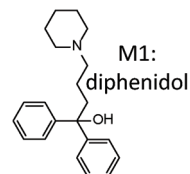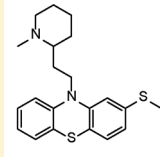# Chemical Structural Novelty: On-Targets and Off-Targets

Emmanuel R. Yera, Ann E. Cleves, and Ajay N. Jain*

University of California, San Francisco, Department of Bioengineering and Therapeutic Sciences, Helen Diller Family Comprehensive Cancer Center, San Francisco, California 94158, United States
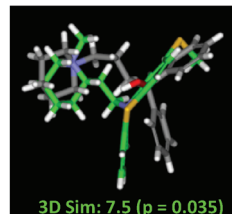
ABSTRACT: Drug structures may be quantitatively compared based on 2D topological structural considerations and based on 3D characteristics directly related to binding. A framework for combining multiple similarity computations is presented along with its systematic application to 358 drugs with overlapping pharmacology. Given a new molecule along with a set of molecules sharing some biological effect, a single score based on comparison to the known set is produced, reflecting either 2D similarity, 3D similarity, or their combination. For predic-

Predicted secondary target: Muscarinic M1 (3D log-odds 39)

Thioridazine

M1: diphenidol

2D Sim: 0.026 (p = 0.493)   3D Sim: 7.5 (p = 0.035)

tion of primary targets, the benefit of 3D over 2D was relatively small, but for prediction of off-targets, the added benefit was large. In addition to assessing prediction, the relationship between chemical similarity and pharmacological novelty was studied. Drug pairs that shared high 3D similarity but low 2D similarity (i.e., a novel scaffold) were shown to be much more likely to exhibit pharmacologically relevant differences in terms of specific protein target modulation.

INTRODUCTION

We have previously examined the relationship between drug pharmacology and structural similarity in the context of demonstrating that the human design process has a strong 2D reasoning bias.[1] Using a deeply annotated database of drug structures linked with their primary (desired) targets and secondary ones ("off-targets" generally responsible for side-effects),[2] we identified drug pairs that shared primary targets (primary target pairs) and those where the primary target of one drug was a *secondary* target of another (side-effect pairs). Among side-effect pairs, 2D similarity was extremely low when compared with primary target pairs. That is, when making an intentional design (developing a new drug for a particular indication where other drugs exist), we observed much higher 2D structural bias than when a pharmacological effect was unintentional. Apart from quantifying the 2D bias in human design, the study provided ample support for the proposition that molecules that appear to share little structural similarity by eye often share pharmacologically important effects.[1]

The economic incentives underlying the discovery process appeared to be a key driver of incremental design strategies. Among drug pairs patented close together in time, 2D structural similarity was much higher than for drug pairs patented at distant times. In cases where on-patent therapeutics exist for an indication, introduction of a patentable close analogue can still be profitable. However, novelty in pharmacological action is clearly more important when competing against cheaply priced off-patent drugs. We speculated that structural novelty, measured by lower 2D similarity, leads to increasing novelty of pharmacologic action in the whole human organism. Figure 1 shows a typical example that illustrates these points. The 2D structures show imipramine (the first serotonin reuptake inhibitor), amitriptyline (a fast follow-on drug), and citalopram (a much more selective serotonin reuptake inhibitor). The 3D overlay shows that, while citalopram exhibits significant structural novelty at the 2D level,

consideration of its similarity to imipramine in 3D shows high congruence.

Our previous work considered what was true about the similarities of drug pairs given that one knew about the pharmacology of both of the drugs for the pairs in question. The present study asks the converse question. What is true about the molecular pharmacology of a new molecule given its similarity to a molecule or sets of molecules with known pharmacology? The question takes two forms. One is formulated as the task of prediction of primary and secondary targets of an as yet uncharacterized molecule. This is an important operational issue: identifying potential off-targets early in drug discovery. The other question asks how much novelty in pharmacological action is expected to arise from structural novelty in a new drug. This question revolves around me-too drugs. It is primarily a strategic issue for pharmaceutical development and a public policy issue for regulatory bodies. Both the prediction and novelty questions hinge on differences between 2D and 3D molecular similarity approaches because their underlying biases are different.

The present study establishes a framework in which 2D and 3D similarity computations can be directly compared and also combined. Given this framework, we studied the similarity patterns exhibited by 358 marketed small molecule drugs linked through partially shared molecular pharmacology and addressed two broad questions. First, we quantified the degree to which primary and secondary targets could be predicted using 2D similarity, 3D similarity, or a combination of both by making use of sets of drugs whose targets were known. Second, we quantified the likelihood, based purely on molecular similarity, that drug pairs would exhibit different levels of overlap in terms of their detailed molecular pharmacology. The specific methods used were Surflex-Sim, and
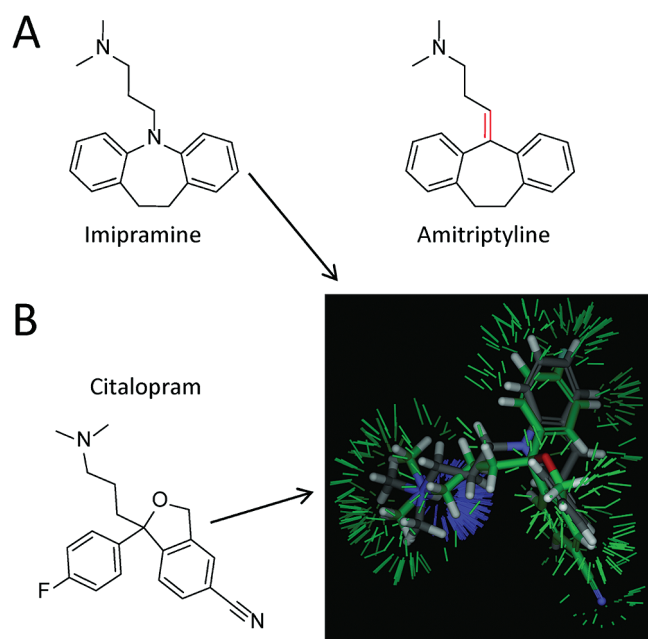
Figure 1. Shown are three 5HT reuptake transporter ligands: imipramine (the first in its class), amitriptyline (a fast follow-on compound), and citalopram (a later generation SSRI). (A) Minor structural differences between imipramine and amitriptyline (highlighted in red). (B) Surflex-Sim's 3D similarity overlay of citalopram (green carbons) and imipramine (atom colors). The significant regions of similarity within the molecule pair are illustrated with sticks, green (hydrophobic), blue and red (polar).

the 2D GSIM computation implemented within the Surflex platform.[1]

With respect to the first question, the results were expected, but striking as to degree. The performance of the methods for predicting target annotations was 2D < 3D < 2D + 3D. For primary target prediction, at a conservative threshold at which to assign annotations, true positive rates were 37%, 46%, and 59%, respectively. Consistent with our prior observations, 3D similarity did not yield a dramatic gain over 2D for primary targets due to the historical design bias problem: molecules designed to hit target X are often made specifically to look (in 2D) very much like molecules that are already marketed to modulate X. However, for off-target predictions, 2D yielded a 37% true positive rate, but 3D yielded 61%, with the combination yielding 68%. So, for off-targets, we observed a dramatic improvement over 2D in the ability of 3D molecular similarity to identify relevant pharmacological effects.

With respect to the second question, the primary broad observation is that a drug that shares high 2D and 3D structural similarity with another drug is likely to have indistinguishable pharmacological effects at the level of biochemically characterized modulation of protein targets. If, on the other hand, a drug shares little 2D similarity to existing drugs for the same cognate target but has high 3D similarity, there is greater likelihood to obtain a novel pharmacological effect. Specifically, drug pairs with high 3D and high 2D similarity showed identical biochemical targets four times more frequently than did pairs with high 3D similarity but low 2D similarity. This second case is particularly important from a design perspective because it represents a tractable one: where computational approaches making use of knowledge of existing therapeutics can guide design by

mimicking 3D surface shape and electrostatics. It is exemplified in Figure 1 with imipramine and citalopram, the latter bringing both structural novelty and greater target specificity. Shared targets occurred very rarely when no structural similarity existed between pairs of molecules: 94% of the time there was no overlap at all in biological targets, with overlap in primary targets occurring just 2% of the time.

The work reported here introduces a new methodological approach for data fusion, demonstrated with 2D and 3D molecular similarity. Given other recent reports of methods for data fusion and off-target prediction,[3−6] the differentiating features of what can be concluded based upon 2D and 3D molecular similarity is important to understand. This analysis of roughly one-third of small molecule drugs with extensively overlapping pharmacology has implications for both practical and strategic aspects of molecular design for therapeutic intervention.

■ METHODS AND DATA

The following describes the molecular data sets, computational methods, and specific computational procedures. Data and protocols are available for download (see http://www.jainlab.org for details).

Molecular Data Sets. Beginning with a set of nearly 1000 drugs for which we have curated primary and secondary target information, we identified a set of drugs with overlapping pharmacology. This was done by considering the drugs as nodes of a graph. Edges existed between pairs of nodes when the drugs representing each node shared either a primary or secondary target. The connected subgraph that contains the molecules shown in Figure 1 consisted of 358 small molecule drugs (approximately one-third of those marketed in North America), which formed a connected web through their established biological targets.

The database of annotations of primary and secondary drug targets has been described previously,[1,2] and only the specific salient aspects will be described here. In particular, all target annotations for each drug have been established well enough to identify a particular binding site on a particular protein assembly. So, benzodiazepines and barbiturates, both of which target the GABA$_A$ receptor, are distinguished because they bind different sites. This is a key distinction between our database and resources such as DrugBank[7] or commercial databases such as MDDR. DrugBank records annotations at the level of formal names (as we do, using HUGO[8] gene nomenclature conventions), but the distinction between different sites within an assembly of proteins is not made. Databases such as MDDR are even less specific in many cases, with biological activity defined by broad activity classes. This distinction is crucial because inferences may be drawn based on the comparison of a molecule to a set of drugs that share an annotation. If the set binds a single receptor assembly but contains two groups that bind different sites, inferential power of a similarity-based scheme will be diminished because one expects a new molecule might be similar to one set but not both. Machine-learning approaches can rely upon such data because they can form disjunctive models relating different chemical features to the same activity class, as was done by Nidhi et al.[9] with multicategory Bayesian models. Others have used data in this form to analyze aspects of the network topology of drugs in the context of biological pathways, which is also appropriate because the focus is on the relationship of target modulation to biological effect but does not depend upon site of action.[10]

The second important difference between our data set and other resources is that a distinction is made between *primary* targets, thought to modulate the therapeutically beneficial effects of each drug, and *secondary* targets, which mediate pharmacologically relevant (but undesirable) effects. Drugs may have multiple primary and multiple secondary targets. Annotation of this distinction varies among public resources, with DrugBank explicitly indicating whether a target is responsible for desirable

pharmacological action, but with resources such as BindingDB[11] not making such a distinction (instead focusing on careful curation of biochemical assay results). One important aspect of our database is that a *missing* annotation between a drug and a target cannot be interpreted to mean that the drug *does not* hit the target. It is simply a lamentable fact that comprehensive biochemical profiling of marketed drugs has not been carried out and made public.

While detailed information regarding a drug's effects, side-effects, dosage, and administration is available, systematized data systems that accurately capture this natural-language information in a form suitable for query and analysis are only beginning to become available. The SIDER[12] resource, for example, links drugs to a formal library of side-effect terms, but automated extraction of information from free-form documents is extremely challenging and can limit the utility of such resources for quantitative performance assessments. In our curation effort, we make use of resources such as DrugBank and SIDER, but annotations within our data set are always linked to primary literature, texts that themselves cite primary literature, or drug package-insert information. Annotations of primary and secondary targets or of specific side-effects are only made after direct examination of the evidence by an expert.

The set of 358 drugs identified through the primary and secondary target graph analysis was especially rich in targeting aminergic GPCRs (194 drugs) and ion channels and transporters (90) because drugs within these classes have long been known to have promiscuous and overlapping activity profiles.[13−15] The remaining drugs were roughly evenly split among pathogenic and human enzyme inhibitors. A total of 67 different biochemical targets were represented among the 358 drugs. Of the 67 targets, 44 were annotated to be modulated by 5 or more drugs.

Computational Methods. The core computational methods for computation of molecular similarity have been published and will be briefly summarized here.

3D Similarity Computation with Surflex-Sim. The Surflex-Sim 3D molecular similarity method and its use for virtual screening has been described at length in multiple publications.[1,2,16−18] Briefly, the method uses a molecular similarity function that computes, given two molecules in specific poses, a value from 0 to 1 that reflects the degree to which their molecular surfaces are congruent with respect to both shape and polarity. The function is computed based on the differences in distances from observer points surrounding the molecules to the closest points on their surfaces, including both the closest hydrophobic surface points and the closest polar surface points. So, two molecules that may have very different underlying scaffolds may exhibit nearly identical surfaces to the observer points, which are intended to be analogous to a protein binding pocket, which also "observes" ligands from the outside.

A command has been added to Surflex-Sim for the specific computation of "agnostic" similarities between two molecules, called "simcover". For each molecule pair A and B, each is sampled to yield a diverse set of conformers (default 20). Molecule B optimized in terms of both conformation and alignment to each of the conformations of A, and scores are recorded for each such optimization. The same is done for A matched to the conformations of B. Both the average bidirectional similarity and maximum are reported for each pair. For this work, the maximum similarity was used, but little difference is apparent making use of the average instead. Scores range from 0.0 to 10.0. Figure 1 illustrates the optimal mutual superimposition of imipramine and citalopram, which exhibits excellent concordance in terms of overall surface shape and correspondence of charge and directionality of the amines. The corresponding similarity score was 8.2.

GSIM-2D. The GSIM method was implemented as a strawman approach to identify ligand-retrieval problems that are relatively un-challenging.[1] However, due to the presence of structural analogues within many activity classes, it is frequently effective. Given two molecules A and B as input, it identifies all subgraphs of molecule A of depth 1, 2, and 3 at each heavy atom. For each subgraph containing 3 or more heavy atoms, existence of the subgraph is checked in molecule B. If it
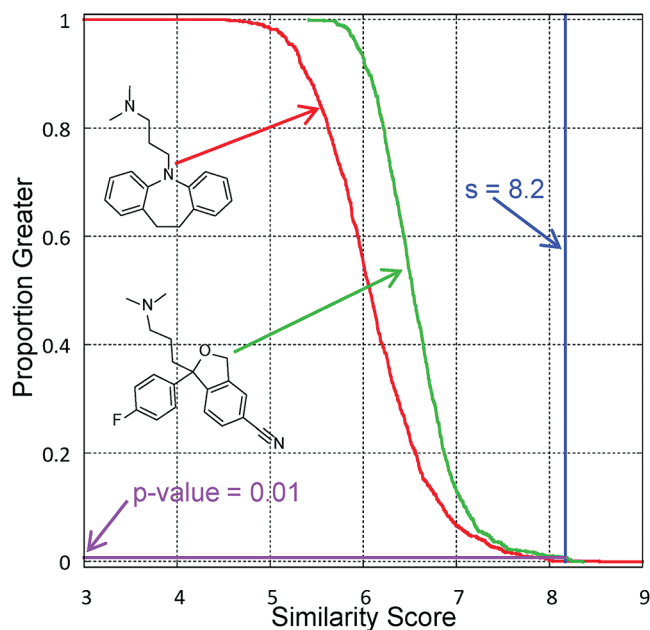

Figure 2. To normalize the similarity score s for a pair of molecules, x and y, we compare the magnitude of *s* to the empirically computed distributions of scores for x and y against a background molecule set. Here, the molecules are imipramine and citalopram, with similarity score 8.2, and similarity distributions are shown in red and green, respectively. The blue line marks the similarity score between the two. For both molecules, all background scores are greater than 4.0, but scores significantly greater than 6.0 become rare quickly. Imipramine had the higher proportion of background similarities >8.2, resulting in a *p*-value of 0.008.

exists in molecule B, the score is incremented based on the number of subgraph atoms and whether the root atom is a carbon or not (non-carbon rooted subgraphs are weighted 5.0, else 1.0). We repeat the procedure for molecule B looking for its subgraphs within molecule A. The two scores are normalized to the interval [0,1] based on the maximum possible score in each direction. The minimum of the ratio of number of heavy atoms in molecule A to molecule B and vice versa is computed, and the overall similarity is the minimum of the two scores multiplied by the minimum heavy atom ratio. The overall effect is that to yield high similarity, molecules A and B must be roughly the same size and have contain subgraphs, especially those rooted at heteroatoms.

Statistical Framework for Unifying 2D and 3D Comparisons. Both of the similarity approaches just described yield scores in arbitrary units that have no fundamental physical meaning except at the level of perfect identity. To compare the approaches directly, or to combine the computed similarities for a single molecule against multiple molecules, some normalization is required. Figure 2 shows the inverse cumulative histogram of similarity scores for imipramine and citalopram computed against a random set of 1000 screening molecules from the ZINC database, the same set we have used previously as a decoy set in assessing docking and similarity virtual screening performance.[1,19,20] Both range from approximately 4.0 to 8.0 overall. The distributions are nearly perfectly normal, with respective means of 6.6 and 6.1 and respective standard deviations of 0.58 and 0.43.

Given that the background molecule set was chosen randomly, the chances that any particular molecule within the 1000 is related to a given ligand are very small. Consequently, we can make use of these distributions to estimate *p*-values for particular levels of similarity for a given compound. In the comparison shown in Figure 2, the more pessimistic *p*-value based on the distributions came from citalopram because it had a
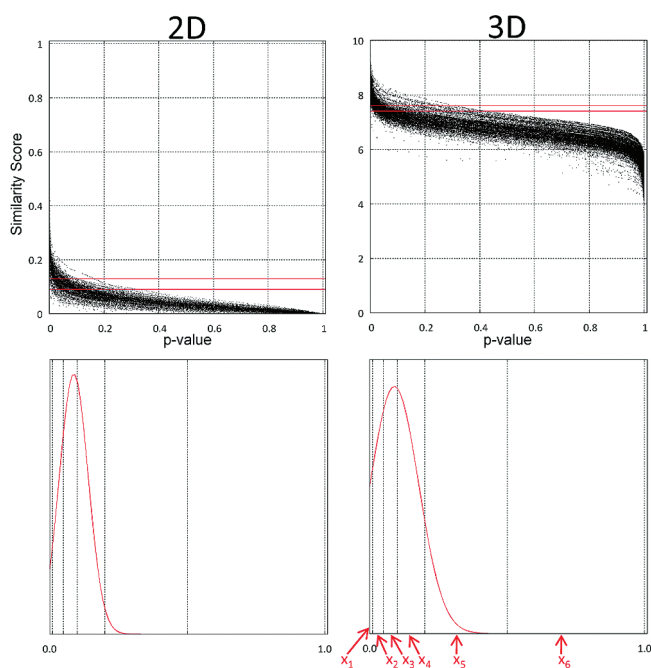
Figure 3. The top plots show the relationship between 2D and 3D similarity scores and corresponding p-values for all 63903 drug pairs. Higher 2D and 3D similarity scores generally lead to low p-values (i.e., are more significant). However, even within narrow ranges of *raw* similarity scores, the p-values varied substantially. The bottom plots show histograms of the p-values for 2D similarity scores between 0.09 and 0.13, and 3D similarity scores between 7.4 and 7.6. Numerically close similarity scores can give rise to very different p-values depending on the molecules in question. The labels x_{1-6} identify different qualitative bins of significance at p-value boundaries of 0.01, 0.05, 0.1, 0.2, and 0.5 (see Figure 4 for their use in the multinomial distribution computation).

larger number of similarities to random molecules that met or exceeded 8.2. The corresponding p-value was 0.008. More generally, given any molecular similarity method that yields some score S when comparing molecules A and B, a p-value can be computed by assessing the proportion of random molecules that have equal or greater similarity to S than to each of A and B and then taking the larger of the two proportions.

Figure 3 shows the relationship between raw similarity scores and p-values for both the 2D and 3D computations. Note that even within narrow bands of numerical similarity, both for 2D and 3D, very different p-values obtain. Between 3D similarity of 7.4 and 7.6, we observed p-values between 0.00 and 0.45, corresponding to highly improbable and clearly significant all the way to clearly random. The relationship of molecular structure to changes in distributional character will be discussed later, but the salient feature from a computational perspective is that while there is a global relationship between a similarity score and p-value irrespective of the molecule under consideration, there is enough variability to warrant normalization on a per-molecule basis. Because the specific p-values will depend on the exact composition of the decoy set used, we have chosen to bin the values, which also leads to a simple means to *combine* multiple p-values into a single value for the purpose of data fusion.

Given some new molecule and a collection of molecules known to share some activity, we wish to be able to combine similarities to the set of knowns, potentially using multiple similarity methods, into a single value that reflects the combined information. We make use of the multinomial distribution for this purpose, as shown in Figure 4. The basic idea is simple. Given some set of k different outcomes, each with associated prior probability p_i, and counts of each outcome x_i, M gives the

$$M(x_i,...,x_k,n,p_1,...,p_k) = \frac{n!}{x_1!...x_k!} p_1^{x_1}...p_k^{x_k} \quad (1)$$

$$S = -\log\left(\frac{M(\vec{x},n,\vec{p})}{M^*(\vec{x},n,\vec{p})}\right) \quad (2)$$

Figure 4. We compute a single score S to reflect the likelihood that a test molecule shares an annotation with a set of molecules known to have a particular activity. The similarity p-values are determined for the test molecule the set of known ligands. We compute the likelihood that the observed set of similarity p-values is extreme using the multinomial distribution. The p-values for the set of similarity comparisons are binned (top), and the bin counts are computed (x_{1-6}). M is the likelihood of having observed such a set of p-values, and M^* is the same computation using the converse probabilities. The Log Odds score S combines the two. Positive S indicates that it is more likely that the molecule in question shares the annotation with the ligand set than that it does not.

probability of observing the set of outcomes that gave rise to the counts observed. For application to the molecular data fusion problem, because we have converted each similarity score into a probability, we simply count the number of occurrences within each p-value bin, and the prior probability of each is simply the bin size itself. We make the computation using the computed probabilities for a set of similarities *and* using the converse probabilities, which yields symmetrical treatment for high similarity and for high dissimilarity. The final step in computing the score S is taking the log of the ratio of the two probabilities M and M^* and inverting the sign. So, in a case where the similarities of a particular molecule to a collection that share some annotation are low, S will be high. In a case where they are evenly spread from low to high, S will be zero. And if similarity values skew low, S will be negative.

The interpretation of a log-odds score of 2.0 is that it is 100 times more likely that the molecule in question shares high similarity with the annotated set of molecules than that it does not. To the extent that the similarity approaches that underpin the probability computations are related to biological effects, the log-odds score S may be used as a predictor of such effects (see Figure 5 for an example).

Computational Procedures. Detailed scripts for generating the results presented here are available in the data archive associated with this paper. Briefly, all 3D molecular similarity computations between drugs were made using default parameters for Surflex-Sim: sf-sim simcover drug-list drug-list log-file-3d. Similarly, all 2D molecular similarity computations between drugs were made as follows: sf-sim gsimcover drug-list drug-list log-file-2d. To compute the background distributions to normalize the similarity values, the analogous computations were done using the 1000 molecule ZINC decoy set: sf-sim [g]simcover drug-list zinc-list log-file-[3d][2d]-norm.

■ RESULTS AND DISCUSSION

The primary results of the study fall into two basic categories. The first relate to what can be predicted about a compound's potential biological effects based purely on molecular similarity and whether the use of 2D or 3D similarity methods influences the types of inferences that can be made. The second involve a census of the pharmacological congruence between pairs of drugs, where the pairs have been defined based upon the
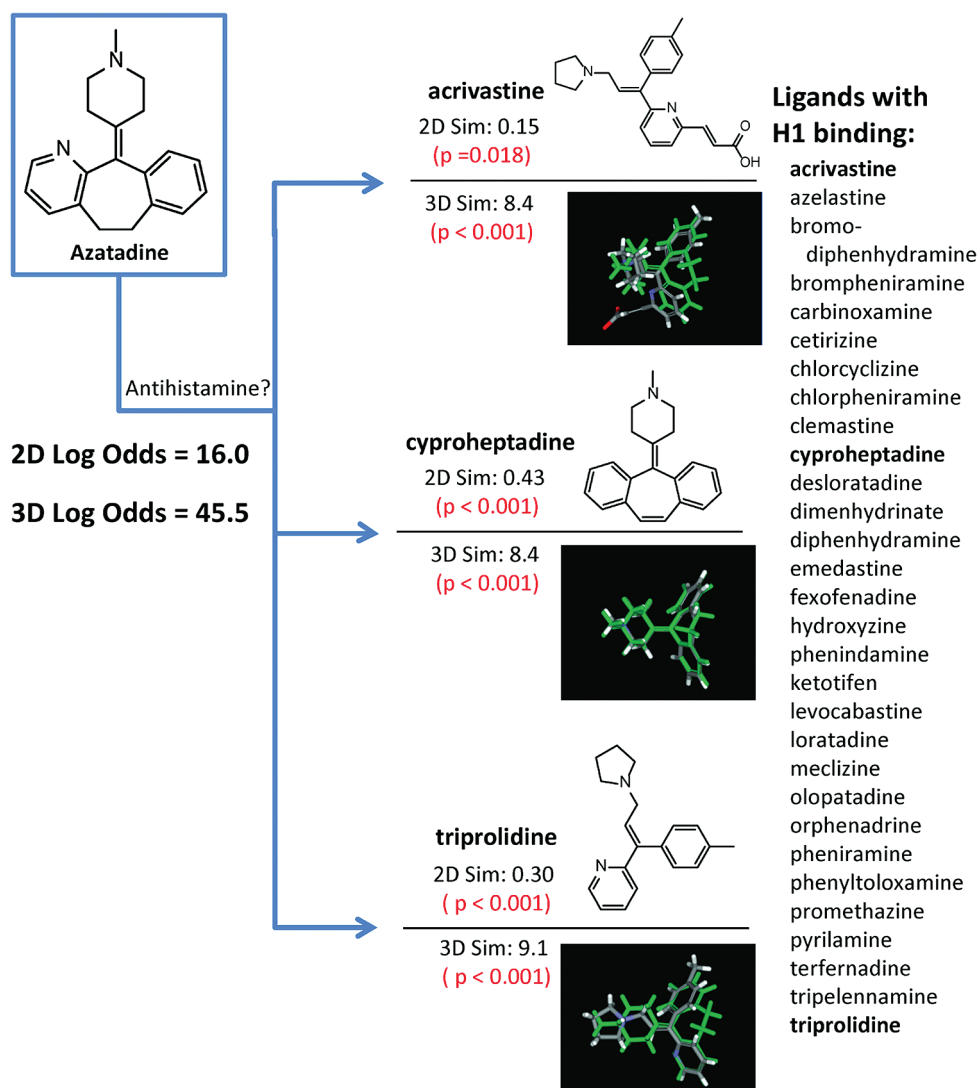
Figure 5. Log Odds calculation example. To test if a new molecule, in this case azatadine, is a ligand of a target of interest such as the histamine H1 receptor, we compute 2D and 3D similarity *p*-values for azatadine to known ligands of the target. Shown above are the 2D structures, 3D similarity overlays, and corresponding *p*-values of azatadine with three ligands. To calculate the Log Odds, similarity *p*-values were determined for the complete list of 30 histamine receptor ligands shown on the right. Both the 2D and 3D Log Odds scores are high and indicate a much higher probability that azatadine binds the histamine H1 receptor rather than it does not.

characteristics of their 2D and 3D similarities. The principal observation from the first category is that both 2D and 3D similarity (and their combination) are able to predict biological targets, but 3D similarity is more likely to identify effects that are not obvious from knowledge of pre-existing molecular pharmacology. The principal observation from the second category is that molecules sharing high similarity in both a 2D and 3D sense are much more likely to exhibit highly similar target profiles than those molecules that exhibit topological variation but retain high 3D similarity.

Predicting On- and Off-Target Effects. For each of the 358 drugs (see Methods and Data), we asked what their computed log-odds score was for each of the 44 targets that had 5 or more annotated drugs as either primary or secondary modulators to serve as positive examples. Figure 6 shows the results of the computation for both primary targets (top) and secondary targets (bottom). Nearly all computed log-odds scores were positive (about 90% for all methods), indicating greater similarity than

dissimilarity to example sets of compounds for either 2D, 3D, or 2D + 3D log-odds computations. This was the desired result because the definitions of targets differentiated between different binding sites on the same protein assemblies, so ligands within a given set of modulators were known to bind competitively.

For primary target predictions, performance of the methods was 2D < 3D < 2D + 3D, but the degree of improvement in moving beyond 2D ranged from 9 percentage points at a log-odds threshold of 6.0—15 percentage points at a log-odds threshold of 20.0. The added value of combining the two similarity approaches yielded typical gains of 10 percentage points over a broad range of log-odds values. At a threshold of 6.0, the combination of 2D + 3D similarity methods was able to identify a majority (59%) of all primary target annotations. As mentioned earlier, and as we have previously reported, the relatively limited gains of 3D over 2D are explained directly by human design bias.[1] The new observation here is that the effect holds in the forward predictive direction: when one has a set of ligands with known

Primary Targets

Secondary Targets
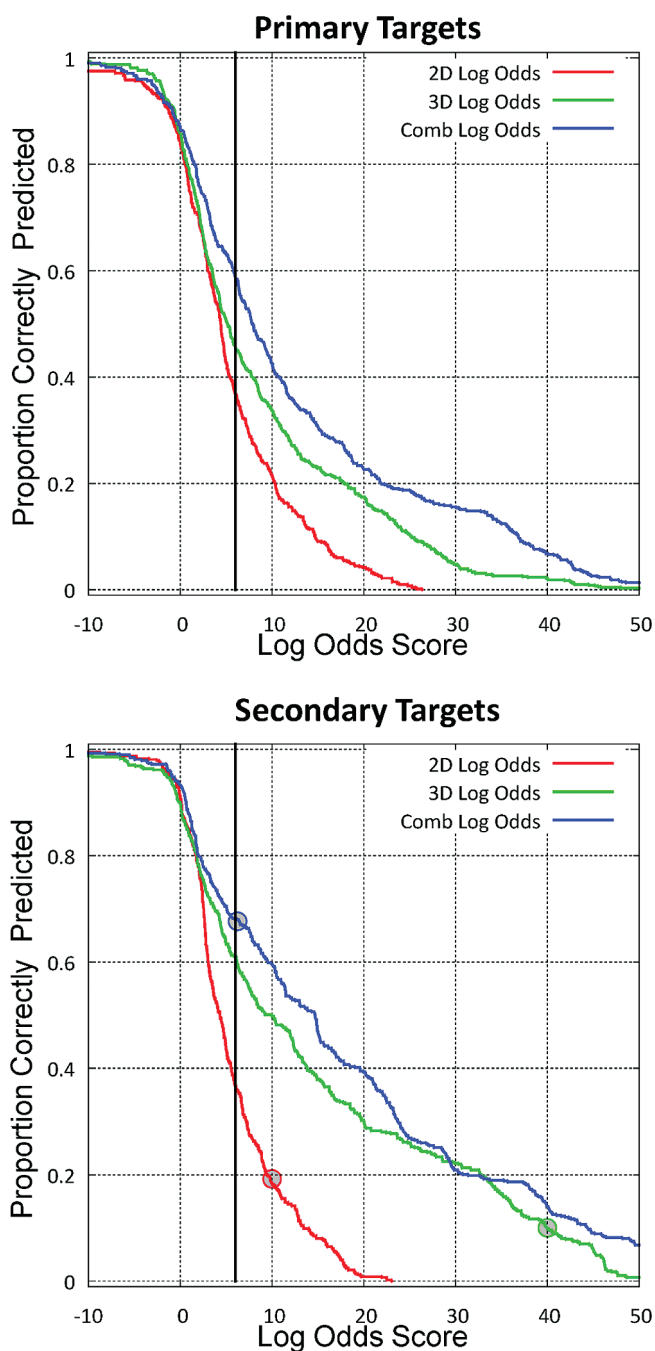
Figure 6. Proportion of drug targets correctly predicted. These plots indicate the proportion of drug targets correctly predicted by the three similarity methods at various Log Odds thresholds. Using 2D (red line), 3D (green line), or a combination of the similarity methods (blue line), we determined the Log Odds for both primary and secondary targets of the 358 drug set against 67 targets. At a Log Odds threshold of 6.0 (black vertical line), prediction success rates for primary targets for 2D, 3D, and combination were 37%, 46%, and 59%, respectively. For secondary targets, success rates were 37%, 61%, and 68%. Examples will be shown of predictions highlighted by the gray circles in the bottom plot.

activity, 2D similarity works quite well in assigning primary targets to new molecules. For secondary target prediction, the same qualitative performance was observed, but the performance gains for 3D over 2D were 24 percentage points (log-odds threshold of 6.0) to 30 points or greater for log-odds thresholds of 10.0 or more. The
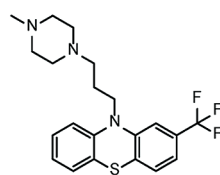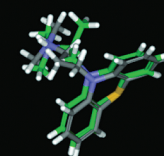
2D → Secondary Target
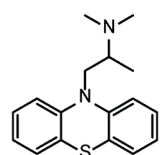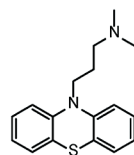
Primary: H1 receptor
Predicted Secondary:
Dopamine receptors
Log Odds:
- 2D   10
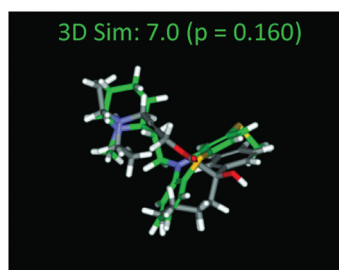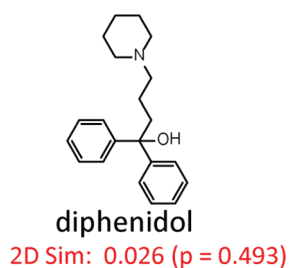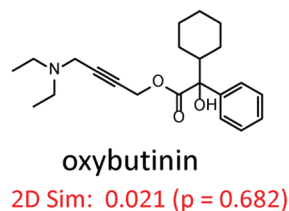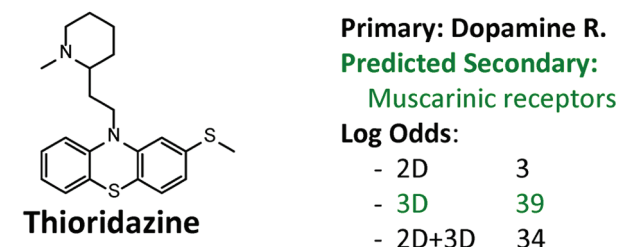- 3D   15
- 2D+3D   22
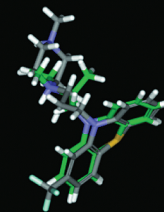
Figure 7. A 2D similarity method can sometimes correctly predict a target. Shown above are the 2D structures and 3D similarity overlays for the drug promethazine (a histamine H1 receptor antagonist) compared to two dopamine receptor antagonists, promazine and trifluoperazine. Although promethazine is prescribed as an antihistamine (Phenergan), it is a branched derivative of the antipsychotic phenothiazines and is known to have about 1/10* the dopamine receptor antagonistic activity. Our 2D method alone was sufficient to identify the dopamine receptors as an off-target of promethazine. As shown above, the Log Odds predictions that promethazine is a dopamine D2 receptor ligand were 2D = 10, 3D = 15, and the combination of 2D and 3D = 22. Note that at a log-odds score of 10, just 20% of secondary targets are identified by 2D alone, with 50% and 60% identified by 3D and the combination of 2D + 3D, respectively.

combination of methods yielded only a marginal gain, as with primary targets, of typically 10 percentage points or less, identifying 68% of the secondary targets at a log-odds threshold of 6.0.

The highlighted circles from Figure 6 (bottom) provide examples of specific secondary target predictions shown in Figures 7, 8, and 9. Figure 7 shows the case of promethazine, whose primary target is the H1 receptor, and whose off-targets include multiple dopamine receptor subtypes. The drugs promazine and trifluoperazine are examples of the degree of structural concordance that can occur, allowing for predictions of targets by essentially any method for computing molecular similarity. For these two molecules compared to promethazine, both 2D and 3D approaches produced *p*-values less than 0.01 (the most extreme bin from Figure 4), and combined with *p*-values from 33 other dopamine D2 receptor drug comparisons, yielded log-odds scores of 10, 15, and 22 for 2D, 3D, and 2D + 3D, respectively. These drugs were all synthesized as part of the medicinal investigation of what were then
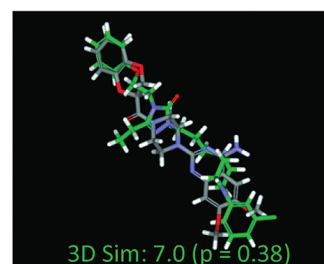
3D → Secondary Target

Thioridazine

Primary: Dopamine R.
Predicted Secondary:
  Muscarinic receptors
Log Odds:
  - 2D   3
  - 3D   39
  - 2D+3D   34

3D Sim: 7.0 (p = 0.160)

oxybutinin
2D Sim: 0.021 (p = 0.682)

diphenidol
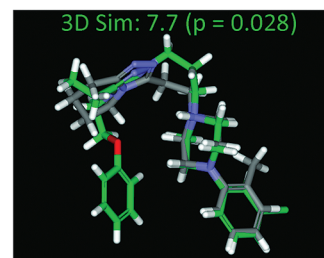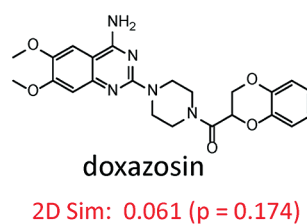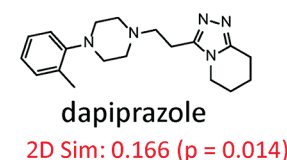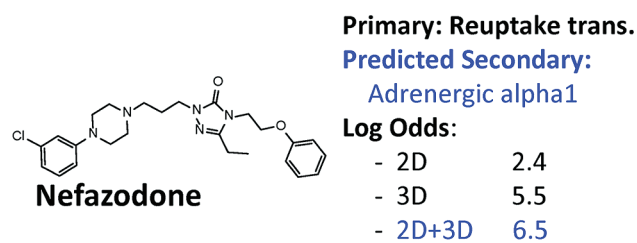2D Sim: 0.026 (p = 0.493)

3D Sim: 7.5 (p = 0.035)
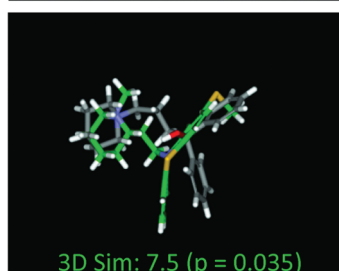
Figure 8. Our 3D similarity method more accurately predicts off-targets. Shown above are the 2D structures and 3D similarity overlays for the drug thioridazine (a dopamine receptor antagonist) compared to two muscarinic receptor antagonists, oxybutinin and diphenidol. Although thioridazine is prescribed as an antipsychotic (Mellaril), it is well-known to have antimuscarinic effects such as dry-mouth and blurred vision. Our 2D method did not identify the muscarinic receptors as an off-target of thioridazine (Log Odds for M1 = 3). In contrast, our 3D method strongly predicted the off-target, with an M1 Log Odds score approaching 40. Whereas over 10% of secondary target annotations are captured at this level by 3D similarity, none are captured with such a high level of confidence by 2D.

2D+3D → Secondary Target

Nefazodone

Primary: Reuptake trans.
Predicted Secondary:
  Adrenergic alpha1
Log Odds:
  - 2D   2.4
  - 3D   5.5
  - 2D+3D   6.5

3D Sim: 7.7 (p = 0.028)

dapiprazole
2D Sim: 0.166 (p = 0.014)

doxazosin
2D Sim: 0.061 (p = 0.174)

3D Sim: 7.0 (p = 0.38)

Figure 9. A combination of 2D and 3D similarity methods makes a small improvement over 3D alone. Shown above are the 2D structures and 3D similarity overlays for the drug nefazodone (a 5HT reuptake transporter inhibitor) compared to two α1 adrenergic receptor antagonists, dapiprazole and doxazosin. Although nefazodone is prescribed as an antidepressant (Serzone), it causes postural hypotension, a known side effect of α1 blockers. Neither the 2D or 3D methods alone identified the α1 adrenergic receptors as an off-target of nefazodone at the conservative threshold of 6.0 (Log Odds for α lA = 2.4 and 5.5, respectively). In contrast, a combination of the 2D and 3D methods predicted the off-target. Close to 70% of all off-targets were correctly identified at Log Odds = 6.0 using the combination approach.

termed "anti-histaminic phenothiazines," many of which had antipsychotic properties.[21] These properties were due to a host of effects on different brain receptors but are thought to primarily derive from modulation of dopamine receptors of multiple subtypes. Relatively subtle changes in structure (e.g., from promazine to promethazine) yield sufficient different in target potencies to shift primary indication from antipsychotic for promazine to antihistamine for promethazine. However, the shifts in potency are not so dramatic as to abrogate the multiple target effects entirely.

Figure 8 shows another example of a phenothiazine antipsychotic whose primary effects derive from dopamine receptor modulation. Here, however, some of the more significant side-effects are those modulated by muscarinic antagonism, including dry-mouth and blurred vision. In this case, the 2D log-odds was just 3. Whereas 2D similarity did not produce a low *p*-value when comparing thioridazine to either oxybutynin or diphenidol (two potent antimuscarinics), 3D similarity yielded much lower *p*-values. Coupled with those derived from comparisons to 61 other M1

drugs, the 3D log-odds score was 39, allowing very confident assignment of muscarinic targeting to thioridazine. In this case, the addition of 2D similarity to 3D produced a slight *reduction* in computed log-odds score. Log-odds scores are not additive; additional observations affect the combinatorics such that a collection of *p*-values which alone yield a marginally positive log-odds score may diminish the score derived from a collection of *p*-values that produced a high score.

Figure 9 shows a case where the combination of 2D and 3D similarity produced a log-odds score greater than 6.0 where neither alone met that threshold. Nefazodone yields its primary effects through modulation of multiple reuptake transporters, but it has a significant side-effect of postural hypotension deriving from modulation of α-adrenergic receptors. In this case, for α-1A receptor, 2D alone yielded log-odds of 2.4, with 3D yielding 5.5. Comparisons to dapiprazole and doxazosin produced no extreme *p*-values using either 2D or 3D, but all four scores leaned in favor of similarity to nefazodone. Along with 50 other drug comparisons, the combined log-odds for adrenergic effects was 6.5.

Excess Targets: False Positive Predictions. The framework we have developed allows for the combination of multiple sources of information to yield a single scalar value associated with a class prediction. In such a situation, it is both customary and desirable to make an estimate not only of true positive success rates but also of the corresponding false positive rates (e.g., with a receiver-operator characteristic (ROC) analysis). Here, we were able to identify primary and secondary targets about 60−70% of the time at a combination log-odds score threshold of 6.0. However, at that threshold, there are targets suggested for drugs for which *no annotation* is known. At a log-odds threshold yielding a true positive rate of 60%, the typical ratio of *excess* predicted targets relative to the total number of known primary and secondary targets was roughly 2−3, depending on the class of drugs involved. Larger numbers of excess targets were observed for drugs whose primary targets were among the aminergic GPCRs. The difficulty in interpreting this observation is that public data do not exist that systematically profile small molecule drugs in biochemical assays.

As a surrogate for biochemical data in unknown drug−target relationships, we manually assessed package insert and related information to make a determination of whether muscarinic side-effects were both present and drug-related. These included dry-mouth, urinary retention, blurred vision, drowsiness, mydriasis, and other effects. For the 358 drugs where we had no formal annotations of muscarinic target effects, which totaled 294 compounds, we surveyed a random subset of slightly more than half of them (180 drugs total), resulting in 84 with muscarinic side-effects and 96 without. We also surveyed 29 of the 64 drugs that we had previously annotated as binding muscarinic receptors. All 29 of the previously annotated muscarinic modulators showed clear, drug-related side-effects (90% exhibited dry mouth effects, 69% drowsiness, and a majority also showed urinary retention, blurred vision, and dizziness). For the 64 drugs with annotated muscarinic target effects, the mean log-odds score for muscarinic receptors was 25.8, with 92% scoring higher than 6.0.

Overall, using the side-effect assessments as a binary class label for the 180 surveyed drugs that *had not been annotated* as muscarinic modulators, the log-odds score produced an ROC area of 0.88 (95% confidence interval of 0.83−0.93). The enrichment for drugs with muscarinic side-effects among the top 1% log-odds scores was 19-fold. Of the surveyed drugs, 90% of those with a log-odds score of 6.0 or greater showed classic muscarinic side-effects (38/42 surveyed drugs). Even at a threshold of just 2.0, 85% were positive (55/65). Above a threshold of 26.0, all surveyed drugs (16 total) showed such side-effects. Conversely, below a log-odds threshold of −6.0, just 6% (3/47 surveyed) showed potentially muscarinic effects. Below a threshold of −16.0, no drugs showed such effects (26 total).

Three examples of drugs that had lacked muscarinic annotations are particularly informative. Amoxapine, an antidepressant working primarily through the norepinephrine reuptake transporter, received a log-odds score of 4.8. Prescribing information indicated that the most frequent side-effects included dry mouth, constipation, and blurred vision. It has also been shown biochemically to bind muscarinic receptors.[22−24] Orphenadrine, an antihistamine prescribed to relieve muscular pain, received a score of 42.9. Prescribing information indicates that "dryness of the mouth is usually the first adverse effect to appear." The drug has been shown to antagonize muscarinic receptors with a K_i of 100 nM.[25] Mesoridazine received a score of 37.1, had clear muscarinic side-effects, and also has a K_i of 69 nM against the M1

receptor.[22] Notably, it received a log-odds score of 6.4 against the HERG potassium channel, although it had not been annotated for such activity. It was withdrawn from the U.S. market in 2004 due to HERG-mediated cardiac side-effects.[26]

This survey of muscarinic side-effects among previously unannotated drugs makes three points. First, the empty cells of the annotation matrix of drug to target interactions *cannot* be thought of as indicating no effect. Second, the log-odds scores were both sensitive and specific with respect to muscarinic target annotation. Third, the lack of systematic profiling of drugs for which ample human data exist represents a large gap in our knowledge. Manual curation of this depth requires on the order of 30 min to 1 h per drug per side-effect, *after* establishing the relationship between a particular target and the relevant human pharmacology down to specific terms and variations. We are exploring automated means to consider databases of side-effect terms and their relationship to predicted on- and off-targets in order to carry out a more comprehensive study.

Approaches for semiautomatic curation such as SIDER[12] are challenged by variations in language such as "dryness of the mouth" instead of "dry mouth." The MedDRA dictionary,[27] for example, lists the latter as a defined medical term (but not the former), and relatively sophisticated language parsing is required to relate the two together. In the case of orphenadrine, one of the 84 drugs with clear muscarinic side-effects, SIDER misses the dry mouth effect, which is clinically the most prominent. Even with much more extensive synonym mapping, cases exist where side-effects are listed as *not* being present or are listed as being present but then dispensed with as not different from placebo, which is challenging to assess without expert manual curation.

Relationship to Other Methods. Two relatively recent approaches to data fusion involving molecular similarity are particularly relevant to our log-odds scoring approach. Muchmore and Hadjuk's belief theory approach[5] and the similarity ensemble approach (SEA)[3] introduced by Shoichet's group both offer the means to make predictions about a given molecule's activity based upon its relationships to other molecules.

The belief theory approach makes use of Hooper's Rule, which was devised in the late 1600s by George Hooper, predating the Bayesian belief approaches later popularized by Laplace and his adherents.[28] The rule was devised to address the credibility of a report of some fact when simultaneously attested by N reporters, each with credibility p (high p implying high credibility). This rule formalizes the notion that multiple partially credible sources strengthen one another's credibility. In the original report applying this rule to predictions of molecular activity based on similarity, the definitions of positive pairs of molecules and negative pairs differed from the current work, with molecule pairs considered as positive sharing not only a target but similar potency against the target. Similarity descriptors were converted into probability functions by considering a large set of positive and negative pairs and counting the number of times that a pair with some level of similarity was a positive example. Evidence from multiple similarity methods concerning pairs of molecules was combined using Hooper's Rule. A key distinction with our log-odds approach is that the belief theory formalism *always* increases belief, no matter how marginal an additional source's belief may be. In the log-odds approach, N very low p-values coupled with N symmetrically high p-values yield a log-odds of associating a target to a ligand of zero. The belief theory approach treats the cases of very low similarity as attesting *in favor* of the proposition that the query molecule will hit the target in question, but with low belief.

One might argue that the interpretation of such a value is more akin to a reporter attesting *against* a fact rather than giving it marginal support, making the application of Hooper's Rule a matter of empirical choice rather than purely logical.

The SEA method[3] uses a framework for estimating probabilities that is similar to that used for comparing sequence similarity of proteins, with likelihoods represented as E-values (a p-value multiplied by a large, arbitrary constant representing a database size). SEA makes use of 2D topological similarity to compute pairwise similarities between sets of molecules. By choosing a threshold below which to ignore similarity values, the pairwise sum of all similarities between two sets of unrelated molecules was shown to fit an extreme value distribution. So, to compare one (or several) molecules against a set with known activity, the magnitude of the raw similarity set comparison score is compared with that expected from unrelated sets, a probability is derived, and an E-value is produced. In contrast with the belief theory approach, in this formulation, the presence of poor similarity values yields poorer E-values.

Both the belief theory and SEA approaches treat raw similarity values as being equivalent regardless of the specific molecules or molecule sets in question. Our observation of both the GSIM and Surflex-Sim methods, which we believe will also hold for other methods such as ROCS[29] (3D) and Daylight fingerprint-based similarity[30] (2D), is that the probability of observing some raw value varies depending on the particular structure involved. As seen in Figure 3, p-values associated with narrow similarity ranges included extremely significant values as well as clearly random ones. For similar molecules, the distributions of observed similarities to the background set tend to be close (see Figure 2 for an example). However, for a small and simple molecule, such as acetaminophen, the required similarity score to reach a p-value of 0.01 is higher (8.3) than that for more complex molecules, such as azithromycin, where the required similarity is lower (5.8). Clearly, the particular values depend on the composition of the background molecule set, but we do not believe it is possible to construct a nondegenerate background set against which all molecules will exhibit congruent similarity distributions. By assessing differences in likelihood of observing different similarity levels within the context of each specific molecule pair, it is likely that the associated log-odds scores better reflect the underlying similarity relationships than approaches that take a coarser-grained approach. Of course, it is also possible to make use of global similarity distributions with the log-odds approach, but it is difficult to justify doing so.

Quantitative Comparison to Other Approaches. The muscarinic side-effect prediction task offers the opportunity for direct comparison of our approach to belief theory and to SEA. We computed joint beliefs regarding muscarinic activity for the 180 drugs and used these beliefs to assess ROC area. Recall that the 180 drug set consisted of 86 positives and 94 negatives based on the presence of side-effects, with similarities for each computed against 64 known muscarinic modulating drugs. For belief theory, the formula for combining evidence is given by $B = 1 - (1 - B_1)^*(1 - B_2)^*...^*(1 - B_N)$, where $B_1...B_N$ are the separate beliefs associated with the assertion that a given molecule has a particular activity. The most direct comparison to Muchmore and Hadjuk's formulation is made by setting each $B_i = (1 - p_i)$, with each p_i derived from the 3D similarity computations used above for the log-odds approach. Using a single global distribution to obtain p-values from the similarities, we observed an ROC area of 0.61 ± 0.05 (95% confidence

interval), which was significantly worse than for the log-odds approach (0.88 ± 0.05). Using empirically determined p-values for each molecular comparison (as the log-odds approach does), the performance improved to 0.72 ± 0.05 but was still significantly worse than the log-odds result.

Note, however, that the ROC area comparisons are somewhat misleading due to the degeneracy in the belief theory evidence rule. If a single belief is 1.0 (p-value of 0.0), the overall joint belief will be 1.0 no matter what the other belief values may be. For the muscarinic side-effect prediction task, this results in a large proportion of joint beliefs for the 180 drugs to be exactly 1.0. This degeneracy stems from the definition of Hooper's Rule, but its effect can be ameliorated by scaling down all beliefs by a constant factor. The best result we were able to obtain for belief theory was an ROC area of 0.85 ± 0.05 (nominally indistinguishable from log-odds), using $B = 0.5(1 - p)$, with empirically determined p-values for each pairwise molecular comparison. Even with this augmentation, there were a significant number of tied values of high belief, covering nearly 10% of the 180 ligands. The maximal enrichment for belief theory, in this most favorable (and artificial) formulation, was 6.4, corresponding to a true-positive (TP) rate of 55% and false-positive (FP) rate of 9%. Much better early enrichment was possible with the log-odds approach because there is no multiplicative degeneracy involving strict interpretation of p-values. We obtained maximal enrichment of 20-fold at a false-positive rate of just 1% using 3D log-odds.

For the SEA approach, a direct performance comparison (with the same set of 64 annotated muscarinic ligands used here) was not possible using the web-based SEA interface (sea.bkslab.org). However, the annotations underpinning SEA predictions are far more extensive than those used here, with over 1000 ligands having muscarinic target activity (including exact matches for 50% of the 64 used here, and close analogues for over 85%). We queried the 180 drugs for SEA predictions, which were reported for target predictions with E-values <10.0 (recall that such E-values are generally thought to be significant when less than $10^{-10.0}$). For each drug, we recorded the most extreme E-values against any muscarinic subtype. Those molecules with no predicted muscarinic targets were assigned an E-value of 100.0. The corresponding ROC area was 0.57 ± 0.05, significantly worse than the log-odds approach. As with the belief theory approach, interpretation of ROC areas is problematic due to tied values. With SEA, the tied values were at the low end of the ranking because the majority of drugs received no muscarinic target predictions at all. Maximal enrichment for the SEA approach occurred within the nontied value range at an E-value cutoff of $10^{-1.2}$, allowing for a direct comparison. Maximal enrichment was 3.8-fold. This corresponded to an FP rate of 4% and a TP rate of 15%. Three direct comparisons between the log-odds approach and SEA are particularly meaningful: (1) the maximal enrichment, which was 20-fold for log-odds vs 4-fold for SEA, (2) corresponding TP rates at the same 4% FP rate, 48% vs 15%, respectively, and (3) corresponding FP rates at the same 15% TP rate, 0% vs 4%.

We believe that the inherent degeneracy in Hooper's rule favoring high beliefs makes it inappropriate to use in a situation where belief values cannot be fully trusted. Given a single spurious annotation or a single similarity method yielding an inappropriately high confidence in a single molecular comparison, belief theory will produce incorrectly high belief in a prediction. In the case of SEA, we believe that the fundamental divergence of 2D similarity methods from the direct biophysical underpinnings of
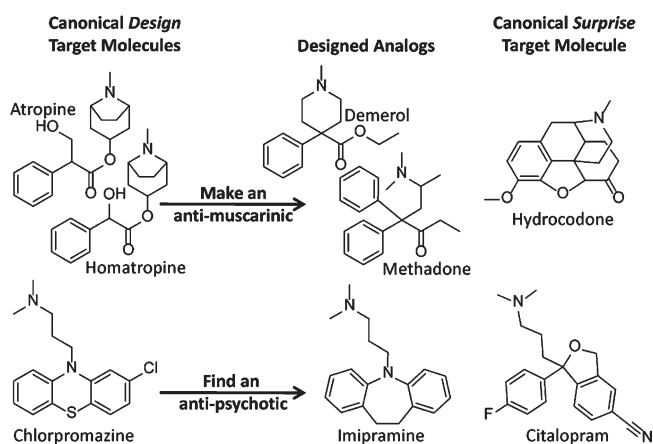
Figure 10. The design intention and surprising effects of some older drugs are known. Methadone and demerol were synthesized in an effort to make synthetically scalable anticholinergics as nerve gas antidotes by the Nazis in WWII. Their opioid effects were discovered serendipitously in a live cat assay. Imipramine was synthesized as an analogue of other tricyclic antipsychotics. Its surprising antidepressive effects were discovered through direct human experimentation, later identified to result from inhibition of the serotonin reuptake transporter. Using 3D similarity, one can correctly associate the *surprising* target effects with the drugs, but using 2D similarity, one cannot.

molecular activity limit the degree to which one can identify surprising off-target effects with high specificity.

Off-Target Prediction: Detection of Surprising Effects. The distinctions among different methods for data fusion, while clearly important, are not as critical as the distinctions among similarity methods that provide information to the data fusion computations themselves. Those similarity approaches whose scores are derived from directly relevant biophysical features (like surface shape and electrostatics) will yield different inferences than those that are less directly related to physical characteristics but which may be closely related to design ancestry. Two particularly telling examples of the distinction involve methadone and imipramine, compounds whose long history allows us to understand not only *what* the compounds do pharmacologically but also *why* they were synthesized and tested to begin with.

Figure 10 illustrates the historic context of the synthesis and testing of methadone and imipramine. Methadone was synthesized during WWII as part of an effort to develop anticholinergics for use as nerve gas antidotes[21] due to the limited availability of the natural product atropine (nerve gas results in an accumulation of acetylcholine by inhibition of acetylcholinesterase, leading to spasm and death). On testing in animals, the surprising finding was that methadone (and demerol as well) produced the Straub-tail effect, indicative of opioid analgesic activity. In a similar serendipitous story,[31] the compound G-22,355, which became known as imipramine, was selected for testing as an antipsychotic. Roland Kuhn, a psychiatrist at the Cantonal Mental Hospital of Münsterlingen, and Robert Domenjoz, a medicinal chemist at Geigy Pharmaceuticals, identified it as being structurally similar to chlorpromazine. Kuhn tested the compound with no success on psychotic patients, but prior to returning the supply, it was tested on a small number of depressive patients. The effects were sufficiently dramatic after just three patients to suggest the compound had unique properties and warranted further testing. Imipramine established a new class of drugs,[32] which ultimately came to be understood as acting primarily through the serotonin reuptake transporter.

Methadone's surprising on-target activity could have been predicted by the 3D log-odds approach based on the structures of morphinan-based opioids such as hydrocodone and codeine that had been identified well before methadone's synthesis. These had very low *p*-values using the Surflex-Sim approach. For hydrocodone, codeine, morphine, and oxycodone, the 3D *p*-values were, respectively, 0.007, 0.048, 0.057, and 0.060. The 2D GSIM *p*-values were, respectively, 0.35, 0.35, 0.35, and 0.63. Clearly, to predict the opioid activity, 3D structural comparisons would be required. The case of imipramine cannot be considered in this pseudoprospective fashion because its synthesis and testing led *subsequently* to the identification of both its primary biological mechanism of action as well as to the line of chemical inquiry that produced selective agents such as citalopram. However, if we consider citalopram's relationship to the eight serotonin-reuptake inhibitors that predated it from our set of 358 (imipramine, clomipramine, trimipramine, amitriptyline, trazodone, paroxetine, fluvoxamine, and fluoxetine), we see that 3D similarity yielded *p*-values ≤0.05 for all eight, but 2D similarity yielded *p*-values ≤0.05 for only three.

Overall, for known secondary targets (most of which can be considered surprises to some degree), the 3D log-odds scores were, on average, 9.3 log units higher than the 2D scores. For known primary targets (where relatively fewer can be considered surprises), the difference was 4.0 in favor of 3D log-odds over 2D. Relationships that can be deduced through 3D molecular similarity include those that genuinely are *surprising*, not just those that would be obvious to someone knowledgeable of molecular pharmacology in a particular area.

Recent Off-Target Predictions. Given these anecdotes, there clearly can be differences between the types of inferences that can be drawn from 2D and 3D molecular similarity methods. The supporting information behind predictions such as these is important because the natural application of computational methods for predicting off-target effects is to identify those that someone intimately involved in a particular pharmacological area could not reasonably guess. What we have seen is that in cases where we are able to understand both the reasoning behind molecular design and the serendipitous discoveries about activity, it is the province of 2D methods to uncover effects related to historical *reasoning* that anticipated the effects but 3D methods to *also* find the surprises.

In 2006, we observed that methadone, based on 3D molecular similarity, cosegregated with muscarinic and histamine receptor antagonists, echoing its genesis more than 60 years earlier.[2] We did not show biochemically that methadone was a muscarinic antagonist, but we pointed out that its side-effects included those associated with muscarinic antagonism: dry mouth, urinary retention, sweating, and reduced bowel motility. Subsequently, a biochemical assay showed that methadone has a K_i of 1.0 μM for the M3 receptor.[3] Using Surflex-Sim 3D similarity, methadone could be properly associated with the μ opioid receptor. What our study lacked was the perspective that 2D similarity provides as to what should have been considered *obvious* in this case: the basic reasoning behind synthesis of methadone was topological analogy to atropine and its analogues. Keiser et al.[3] directly showed that the SEA 2D similarity approach could reveal the off-target muscarinic effect of methadone (but not the on-target opioid effect). The 2D SEA approach successfully detected the association between methadone and the muscarinic receptor because attempts to create antimuscarinics from 2D analogy to atropine eventually succeeded, resulting in compounds such as
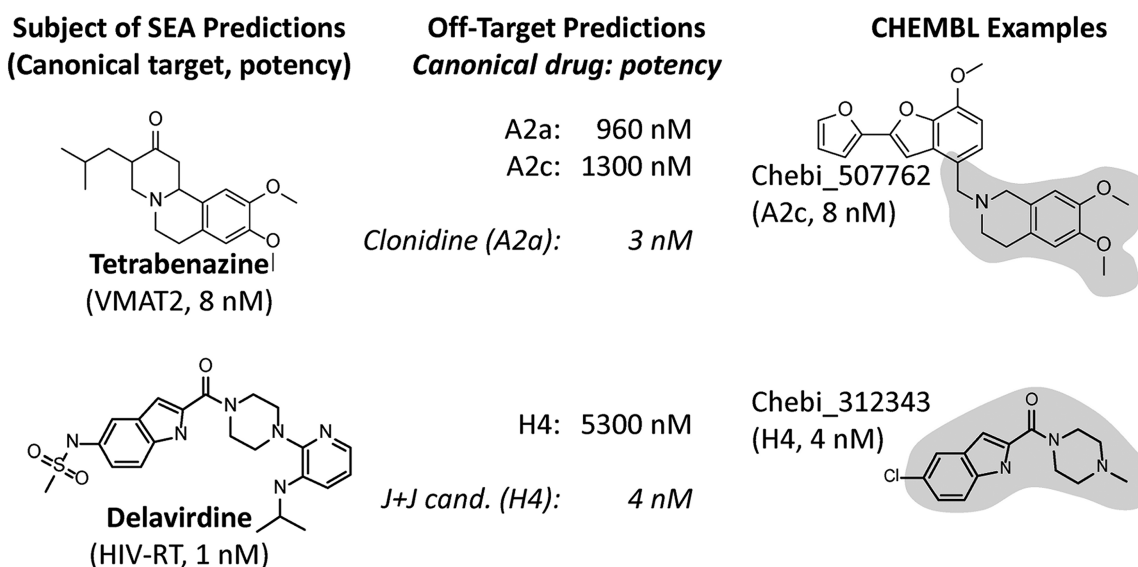
Figure 11. At left, two drugs are shown for which off-targets were identified through application of the SEA 2D similarity approach. Potencies for off-target effects were much weaker than for the on-target drug effects (shown below the drug names). The off-target effects were also much weaker than for primary modulators of the predicted targets. Molecular series containing common core substructures had been actively investigated as desirable scaffolds for the predicted targets.

adiphenine, diphenidol, tolterodine, oxybutynin, dicyclomine, and many others with a clear 2D similarity to methadone.

We observed this same pattern involving scaffold ancestry in a recently published application of the SEA approach.[4] In it, a set of predictions were correctly made for four drugs, where each of the predicted off-targets was unrelated by sequence or structure to the primary targets of the drugs. Figure 11 shows two of the drugs, primary canonical targets, predicted off-targets, and an example of a previously published[33−38] high-affinity ligand of each *off-target* protein that shares a scaffold with each drug. In each case, the scaffold in question had been actively probed in medicinal chemistry exercises for the predicted off-target effect. The specificity of the highlighted scaffold for the off-target in question among CHEMBL[39] annotations was over 40-fold for tetrabenazine, and the highlighted scaffold for the delavirdine prediction was over 1000-fold greater for H4 compared with any other target. Two other sets of predictions were made on drugs which target the NMDA receptor: ifenprodil and a simple analogue thereof. The predicted and verified activities included reuptake transporters (5HTT and NET), opioid receptors (μ and κ), and the D4 receptor. These activities shared the same pattern as those in Figure 11 with respect to the presence of previously published high-affinity analogues against the predicted targets (data not shown). The more general point relates to experimental molecular pharmacology. In 1991, ifenprodil was investigated for activity in addition to the NMDA and adrenergic ones already known,[15] and potent activity was reported for the σ and 5HT1a receptors. Established pharmacological crosstalk among ligands of σ receptors and the opioid μ, δ, and κ subtypes[13,40] anticipated weak opioid activities for ifenprodil and its analogue. Crosstalk between ligands of the adrenergic and 5HT receptors and reuptake transporters[14] anticipated these activities as well. Complex specificity patterns across multiple reuptake transporters and multiple receptor subtypes of σ, NMDA, opioid, adrenergic, and serotonin have been probed for many years.

The presence of many published ligand/target relationships provides data for computational inferences that parallel pharmacological
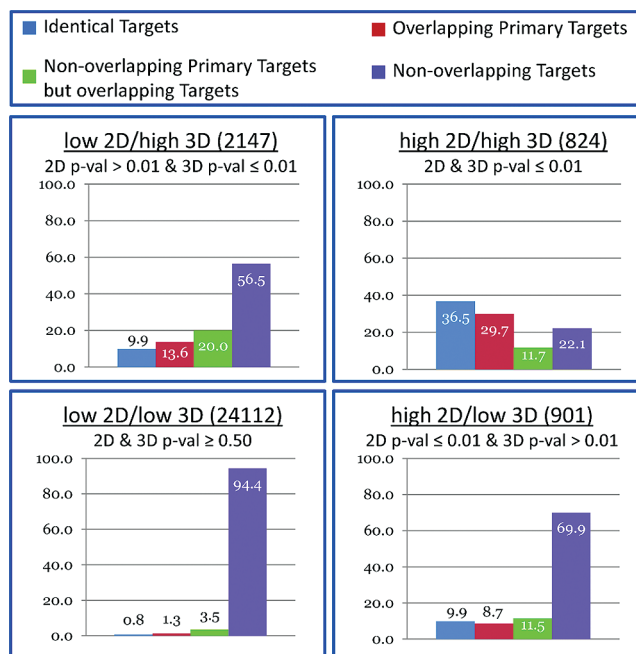
Figure 12. Drug pairs were segregated based on 2D and 3D similarity p-values into the 4 quadrants shown above (number of pairs per quadrant shown in parentheses). Conservative structural modifications are much more likely to yield highly similar pharmacology. Drug pairs with high2D/high3D similarity are about 4 times more likely to exhibit identical pharmacology relative to drug pairs with low2D/high3D similarity. Also, drug pairs with overlapping primary targets are 2 times more likely to have high2D/high3D similarity compared to those with low2D/high3D similarity. As expected, most drug pairs with high2D/low3D or low2D/low3D similarity have no target overlap.

knowledge. For predictions to have high practical utility, they must identify off-target effects for drugs automatically, reliably, and with high specificity, and ideally they must identify effects that
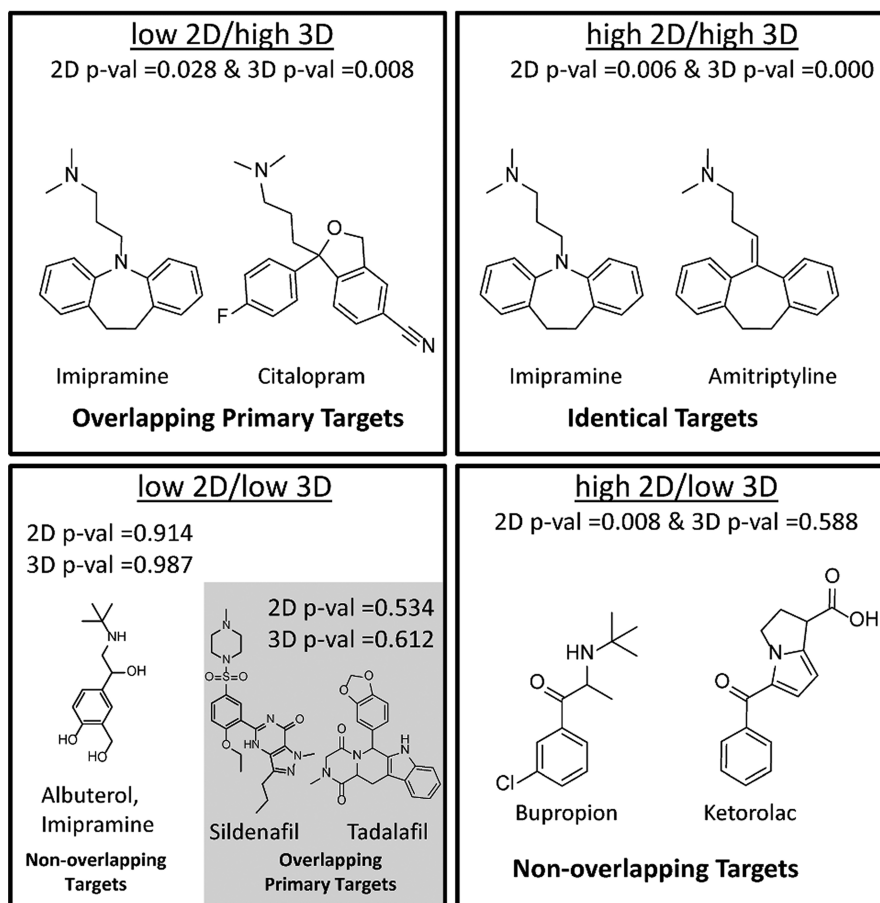
Figure 13. Imipramine and amitriptyline differ by only 1 atom, share identical on- and off-targets, and were approved as antidepressants in 1959 and 1961, respectively. In contrast, citalopram has significantly lower 2D similarity to imipramine, has fewer off-targets, and thus a more favorable pharmacological profile. Citalopram was approved as an antidepressant in 1998. For bupropion and ketorolac, the high 2D similarity is apparent on inspection, leading to a very low 2D *p*-value, but the 3D similarity correctly suggests that the compounds are unrelated in effects. Albuterol and imipramine are aminergic GPCR ligands with no target overlap and very low similarity. Sildenafil and tadalafil are a rare drug pair within the low 2D/3D quadrant that share a primary target (PDE5).

are truly surprising. Evaluating computational methods is challenging because even nominally prospective predictions can be driven by the evolutionary history of drugs. One can "predict" an activity for a ligand based on the fact that someone thought of the activity in connection with the ligand's scaffold before, causing analogues to be developed and probed for that activity. In such cases, tools that ferret out such information will be useful only to the extent that they are either more effective than someone knowledgeable in molecular pharmacology or that they are facile to apply automatically and have a low rate of false predictions. Developers of predictive methods should disclose the reasons why a method made a particular prediction. Usually this requires only the provision of typical molecular structures that underpinned an inference. Special care must be taken in the case of methods for predicting off-target effects because the goal is to identify those effects that might otherwise derail a clinical candidate, and it is reasonable to believe that the more obvious potential effects would have been extensively investigated.

Relationship of Structural Novelty to Pharmacology. From the foregoing discussion and our previous work,[1] it is clear that the drug design process shows a clear component of design relating directly to topological reasoning about the biological activity

expected from a particular molecular structure. It is also clear that clinically relevant surprises occur both with respect to primary targets as well as secondary ones. To assess the degree to which chemical structural novelty was directly related to novelty in pharmacological effect, we computed the pairwise similarity of all 358 drugs and split them into four groups: pairs with high 2D and high 3D similarity, low 2D but high 3D, low 3D but high 2D, and low 2D and low 3D. Figure 12 shows the proportions of molecule pairs within each group that had identical annotated targets (blue bars), overlapping primary targets but including some differences as to overall target effects (orange), nonoverlapping primary targets but some overlap among secondary targets (green), and completely nonoverlapping targets (purple). It is important to understand that the annotation of target effects include only those where sufficient experimentation exists in order to localize an effect to a specific binding site on a particular protein assembly. So, as we saw above with the analysis of muscarinic side-effects, many unannotated drug–target relationships may well exist.

In the case of high 3D and high 2D similarity (upper right), nearly 80% of drug pairs show some degree of target overlap, with nearly 40% having identical targets and nearly 70% sharing primary targets. With the same level of 3D similarity but with low 2D similarity (upper left), slightly less than half of the drugs

share targets, and just 10% have identical targets. The converse case (high 2D, low 3D, bottom right) produces somewhat similar proportions but with 70% having no common targets. As expected, molecules sharing no molecular similarity shared no targets nearly 95% of the time. Figure 13 shows examples from each quadrant. The case of imipramine and its fast follow-on compound amitriptyline fell into the identical target set; indeed, they have very little to differentiate them in terms of pharmacology even beyond specific targets.[41] However, the structural creativity shown by citalopram relative to imipramine (high 3D, low 2D) produced much more specificity with respect to the serotonin reuptake transporter, and citalopram along with other SSRIs came to dominate antidepressant therapy. A typical case for low 3D but high 2D similarity is the pair bupropion and ketorolac, which share no targets. The overwhelmingly common case for low 2D and low 3D similarity is exemplified by albuterol and imipramine, again sharing no common targets. About 2% of the time, drugs with some overlapping targets share no similarity at all. The case of sildenafil and tadalafil are a particularly striking example, both binding PDE5 within the same volume, but exhibiting no molecular similarity, either by eye or through computational means.[1] Note, however, that while the annotated targets were identical for the pair, their detailed pharmacology is significantly different, particularly with respect to half-life.

These findings are not surprising in a qualitative sense. It should be the case that nearly identical molecules will more frequently share very similar biological effects than those that start to differ. We believe that the *degree* of deviation in effects is striking. There is a 4-fold difference in the a priori likelihood that two drugs will share identical pharmacological targets when shifting from high 2D and high 3D similarity to a case that shares only high 3D congruence. Consider the case of designing a new drug with knowledge of the structures of existing drugs within a therapeutic category. In the modern research environment, it is likely that one will be able to guarantee that the desired target be among those that will exhibit pharmacological effects, but one cannot know that these effects will be the dominating ones. From Figure 12, we will consider the molecule pairs that share *some* targets in this analysis. By designing a me-too analogue (high 2D and 3D similarity to existing drugs), one has about a 47% chance of showing the same target profile as the incumbent compound versus showing either a difference in secondary targets or overlapping targets with different primary effects. By designing a structurally novel compound (high 3D but low 2D), one has a 23% chance of showing the same target profile. In the me-too case, chances are even (53%:47%) in terms of seeing novelty at the level of target specificity, but in the case of a structurally novel drug scaffold, the chances are 3:1 in favor of novelty (77%:23%). Two things are worth noting: First, with modern 3D molecular similarity and 3D QSAR methods, design of such compounds is tractable. Second, the development risks associated with novelty are almost certainly higher because one cannot know a priori with full confidence what the precise biological effect differences might be, only that one is more likely to encounter them.

■ CONCLUSIONS

We have reported a new method for combining information from molecular similarity computations in order to effectively make inferences based on the known activity of sets of molecules. The approach is general and can be applied to any similarity method, offering a unified means to fuse the output from multiple

sources to produce a single log-odds score. Two aspects are particularly important: (1) mapping of similarity values to p-values in a *context-specific* way because raw similarity values have different significance depending on the complexity of the molecules being compared, and (2) offering a means of data fusion that balances evidence in favor of an assertion with evidence against the assertion while avoiding degeneracies that arise from literal interpretation of empirically estimated p-values.

By comprehensively applying the method to a large set of drugs with overlapping pharmacological effects, we were able to identify differences in the predictive ability of 2D similarity methods compared with 3D ones. In assessing the predictive value of the approach, the most comprehensive analyses were done considering recovery of known primary and secondary targets. Particularly for the prediction of unanticipated off-target effects, 3D performed much better than 2D, although the combination of both was beneficial. Comprehensive analysis of false positive predictions was impossible due to the lack of systematic profiling of drugs against large panels of targets. However, using the well-known muscarinic side effects as a surrogate for direct muscarinic modulation, we assessed the behavior of drugs lacking explicit annotation. Within this set, we established excellent separation of drugs with muscarinic side-effects from those with none apparent. Alternative approaches such as belief theory and SEA performed less well.

We also considered the relationship between chemical structural novelty and pharmacological novelty. The key finding was that for a me-too drug pair (exhibiting both high 2D and 3D similarity), the chances were essentially even of observing modulation of identical sets of biological targets compared with non-identical sets. However, in the case where the drugs show high 3D similarity but show differences at the topological level, these odds shifted to roughly three-to-one in favor of observing novel effects at the gross level of target modulation. Clearly, even very subtle changes in chemical structure can yield sufficient differences in potency, selectivity, or ADME/toxicity characteristics to make for novelty in pharmacological action that can provide benefits to patient populations over existing therapies. This is especially true for therapeutics that target rapidly evolving pathogens, where minor structural modifications can overcome resistance. But it is clear that introduction of a novel scaffold brings significantly higher risk and potentially higher reward in terms of novelty that might be beneficial for patients.

While we believe that more detailed study is warranted, consideration should be given to regulatory changes that better balance the risk/reward equation for drug discovery. One possibility would be to *always* require a head-to-head clinical comparison against an approved close analogue in cases of me-too drug candidates seeking regulatory approval (excepting those that target pathogens). For drug candidates exhibiting novel structures, such requirements might still be imposed as they are currently on an ad hoc basis but would not be presumptively required. We believe that a disproportionate amount of research and clinical development effort is currently spent on drugs that have relatively little chance of providing significant new benefits to patients. Additional study will be required, however, in order to quantify this to the extent desirable for a regulatory modification.

An area for future research will involve systematic projection of phenotypic side-effects shared by drugs onto potential biological targets. As shown by Scheiber et al.,[42] integration of such projections with information about biological network structures can help to identify the specific molecular basis for clinically

important adverse drug reactions. Use of more sophisticated methods for the underlying chemical to target inferences, as presented here, should serve to make such exercises more effective in identifying the causative factors.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (415) 502-7242. Fax: (650) 240-1781. E-mail: ajain@jainlab.org.

■ ACKNOWLEDGMENT

■ ABBREVIATIONS USED

(ROC), receiver-operator characteristic; SEA, similarity ensemble approach; TP, true-positive; FP, false-positive

■ REFERENCES

(1) Cleves, A. E.; Jain, A. N. Effects of inductive bias on computational evaluations of ligand-based modeling and on drug discovery. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 147–159.

(2) Cleves, A. E.; Jain, A. N. Robust ligand-based modeling of the biological targets of known drugs. *J. Med. Chem.* **2006**, *49*, 2921–2938.

(3) Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating protein pharmacology by ligand chemistry. *Nature Biotechnol.* **2007**, *25*, 197–206.

(4) Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175–181.

(5) Muchmore, S. W.; Debe, D. A.; Metz, J. T.; Brown, S. P.; Martin, Y. C.; Hajduk, P. J. Application of belief theory to similarity data fusion for use in analog searching and lead hopping. *J. Chem. Inf. Model.* **2008**, *48*, 941–948.

(6) Taboureau, O.; Nielsen, S. K.; Audouze, K.; Weinhold, N.; Edsgärd, D.; Roque, F. S.; Kouskoumvekaki, I.; Bora, A.; Curpan, R.; Jensen, T. S. ChemProt: A disease chemical biology database. *Nucleic Acids Res.* **2011**, *39*, D367–D372.

(7) Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A. C.; Wishart, D. S. DrugBank 3.0: a comprehensive resource for "omics" research on drugs. *Nucleic Acids Res.* **2011**, *39*, 1035–1041.

(8) Eyre, T. A.; Ducluzeau, F.; Sneddon, T. P.; Povey, S.; Bruford, E. A.; Lush, M. J. The HUGO Gene Nomenclature Database, 2006 Updates. *Nucleic Acids Res.* **2006**, *34*, 319–321.

(9) 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.* **2006**, *46*, 1124–1133.

(10) Yildirim, M. A.; Goh, K. I.; Cusick, M. E.; Barabasi, A. L.; Vidal, M. Drug-target network. *Nature Biotechnol.* **2007**, *25*, 1119–1126.

(11) Chen, X.; Lin, Y.; Liu, M.; Gilson, M. K. The Binding Database: data management and interface design. *Bioinformatics* **2002**, *18*, 130–139.

(12) Kuhn, M.; Campillos, M.; Letunic, I.; Jensen, L. J.; Bork, P. A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* **2010**, *6*, 343.

(13) Largent, B. L.; Wikstrom, H.; Gundlach, A. L.; Snyder, S. H. Structural determinants of sigma receptor affinity. *Mol. Pharmacol.* **1987**, *32*, 772–784.

(14) Silvestrini, B. Beyond the limits of typical antidepressants. *Psychopharmacology (Berlin)* **1988**, *95* (Suppl), S1–S2.

(15) Chenard, B. L.; Shalaby, I. A.; Koe, B. K.; Ronau, R. T.; Butler, T. W.; Prochniak, M. A.; Schmidt, A. W.; Fox, C. B. Separation of alpha 1 adrenergic and N-methyl-D-aspartate antagonist activity in a series of ifenprodil compounds. *J. Med. Chem.* **1991**, *34*, 3085–3090.

(16) Jain, A. N. Morphological similarity: a 3D molecular similarity method correlated with protein−ligand recognition. *J Comput.-Aided Mol. Des.* **2000**, *14*, 199–213.

(17) Jain, A. N. Ligand-based structural hypotheses for virtual screening. *J. Med. Chem.* **2004**, *47*, 947–961.

(18) Nicholls, A.; McGaughey, G. B.; Sheridan, R. P.; Good, A. C.; Warren, G.; Mathieu, M.; Muchmore, S. W.; Brown, S. P.; Grant, J. A.; Haigh, J. A.; Nevins, N.; Jain, A. N.; Kelley, B. Molecular shape and medicinal chemistry: a perspective. *J. Med. Chem.* **2010**, *53*, 3862–386.

(19) Pham, T. A.; Jain, A. N. Parameter estimation for scoring protein−ligand interactions using negative training data. *J. Med. Chem.* **2006**, *49*, 5856–5868.

(20) Pham, T. A.; Jain, A. N. Customizing scoring functions for docking. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 269–286.

(21) Sneader, W. *Drug Discovery: A History*; John Wiley & Sons Inc: New York, 2005.

(22) Horn, F.; Vriend, G.; Cohen, F. E. Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems. *Nucleic Acids Res.* **2001**, *29*, 346–349.

(23) Coupet, J.; Fisher, S. K.; Rauh, C. E. Interaction of amoxapine with muscarinic cholinergic receptors: an in vitro assessment. *Eur. J. Pharmacol.* **1985**, *112*, 231–235.

(24) Lin, S. C.; Olson, K. C.; Okazaki, H.; Richelson, E. Studies on muscarinic binding sites in human brain identified with [3H] pirenzepine. *J. Neurochem.* **1986**, *46*, 274–279.

(25) Syvalahti, E. K.; Kunelius, R.; Lauren, L. Effects of antiparkinsonian drugs on muscarinic receptor binding in rat brain, heart and lung. *Pharmacol. Toxicol.* **1988**, *62*, 90–94.

(26) Su, Z.; Martin, R.; Cox, B. F.; Gintant, G. Mesoridazine: an open-channel blocker of human ether-a-go-go-related gene K^+ channel. *J. Mol. Cell. Cardiol.* **2004**, *36*, 151–160.

(27) Brown, E. G.; Wood, L.; Wood, S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf.* **1999**, *20*, 109–117.

(28) Shafer, G. The combination of evidence. *Int. J. Intell. Syst.* **1986**, *1*, 155–179.

(29) Rush, T. S., III; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein−protein interaction. *J. Med. Chem.* **2005**, *48*, 1489–1495.

(30) Flower, D. R. On the properties of bit string-based measures of chemical similarity. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 379–386.

(31) Healy, D. Obituary: Roland Kuhn (1912−2005). *Hist. Psychiatry* **2006**, *17*, 253–255.

(32) Kuhn, R. The treatment of depressive states with G 22355 (imipramine hydrochloride). *Am. J. Psychiatry* **1958**, *115*, 459–464.

(33) Hagihara, K.; Kashima, H.; Iida, K.; Enokizono, J.; Uchida, S.; Nonaka, H.; Kurokawa, M.; Shimada, J. Novel 4-(6,7-dimethoxy-1,2,3,4-tetrahydroisoquinolin-2-yl)methylbenzofuran derivatives as selective alpha(2C)-adrenergic receptor antagonists. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1616–1621.

(34) Jablonowski, J. A.; Grice, C. A.; Chai, W.; Dvorak, C. A.; Venable, J. D.; Kwok, A. K.; Ly, K. S.; Wei, J.; Baker, S. M.; Desai, P. J.; Jiang, W.; Wilson, S. J.; Thurmond, R. L.; Karlsson, L.; Edwards, J. P.; Lovenberg, T. W.; Carruthers, N. I. The first potent and selective non-imidazole human histamine H4 receptor antagonists. *J. Med. Chem.* **2003**, *46*, 3957–3960.

(35) Efange, S. M.; Kamath, A. P.; Khare, A. B.; Kung, M. P.; Mach, R. H.; Parsons, S. M. N-Hydroxyalkyl derivatives of 3 beta-phenyltropane and 1-methylspiro[1H-indoline-3,4′-piperidine]: vesamicol analogues with affinity for monoamine transporters. *J. Med. Chem.* **1997**, *40*, 3905–3914.

(36) Wright, J. L.; Downing, D. M.; Feng, M. R.; Hayes, R. N.; Heffner, T. G.; MacKenzie, R. G.; Meltzer, L. T.; Pugsley, T. A.; Wise, L. D. Identification, characterization and pharmacological profile of three metabolites of (R)-(+)-1,2,3,6-tetrahydro-4-phenyl-1-[(3-phenyl-cyclohexen-1- yl)methyl]pyridine (CI-1007), a dopamine autorecep-tor agonist and potential antipsychotic agent. *J. Med. Chem.* **1995**, *38*, 5007–5014.

(37) Thomas, J. B.; Mascarella, S. W.; Rothman, R. B.; Partilla, J. S.; Xu, H.; McCullough, K. B.; Dersch, C. M.; Cantrell, B. E.; Zimmerman, D. M.; Carroll, F. I. Investigation of the N-substituent conformation governing potency and mu receptor subtype-selectivity in (+)-(3R,4R)-dimethyl-4-(3-hydroxyphenyl)piperidine opioid antagonists. *J. Med. Chem.* **1998**, *41*, 1980–1990.

(38) Zimmerman, D. M.; Leander, J. D.; Cantrell, B. E.; Reel, J. K.; Snoddy, J.; Mendelsohn, L. G.; Johnson, B. G.; Mitch, C. H. Structure—activity relationships of *trans*-3,4-dimethyl-4-(3-hydroxyphenyl)piperi-dine antagonists for mu- and kappa-opioid receptors. *J. Med. Chem.* **1993**, *36*, 2833–2841.

(39) De Matos, P.; Alcántara, R.; Dekker, A.; Ennis, M.; Hastings, J.; Haug, K.; Spiteri, I.; Turner, S.; Steinbeck, C. Chemical entities of biological interest: an update. *Nucleic Acids Res.* **2010**, *38*, D249–D254.

(40) Whittemore, E. R.; Ilyin, V. I.; Woodward, R. M. Antagonism of N-methyl-D-aspartate receptors by sigma site ligands: potency, subtype-selectivity and mechanisms of inhibition. *J. Pharmacol. Exp. Ther.* **1997**, *282*, 326–338.

(41) Hollister, L. E. O. E.; Overall, J. E.; Johnson, M.; Pennington, V.; Katz, G.; Shelton, J. Controlled comparison of amitriptyline, imipramine and placebo in hospitalized depressed patients. *J. Nerv. Ment. Dis.* **1964**, *139*, 370–375.

(42) Scheiber, J.; Chen, B.; Milik, M.; Sukuru, S. C.; Bender, A.; Mikhailov, D.; Whitebread, S.; Hamon, J.; Azzaoui, K.; Urban, L.; Glick, M.; Davies, J. W.; Jenkins, J. L. Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis. *J. Chem. Inf. Model.* **2009**, *49*, 308–317.