Relating protein pharmacology by ligand chemistry

Michael J Keiser[1,2], Bryan L Roth[3,4], Blaine N Armbruster[4], Paul Ernsberger[3], John J Irwin[1] & Brian K Shoichet[1]

The identification of protein function based on biological information is an area of intense research. Here we consider a complementary technique that quantitatively groups and relates proteins based on the chemical similarity of their ligands. We began with 65,000 ligands annotated into sets for hundreds of drug targets. The similarity score between each set was calculated using ligand topology. A statistical model was developed to rank the significance of the resulting similarity scores, which are expressed as a minimum spanning tree to map the sets together. Although these maps are connected solely by chemical similarity, biologically sensible clusters nevertheless emerged. Links among unexpected targets also emerged, among them that methadone, emetine and loperamide (Imodium) may antagonize muscarinic M3, α2 adrenergic and neurokinin NK2 receptors, respectively. These predictions were subsequently confirmed experimentally. Relating receptors by ligand chemistry organizes biology to reveal unexpected relationships that may be assayed using the ligands themselves.

It is a curious pharmacological fact that related drugs and biological messengers can bind to receptors that appear unrelated by many bioinformatics metrics. For instance, serotonin and serotonergic drugs bind to G-protein coupled receptors (GPCRs) such as the 5-hydroxytryptamine subtypes 1, 2 and 4–7 (5-HT$_{1,2,4–7}$), but also to an ion channel, the 5-HT$_{3A}$ receptor[1,2]. Ionotropic and metabotropic 5-HT receptors are unrelated by sequence and structure, yet both are involved in the pharmacological effects of serotonergic drugs. Similarly, the well-known opioid methadone binds not only to the μ-opioid receptor, a GPCR, but also to the *N*-methyl-D-aspartic acid (NMDA) receptor[3], an ion channel, and both are thought to be involved in the drug's biological activity[4]. Benzodiazepines affect

mitochondrial proteins in addition to their primary therapeutic actions on ion channels[5]. The enzymes thymidylate synthase (TS), dihydrofolate reductase (DHFR) and glycinamide ribonucleotide formyltransferase (GART) all recognize folic acid derivatives and are inhibited by antifolate drugs. Despite this, the three enzymes have no substantial sequence identity and are structurally unrelated. This disregard for typical biological categories on the part of small molecules can lead to infamous side effects—although cisapride stimulates 5-HT$_4$ receptors and astemizole inhibits histamine H$_1$ receptors, both also inhibit the hERG ion channel, leading to unexpected cardiac pathologies[6]. The ability of chemically similar drugs to bind proteins without obvious sequence or structural similarity can confound a purely biological logic to understanding and categorizing their action.

A chemo-centric approach to this problem is to compare not the biological targets themselves but rather the chemistry of their ligands[7]. The motivating hypothesis is that two similar molecules are likely to have similar properties[8], and will bind to the same group of proteins. Whereas this hypothesis may be violated in specific cases—a small change in chemical structure can dramatically change binding affinity—chemical similarity is often a good guide to the biological action of an organic molecule[9]. Indeed, chemical similarity is a central principle in ligand design[10], and an extensive chemoinformatic literature explores many methods to compare pairs of ligands for such similarity[11]. Recently, Hopkins and colleagues found that using the simplest form of chemical similarity—full chemical identity among ligands shared by two or more receptors—linkage maps can be calculated to relate targets[12]. Vieth and colleagues, using a different approach, have used dendrograms of inhibitors to organize the selectivity relationships among kinases[13]. Izrailev and Farnum have also linked ligand sets by focusing on the most similar molecules between them[14]. These and recent efforts in predicting pharmacologic profiles[15–19] have led to the development of probabilistic models to predict polypharmacology and assess the 'druggability' of protein targets.

Here we investigate techniques to relate receptors to each other quantitatively based on the chemical similarity among their ligands. In this method, which we call the Similarity Ensemble Approach (SEA), two sets of ligands are often judged similar even though no single identical ligand is shared between them. We use a collection of about 65,000 ligands annotated for drug targets, where most annotations contain hundreds of ligands. To compare sets without size or chemical composition bias, we introduce a technique that corrects for the chemical similarity we might expect between ligand sets at random, using a model resembling that of BLAST[20–22]. This technique enables

[1]Department of Pharmaceutical Chemistry, University of California San Francisco, 1700 4th St, San Francisco California 94143-2550, USA. [2]Biological and Medical Informatics, University of California San Francisco, 1700 4th St., San Francisco, California 94143-2550, USA. [3]Departments of Biochemistry and Nutrition and National Institute of Mental Health Psychoactive Drug Screening Program, Case Western Reserve University Medical School, 2109 Adelbert Road, Cleveland, Ohio 44106, USA. [4]Department of Pharmacology and Division of Medicinal Chemistry and Natural Products (BLR), The University of North Carolina Chapel Hill Medical School, Chapel Hill, North Carolina 27705, USA. Correspondence should be addressed to B.K.S. (shoichet@cgl.ucsf.edu) or J.J.I. (irwin@cgl.ucsf.edu).
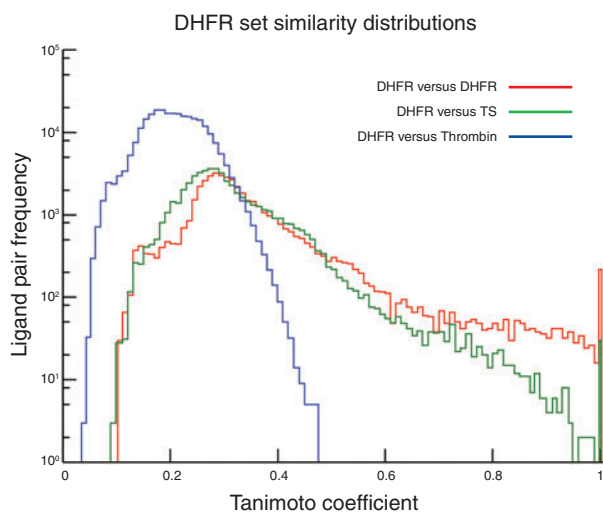
DHFR set similarity distributions

Figure 1 Comparing similar and dissimilar ligand sets to that of DHFR. Log-scale distributions of ligand-ligand similarity for different ligand sets: DHFR ligands compared to themselves (red), DHFR ligands compared to the related thymidylate synthase (TS) ligands (green), and DHFR ligands compared to the unrelated thrombin ligands (blue). The Tc ranges from 0 (complete dissimilarity) to 1 (identity). The ligand sets were derived from MDDR annotations.

us to link hundreds of ligand sets—and correspondingly the protein targets—together in minimal spanning trees. Whereas these trees are calculated by chemical similarity, recognizable clusters of biologically related proteins emerge from them. We consider the origins and possible significance of both the recognized and unexpected relationships, and their use for uncovering side effects and polypharmacology of individual chemical agents. We test several such unexpected relationships in biochemical and cell-based assays.

RESULTS

Similarity scores between ligand sets. We used a 246-receptor subset of the MDL Drug Data Report (MDDR), which annotates ligands according to the receptor whose function they modulate. Each ligand in each set was compared to each ligand in every other set. Overall, 246 versus 246 set comparisons were made, involving 65,241 unique ligands and 5.07×10^9 total ligand pairs. Tanimoto coefficients (Tc) of chemical similarity were calculated for each pair of ligands. For most ligand pairs the Tc was low, in the 0.2 to 0.3 range, which is typically considered insubstantial similarity. This was true even when comparing a set to itself. For instance, when comparing the 216 ligands of the antifolate enzyme DHFR to themselves, 80.4% of the pairs had a Tc in the 0.1 to 0.4 range, with only 4.7% having more substantial scores in the 0.6–1.0 range and only 0.5% having a Tc of 1.0 (only 216 ligands are, after all, identical) (**Fig. 1**). This pattern was also observed in comparing the 253 ligands of the antifolate enzyme TS to the DHFR ligands. Here only 0.06% of ligand pairs were identical (Tc of

1.0), 1.6% of pairs had Tc values of 0.6 to 1.0 and 85.5% had Tc values between 0.1 and 0.4. When the set of 1,226 ligands for the protease thrombin was compared to that of DHFR, a peak containing 97.1% of all pairs was observed between Tc values of 0.1 to 0.4, but no identical pairs were observed nor were there any ligand pairs that had Tc values >0.5. The raw similarity score, which is the sum of ligand pair Tcs over all pairs with Tc ≥ 0.57, between the DHFR and thrombin ligand sets was therefore 0; the raw score between DHFR and TS ligand sets was 772.25, whereas that of the DHFR set against itself was 1,931.60. This is consistent with the lack of similarity between the ligand sets of thrombin and DHFR and with the considerable similarity between the sets of TS and DHFR, both of which contain related antifolate drugs and their analogs.

Patterns of similarity. Most pairs of ligand sets resembled the TS versus thrombin comparison and had no raw score similarity. Of the 60,516 set pairs, 70.8% had raw scores of 0. As the size of the sets grew, however, the likelihood that two would have pairs of ligands with Tc ≥ 0.57 also grew. Indeed, there was a linear relation between the raw score and the number of ligands in the sets being compared (see **Supplementary Fig. 1** online). To compare the significance of the set similarity raw scores across sets of different sizes, we developed a statistical model of the similarity we would expect at random for sets drawn from the same large but finite database of ligands. This allowed us to calculate Z-scores and expectation values for any raw score for ligand sets of any size, such that the background fit an extreme value distribution (see **Supplementary Fig. 1c** online). As far as we know, a statistical model for random set similarity has not been previously used in chemoinformatics (although Z-scores have been used for comparisons of individual compounds[23,24]). As in sequence comparisons, the expectation values that such a model allows are critical for unbiased and quantitative comparison of multiple ligand sets. As would be expected, 95.2% of set-to-set comparisons had expectation values >1. The similarity of the overwhelming majority of ligand sets was thus no greater than what one would expect at random. Returning to the comparison of DHFR, TS and thrombin, the DHFR set versus itself had a Z-score of 333.4 and an expectation value of 7.07×10^{-182} (**Table 1**), suggesting very high similarity, whereas DHFR versus TS had a Z-score of 117.6 and an E-value of 1.11×10^{-61}. As DHFR versus thrombin did not yield a

Table 1 MDDR activity classes resembling MDDR "Dihydrofolate Reductase Inhibitor"

Rank	Activity class	E-value	Example molecule
1	DHFR inhibitor	7.07×10^{-182}	
2	Glycinamide ribonucleotide formyltransferase inhibitor	3.97×10^{-100}	
3	Folylpolyglutamate synthetase inhibitor	4.59×10^{-62}	
4	TS inhibitor	1.11×10^{-61}	

raw score >0, no Z-score was calculated and the comparison was unranked.

With a model of random similarity, we could compare statistically weighted versions of the raw scores for all pairs of sets. Even fewer sets had statistically significant similarity after correction for random expectation. On average, any given receptor was similar to only 5.8 other receptors with an expectation value $<10^{-10}$. Further down the rank-ordered list, the expectation values among targets fell off steeply, and within a few targets the similarity typically fell to insignificance. For example, the set of α-amino-5-hydroxy-3-methyl-4-isoxazole propionic acid (AMPA) receptor antagonists was highly similar to two other ligand sets: kainic acid antagonists and NMDA antagonists, with E-values of 5.28×10^{-80} and 3.08×10^{-63}, respectively. The third most significant ligand set was the anaphylatoxin receptor antagonists, with an E-value of 3.81×10^{-4}, and by the sixth ranked target the similarity was insignificant (E-value 1.00×10^{-1}, **Table 2**; for more detail see **Supplementary Table 1** online). Correspondingly, few targets were unrelated to any others; only 18 such orphans were found (see **Supplementary Table 2** online). A few targets were relatively promiscuous, with 14 being related to more than 10 other targets with expectation values $<10^{-50}$.

The similarity of ligand sets to small archipelagos of other ligand sets allowed us to calculate maps connecting almost all sets together through sequential linkage (**Fig. 2a**). In this map and in the sparser minimal spanning tree, where we connect only the most similar neighbors (**Fig. 2b**), clusters of biologically related targets may be observed as an emergent property, as no explicit biological information, only ligand information, is used to calculate the cross-target similarity. Thus, the glutamate receptors group together (**Fig. 2b**), and the steroids localize around androgen- and estrogen-receptor ligands (**Fig. 2b**, iv). Likewise, the folate, phosphodiesterase and β-lactam sets each colocalize and intraconnect (**Fig. 2b**). Conversely, whereas the serotonin metabotropic receptors cluster together, and ionotropic ligand receptors do so as well, the two receptor subtypes are distinct (**Fig. 2b**, ii and iii). Similar clustering may be observed in other regions of the map.

For this method to have wide utility, it is important that sets of ligands from different sources – for instance, not just from within the MDDR – can be compared. To test this, we built 23 ligand sets from 1,421 compounds in PubChem Compound that were not in the MDDR, organized by their MeSH Pharmacological Actions. We then queried these sets against our collection of 246 MDDR activity classes and ranked them by ligand-set pharmacological similarity (**Table 3**). Of the 23 PubChem query sets, 17 found a matching MDDR

Table 2 MDDR activity classes resembling five example MDDR activity classes

Query	Rank	Size	Similar activity classes	E-value	Tc 1.0	Max Tc
AMPA receptor Antagonist	1	569	AMPA receptor antagonist	2.45×10^{-219}	577	1.00
	2	75	Kainic acid receptor antagonist	5.28×10^{-80}	74	1.00
	3	1485	NMDA receptor antagonist	3.08×10^{-63}	181	1.00
	4	22	Anaphylatoxin receptor antagonist	3.81×10^{-4}	0	0.70
	5	130	μ agonist	1.69×10^{-3}	0	0.83
	6	99	Ribonucleotide reductase inhibitor	1.00×10^{-1}	0	0.73
Carbacephem	1	98	Carbacephem	0^a	106	1.00
	2	1614	Cephalosporin	1.11×10^{-222}	14	1.00
	3	35	Isocephem	2.30×10^{-17}	0	0.64
	4	257	Penem	2.43×10^{-4}	0	0.68
	5	13	Oxacephem	8.38×10^{-3}	0	0.69
	6	39	Lactam (β) antibiotic	2.62×10^{-2}	0	0.62
	7	223	Lactamase (β) inhibitor	6.58×10^{-1}	1	1.00
	8	116	Monocyclic β-lactam	3.18×10^{2}	0	0.61
Androgen	1	50	Androgen	0^a	138	1.00
	2	577	Aromatase inhibitor	6.87×10^{-307}	0	0.88
	3	43	Antiglucocorticoid	2.30×10^{-102}	0	0.89
	4	6	Cytochrome P450 oxidase inhibitor	4.01×10^{-93}	0	0.92
	5	179	Estrogen	9.97×10^{-89}	0	0.91
	6	86	Antiestrogen	2.18×10^{-76}	0	0.84
	7	936	Steroid (5α) reductase inhibitor	1.58×10^{-72}	0	0.80
	8	103	Antiandrogen	1.14×10^{-70}	0	0.99
	9	86	17α-hydroxylase/C17-20 lyase inhibitor	7.88×10^{-66}	0	0.76
	10	164	Progesterone antagonist	3.26×10^{-44}	0	0.89
	11	62	Prostaglandin	1.93×10^{-38}	0	0.75
5 HT1F Agonist	1	111	5 HT1F agonist	6.72×10^{-187}	113	1.00
	2	621	5 HT1D agonist	8.08×10^{-38}	0	0.95
	3	51	5 HT1B agonist	2.96×10^{-10}	0	0.95
	4	65	5 HT1 agonist	3.03×10^{-8}	0	0.81
	5	670	Dopamine (D4) antagonist	1.90×10^{-6}	0	0.79
	6	565	5 HT1A antagonist	8.64×10^{-1}	0	0.71
	7	33	5 HT2 antagonist	8.78×10^{-1}	0	0.65
	8	705	5 HT2A antagonist	1.47	0	0.73
Adrenergic (β1) Agonist	1	8	Adrenergic (β1) agonist	3.85×10^{-241}	10	1.00
	2	305	Adrenergic (β) agonist	9.50×10^{-34}	0	0.81
	3	67	Adrenergic (β1) blocker	4.99×10^{-32}	0	0.64
	4	563	Adrenoceptor (β3) agonist	2.98×10^{-24}	0	0.72
	5	212	Adrenergic (β) blocker	3.96×10^{-13}	0	0.78
	6	13	Adrenergic, ophthalmic	2.77×10^{-7}	0	0.70
	7	518	Adrenergic (α1) blocker	6.84×10^{-5}	0	0.73
	8	124	Melatonin agonist	1.04×10^{-1}	0	0.63
	9	76	Dopamine (D1) agonist	2.18×10^{-1}	0	0.71
	10	102	Adrenergic (α2) agonist	4.72×10^{-1}	0	0.66

^{a}E-value $< 10^{-320}$.

activity class as the top-ranked hit. When repeated using the mean pair-wise similarity (MPS)[14,25,26] of the sets instead of the statistically-corrected expectation values, only nine of the queries found a matching top-ranked hit. On average, a matching MDDR hit was found within the top 1.4 ranks of the PubChem queries' hit lists using pharmacological similarity (SEA), compared to within the top 8.2 ranks when ranked by MPS (see **Supplementary Table 3** online). This attests to the importance of a statistical control for similarities expected at random.
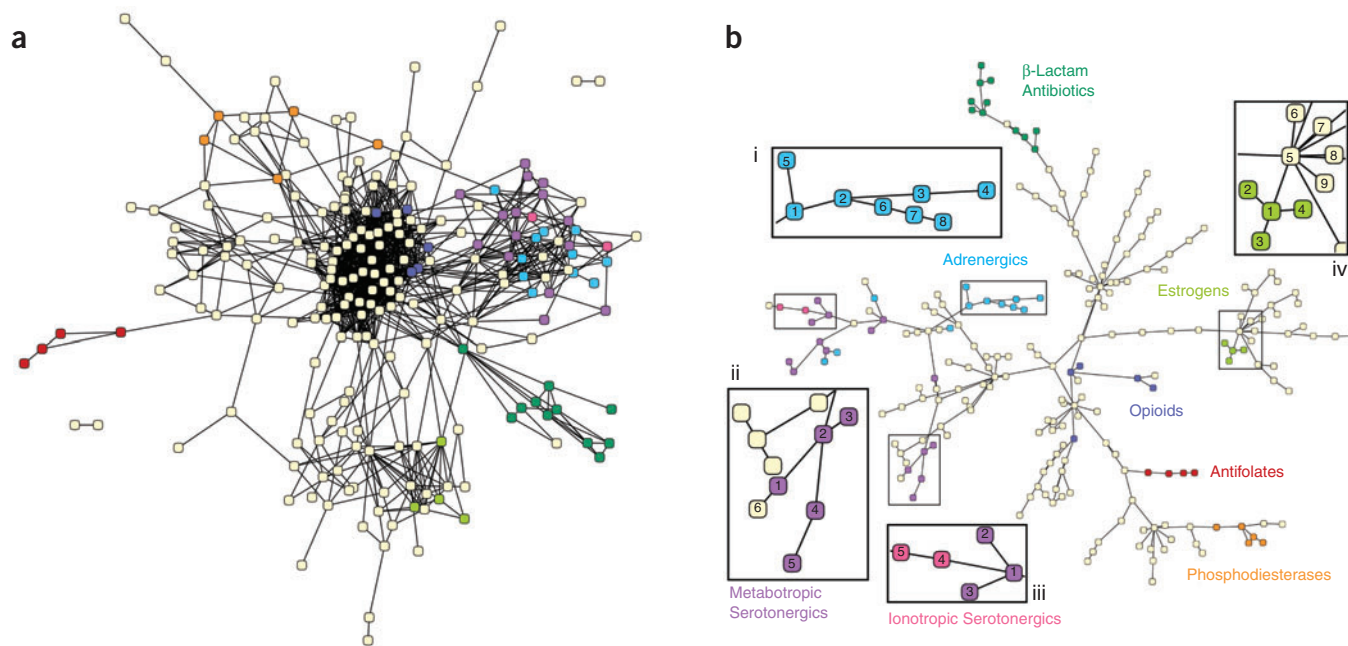
Figure 2 Similarity maps for 246 enzymes and receptors. (**a**) Network view of pharmacological space, in which each node represents a particular target in the MDDR. The nodes are colored for several pharmacologically related targets: antifolates (red), phosphodiesterases (orange), opioids (blue), β-lactam antibiotics (dark green), metabotropic serotonergics (violet), ionotropic serotonergics (pink), adrenergics (cyan) and estrogen modulators (light green). This network is a naive threshold graph that includes only edges that have expectation values <1. (**b**) A tree view of pharmacological space. This is an alternate view of the same network as in **a**, over which we have calculated a minimal spanning tree. This approach connects all nodes (protein targets) using only the most significant connections. The node coloring is the same as that in **a**. (i) Detailed view of adrenergics: β adrenergic agonists (1), β_1 adrenergic agonists (2), β_1 adrenergic blockers (3), β adrenergic blockers (4), β_3 adrenoceptor agonists (5), ophthalmic adrenergics (6), α_2 adrenergic agonists (7) and α_1 adrenoceptor agonists (8). (ii) Detailed view of metabotropic serotonergics subset: 5-HT_{1F} agonists (1), 5-HT_{1D} agonists (2), 5-HT_1 agonists (3), 5-HT_{1B} agonists (4) and 5-HT_{1D} antagonists (5). (iii) Detailed view of ionotropic (5-HT_3) serotonergics: 5-HT_4 agonists (1), 5-HT_4 antagonists (2), 5-HT_2 antagonists (3), 5-HT_3 antagonists (4), and 5-HT_3 agonists (5). (iv) Detailed view of steroids: estrogens (1), antiestrogens (2), estrone sulfatase inhibitors (3), estrogen receptor modulators (4), androgens (5), HMG-CoA reductase β-inhibitors (6), antiandrogens (7), aromatase inhibitors (8) and glucocorticoids (9).

Comparison to sequence similarity. The statistical model for ligand set similarity allowed us to directly compare the resulting E-values with those derived from sequence comparison. We mapped 193 MDDR activity classes to their protein target sequences and determined the sequence similarity among them using PSI-BLAST[27]. We then computed a heat map highlighting the differences between pharmacological similarity and sequence similarity among these targets (**Fig. 3a**). In this heat map, many ligand sets with enzyme targets were pharmacologically similar but sequence dissimilar. Examples include folate-recognition enzymes and adenosine-binding enzymes (**Fig. 3b**). By comparison, many neurological receptors had stronger sequence, than pharmacological, similarity (**Fig. 3c**).

Predicting and testing drug promiscuity. We were interested in exploring the behavior of single agents that were known to have either promiscuous or off-target actions. An example of the latter was methadone, known to have dual specificity for NMDA and μ-opioid receptors. Methadone is an unusual chemotype for μ-opioid agonists, one that is not represented in the MDDR, although it and several congeners can be found in PubChem. Because of this, when the methadone ligand set was queried against all 246 MDDR targets, the μ-opioid ligands were only found as the third-ranking hit. Unexpectedly, the set of methadone and its analogs was found by this method to be far more similar to the antimuscarinics activity class, particularly the M3 receptor antagonists (**Table 4**). This attests to the MDDR's known false-negative problem[28], but more provocative was the predicted M3

antagonism, as methadone is not known to have muscarinic activity. To test this possibility experimentally, we measured the affinity and activity of methadone on M3 muscarinic receptors by direct binding and a cell-based functional assay. Methadone was observed to have a K_i of 1.0 μM (**Fig. 4a**) and to antagonize activation of M3 receptors, consistent with the prediction (**Fig. 4b**).

We then looked for other single compounds with novel off-target effects. To increase the chance of novel action, we screened PubChem compounds—many of which are not in the MDDR database—against 246 MDDR targets. Over 12,000 PubChem compounds with annotated activities were compared to the MDDR ligand sets, using an automated procedure, looking for those where the target annotated in PubChem differed from that of the highest scoring MDDR set, using SEA. For the vast majority of the resulting 6,000 high-scoring hits, the annotations differed only trivially and could be rapidly excluded by post-filtering (e.g., "androgen antagonist" is formally different from "steroid antagonist," but not in a pharmacologically interesting way). There were, however, 30 PubChem compounds that had very low (good) expectation values against genuinely unrelated MDDR categories. Two stood out by visual examination of their structures and by our ability to actually acquire and test them in the appropriate assay. These were the drugs emetine and loperamide, which were predicted to antagonize adrenergic α2 and neurokinin NK2 receptors, respectively, based on set similarities (**Table 4**). Both predictions were tested by functional assay: 10 μM emetine was observed to induce 10.6- and 27.5-fold increases in the EC_{50} of the α2-agonist

Table 3 Comparing ligands from different sources: 23 PubChem pharmacological action sets versus 246 MDDR activity classes

	Size	MeSH pharmacological action	Pharmacological similarity top hits		Mean pair-wise similarity top hits	
			MDDR activity class	E-value	MDDR activity class	MPS
1	131	Adrenergic α-antagonists	Adrenergic (α) blocker	1.18×10^{-22}	Somatostatin analog	0.287
2	138	Adrenergic β-agonists	Adrenergic (β1) agonist	1.54×10^{-203}	Adrenergic (β1) agonist	0.395
3	132	Adrenergic β-antagonists	Adrenergic (β1) blocker	6.65×10^{-77}	Adrenergic (β1) agonist	0.370
4	30	Androgen antagonists	Androgen	4.54×10^{-125}	Androgen	0.300
5	21	Androgens	Androgen	0	Androgen	0.551
6	10	Aromatase inhibitors	Androgen	4.36×10^{-108}	Androgen	0.226
7	29	Carbonic anhydrase inhibitors	Carbonic anhydrase inhibitor	1.24×10^{-152}	Carbonic anhydrase inhibitor	0.269
8	11	Cholinergic antagonists	Anticholinergic	4.80×10^{-155}	Anticholinergic	0.396
9	91	Cholinesterase inhibitors	Acetylcholinesterase inhibitor	1.87×10^{-70}	Melatonin agonist	0.207
10	98	Cyclooxygenase inhibitors	Androgen	4.50×10^{-58}	3-Hydroxyanthranilate oxygenase inhibitor	0.249
11	111	Dopamine agonists	Dopamine agonist	5.50×10^{-120}	Adrenoceptor (α2) antagonist	0.306
12	52	Estrogen antagonists	Antiestrogen	3.56×10^{-112}	Antiestrogen	0.281
13	20	Estrogens	Estrogen	0	Estrogen	0.401
14	80	Glucocorticoids	Glucocorticoid	0	Glucocorticoid	0.506
15	34	Histamine H2 antagonists	H2 antagonist	1.47×10^{-53}	H2 antagonist	0.248
16	20	HIV protease inhibitors	HIV-1 protease inhibitor	8.41×10^{-108}	Somatostatin analog	0.378
17	28	Lipoxygenase inhibitors	Lipoxygenase inhibitor	2.05×10^{-16}	Melatonin agonist	0.245
18	106	Muscarinic antagonists	Anticholinergic	2.67×10^{-151}	Anticholinergic	0.343
19	22	Nicotinic agonists	Nicotinic agonist	3.00×10^{-22}	Anaphylatoxin receptor antagonist	0.297
20	94	Phosphodiesterase inhibitors	Phosphodiesterase I inhibitor	8.33×10^{-25}	Anticholinergic, ophthalmic	0.227
21	86	Protease inhibitors	Renin inhibitor	2.25×10^{-78}	Anaphylatoxin receptor antagonist	0.334
22	65	Reverse transcriptase inhibitors	Thymidine kinase inhibitor	1.63×10^{-145}	Thymidine kinase inhibitor	0.333
23	12	Trypsin inhibitors	Trypsin inhibitor	3.14×10^{-19}	3-Hydroxyanthranilate oxygenase inhibitor	0.346

clonidine for α2a and α2c adrenergic receptors, respectively, and 10 μM loperamide induced a 7.5-fold increase in the EC_{50} of the NK2 agonist [β-Ala8]-neurokinin (**Fig. 4c,d,e**, see **Supplementary Table 4** online). Assuming competitive binding, these results put the affinity of emetine for the adrenergic receptors in the 400-nM to 1-μM range, and the affinity of loperamide for NK2 receptors in the 1- to 2-μM range.

Discussion

We have shown that protein targets may be quantitatively related by their ligands. SEA reveals both expected and unexpected similarities that may be tested by examining the 'off-target' activities of the ligands themselves. Three aspects of these similarities merit particular emphasis. First, most ligand sets are highly related to only a few others; the vast majority of ligand sets are unrelated. Second, there are
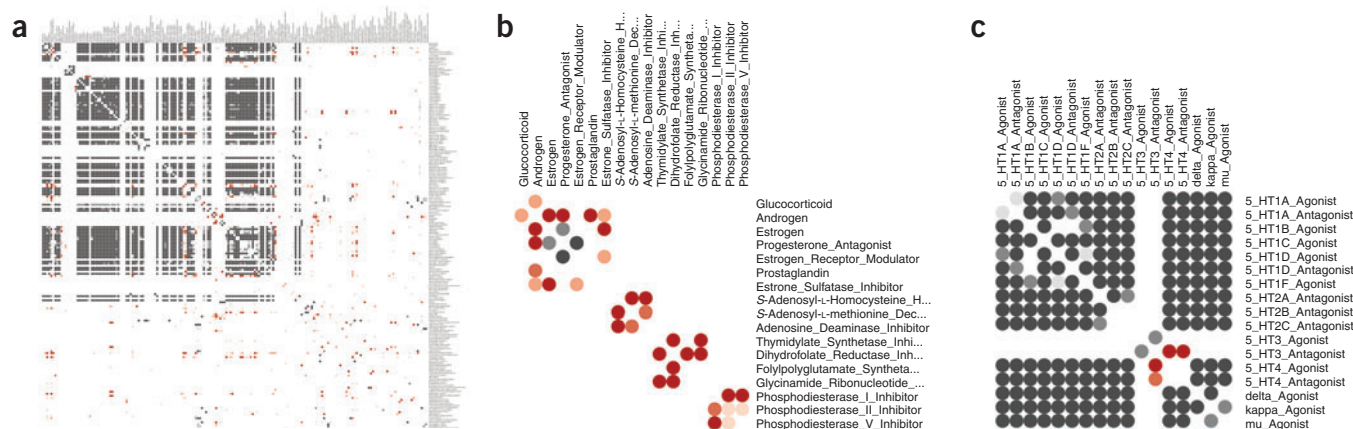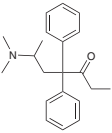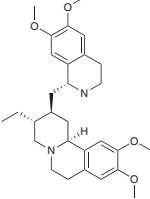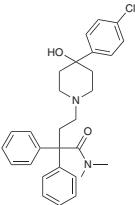

Figure 3 Comparison of sequence and ligand-based protein similarity. (**a**) In difference heat map, red, red ellipses mark activity class pairs with strong ligand-set similarity but weaker sequence similarity. (**b**) Enzyme activity classes often fall into this category. Dark gray regions mark target pairs with strong sequence similarity but comparatively lower ligand-set similarity. (**c**) This region includes many GPCRs, ion channels and nuclear hormone receptors; such receptors may share evolutionary history but have often diverged in terms of pharmacological function. The white regions mark cases where pharmacological and sequence similarity approaches agree. This heat map was calculated by taking the difference of the two log-space heat maps available in **Supplementary Figures 6** and **7** online.

Table 4 Novel target selectivity predictions for three existing drugs

Query	Rank	Size	Activity class	E-value	Max Tc
Methadone[a]	1	188	Antimuscarinic	4.45×10^{-50}	0.77
	2	266	Muscarinic M3 antagonist	1.22×10^{-11}	0.67
	3	68	Opioid agonist	1.84	0.61
	4	1485	NMDA receptor antagonist	9.04	0.67
	5	975	Muscarinic (M1) agonist	61.9	0.60
	6	717	Cyclooxygenase inhibitor	12.1	0.61
Emetine	1	277	Adrenergic (α2) blocker	4.34×10^{-118}	0.85
	2	564	Dipeptidyl aminopeptidase IV inhibitor	6.50×10^{-17}	0.94
	3	180	Dopamine (D1) antagonist	1.23×10^{-10}	0.74
	4	1820	Substance P antagonist	25.8	0.64
	5	288	Dopamine (D3) antagonist	179	0.61
	6	212	Neurokinin NK3 antagonist	2.76×10^{4}	0.60
Loperamide	1	462	Neurokinin NK2 antagonist	1.55×10^{-20}	0.75
	2	1820	Substance P antagonist	2.12×10^{-15}	0.75
	3	212	Neurokinin NK3 antagonist	2.63×10^{-14}	0.66
	4	518	Adrenergic (α1) blocker	1.64×10^{-10}	0.72
	5	583	Protein kinase C inhibitor	1.45×10^{-1}	0.63
	6	266	Muscarinic M3 antagonist	2.42	0.59

No query compound was already present in the reference 246 MDDR activity classes, and thus the Tc 1.0 (identity) column is omitted. [a]Although methadone was compared as a set of analogs, only the structure for methadone itself is displayed for clarity.

nevertheless enough connections among them to link almost all sets together, through sequential linkages, in coherent maps of pharmacologically interesting chemical space. Third, biologically related targets cluster in these maps. No biological information was used to make these connections, only ligand chemistry, and such clustering is an emergent property of this technique. It is also an imperfect property, in that the clusters of targets can differ from those expected from biological information alone. Both the expected and unexpected connections among the ligand sets have implications for understanding the effects of bioactive molecules, and lead to testable hypotheses.

The similarity of the ligand sets to only a few others owes to the intrinsic chemical differences between most sets and to the statistical model's discrimination between significant (e.g., E-value $< 1 \times 10^{-10}$) and insignificant (e.g., E-value > 1.0) similarity. In the case of DHFR inhibitors, for instance, the three most related target sets are the folate recognition enzymes glycinamide ribonucleotide formyltransferase, folylpolyglutamate synthetase (FPGS) and TS, with expectation values ranging from 3.97×10^{-100} to 1.11×10^{-61}; that is, highly significant. The next most related set had no measurable similarity and the other 241 are even less related (**Table 1**). Likewise, AMPA receptor antagonists score strongly against both kainic acid receptor and NMDA receptor antagonists (**Table 2**); all three are ionotropic glutamate receptors traditionally subdivided into NMDA and non-NMDA types[29]. A key point is that many related targets would be missed if ligand identity was substituted for chemical similarity between sets, that is, if we only related sets that shared common ligands (the flip side of this is that many large ligand sets would be related artifactually if we did not control for similarity expected at random). For instance, the antiglucocorticoids, estrogen agonists, estrogen antagonists, progesterone antagonists and prostaglandins all rank as highly similar to the androgen agonists, as is sensible (**Table 2** and **Fig. 2b**, iv). Yet not

one of these sets shares a single ligand with the androgens (**Table 2**). Correspondingly, serotonergic 1F agonists closely resemble serotonergic 1B, 1D and 5-HT$_1$ agonists and D$_4$-dopamine receptor antagonists without sharing a single ligand in common (**Fig. 2b**, ii, and **Table 2**); the same is true for the relationship of β$_1$ adrenergic receptor agonists to other β-receptor agonists and antagonists (**Fig. 2b**, i).

Related by chemical similarity, almost all of the 246 receptors may be mapped, through intermediate receptors, to all others. We found it convenient to interrogate this map interactively: one may click on any node to display a table of all the nearest ligand set neighbors, including the molecules that make up any given set (http://sea.docking.org). Thus, different classes of β-lactam antibiotics cluster together in this map, as do the several classes of phosphodiesterase inhibitors (**Fig. 2**). The serotonergics form their own branch of the tree, with the ionotropic (5-HT$_3$) agents isolated (**Fig. 2b**, iii), just as the androgens and estrogens group closely but separately (**Fig. 2b**, iv).

Another way to view such clustering is through a heat map that compares ligand-set with sequence similarities between the same targets (**Fig. 3a**). When the ligand-set and sequence similarities agree, as with μ-receptor agonists versus δ-receptor agonists (**Fig. 3c**) and neurokinin NK2 antagonists versus NK3 antagonists, the matrix element in the heat map is white (it will also be white when there is neither sequence nor ligand-set similarity). Such correspondences are comforting, but more interesting are those targets for which the chemoinformatic and bioinformatic techniques disagree. Many target sequences are more similar than their ligand sets (dark gray matrix elements). For instance, the serotonin 5-HT$_{1A-C}$ subtypes are highly related by sequence but less so by ligand sets (**Fig. 3c**), although the latter are not dissimilar. However, the serotonergics are also highly similar to the opioids by sequence, yet the ligands are different (**Fig. 3c**); much of this similarity arises from non-ligand-binding regions. Conversely, some targets unrelated by sequence are closely related by ligand sets (red matrix elements in **Fig. 3**). Thus, the antifolates cluster together even though DHFR, GART, TS and FPGS are dissimilar by sequence (**Fig. 3b**). The differences between the chemoinformatic and bioinformatic views have several sources, among them that sequence similarity arises from evolutionary history, but chemoinformatic similarity and dissimilarity arise from the state of the art of medicinal chemistry. Indeed, designing the specificity necessary to pharmacologically distinguish receptor subtypes, such as, 5-HT$_{1A}$, 1B and 1C, is a longstanding goal of medicinal chemistry, one executed in the teeth of their evolutionary relationships. Both the similarities and dissimilarities between the chemoinformatic and bioinformatic views lead to testable hypotheses.

Perhaps the most compelling result of this study is the experimental testing of three different drugs against targets to which they were not previously known to bind. We looked for candidate drugs based on known polypharmacology or on ligand-set similarities between targets with no clear precedence for cross-reactivity in the literature. Methadone attracted us because of its well-known polypharmacology,

modulating both NMDA and μ-opioid receptors. Surprisingly, methadone most resembled the ligand-set of M3 muscarinic receptor antagonists (**Table 4**). Both by direct binding and by functional assay, we find that methadone is a 1 μM antagonist of the M3 receptor, consistent with prediction (**Fig. 4a,b**). As far as we know, methadone's action on M3 muscarinic receptors has not been reported previously, although a pharmacophore model that may be related to its promiscuity has very recently appeared[30]. Intriguingly, its affinity for the M3 receptor is consistent with some of the side effects of this drug[29,31], which reaches micromolar steady-state concentrations in patients[32]. Emetine and loperamide are further examples of drugs that resemble, by SEA, target classes that they are not known to modulate. Emetine is an amebicide that inhibits polypeptide chain elongation in parasites[33]. By SEA, it has striking similarities to the adrenergic α2-blocker ligand-set, with an expectation value of 4.3×10^{-118} (**Table 4**). Consistent with that similarity, we find that emetine antagonizes α2 receptors in the micromolar and possibly sub-micromolar range (**Fig. 4d,e**, and **Supplementary Table 4** online). Although this activity has not, to our knowledge, been previously reported, it is consistent with the known side effects of this drug, which can lead to hypotension, tachycardia, dyspnea, myocarditis and congestive heart failure. Loperamide is an opioid that is used for relief of diarrhea through action on μ-opioid receptors in the gut[29] (**Table 4**). The drug closely resembles the neuro-

kinin NK2 antagonist ligand-set, when compared by the SEA method (**Table 4**). Consistent with that prediction, we find that loperamide antagonizes NK2 receptors in the micromolar concentration range (**Fig. 4c** and **Supplementary Table 4** online). Intriguingly, loperamide has been observed to modulate neurokinin NK3-receptor-triggered serotonin release, though this has been thought to be through its action on opioid receptors[34]. The results of this study suggest that the drug also has a direct effect on neurokinin receptors.

The polypharmacology of drugs and bioactive molecules emerges at the confluence of two currents: medicinal chemistry's elaboration of new molecules and the molecular evolution of biological function. Fortuitously, this channeled elaboration relates receptors and enzymes frequently enough to link almost all targets together in a single map of chemically relevant biology with sufficient specificity, when the background of random possibilities is controlled for, to distinguish the significant links from a stochastic sea of possibilities. In the minimum spanning trees that are one result of this analysis, many proteins with related functions cluster together. Thus, ion channels and GPCRs that have no obvious sequence or structure similarity are linked quantitatively based on their bioactive ligands. An advantage of this way of relating biological receptors is that it is articulated through the very agents used to probe biology experimentally—drugs and related reagents. The hypotheses that emerge from this analysis
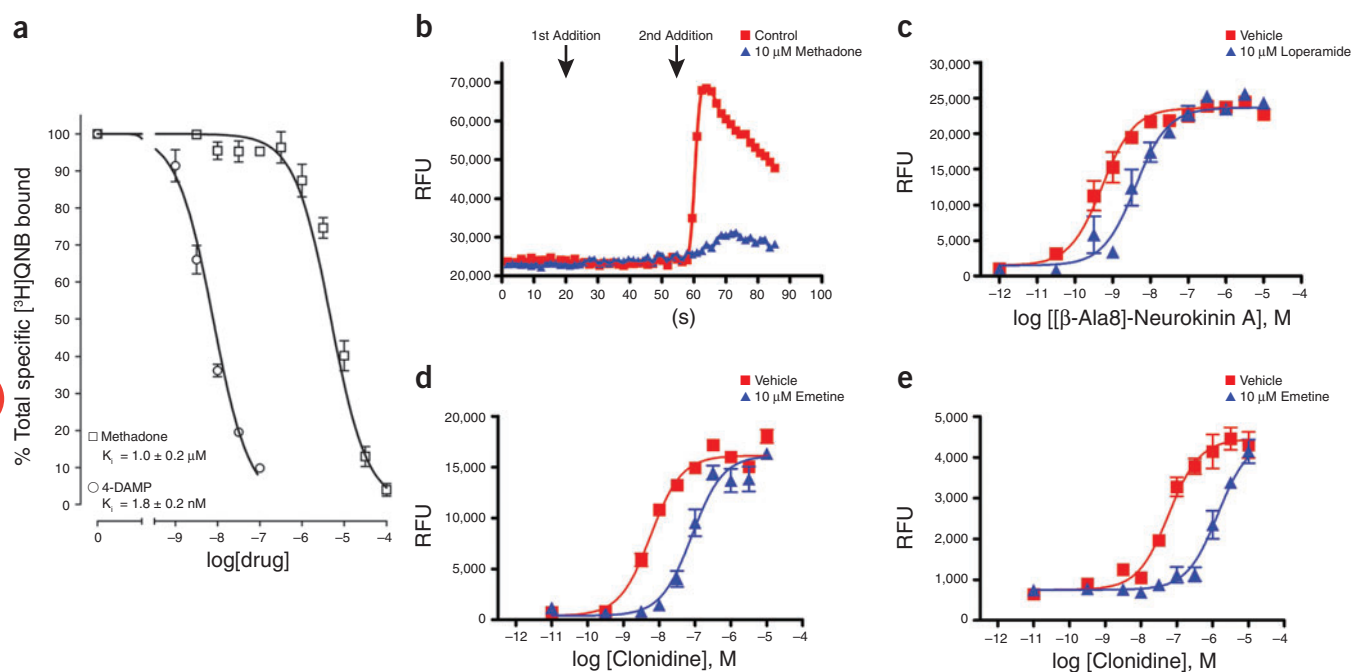

Figure 4 Testing the off-target activities of methadone, loperamide, and emetine. (**a**) Antagonism of M3 muscarinic receptors by the μ-opioid agonist methadone in a direct binding assay. Competition binding curves with [³H]quinuclidinyl benzilate in membrane fractions from CHO cells stably transfected with the human M3 muscarinic receptor. Each data point represents the mean and standard error of 4 conducted in duplicate or quadruplicate. Competition curves represent the best fit to a single-component logistic equation (GraphPad Prism 4.0). Two-site models did not yield a better fit. Membranes were incubated for 60 min at 25 °C with 0.5 nM [³H]quinuclidinyl benzilate and increasing concentrations of competing drug. Incubations were terminated by rapid vacuum filtration. Nonspecific binding was defined in the presence of 1.0 μM atropine and represented less than 10% of total binding. (**b**) Methadone antagonism of M3 muscarinic receptors by functional assay. Either methadone (10 μM final concentration) or vehicle was added at $T = 20$ s (1st addition), and then at $T = 50$ s (2nd addition) 1 μM carbachol was added to CHO-M3 cells and intracellular Ca^{2+} mobilization was measured, as previously described[34]. Dose-response curves (not shown) indicated that methadone was a competitive antagonist at M3-muscarinic receptors. (**c**) Loperamide antagonism of neurokinin NK2 receptors. Dose responses of CHO cells expressing Neurokinin NK2 receptors treated with [β-Ala8]-Neurokinin A were measured following administration of either DMSO vehicle or 10 μM loperamide. (**d,e**) Emetine antagonism of adrenergic receptors. Dose response of clonidine treatment of MDCK cells expressing either (**d**) alpha 2a adrenergic or (**e**) alpha 2c adrenergic receptors after incubation with DMSO vehicle or 10 μM emetine. Shown are representative curves, mean values ± s.e.m., of intracellular calcium release experiments performed in quadruplicate for each drug concentration per pretreatment condition as described in Methods.

thus may be subjected to experiment, and to this end we have made the relationships and linkage maps among the targets studied here publicly available (http://sea.docking.org/). The predictions and subsequent experimental observations that methadone, emetine and loperamide act as muscarinic M3, adrenergic α2 and neurokinin NK2 antagonists suggest that at least some of the predicted relationships merit investigation.

METHODS

Ligand sets. We extracted ligands from compound databases that annotate molecules by therapeutic or biological category. Multiple ligands in any annotation defined a set of functionally related molecules. As a source of ligands we used the 2006.1 MDDR[35], a compilation of about 169,000 drug-like ligands in 688 activity classes. We focused on a subset of this database, based on an ontology[36] that maps Enzyme Commission (EC)[37] numbers, GPCRs, ion channels and nuclear receptors to MDDR activity classes. Only sets containing five or more ligands were used. Salts and fragments were filtered, ligand protonation was normalized and duplicate molecules were removed. Of the 688 targets in the MDDR, 97 were excluded as having too few ligands (<5), and another 345 targets were excluded as being nonmolecular targets (e.g., the annotation "Anticancer" was not used). This left 246 targets, made up of a total of 65,241 unique ligands, with a median and mean of 124 and 289 ligands per target. The ligand set for methadone and 14 of its analogs was manually populated by querying "methadone" in PubChem Compound (http://pubchem.ncbi.nlm.nih.gov/). Ligand structures for emetine and loperamide were likewise acquired from PubChem Compound. All ligands were represented as SMILES[38] strings.

Quality of ligand set annotations. The activity class annotations available from the MDDR do not include explicit ligand-target affinity values and were primarily derived from the patent literature. Any given set may thus contain compounds with a wide range of affinities to the intended target. Although Hopkins and colleagues have recently found it useful to restrict the compounds annotated to a particular target to a limited affinity range[12], we have found our methods robust to the number of analogs present and the particular identities of the analogs used. We address this in two experiments, wherein we (i) prefilter the MDDR for unique chemotypes at 0.90 and 0.85 Tc distances to test robustness against analog redundancy (**Supplementary Fig. 2** online), and (ii) delete randomly chosen subsets of the ligand sets to test robustness against the particular choice of analogs present (**Supplementary Fig. 3** online). However, as noted by Sheridan et al., 'false inactives' remain a limitation of patent-based databases such as the MDDR, as any given compound may be tested only for one or two of its potential activities[28].

Set comparisons. All pairs of ligands between any two sets were compared by a pair-wise similarity metric, which consists of a descriptor and a similarity criterion. For the similarity descriptor, we computed standard two-dimensional topological Daylight fingerprints[38] using default settings of 2,048-bit array lengths and path lengths of 2–7 atoms. The similarity criterion was the widely used Tc[39–41]. For set comparisons, all pair-wise Tcs between elements across sets were calculated (**Fig. 1**), and those above a threshold were summed, giving a raw score for the two sets. The threshold was chosen so that the resulting statistics best fit an extreme value distribution (below).

Statistical model. A model for the random chemical similarity of the raw scores, motivated by BLAST[22] theory, was developed and empirically fit. We compared 300,000 pairs of molecule sets, randomly populated from the filtered full MDDR, across logarithmic set size intervals in the range of 10 to 1,000 molecules. This range reflected the set sizes we expected to encounter, though the procedure appears robust over any reasonable range of set sizes.

The raw score for each set comparison was plotted against the total number of ligand pairs in the two sets being compared, and was observed to depend linearly on the product of the number of ligands in the two sets (**Supplementary Fig. 1a** online). The s.d. of the raw scores was fit nonlinearly against this product of the set sizes (**Supplementary Fig. 1b** and **Supplementary Table 5** online). Both fits were determined with the SciPy[42] linear least-squares optimizer.

Set comparison Z-scores were calculated as a function of the set raw scores, expected raw scores and s.d. The histogram of Z-scores of the random sets conformed to an extreme value distribution (**Supplementary Fig. 1c** online). This distribution also underlies BLAST comparisons of protein and DNA sequences[21,22]. The probability of the score being achieved by random chance alone, given the Z-score, was converted to an expectation value (E-value) (**Supplementary Methods** online). The combination of set comparisons with the described statistical model is referred to as SEA. The ability of SEA E-values to correctly discriminate matching MDDR activity classes was tested against three simpler scoring metrics in **Supplementary Figure 4** online.

There is no formal justification for choosing a cutoff for the Tc value between ligands. One criterion that had the virtue of consistency was to insist on a Tc value for which the background Z-scores were best fit by an extreme value distribution (**Supplementary Figure 1c** online). We calculated Z-score distributions for all Tc thresholds in the range 0.00 to 0.99, with step size 0.01. For each such distribution, we plotted the normalized chi-square of their best fit to both normal and extreme value distributions (**Supplementary Fig. 5** online). This led to a Tc threshold of 0.57 (**Supplementary Table 5** online), which is low compared to accepted cutoffs for comparing individual pairs of ligands, emphasizing our different goal here: comparing ligand sets to inform us on the targets.

Similarity maps. All annotations in a given database were exhaustively compared against all others, resulting in a matrix of SEA E-values among the ligand sets (the full matrix is available in **Supplementary Data** online). This matrix defined a strongly connected graph. In one approach, we filtered the graph by removing all edges with significance less than an E-value cutoff of 1.0; this is a threshold graph. We also constructed a minimum spanning tree over the original strongly connected graph with Kruskal's algorithm[43]. We refer to this tree as a similarity map. The final images were rendered with Cytoscape[44].

Difference heat map. Protein sequences for the targets of 193 of the 246 activity classes were obtained, 77 of which were derived from the MDDR-to-EC number mapping provided by Schuffenhauer et al.[36]. The remaining 117 sequences were acquired from PubMed Protein searches. The resulting mapping of MDDR activity class to GI number is available in **Supplementary Data** online. We computed the sequence comparison matrix with PSI-BLAST[27], as implemented in the blastpgp binary available from NCBI. The maximum final E-value displayed was 1×10^5, with low-complexity region filtering enabled, and a maximum of ten iterations computed before convergence. **Supplementary Figure 6** online shows a heat map of the 193×193 PSI-BLAST matrix, created with matrix2png[45]

The unfiltered SEA E-value matrix described in similarity maps is shown as a heat map in **Supplementary Figure 7** online. This matrix was compared against the sequence-comparison E-value matrix built above by taking the difference of the natural logarithms of each E-value pair. To avoid math range errors, both E-values were first confined within the range of 1×10^{-50} to 1×10^5. A smaller E-value cap would allow for greater resolution of high-end E-values (e.g., 1×10^{-250} versus 1×10^{-200}), but this would be at the expense of differentiating from insignificant similarity (e.g., 1×10^{-45} versus 1×10^5). As a cutoff of 1×10^{-50} or better appears necessary for reliable transfer[46], no larger E-value cap was used.

PubChem out-group analysis. All compounds with annotated MeSH (http://www.nlm.nih.gov/mesh/) "Pharmacological Actions" were downloaded from PubChem and filtered as previously described. Any compound already present in the MDDR was removed, resulting in 10,557 unique nonoverlapping structures organized into 352 unique annotated 'action sets'. Of these, 23 action sets could be specifically mapped to a MDDR 'activity class', with mean 62 and median 52 compounds per set. These sets were then ranked by SEA E-values against all 246 MDDR activity classes.

Choice of compounds for novel selectivity prediction. Methadone and 14 analog structures from PubChem Compound were compared as a set against the MDDR to recapitulate known polypharmacology. Instead, novel selectivity was predicted, deemed plausible and ultimately tested. Subsequently, an automated system was developed to compare individual PubChem Compound molecules with annotated pharmacological actions against the

MDDR. All activity class hits resembling known actions were discarded, leaving 30 PubChem compounds with very low (good) expectation values against genuinely unrelated MDDR categories. Among these molecules, we targeted those that we could acquire and actually test, and whose structures resembled members of the novel target to which they were assigned by SEA (that is, there was a human filter on the compounds before assays were developed and compounds tested). The drugs emetine and loperamide met both criteria. We note that neither compound was present in the MDDR, nor was any a close congener. For emetine this reflects the lack of that family of amebicides in the MDDR, whereas loperamide is a nonclassical μ-opioid antagonist whose chemotype happens to be unrepresented among that MDDR ligand set. Thus neither of the classic targets of either drug was found by SEA, simply because the chemical structures were absent or unannotated or both.

Cell lines and functional calcium assay. Radioligand and functional assays were performed as previously detailed using the resources of the National Institute of Mental Health's Psychoactive Drug Screening Program[47,48] using cloned, human M3-muscarinic receptors expressed in Chinese hamster ovary (CHO) cells also, as previously described[49]. Neurokinin 2 receptor stably expressed in CHO cells[50] and alpha 2a and alpha 2c adrenergic receptors stably expressed in Madin-Darby canine kidney (MDCK) II cells[51] were carried in DMEM supplemented with 10% FBS, 1% penicillin-streptomycin, 1 mM sodium pyruvate and 600 μg/ml G418. Cells were plated onto uncoated or poly-L-lysine coated in 96-well plates in DMEM supplemented with 5% dialyzed FBS and 1% penicillin-streptomycin. The following day, media was replaced with 30 μl/well of Calcium Assay Kit Component A Dye (Molecular Devices) dissolved in 28 ml/bottle of assay buffer (2.5 mM probenecid, 20 mM HEPES and 1× HBBS (Gibco) (138 mM NaCl, 5.3 mM KCl, 1.3 mM $CaCl_2$, 0.49 mM $MgCl_2$, 0.41 mM $MgSO_4$, 0.44 mM KH_2PO_4, 0.34 mM Na_2HPO_4) pH 7.4. Plates were incubated in the dye for 1 h at 37 °C. Drugs predicted to be antagonists were diluted in assay buffer to a concentration of 30 μM and 30 μl of solutions were added to 96-well plates for ~15 min before reading. Fluorometric imaging was performed using a FlexStation II plate reader (Molecular Devices) reading the plate at 1.5 s intervals for 1 min. After establishing a fluorescent baseline (excitation at 485 nM and emission at 525 nM, using a 515 nM cutoff), 30 μl of agonist was transferred to assay plates at the 20 s time point with reading for another 40 s. Peak relative fluorescence units (RFU) were subtracted from baseline RFUs using SoftMax Pro (Molecular Devices) and data were then analyzed by nonlinear regression to obtain pEC_{50} values using GraphPad Prism version 4.03 (GraphPad Software). Statistical significance between pEC_{50} values obtained from vehicle and predicted antagonist pretreatment were analyzed by two-tailed t-test ($P < 0.05$) using GraphPad Prism.

AUTHOR CONTRIBUTIONS
J.J.I., B.K.S. and M.J.K. developed the ideas for SEA, M.J.K. wrote the SEA algorithms and undertook the calculations reported here, with some assistance from J.J.I. B.L.R. and P.E. performed the methadone assays, B.N.A. performed the emetine and loperamide assays, and B.K.S. and M.J.K. wrote the manuscript with editorial review from J.J.I. and B.L.R.

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

1. Roth, B.L., Sheffler, D.J. & Kroeze, W.K. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat. Rev. Drug Discov.* **3**, 353–359 (2004).
2. Kroeze, W.K., Kristiansen, K. & Roth, B.L. Molecular biology of serotonin receptors structure and function at the molecular level. *Curr. Top. Med. Chem.* **2**, 507–528 (2002).
3. Ebert, B., Andersen, S. & Krogsgaard-Larsen, P. Ketobemidone, methadone and pethidine are non-competitive *N*-methyl-D-aspartate (NMDA) antagonists in the rat cortex and spinal cord. *Neurosci. Lett.* **187**, 165–168 (1995).
4. Callahan, R.J., Au, J.D., Paul, M., Liu, C. & Yost, C.S. Functional inhibition by methadone of N-methyl-D-aspartate receptors expressed in *Xenopus* oocytes: stereospecific and subunit effects. *Anesth. Analg.* **98**, 653–659 (2004).
5. Krueger, K.E. Peripheral-type benzodiazepine receptors: a second site of action for benzodiazepines. *Neuropsychopharmacology* **4**, 237–244 (1991).
6. Finlayson, K., Witchel, H.J., McCulloch, J. & Sharkey, J. Acquired QT interval prolongation and HERG: implications for drug discovery and development. *Eur. J. Pharmacol.* **500**, 129–142 (2004).
7. Schreiber, S.L. Small molecules: the missing link in the central dogma. *Nat. Chem. Biol.* **1**, 64–66 (2005).
8. Johnson, M.A. & Maggiora, G.M. *Concepts and applications of molecular similarity.* (Wiley, New York; 1990).
9. Matter, H. Selecting optimally diverse compounds from structure databases: a validation study of two-dimensional and three-dimensional molecular descriptors. *J. Med. Chem.* **40**, 1219–1229 (1997).
10. Whittle, M., Gillet, V.J., Willett, P., Alex, A. & Loesel, J. Enhancing the effectiveness of virtual screening by fusing nearest neighbor lists: a comparison of similarity coefficients. *J. Chem. Inf. Comput. Sci.* **44**, 1840–1848 (2004).
11. Willett, P. Searching techniques for databases of two- and three-dimensional chemical structures. *J. Med. Chem.* **48**, 4183–4199 (2005).
12. Paolini, G.V., Shapland, R.H.B. & v Hoorn, W.P. Mason, J.S. & Hopkins, A.L. Global mapping of pharmacological space. *Nat. Biotechnol.* **24**, 805–815 (2006).
13. Vieth, M. *et al.* Kinomics-structural biology and chemogenomics of kinase inhibitors and targets. *Biochim. Biophys. Acta* **1697**, 243–257 (2004).
14. Izrailev, S. & Farnum, M.A. Enzyme classification by ligand binding. *Proteins* **57**, 711–724 (2004).
15. Bender, A. *et al.* "Bayes affinity fingerprints" improve retrieval rates in virtual screening and define orthogonal bioactivity space: when are multitarget drugs a feasible concept? *J. Chem. Inf. Model.* **46**, 2445–2456 (2006).
16. Nidhi, Glick, M., Davies, J.W. & Jenkins, J.L. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J. Chem. Inf. Model.* **46**, 1124–1133 (2006).
17. Steindl, T.M., Schuster, D., Laggner, C. & Langer, T. Parallel screening: a novel concept in pharmacophore modeling and virtual screening. *J. Chem. Inf. Model.* **46**, 2146–2157 (2006).
18. Schuffenhauer, A., Floersheim, P., Acklin, P. & Jacoby, E. Similarity metrics for ligands reflecting the similarity of the target proteins. *J. Chem. Inf. Comput. Sci.* **43**, 391–405 (2003).
19. Horvath, D. & Jeandenans, C. Neighborhood behavior of in silico structural spaces with respect to in vitro activity spaces-a novel understanding of the molecular similarity principle in the context of multiple receptor binding profiles. *J. Chem. Inf. Comput. Sci.* **43**, 680–690 (2003).
20. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
21. Karlin, S. & Altschul, S.F. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA* **87**, 2264–2268 (1990).
22. Pearson, W.R. Empirical statistical estimates for sequence similarity searches. *J. Mol. Biol.* **276**, 71–84 (1998).
23. Sheridan, R.P. & Miller, M.D. A Method for Visualizing Recurrent Topological Substructures in Sets of Active Molecules. *J. Chem. Inf. Comput. Sci.* **38**, 915–924 (1998).
24. Bradshaw, J. & Sayle, R.A. Some thoughts on significant similarity and sufficient diversity. Presented at the 1997 EuroMUG meeting, 7–8 October 7–8, 1997, Verona, Italy. <http://www.daylight.com/meetings/emug97/Bradshaw/Significant_Similarity.html>.
25. Hert, J. *et al.* Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures. *J. Chem. Inf. Comput. Sci.* **44**, 1177–1185 (2004).
26. Hert, J. *et al.* New methods for ligand-based virtual screening: use of data fusion and machine learning to enhance the effectiveness of similarity searching. *J. Chem. Inf. Model.* **46**, 462–470 (2006).
27. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
28. Sheridan, R.P. & Kearsley, S.K. Why do we need so many chemical similarity search methods? *Drug Discov. Today* **7**, 903–911 (2002).
29. Goodman, L.S., Gilman, A., Brunton, L.L., Lazo, J.S. & Parker, K.L. *Goodman & Gilman's The Pharmacological Basis Of Therapeutics,* edn. 11 (McGraw-Hill, New York; 2006).
30. Cleves, A.E. & Jain, A.N. Robust ligand-based modeling of the biological targets of known drugs. *J. Med. Chem.* **49**, 2921–2938 (2006).
31. DRUGDEX (see Methadone) (Thomson Micromedex, Greenwood Village, Colorado, 2006). <http://www.thomsonhc.com>.
32. de Vos, J.W., Geerlings, P.J., van den Brink, W., Ufkes, J.G. & van Wilgenburg, H. Pharmacokinetics of methadone and its primary metabolite in 20 opiate addicts. *Eur. J. Clin. Pharmacol.* **48**, 361–366 (1995).
33. DRUGDEX (see Emetine) (Thomson Micromedex, Greenwood Village, Colorado; 2006). <http://www.thomsonhc.com>.
34. Kojima, S., Ikeda, M. & Kamikawa, Y. Loperamide inhibits tachykinin NK3-receptor-triggered serotonin release without affecting NK2-receptor-triggered serotonin release from guinea pig colonic mucosa. *J. Pharmacol. Sci.* **98**, 175–180 (2005).
35. MDL Drug Data Report, 2006.1 (MDL Information Systems Inc., San Leandro, CA, 2006).
36. Schuffenhauer, A. *et al.* An ontology for pharmaceutical ligands and its application for in silico screening and library design. *J. Chem. Inf. Comput. Sci.* **42**, 947–955 (2002).
37. International Union of Biochemistry and Molecular Biology, Nomenclature Committee

& Webb, E.C. Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union Of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes (Academic Press, San Diego; 1992).

38. James, C., Weininger, D. & Delany, J. *Daylight Theory Manual* (Daylight Chemical Information Systems Inc., Mission Viejo, CA; 1992–2005).

39. Willett, P. *Similarity and Clustering in Chemical Information Systems* (Research Studies Press; Wiley, Letchworth, Hertfordshire, England; New York; 1987).

40. Brown, R.D. & Martin, Y.C. Use of structure Activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sc.i* **36**, 572–584 (1996).

41. Chen, X. & Reynolds, C.H. Performance of similarity measures in fragment-based similarity searching: comparison of structural descriptors and similarity coefficients. *J. Chem. Inf. Comput. Sci.* **42**, 1407–1414 (2002).

42. Jones, E., Oliphant, T. & Peterson, P. SciPy: Open Source Scientific Tools for Python. (2001). <http://www.scipy.org/>.

43. Kruskal, J. On the shortest spanning subtree and the traveling salesman problem. *Proc. Am. Math. Soc.* **7**, 48–50 (1956).

44. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

45. Pavlidis, P. & Noble, W.S. Matrix2png: a utility for visualizing matrix data. *Bioinformatics* **19**, 295–296 (2003).

46. Rost, B. Enzyme function less conserved than anticipated. *J. Mol. Biol.* **318**, 595–608 (2002).

47. Roth, B.L. *et al.* Salvinorin A: a potent naturally occurring nonnitrogenous kappa opioid selective agonist. *Proc. Natl. Acad. Sci. USA* **99**, 11934–11939 (2002).

48. Davies, M.A., Compton-Toth, B.A., Hufeisen, S.J., Meltzer, H.Y. & Roth, B.L. The highly efficacious actions of N-desmethylclozapine at muscarinic receptors are unique and not a common property of either typical or atypical antipsychotic drugs: is M1 agonism a prerequisite for mimicking clozapine's actions? *Psychopharmacology (Berl.)* **178**, 451–460 (2005).

49. Chelala, J.L., Kilani, A., Miller, M.J., Martin, R.J. & Ernsberger, P. Muscarinic receptor binding sites of the M4 subtype in porcine lung parenchyma. *Pharmacol. Toxicol.* **83**, 200–207 (1998).

50. Takeda, Y. *et al.* Ligand binding kinetics of substance P and neurokinin A receptors stably expressed in Chinese hamster ovary cells and evidence for differential stimulation of inositol 1,4,5-trisphosphate and cyclic AMP second messenger responses. *J. Neurochem.* **59**, 740–745 (1992).

51. Wozniak, M. & Limbird, L.E. The three alpha 2-adrenergic receptor subtypes achieve basolateral localization in Madin-Darby canine kidney II cells via different targeting mechanisms. *J. Biol. Chem.* **271**, 5017–5024 (1996).