frequently occurring DIODs and DCODs

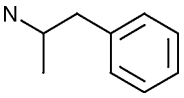
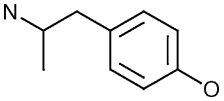
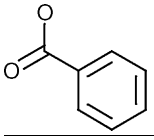
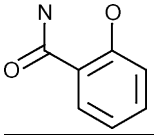
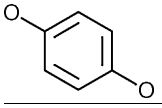
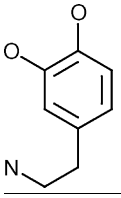
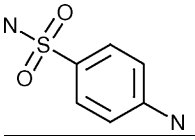
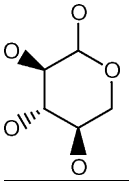
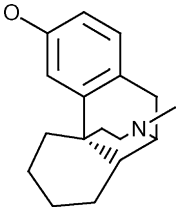
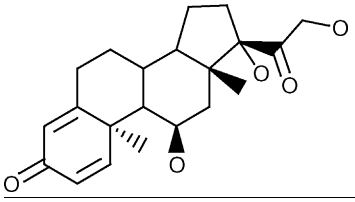
Structure	MW	Name	Number of DIOD–DCOD pairs
(a) DIODs[a]			
	135	Amphetamine sulfate	62
	151	Hydroxyamphetamine hydrobromide	34
	122	Benzoic acid	26
	137	Salicylamide	20
	110	Hydroquinone	15
	153	Dopamine hydrochloride	14
	172	Sulfanilamide	14
	150	Xylose	12
	257	Levorphanol tartrate	11
	360	Prednisolone	11

TABLE 2 (*Continued*)

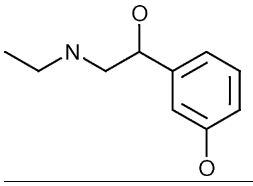
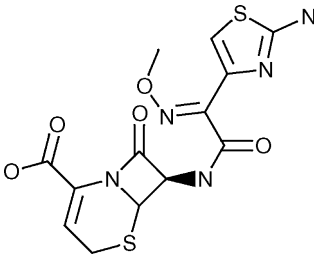
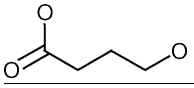
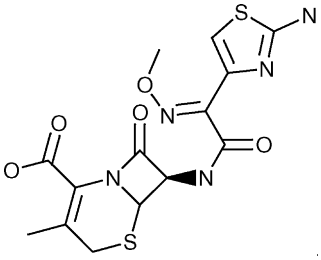
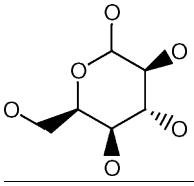
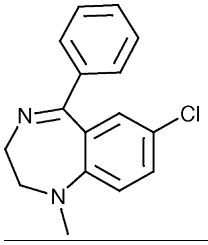
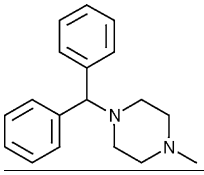
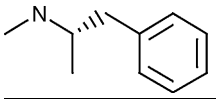
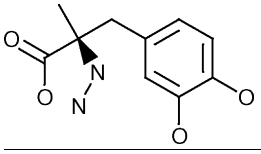
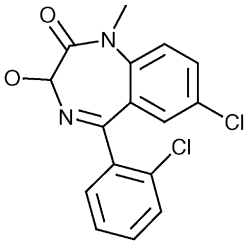
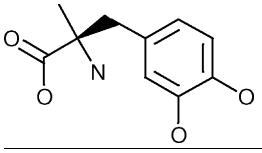
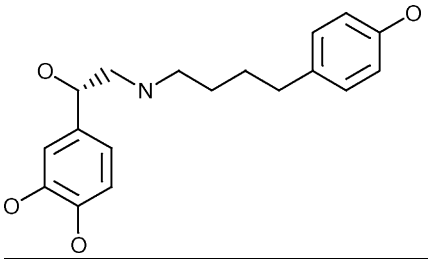
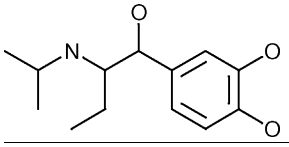
Structure	MW	Name	Number of DIOD–DCOD pairs
	181	Etilefrine	10
	383	Ceftizoxime	9
	104	Sodium oxybate	9
	397	Cefetamet	8
	180	Dextrose	8
	271	Medazepam	8
	266	Cyclizine lactate	7
	149	Methamphetamine hydrochloride	7
(b) DCODs			
	226	Carbidopa	7

TABLE 2 (*Continued*)

Structure	MW	Name	Number of DIOD–DCOD pairs
	335	Lormetazepam	6
	211	Methyldopa	5
	317	Arbutamine hydrochloride	5
	239	Isoetharine hydrochloride	5

[a] Only DIODs that differ from DCODs by less than five fragments were included. Trientine, because it differs by eight or more fragments from many polypeptide DCODs, was omitted.

TABLE 3

Statistics for DIOD–DCOD pairs for drugs broken down by the number of fragments the pairs differ by[a]

Number of fragments	Number of DIOD–DCOD pairs	% of pairs from total of 642	Number of unique DIODs	Number of unique DCODs
1	274	43	167	243
2	154	24	80	136
3	101	16	35	92
4	69	11	21	59
5	16	2	7	13
6	3	0	3	3
7	3	0	1	3
8	20	3	1	20
9	1	0	1	1
12	1	0	1	1

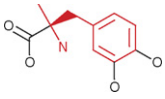
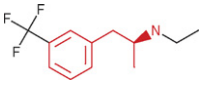
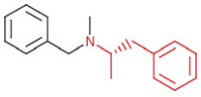
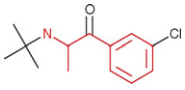
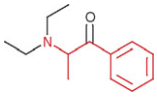
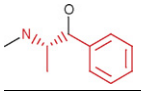
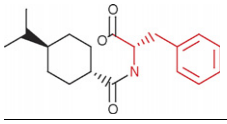
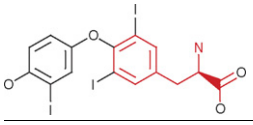
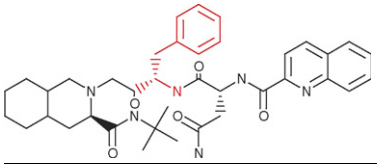
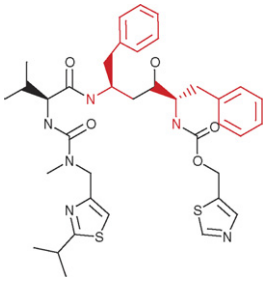
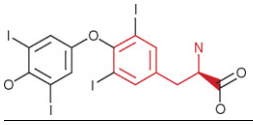
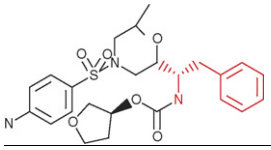
[a] Most of the DIODs with more than four fragment changes from DCODs are of low interest (e.g. trientine accounts for all eight fragment changes and is an uninteresting polypeptide drug substructure).

By contrast, most of the DIOD–DCOD pairs in the high-MW-change category displayed major differences in their biological targets (Table 4b). Taking the example of amphetamine, several DCODs with MW changes >150 Da incorporate amphetamine as a phenylalanine within a protease inhibitor molecule.

Next, we examined structural differences for the most common DIODs in the low- and high-MW-change classes. The most frequently occurring DIODs between these two groups are almost identical. Furthermore, all of the most common DIODs, including amphetamine, benzoic acid, hydroxyamphetamine, oxybate, salicynamide, hydroxyquinone and xylose, have a MW <250 Da. Clearly, several of these DIODs vary dramatically in physical properties from their DCOD counterparts; for example, trientin, a polybasic tetramine, is different in structure, physical properties and pharmacophoric elements from the many peptide drugs in which this substructure is found. Other DIODs such as etidronate display specific pharmacophoric elements crucial for a particular class of drugs (e.g. in this case the bisphosphonates), but they are

TABLE 4

Exemplification of pairs for amphetamine for small (a) and large (b) differences

DCOD structure	DCOD name	Platform and/or indication
(a) Small differences		
	Methyldopa	G-protein-coupled receptor (GPCR)-biogenic amine Hypotensives, sympathomimetics-α
	Dexfenfluramine	Transporter Anorectics, serotoninergics
	Benzphetamine hydrochloride	GPCR-biogenic amine Anorectics, psychostimulants, psychotonics
	Bupropion-hydrochloride	Transporter Antidepressants, psychostimulants
	Amfepramone	Transporter Anorectics
	Pseudoephedrine-hydrochloride	GPCR-biogenic amine Antiasthmatics, bronchodilators, vasoconstrictors, sympathomimetics-β, sympathomimetics-α
(b) Large differences		
	Nateglinide	Ion channel
	Liothyronine	Nuclear hormone receptor (NHR)
	Saquinavir	Protease
	Ritonavir	Protease
	Dextrothyroxine-sodium	NHR
	Amprenavir	Protease

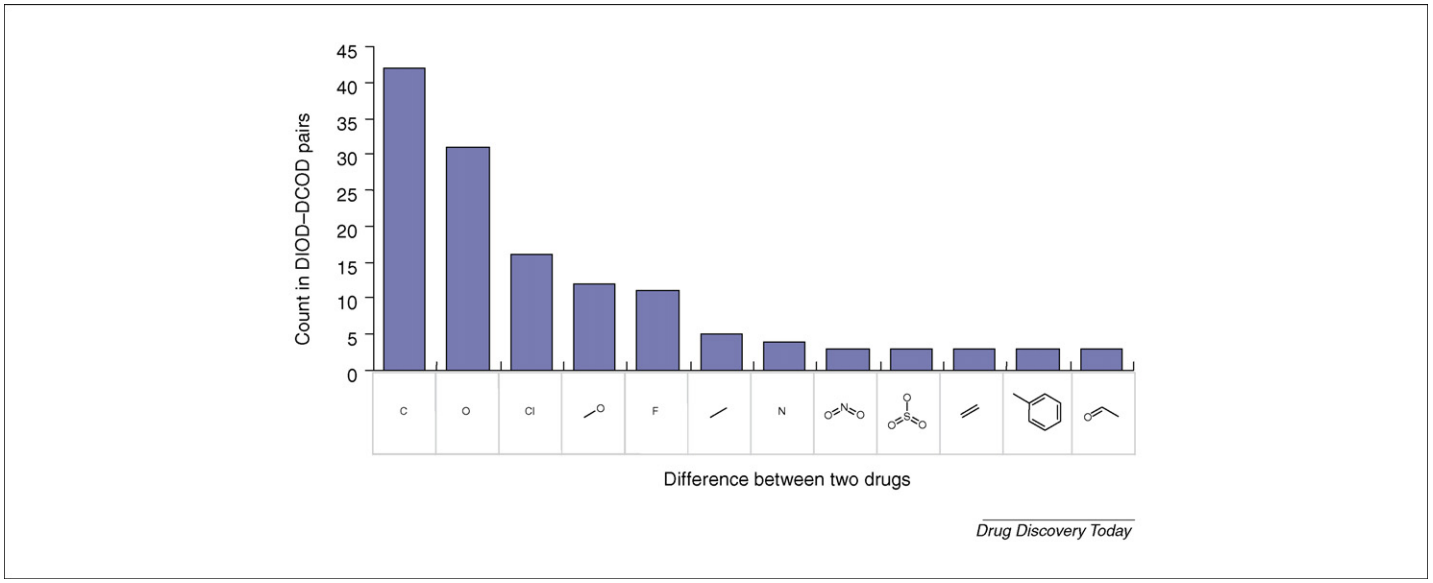


FIGURE 3

Histogram of the most common fragments differentiating DIOD–DCOD pairs that differ by one fragment. Only fragments differentiating four or more pairs are shown. The height of each histogram corresponds to the number of DIOD–DCOD pairs that differ by the fragment pictured directly below each histogram bar.

probably undesirable or irrelevant in a broader drug discovery context. The remaining DIODs, however, seem to be subelements within larger DCOD pharmacophores, and might, therefore, be considered as core entities in fragment-based discovery efforts.

Concluding remarks

We have examined the set of DIOD–DCOD pairs with particular emphasis on structural and therapeutic differences within pairs. Unsurprisingly, many common, structural fragment differences between drugs are small changes, often just a single carbon, oxygen, chlorine or fluorine. These small structural changes frequently lead to drug pairs occurring in the same therapeutic class. In many of these cases, even when the listed target or mechanism of action differs between DIOD and DCOD, the difference can be attributable either to polypharmacology or to an unknown or more-speculative mechanism of action in one or both of the drugs, rather than to a genuine modification of the receptor-binding profile. Larger structural changes in the order of >150 Da, however, are common, and lead more-frequently to drug pairs with genuinely different therapeutic endpoints. Many of these smaller DIOD drugs can, therefore, be considered as biologically active fragments that are capable of producing varied but specific biological activity, where the particular activity observed is dependent on the overall molecular environment of the fragment.

The observations mentioned in this review regarding DIODs can have relevance in fragment-based discovery strategies. The analysis provides another perspective on use of existing drugs to selective optimization of side activities (SOSA) and chemogenomics [23–25] paradigms, and adds merit to the statement, by 1994 Nobel Prize Laureate in Physiology and Medicine Sir James Black, that '. . .the most fruitful basis for the discovery of a new drug is to start with an old drug' [26]. In this study, we only consider cases where one drug is wholly incorporated in another; a less-strict search accounting for atom mutations or partial deletions would certainly produce additional drug pairs for further analysis.

Supplementary material available

DIOD and DCOD pairs are available from the authors by request.

Acknowledgement

We would like to thank Drs Ian Watson and Jibo Wang for providing tools to help in the automatic assignment of differences between drug pairs, and Dr Sutherland for improving the quality of this review. In addition, we want to acknowledge the referees for their constructive comments, which improved the quality of the article.

References

1 Lipinski, C. and Hopkins, A. (2004) Navigating chemical space for biology and medicine. *Nature* 432, 855–861
2 Leeson, P.D. and Davis, A.M. (2004) Time-related differences in the physical property profiles of oral drugs. *J. Med. Chem.* 47, 6338–6348
3 Vieth, M. *et al.* (2004) Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* 47, 224–232
4 Wenlock, M.C. *et al.* (2003) A comparison of physiochemical property profiles of development and marketed oral drugs. *J. Med. Chem.* 46, 1250–1256
5 Vieth, M. and Sutherland, J.J. (2006) Dependence of molecular properties on proteomic family for marketed oral drugs. *J. Med. Chem.* 49, 3451–3453
6 Proudfoot, J.R. (2005) The evolution of synthetic oral drug properties. *Bioorg. Med. Chem. Lett.* 15, 1087–1090
7 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25
8 Lipinski, C.A. (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* 44, 235–249

9 Oprea, T.I. (2002) Current trends in lead discovery: Are we looking for the appropriate properties. *J. Comp-Aided Mol. Des.* 16, 325–334
10 Muegge, I. (2003) Selection criteria for drug-like compounds. *Med. Res. Rev.* 23, 302–321
11 Blake, J.F. (2003) Examination of the computed molecular properties of compounds selected for clinical development. *Biotechniques* 34, S16–S20
12 Lajiness, M. *et al.* (2004) Molecular properties that influence oral drug-like behavior. *Curr. Opin. Drug Discov. Devel.* 7, 400–407
13 Hann, M.M. *et al.* (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* 41, 856–864
14 Hann, M.M. and Oprea, T.I. (2004) Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* 8, 255–263
15 Erlanson, D.A. *et al.* (2004) Fragment-based drug discovery. *J. Med. Chem.* 47, 3463–3482
16 Zartler, E.R. and Shapiro, M.J. (2005) Fragonomics: fragment-based drug discovery. *Curr. Opin. Chem. Biol.* 9, 366–370
17 Rees, D.C. *et al.* (2004) Fragment-based lead discovery. *Nat. Rev. Drug Discov.* 3, 660–672
18 Lewell, X.Q. *et al.* (1998) RECAP–retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* 38, 511–522
19 Vieth, M. and Siegel, M. (2006) Structural fragments in marketed oral drugs. In *Fragment based drug discovery* (Erlanson, J., ed.), pp. 123–135, Wiley-VCH
20 MICROMEDEX, T. (2003) MICROMEDEX(R) healthcare series (http://www.micromedex.com)
21 Daylight chemical information systems, I. (2002) CLOGP (http://www.daylight.com)
22 Paolini, G.V. *et al.* (2006) Global mapping of pharmacological space. *Nat. Biotechnol.* 24, 805–815
23 Wermuth, C.G. (2001) The ''SOSA'' approach - an alternative to high throughput screening. *Med. Chem. Res.* 10, 431–439
24 Wermuth, C.G. (2004) Selective optimization of side activities: another way or drug discovery. *J. Med. Chem.* 47, 1303–1314
25 Caron, P.R. *et al.* (2001) Chemogenomic approaches to drug discovery. *Curr. Opin. Chem. Biol.* 5, 464–470
26 Raju, T.N. (2000) The Nobel chronicles. *Lancet* 355, 1022