JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

3D Molecular Descriptors Important for Clinical Success

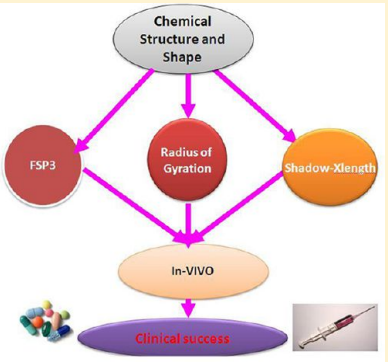
David C. Kombo,* Kartik Tallapragada, Rachit Jain,† Joseph Chewning, Anatoly A. Mazurov, Jason D. Speake, Terry A. Hauser, and Steve Toler

Targacept, Inc., 200 East First Street, Suite 300, Winston-Salem, North Carolina 27101-4165, United States

†Rutgers University, Busch Campus, 98 Brett Rd., Piscataway, New Jersey 08854, United States

Supporting Information

ABSTRACT: The pharmacokinetic and safety profiles of clinical drug candidates are greatly influenced by their requisite physicochemical properties. In particular, it has been shown that 2D molecular descriptors such as fraction of Sp3 carbon atoms (Fsp3) and number of stereo centers correlate with clinical success.[1] Using the proteomic off-target hit rate of nicotinic ligands, we found that shape-based 3D descriptors such as the radius of gyration and shadow indices discriminate off-target promiscuity better than do Fsp3 and the number of stereo centers. We have deduced the relevant descriptor values required for a ligand to be nonpromiscuous. Investigating the MDL Drug Data Report (MDDR) database as compounds move from the preclinical stage toward the market, we have found that these shape-based 3D descriptors predict clinical success of compounds at preclinical and phase1 stages vs compounds withdrawn from the market better than do Fsp3 and LogD. Further, these computed 3D molecular descriptors correlate well with experimentally observed solubility, which is among well-known physicochemical properties that drive clinical success. We also found that about 84% of launched drugs satisfy either Shadow index or Fsp3 criteria, whereas withdrawn and discontinued compounds fail to meet the same criteria. Our studies suggest that spherical compounds (rather than their elongated counterparts) with a minimal number of aromatic rings may exhibit a high propensity to advance from clinical trials to market.



INTRODUCTION

In recent years, significant progress in understanding biomolecular mechanisms underlying drug action coupled with the development of high-throughput screening (HTS) has resulted in thousands of chemical entities available for the treatment of pathological conditions and diseases. HTS hits, however, include false positives that can stem from interfering in binding via aggregates formation,[2−5] protein reactive functional groups,[6,7] and pan assay interference compounds (Pains).[8] Substructural features characterizing these types of promiscuous compounds have been intensely discussed.[2−8] Furthermore, contemporary rational drug design does not ensure development of pharmaceutical agents with minimal side effects. Even recently, several drugs have been withdrawn from the market place or their use has been restricted. In 2000, the reasons for drug candidate failure in clinical stage were estimated as follows: side effects (31%), efficacy (24%), commercial reasons (19%), bioavailability (8%), formulation (4%), and others (14%).[9] Side effects may arise from the fact that compounds can authentically bind to more than one molecular target. This type of promiscuity is the focus of our current work reported herein.

As Paul Ehrlich noticed, successful research still needs the four Gs (in German): Glück, Geduld, Geschick, and Geld; these can be translated in English as follows: luck, patience, skill, and money.[10] Although the last three prerequisites of successful research gained impressive commitments, the prospect of a new chemical entity to be a successful medicine remains remote. The selection of small molecules with optimized physicochemical properties can enhance the viability of drug candidates, reducing the need for "luck".

Since the seminal work of Lipinski et al. on oral bioavailability,[11] and the subsequent refinement of Lipinski's rule of five (RO5), 2D molecular descriptors such as molecular weight (MW), LogP, number of hydrogen-bond acceptors (HBA), number of hydrogen-bond donors (HBD), number of rotatable bonds (NROT), topological molecular surface area (PSA), and number of aromatic rings have been routinely used in screening for "drug-likeness", as compounds move from hit finding to the preclinical stage.[11−14] To address molecular complexity, Lovering et al. have proposed and successfully demonstrated that an increase in saturation as measured by Fsp3 and number of chiral centers in the molecule improves clinical success.[1] Clemons et al. have successfully shown that structural complexitiy of small molecules of different origins correlate with protein-binding profiles.[15] In a follow-up study, Clemons et al. extensively characterized structural properties, physicochemical properties, shape-based 3D descriptors, and performance diversity of these data sets of different origins.[16] This important study paved the way to a paradigm shift in the assessment of biological performance, from an individual context to a collective one.

Received: September 18, 2012
Published: December 17, 2012

Nicotinic ligands are basic amines that bind and modulate the nicotinic acetylcholine receptors (nAChRs). These receptors are ion-channels that have been validated as therapeutic targets for various pathologies of the central nervous system (CNS).[17−23] Examples of therapeutic indications currently being investigated include Alzheimer's disease, Parkinson's disease, cognitive dysfunction in Schizophrenia, smoking cessation, and attention deficit hyperactivity. Basic amines, however, are known to be the most promiscuous class of ligands targeting ion channels on the basis of studies by Hoffman-La Roche[24,25] and Novartis.[25−27] In the present study, we have used nicotinic ligands as a model system and have aimed at answering the following questions: Do saturation and number of stereo centers impact off-target promiscuity? As ligand-binding events and resulting functional response are mainly associated with 3D structure and shape of both the compound and the cognate macromolecular receptor, are there any 3D descriptors that can predict off-target interactions? If yes, what are the critical values that one can use in a high throughput virtual screening endeavor? Could the ability of such descriptors to predict promiscuity be translated to predict clinical success in general? We herein show that shape descriptors such as shadow indices and radius of gyration strongly influence off-target activity. We have determined the threshold values required for lack of promiscuity and have applied our findings to analyze MDDR[28] compounds as they transition through various clinical development stages. Similar to Fsp3, we find that these molecular descriptors correlate well with experimentally observed solubility, which is also known to drive clinical success.

MATERIALS AND METHODS

Chemical Library, Biological Activity, and Off-Target Selectivity. Nicotinic ligands studied include compounds well characterized in the literature such as ABT-894 (compound 12), Varenicline (compound 16), and other compounds described in patent applications as listed below. Pairwise Tanimoto similarity was computed using functional class FCFP-6 molecular fingerprints (with stereochemistry), as implemented within Pipeline Pilot (Accelrys, Inc., San Diego, CA, U.S.A.). The off-target binding profile of these compounds was experimentally determined at Caliper Life Science using their in vitro molecular screening and profiling platform, as described at http://www.caliperls.com. Off-target proteins tested were comprised of neurotransmitter related receptors, ion channels, second messengers receptors, steroids receptors, prostaglandin receptors, growth factors/hormones receptors, brain/gut peptides receptors, cytochrome P450, and other enzymes. Details on the screening assays and targets used are described in the Supporting Information (SI) sheets S1 and S2, which also illustrate the percent inhibition values derived at 10 uM for compounds 12 (ABT-894) and 16 (Varenicline), respectively. The off-target hit rate is defined as the ratio of the number of times the percent inhibition derived at 10 uM is greater or equal to 50% by the total number of off-target receptors assayed. Experimental procedure and details on the binding affinity to nAChRs are also given in the Supporting Information.

Exemplified Structures. Chemical names for representative compounds are as follows: benzofuran-2-yl(3-(pyridin-3-yl)-hexahydro-4,7-ethanopyrrolo[3,2-b]pyridin-1(2H)-yl)-methanone (**1**),[29] 4-bromo-N-[2-(3-pyridylmethyl)quinuclidin-3-yl]benzamide (**2**),[30] [2-(3-pyridylmethyl)quinuclidin-3-yl] N-(4-bromophenyl)carbamate (**3**),[30] 5-methyl-N-[2-(3-pyridylmethyl)quinuclidin-3-yl]thiophen-2-carboxamide (**4**),[31] (1R,5R)-7-(3-pyridyl)-3-azabicyclo[3.3.1]non-6-ene (**5**),[32] (E,2S)-5-[5-(4-chlorophenoxy)-3-pyridyl]-N-methyl-pent-4-en-2-amine (**6**),[33] 3-(3-pyridyl)-6-azabicyclo[3.2.1]oct-3-ene (**7**),[34] (E)-4-(5-ethoxy-3-pyridyl)-N-methyl-but-3-en-1-amine (**8**),[35] 8-(3-pyridyl)-4,8-diazaspiro[4.4]nonane (**9**),[36] 3,6-diazabicyclo[3.1.1]heptane-3-yl(cyclopropyl) methanone (**10**),[37](1R,5S)-7-(3-pyridyl)-3-azabicyclo[3.3.1]non-6-ene (**11**),[32] (1S,5S)-3-(5,6-dichloro-3-pyridyl)-3,6-diazabicyclo[3.2.0]heptanes (**12**),[38] 3-(3-pyridylmethyl)-7-azabicyclo[2.2.1]heptanes (**13**),[39] (5S)-8-(3-pyridyl)-4,8-diazaspiro[4.4]nonane (**14**),[36] 2-(5-ethoxypyridin-3-yl)-azaadamantane (**15**),[40] 7,8,9,10-tetrahydro-6,10-methano-6H-pyrazino(2,3-h)(3)benzazepine (**16**),[41] 5-(quinuclidin-2-yl)-pyridin-3-amine (**17**),[42] 3-(5-fluoro-3-pyridyl)-9-oxa-3,7-diazabicyclo[3.3.1]nonane (**18**),[43] (1R,5R)-3-methyl-7-(3-pyridyl)-3-azabicyclo[3.3.1]non-6-ene (**19**),[32] (E,2S)-5-[5-(1H-indol-5-yloxy)-3-pyridyl]-N-methyl-pent-4-en-2-amine (**20**).[35] furan-2-yl-(hexahydro-pyrrolo[3,4-c]pyrrol-2-yl)-methanone (**21**),[44] (3,7-diaza-bicyclo[3.3.1]non-3-yl)-furan-2-yl-methanone (**22**),[44] (5-chlorofuran-2-yl)-(3,7-diaza-bicyclo[3.3.1]non-3-yl)-methanone (**23**),[44] (hexahydro-pyrrolo[3,4-c]pyrrol-2-yl)-(3-methyl-isoxazol-4-yl)-methanone (**24**),[44] (hexahydro-pyrrolo[3,4-c]pyrrol-2-yl)-(4-methyl-oxazol-5-yl)-methanone (**25**),[44](2S,3R)-N-[2-(pyridin-3-ylmethyl)-1-azabicyclo[2.2.2]-oct-3-yl]-4-fluorobenzamide (**26**),[31,45] (2S,3R)-N-[2-(pyridin-3-ylmethyl)-1-azabicyclo[2.2.2]oct-3-yl]benzo[b]furan-2-carboxamide (**27**),[31,45](2R,3S)-N-[2-(pyridin-3-ylmethyl)-1-azabicyclo[2.2.2]oct-3-yl]benzo[b]furan-2-carboxamide (**28**),[31,45] (2S,3R)-N-[2-(pyridin-3-ylmethyl)-1-azabicyclo[2.2.2]oct-3-yl]benzo[b]isoquinolin-3-carboxamide (**29**),[31,45] (2S,3R)-N-[2-(pyridin-3-ylmethyl)-1-azabicyclo[2.2.2]oct-3-yl]benzo[b]-5-bromofuran-2-carboxamide (**30**),[31,45] (2S,3R)-N-[2-(pyridin-3-ylmethyl)-1-azabicyclo[2.2.2]oct-3-yl]benzo[b]-5-methylthiophen-2-carboxamide (**31**),[31,45] (2S,3R)-N-[2-(pyridin-3-ylmethyl)-1-azabicyclo[2.2.2]oct-3-yl]benzo[b]-5-(pyridin-2-yl)thiophen-2-carboxamide (**32**),[31,45] (2S,3R)-N-[2-(pyridin-3-ylmethyl)-1-azabicyclo[2.2.2]oct-3-yl]-5-chlorobenzo[b]furan-2-carboxamide (**33**),[31,45] and (2S,3R)-N-[2-(pyridin-3-ylmethyl)-1-azabicyclo[2.2.2]oct-3-yl]-5-hydroxybenzo[b]furan-2-carboxamide (**34**).[31,45]

Computation of Molecular Descriptors. The following structural, spatial, thermodynamic, topological, and electronic descriptors, as implemented within Pipeline Pilot, were used: atom polarizability, dipole moment, charge, Kier and Hall molecular connectivity indices (Chi), valence-modified connectivity indices (Chi-v), subgraph count indices (SC), Kier's shape indices (Kappa), E-state keys, radius of gyration, shadow indices, area, molecular volume, density, principal moment of inertia, and Jurs descriptors. LogD was calculated at the physiological pH of 7.4. In addition, we included Fsp3 and number of stereo centers. Following the definitions of Lovering et al.,[1] carbon bond saturation expressed in terms of Fsp3 and number of stereo centers were calculated using Pipeline Pilot 8.5 as follows

$$\text{Fsp3} = (\text{Num of sp3 hybridized carbons}/\text{total carbon count}) \quad (1)$$

dx.doi.org/10.1021/ci300445e | J. Chem. Inf. Model. 2013, 53, 327−342

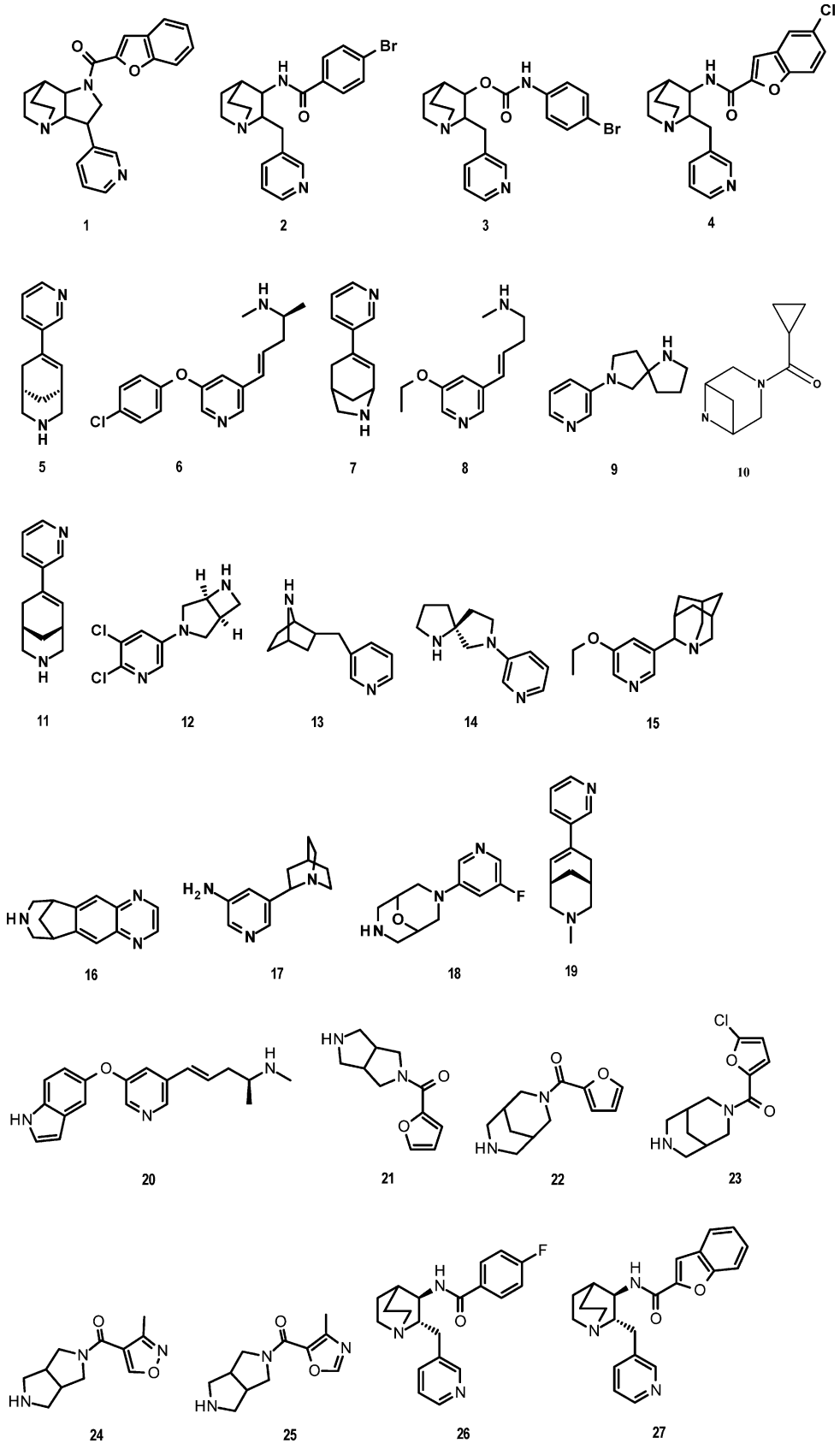


Figure 1. continued

dx.doi.org/10.1021/ci300445e | *J. Chem. Inf. Model.* 2013, 53, 327−342

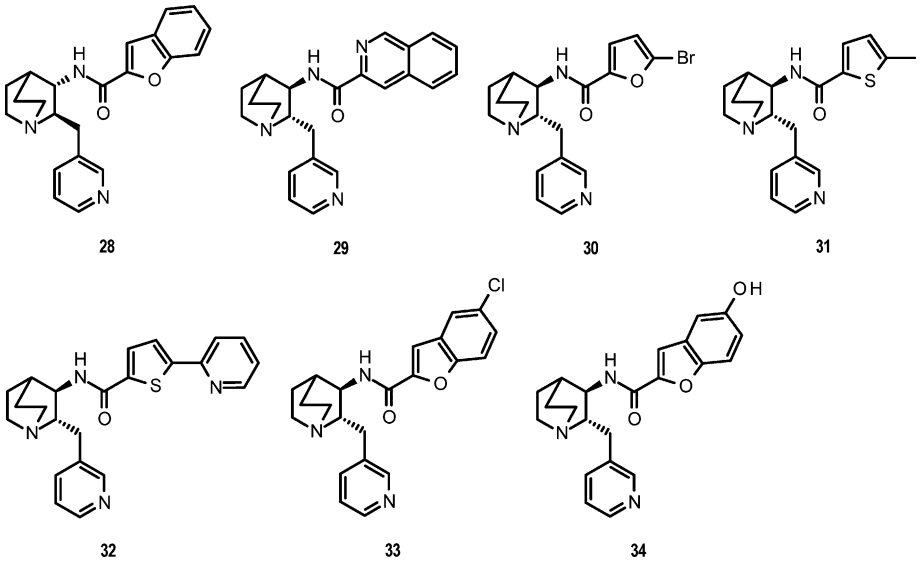


Figure 1. Representative chemical structures of compounds used to calculatee off-target hit rate and molecular descriptors. Chemical names of structures are listed in the Materials and Methods section.

$$\text{Num of stereo centers} = (\text{NumUnknownTrueStereoAtoms} + \text{Num TrueStereoAtoms}) \quad (2)$$

where Num designates number.

A total of 132 descriptors was computed and investigated. Prior to calculating 3D descriptors, 3D models for each structure were energy-minimized using the Clean force field.[46] Compounds containing a basic nitrogen atom were protonated accordingly.

Preclinical, Clinical, and Solubility Data. All of the compounds advancing in the drug development cycle that were investigated herein were retrieved from the bioactivity database MDDR. In order to cross-check the structures and the highest clinical phase reached, we also used the Drug Bank[47−49] and our proprietary competitive Intelligence database. Solubility data were retrieved from the following public Web site: http://modem.ucsd.edu/adme/databases/databases_logS.htm, as provided by Hou et al.[50]

Statistical Analysis. To determine whether or not two data sets differ significantly, statistical analysis was carried out by using the Kolmogorov−Smirnov test (KS-test),[51] a nonparametric statistical test suitable for comparing frequency distributions, which has the advantage of making no assumptions about the distribution of data. A significance level of 0.05 was used throughout the study. The IBM SPSS statistics computer package version 20 (IBM, Armonk, NY) was used to perfom the analysis. Mutliple hypothesis testing correction was carried out by calculating the q-value as an analog of the *p*-value that incorporates false discovery rate estimation.[52]

RESULTS AND DISCUSSION

Data Sets, Off-Target Hit Rate, and Ligand Promiscuity. We have investigated nicotinic ligands that have been tested in at least 45 out of a maximum number of 71 assays commonly used in an off-target panel. Some representative structures used in this study are shown in Figure 1.

Binding affinities of these nicotinic compounds to nAChRs are shown in Table S1 of the Supporting Information. To evaluate compound promiscuity across this whole panel of assays, we calculated the proteomic "off-target hit rate" at 10 uM concentration ($OTHR_{10}$), defined as follows:

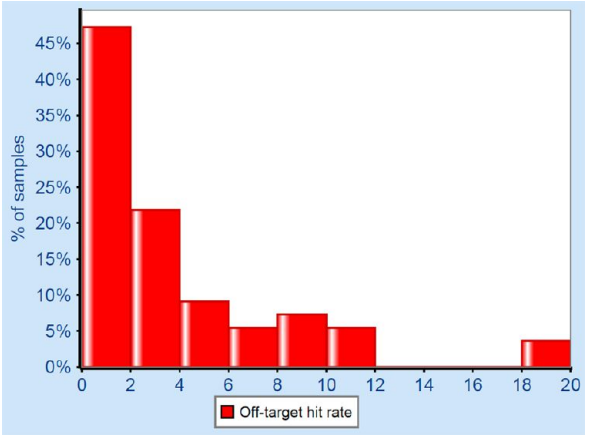


Figure 2. Distribution of compounds by their off-target hit rate derived at 10 uM.

$$OTHR_{10} = (\text{Number of hits})/(\text{Total number of receptors assayed}) \quad (3)$$

where a compound is said to be a hit if its percent inhibition derived at 10 uM is greater or equal to 50%. This definition was originally suggested by Azzaoui et al.[53] and was then termed "target hit rate" in their study.

The distribution of compounds by their $OTHR_{10}$ is shown in Figure 2. The mean value of $OTHR_{10}$ is 4%. We decided to use this value as a threshold for medium promiscuity. Thus, compounds with $OTHR_{10} \leq 4$ were flagged as selective, and those with $OTHR_{10} > 4$ were flagged as medium promiscuous. This classification is consistent with the above-mentioned Azaoui et al. study on compounds promiscuity, which used a larger data set and suggested a selectivity threshold of 5%. We found that about 70% of compounds were selective (S) and 30% were medium promiscuous (MP). Interestingly, Varenicline, launched as Chantix by Pfizer, which is a nicotinic ligand currently on the

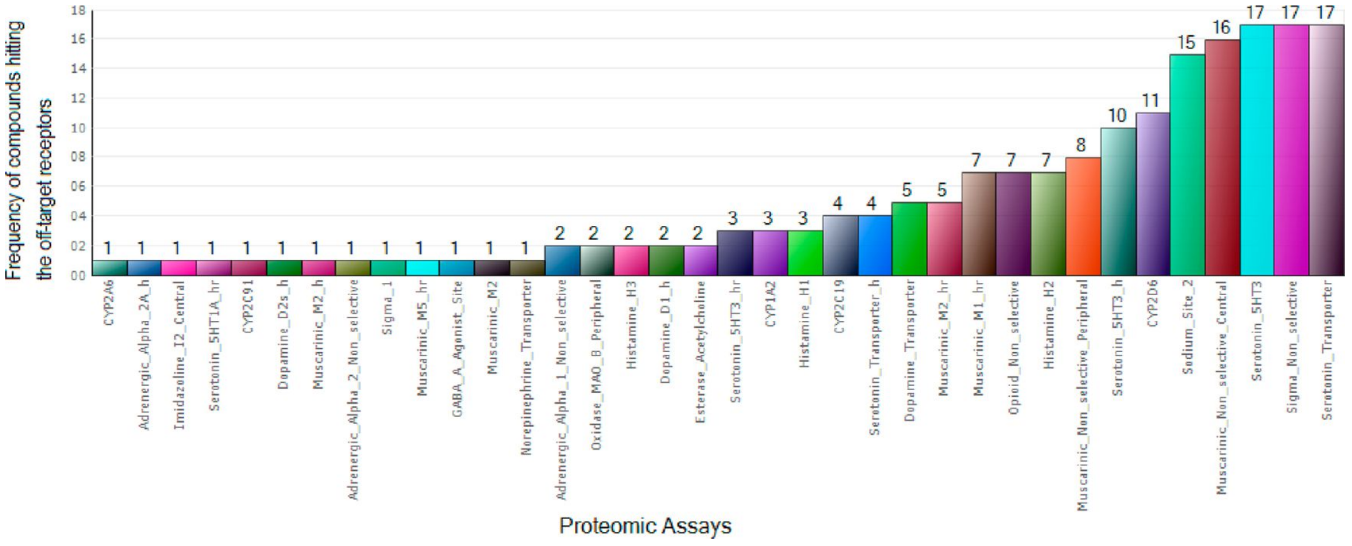


Figure 3. Distribution of number of hits per off-target receptor.

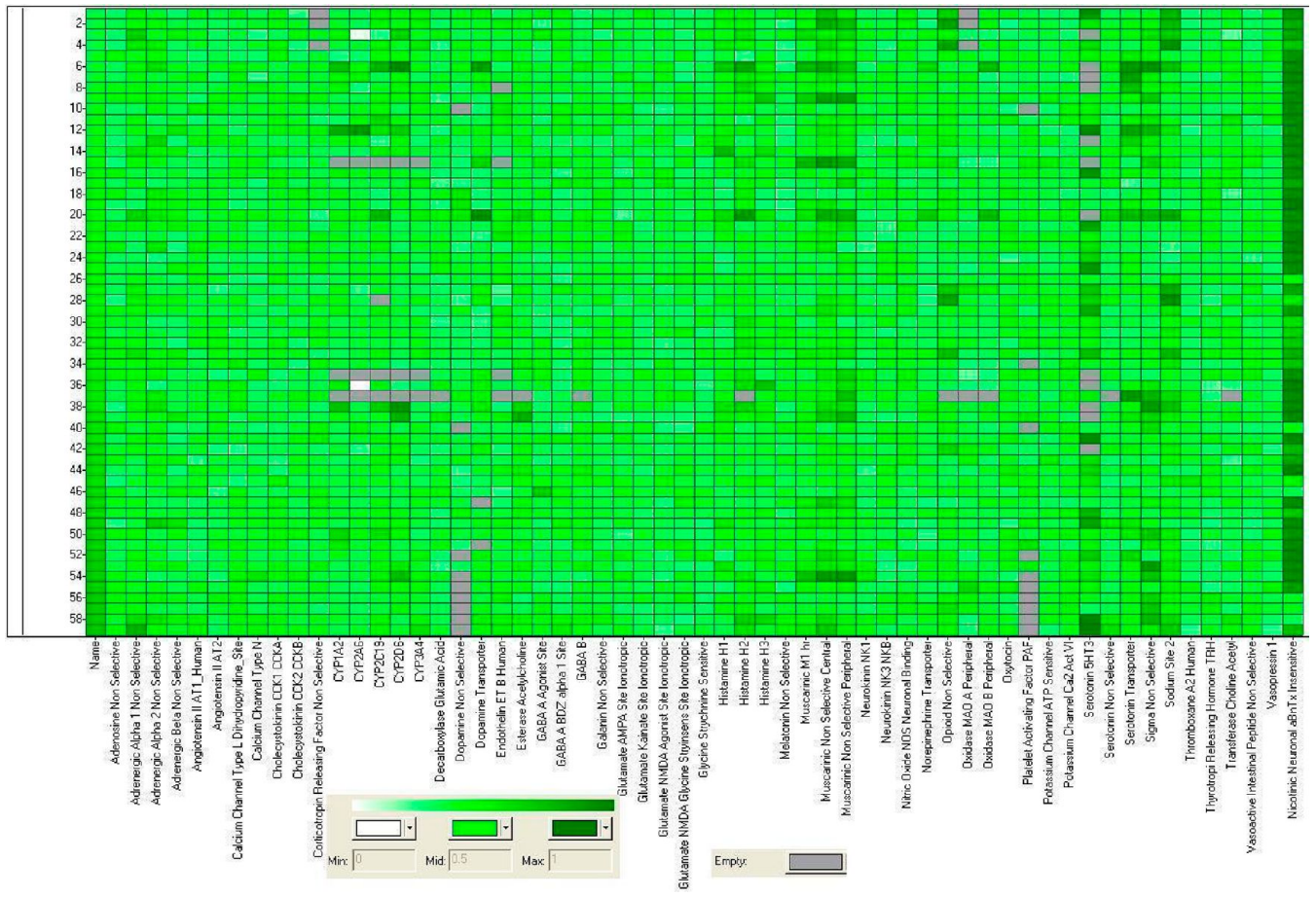


Figure 4. Heatmap representation of percent off-target inhibition values obtained at 10 uM nicotinic ligand concentration. Percent inhibition values were normalized by scaling the data to unit range prior to deriving the heatmap. The color intensity is directly proportional to the percent inhibition, with dark green indicating strong binding and light green indicating poor binding.

market for smoking cessation, exhibits a $OTHR_{10}$ = 3.3% and is thus classified as selective. ABT-894, a compound discontinued at clinical stage by Abbot, exhibits a $OTHR_{10}$ > 4% and is tagged medium promiscuous.

Protein Promiscuity. To investigate about target promiscuity, i.e., the ability of the protein to interact with a diverse set of ligands, the distribution of number of hit compounds per target assayed was determined and is shown in Figure 3. This result indicates that these nicotinic ligands interact with a diverse set of off-targets and preferentially bind to a serotonin transporter (17% of ligands), sigma receptor (17%), and serotonin 5HT3 receptor (17%), followed by a muscarinic receptor (16%) and

dx.doi.org/10.1021/ci300445e | J. Chem. Inf. Model. 2013, 53, 327−342

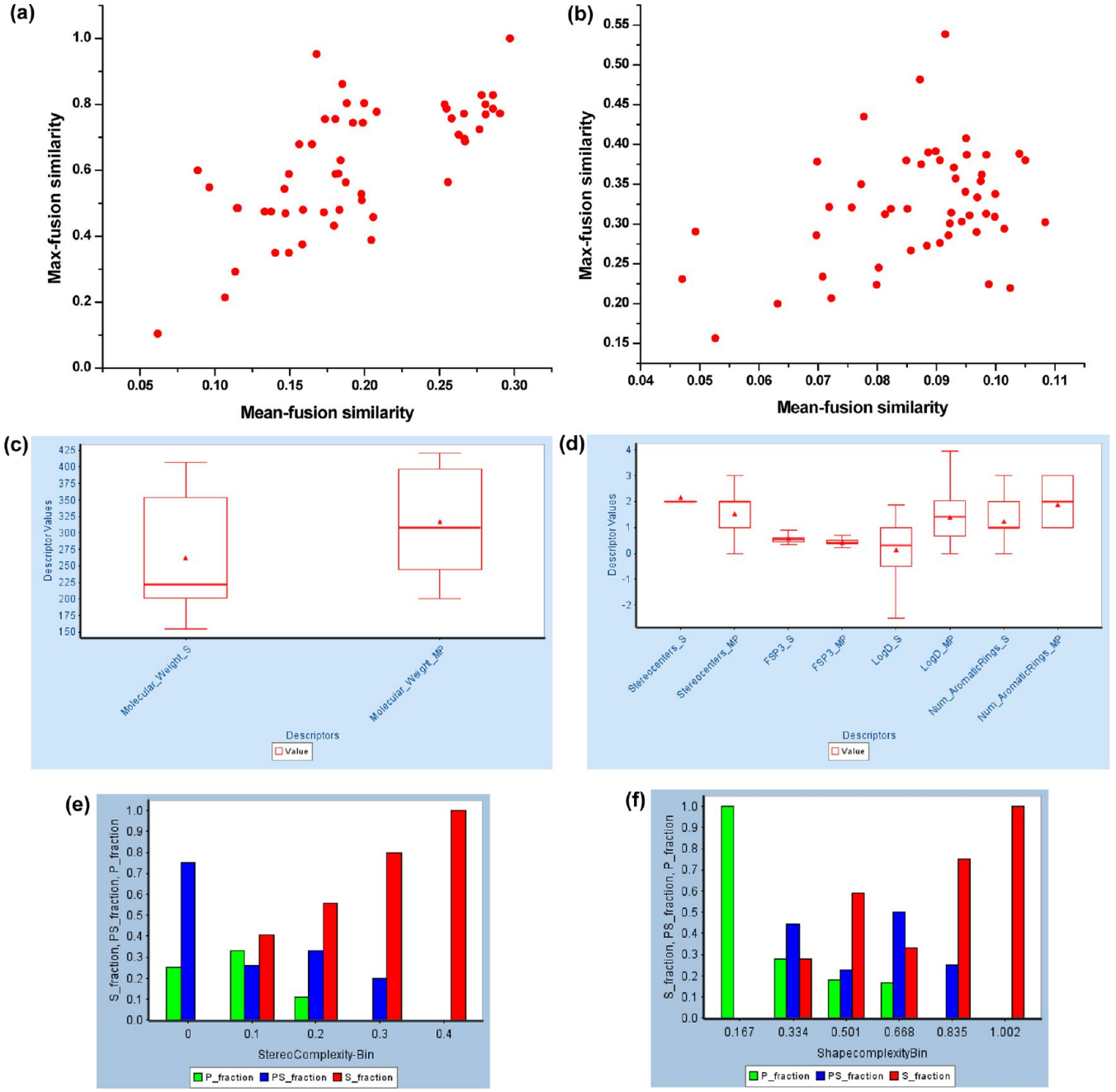


Figure 5. (a) Max-mean multi-fusion similarity map for the nicotinic data set referenced to itself. Pairwise Tanimoto similarity was calculated using functional class fingerprints FCFP-6, with stereochemistry. (b) Max-mean multi-fusion similarity map for the nicotinic data set referenced to MDDR. Pairwise Tanimoto similarity was calculated using functional class fingerprints FCFP-6 with stereochemistry. (c) Box plot of molecular weight for selective vs medium promiscuous. Data points with values within the first and third quartiles are enclosed in the red boxes. Mean and median values or the distribution are shown by a triangle and horizontal line, respectively. Upper and lower adjacent values are shown by the highest and lowest horizontal lines, respectively. (d) Box plot of number of stereo centers, Fsp3, LogD and number of aromatic rings for selective vs medium promiscuous compounds. Data points with values within the first and third quartiles are enclosed in the red boxes. Mean and median values or the distribution are shown by a triangle and horizontal line, respectively. Upper and lower adjacent values are shown by the highest and lowest horizontal lines, respectively. (e) Distribution of selective (S), partially selective (PS), and promiscuous (P) compounds as a function of stereo complexity bins. Using Clemons et al. terminology,[15] a compound is categorized as S, PS, and P when it binds 1 protein, 2−5 proteins, and at least 6 proteins, respectively. Distributions of S, PS, and P compounds are shown in red, blue, and green, respectively. (f) Distribution of selective (S), partially selective (PS), and promiscuous compounds (P) as a function of shape complexity bins. Using Clemons et al. terminology,[15] a compound is categorized as S, PS, and P when it binds 1 protein, 2−5 proteins, and at least 6 proteins, respectively.

sodium channel (15%), etc. In addition, a heatmap of percent inhibition data derived at 10 uM is shown in Figure 4.

Chemical Diversity. To probe the chemical diversity of the nicotinic compounds data set, we calculated intralibrary pairwise Tanimoto similarity using a functional class type of fingerprint (FCFP-6) with stereochemistry taken into account. Results are shown in Figure 5a as a max-mean multi-fusion similarity map.[54,55] Although some high values are observed for the max-

dx.doi.org/10.1021/ci300445e | J. Chem. Inf. Model. 2013, 53, 327−342

Table I. Summary Statistics of $OTHR_{10}$ and Classical Descriptors for the Data Set of Medium Promiscuous (MP) and Selective (S) Nicotinic Compounds

	All		MP		S	
Descriptors	Mean	SD	Mean	SD	Mean	SD
$OTHR_{10}$	4.113	4.022	8.864	4.130	1.987	1.077
LogD	0.517	1.221	1.392	0.981	0.126	1.110
Num_RotatableBonds	2.964	2.115	4.353	2.325	2.342	1.675
Fsp3	0.530	0.187	0.430	0.124	0.575	0.192
Stereo centers	1.945	0.840	1.529	0.848	2.132	0.767
SumNOCount	3.727	1.271	3.941	0.937	3.632	1.385
Num_H_Donors	0.982	0.356	0.941	0.416	1.000	0.324
Num_H_Acceptors	2.927	0.850	3.176	0.381	2.816	0.969
Molecular_PolarSurfaceArea	43.757	16.343	43.490	11.598	43.877	18.065
Num_AromaticRings	1.436	0.949	1.882	0.832	1.237	0.930
Molecular Weight	278.326	84.011	316.051	78.825	261.449	80.711

Table II. Descriptors That Reasonably Predict the off-Target Hit Rate for the Nicotinic Data Set, with Their ROC Score and Pearson and Spearman Rank Correlation Coefficients[a]

Descriptor Model	ROC Accuracy	Best Cutoff	ROC Accuracy	Best Cutoff	Pearson Correlation coefficient	Spearman Rank Correlation Coefficient
$OTHR_{10}$ threshold	4	4	5	5	NA	NA
Shadow_XLength	0.85	≤ 16.09	0.86	14.00	0.51	0.48
Shadow_Nu	0.84	≤ 2.15	0.80	2.24	0.44	0.52
Radius of Gyration	0.82	≤ 4.11	0.82	4.11	0.52	0.50
LogD(@ pH 7.4)	0.80	≤ 1.39	0.86	1.05	0.39	0.36
Kappa_3	0.76	≤ 4.93	0.76	3.97	0.41	0.34
Shadow_XY	0.75	≤ 87.97	0.83	87.97	0.27	0.31
Fsp3	0.72	≥ 0.42	0.68	0.41	−0.38	−0.33
Number of stereo centers	0.69	≥ 2	0.62	2	−0.31	−0.35

[a]The Clean force field was used to energy-minimize the molecular structure. NA: non applicable.

fusion values, we find that their corresponding mean-fusion values are much smaller. The overall mean and standard deviation derived for the pairwise similarity obtained by self-referencing the nicotinic data set were as low as 0.21 and 0.18, respectively. We also derived interlibrary max-mean multi-fusion similarity map, using the nicotinic data set and MDDR as test and reference sets, respectively. Results are shown in Figure 5b. Maps shown in Figure 5a and b indicate that some of the nicotinic compounds are somewhat scattered and removed from the bulk of the nicotinic collection, irrespective of the reference set used. These results suggest that the nicotinic data set is reasonably fairly structurally diverse.

To demonstrate that our nicotinic compounds data set has the same trends as larger databases previously reported (for review see Meanwell[9]) and to further investigate the difference in chemical structural profiles between selective compounds and medium promiscuous ones, we calculated the mean and standard deviations of classical 2D descriptors commonly used in virtual screening. Results are reported in Figure 5c and 5d and Table I.

We found that medium promiscuous compounds exhibit higher values of classical 2D descriptors related to promiscuity (molecular weight, LogD, and number of aromatic rings), than do selective compounds (Figure 5c and d and Table I). Although derived with a relatively small data set of diverse nicotinic compounds, these results are strikingly consistent with the trend reported in the literature (for review see Meanwell[9]). We also found that selective compounds exhibit higher values of 2D descriptors related to stereochemical and shape complexity (number of stereo centers and Fsp3, respectively) than do medium promiscuous compounds (Figure 5d and Table I). Again, these findings are in agreement with results obtained by Clemons et al.[15] using three much larger and more diverse data sets of different origins. To further illustrate this corroboration, we have used one of their binning schemes defined as follows: highly specific compounds (hits binding exactly one protein), partially specific compounds (hits binding 2−5 proteins) and promiscuous compounds (hits binding at least six proteins). We have calculated the corresponding size-independent parameter which they defined to estimate stereochemical and shape complexity, respectively:

$$\text{Stereo-chemical complexity} = (C_{\text{stereogenic}}/C_{\text{total}}) \quad (4)$$

where $C_{\text{stereogenic}}$ and C_{total} denote the number of C atoms that are stereogenic and the total number of C atoms, respectively. To calculate the parameter $C_{\text{stereogenic}}$, we have used eq 2.

$$\text{Shape complexity} = (C_{sp\,3}/[C_{sp\,3} + C_{sp\,2}]) \quad (5)$$

where C_{sp3} and C_{sp2} denote the number of sp^3- and sp^2-hybridized carbons atoms, respectively.

Our results shown in Figure 5e indeed support Clemons et al. findings:[15] the more stereochemically complex the compound is (i.e., $C_{\text{stereogenic}}/C_{\text{total}} > 0.25$), the more selective binder it is. Similar results were also found for shape complexity, (Figure 5f) because a depletion in promiscuous compounds and concomitant enrichment in specific compounds are observed with the increase of shape complexity. In spite of the fact that our study is restricted to basic amine-containing compounds within a single target class (nAChRs), the comparison of the most common 2D descriptors derived from this data set show reasonably good concordance with compounds analyzed from larger and

dx.doi.org/10.1021/ci300445e | J. Chem. Inf. Model. 2013, 53, 327−342

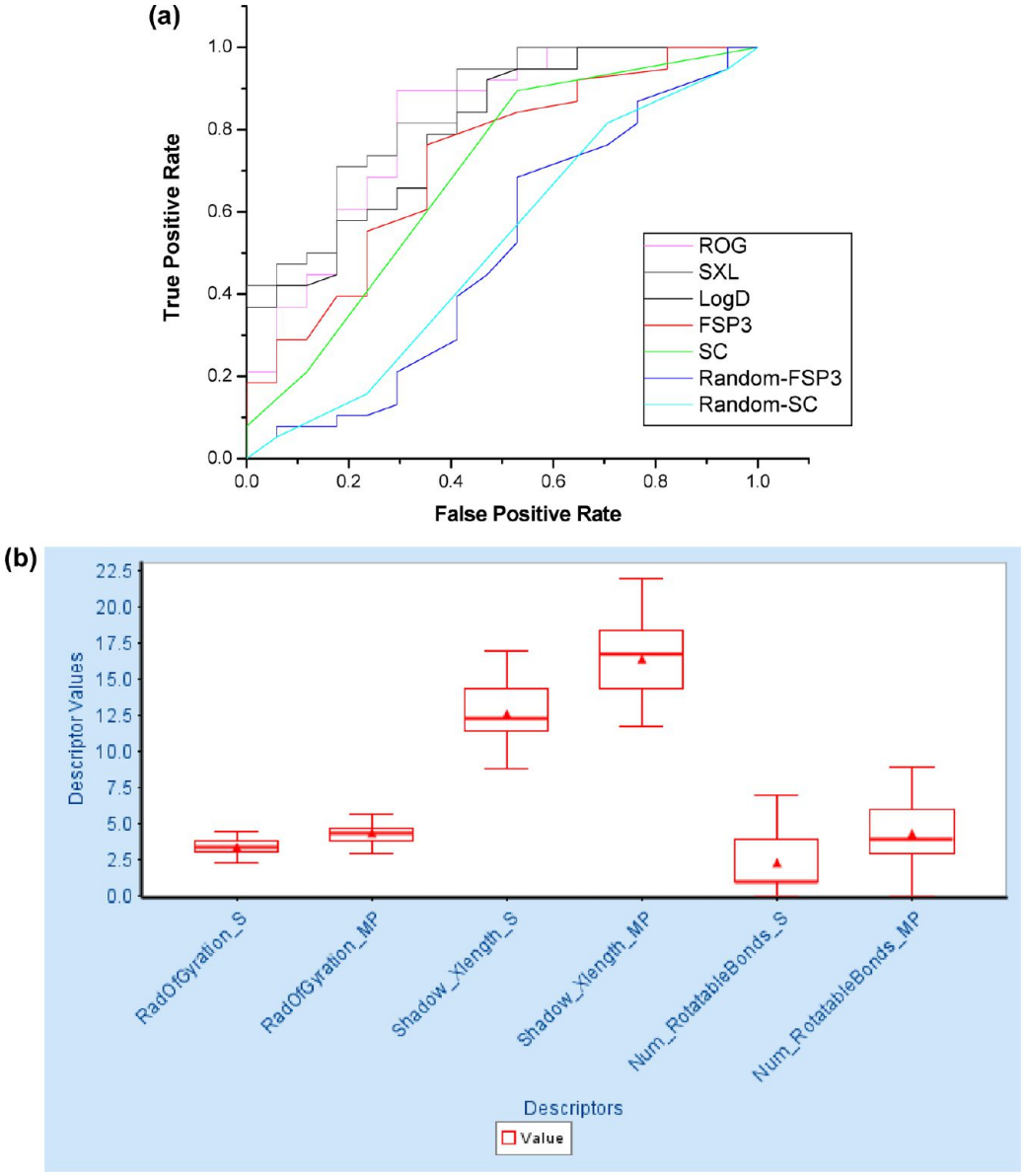


Figure 6. (a) ROC curve for the ability of models solely based on ROG, SXL, LogD, Fsp3, and number of stereo centers (SC) to discriminate between medium promiscuous and selective compounds. A compound is said to be medium promiscuous if $OTHR_{10} > 4\%$ and selective if $OTHR_{10} \leq 4\%$. For comparison, the ROC for random models obtained with SC and Fsp3 are also shown in light and dark blue, respectively. (b) Box plot of ROG, SXL, and NROT for selective vs medium promiscuous compounds. Data points with values within the first and third quartiles are enclosed in the red boxes. Mean and median values or the distribution are shown by a triangle and horizontal line, respectively. Upper and lower adjacent values are shown by the highest and lowest horizontal lines, respectively.

considerably more diverse data sets. Thus, in addition to the diversity in structure (illustrated in Figure 5a and b) and the diversity in pharmacological profile (as shown in Figure 4 and Table S1, SI), these findings further support the notion that our data set is a good representative of a somewhat diverse biochemical space amenable to a detailed study on molecular descriptors.

Molecular Descriptors Related to Off-Target Hit Rate at 10 uM. To further investigate any potential relationship between $OTHR_{10}$ and molecular descriptors, we calculated the Pearson and Spearman correlation matrices of a diverse set of 132 molecular descriptors, listed in the Supporting Information and described in the Materials and Methods section. We then calculated the receiver operating characteristic (ROC) score for the descriptors that exhibited highest Pearson (and Spearman) correlations with $OTHR_{10}$. In this case, a ROC plot can help one understand the tradeoff between the ability of the descriptors to identify true positives, i.e., selective compounds and its ability to avoid false positives, i.e., medium promiscuous compounds, when the descriptor of interest is used as a learned model. The closer the ROC score is to 1.0, the better the descriptor is at distinguishing selective from medium promiscuous samples. By contrast, the correlation coefficient provides a measure of the interrelatedness between the descriptor of interest and $OTHR_{10}$. Table II shows molecular descriptors that highly correlate with off-target hit rate for the nicotinic ligands used in our study, along with their accuracy in discriminating between selective (Off-target hit rate (OTHR) $\leq$ 4%) and medium promiscuous compounds (OTHR > 4%). In particular, shadow-Xlength (SXL) scores the highest, followed by Shadow-nu, radius of

dx.doi.org/10.1021/ci300445e | J. Chem. Inf. Model. 2013, 53, 327−342

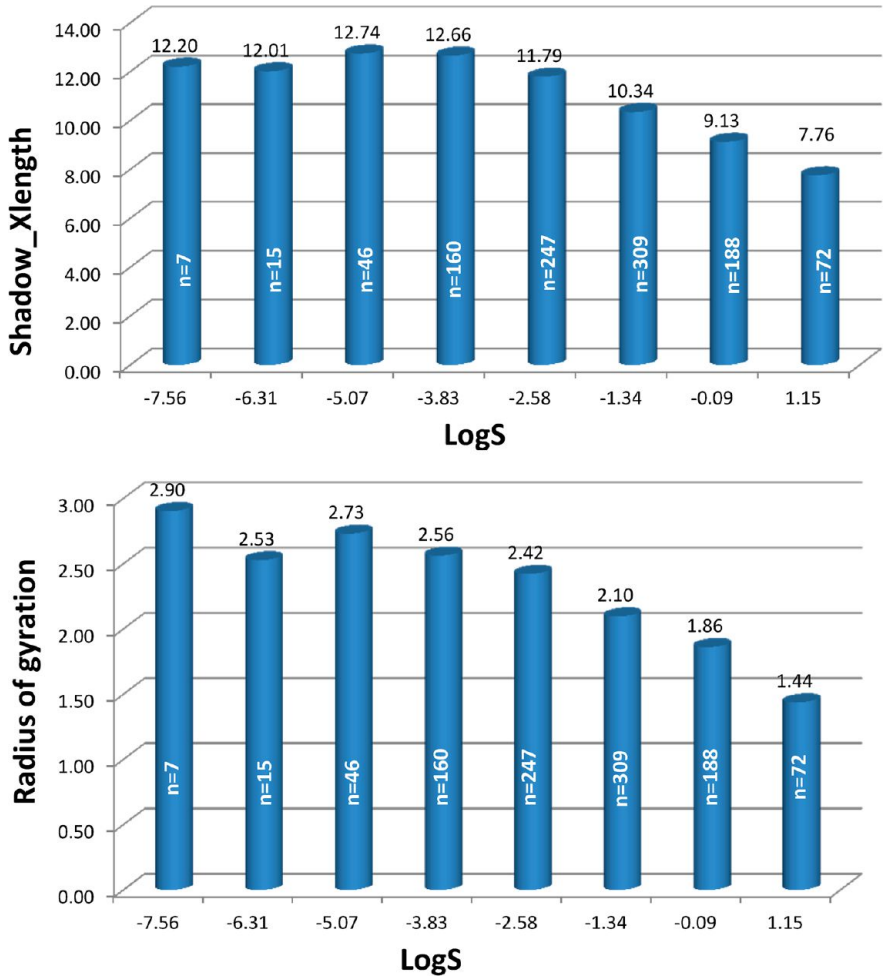


Figure 7. (a) Variations of calculated SXL as a function of experimentally observed LogS. The average values of SXL per LogS bin of size 1.24 are shown on top of the corresponding bars. The number of compounds in each bin is shown within the bar. (b) Variations of calculated ROG as a function of experimentally observed LogS. The average values of ROG per LogS bin of size 1.24 are shown on top of the corresponding bars. The number of compounds in each bin is shown within the bar.

Table III. Size of Compounds Data Set Retrieved from MDDR and Used Throughout This Study, before and after Applying Lipinski's RO5 Filters

Drug Development Stage	Data Set Size Prior to Filtering	Data Set Size after Filtering
Preclinical	11879	7231
Phase I	859	547
Phase II	1188	725
Phase III	307	200
Approved Drugs	1415	994
Withdrawn Drugs	49	33
Discontinued Drugs	219	135

gyration (ROG), and LogD, with a receiver operating characteristic (ROC) curve accuracy of 85%, 84%, 82%, and 80%, respectively. The lowest ROC scores were obtained with Kappa-3 (76%), Fsp3 (72%), and number of stereo centers (69%). For each of these descriptors, the best split that minimizes the number of false positives and false negatives was calculated using a pipeline pilot protocol. The cutoffs thus obtained are also shown in Table II. The ROC scores and cutoffs do not significantly vary when the MMF94 force field[56] is used to minimize the energy of the structure instead of the Clean force field.[46] When the OTHR is changed to 5,[15,53] most of the ROC scores and best cutoff do not vary that much, as shown in Table II. Exceptions are found, however, with LogD, Shadow_XY, and number of stereocenters. ROC curves obtained when the Clean force field was used are shown in Figure.6a. Results indicate that models solely using Fsp3 and number of stereo centers are closer to random than do their SXL, ROG, and LogD counterparts. Furthermore, we found that MP compounds have higher values of SXL, ROG, and number of rotatable bonds than their selective counterparts, as shown in Figure 6b. The difference appears to be more pronounced for SXL, followed by the number of rotatable bonds, and then ROG.

SXL and ROG Correlate with Solubilty. In an attempt to verify whether or not the molecular descriptors SXL and ROG are related to experimentally observed physical properties known to be important to success in drug discovery, we compared their trend with the corresponding variation in experimentally observed solubility data described in the Methods section. Results shown in Figure 7a and b demonstrate that SXL and ROG are both roughly inversely proportional to LogS. By contrast, Fsp3 is directly proportional to LogS, as previously shown by Lovering et al.[1] using the same data set. We conclude that SXL and ROG are related to experimentally observed solubility, which drives clinical success, in an opposite manner as compared to the trend observed for Fsp3 variations.

dx.doi.org/10.1021/ci300445e | J. Chem. Inf. Model. 2013, 53, 327−342

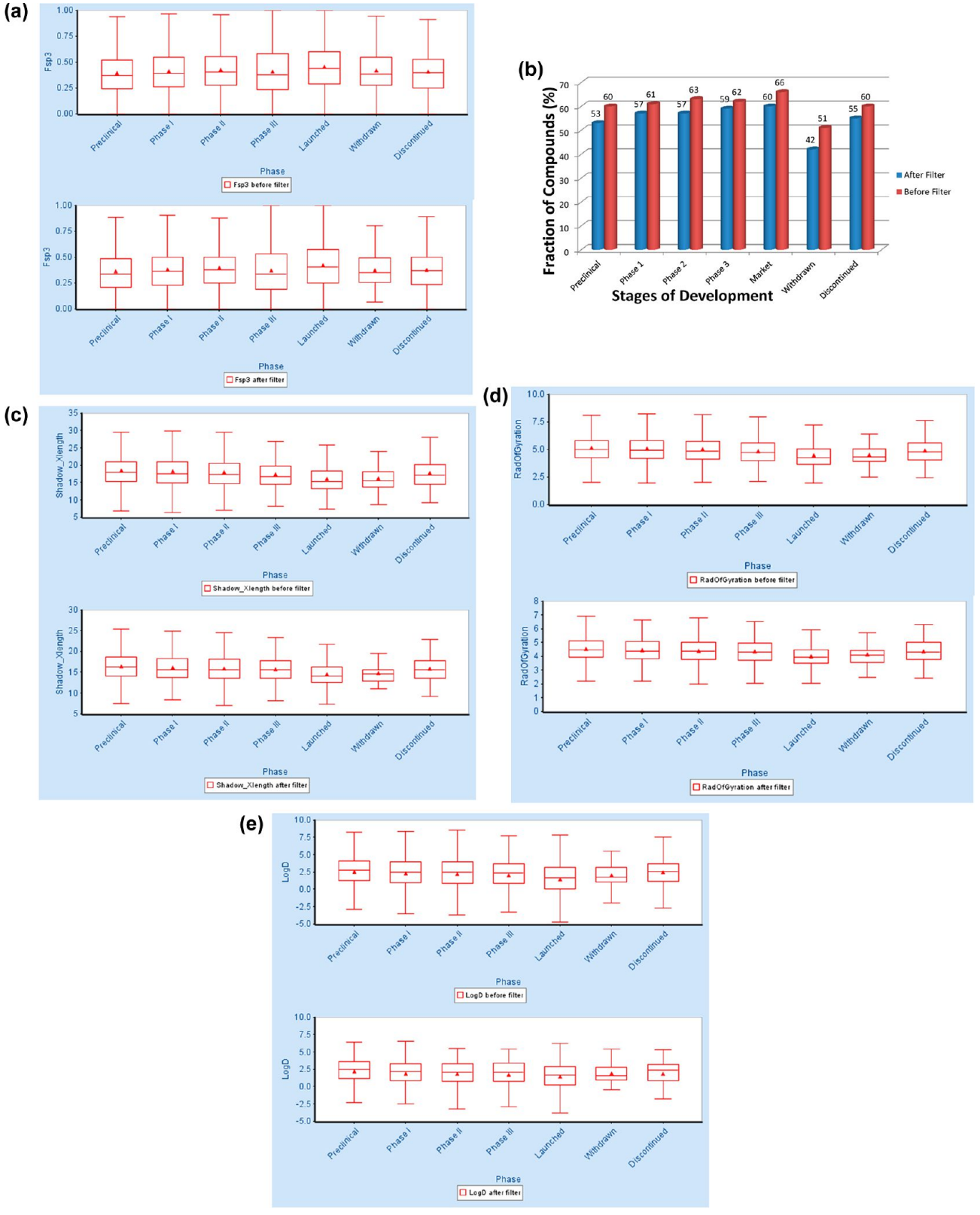


Figure 8. (a) Box plot of Fsp3 as a function of the stage of drug development. Data points with values within the first and third quartiles are enclosed in the red boxes. Mean and median values or the distribution are shown by a triangle and horizontal line, respectively. Upper and lower adjacent values are shown by the highest and lowest horizontal lines, respectively. (b) Fraction of compounds with one or more stereo center increase as MDDR drugs transition from preclinical to market. The average values per phase are shown on top of the corresponding bars. (c) Box plot of Shadow-Xlength as a function of the stage of drug development. Data points with values within the first and third quartiles are enclosed in the red boxes. Mean and median values or the distribution are shown by a triangle and horizontal line, respectively. Upper and lower adjacent values are shown by the highest and lowest horizontal lines, respectively. (d) Box plot of radius of gyration as a function of the stage of drug development. Data points with values within the first and third quartiles are enclosed in the red boxes. Mean and median values or the distribution are shown by a triangle and horizontal line, respectively. Upper

dx.doi.org/10.1021/ci300445e | J. Chem. Inf. Model. 2013, 53, 327−342

Figure 8. continued

and lower adjacent values are shown by the highest and lowest horizontal lines, respectively. (e) Box plot of LogD as a function of the stage of drug development. Data points with values within the first and third quartiles are enclosed in the red boxes. Mean and median values or the distribution are shown by a triangle and horizontal line, respectively. Upper and lower adjacent values are shown by the highest and lowest horizontal lines, respectively.

Table IV. Statistical significance expressed in terms of adjusted p-values for trends in SXL and ROG as compared to Fsp3 and LogD[a]

	q-value			
Drug Development Stages	Radius of Gyration	Shadow-Xlength	FSP3	Log D
1 vs. 2	0.625	0.369	0.448	0.715
1 vs. 3	0.098	**0.026**	0.218	0.712
1 vs. Preclinical	0.051	0.094	0.053	**0.023**
1 vs. Launched	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
1 vs. withdrawn	**0.046**	**0.026**	0.701	0.625
2 vs. 3	0.275	0.098	**0.021**	0.678
2 vs. Preclinical	**< 0.001**	**< 0.001**	**< 0.001**	**0.013**
2 vs. Launched	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
2 vs. withdrawn	0.054	**0.047**	0.701	0.625
3 vs. preclinical	**< 0.001**	**< 0.001**	**0.043**	0.051
3 vs. launched	**< 0.001**	**< 0.001**	**< 0.001**	**0.004**
3 vs. withdrawn	0.239	0.233	0.544	0.712
Preclinical vs. Launched	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
Preclinical vs. withdrawn	**0.009**	**0.004**	0.565	0.302
Launched vs. withdrawn	0.109	0.638	0.411	0.124
Number of statistically significant results	8	10	7	6

[a]. A KS-test was initially used. Correction for multiple hypothesis testing was carried out by calculating q-values,[52] the positive false discovery rate analogues of the *p*-values. Results are said to be significant when q-value <0.05. Significant results are shaded in green.

Variations of SXL and ROG for Compounds at Various Stages of Clinical Development. In our subsequent studies, we focused on comparing Fsp3 and LogD, two well-known 2D physicochemical properties related to clinical success, vs SXL and ROG. It should be noted that both 3D descriptors showed the highest correlation with $OTHR_{10}$ and ROC score values among the best, as shown in Table II. Lovering et al.[1] have shown that Fsp3 and the presence of chiral centers correlate both with solubility and clinical success as compounds transition from discovery to the market. Using the same solubility data set, we have shown that ROG and SXL also correlate with experimentally observed physical solubility. In order to check whether or not SXL and ROG can also correlate with success as compounds move from preclinical to subsequent clinical stages, we have investigated compounds from the MDDR database. As a first control, withdrawn and discontinued compounds were separately studied. As a second control, the trends in Fsp3 and number of stereo centers were also evaluated because the original study conducted by Lovering et al.[1] used a different and larger data set, the GVK Biosciences database (www.gvkbio.com). Lipinski's rule of five for predicting drug permeability[4] and Veber et al.'s [13] rule of NROT ≤10, which is associated with good oral exposure, were used to filter compounds in order to make sure that nondrug like molecules were not overly influencing the results. Specifically, filtering was carried out by removing any compounds that violated any one of the rule of five (HBD ≤ 5, HBA ≤ 10, MW ≤ 500, and LogP ≤ 5) or NROT ≤ 10. The total number of compounds used in each stage of development prior to filtering and after filtering is shown in Table III. Results suggest that about 30% of approved drugs do not satisfy Lipinski's and Veber's rules combined. Figure 8 a and b shows that mean Fsp3 and fraction of compounds with one or more stereo centers tend to increase as compounds retrieved from the MDDR database move forward in subsequent stages of development until they reach the market, as also observed by Lovering et al.[1] On the whole, withdrawn and discontinued drugs do not match the criteria for approved drugs, and the trend observed is the same, regardless of filtering. In particular, prefiltered withdrawn drugs exhibit the lowest value of fraction of compounds with one or more stereo centers (42%), as compared to the finding that about 60% of prefiltered approved drugs contain one or more stereo centers. Likewise, Figure 8 c–e shows that mean SXL, ROG, and LogD decrease as MDDR compounds transition from preclinical to market, respectively. Withdrawn and discontinued compounds do not follow the same trend with approved drugs. The overall results are also independent of filtering on the basis of the above-mentioned drug-likeness rules. Of the four physicochemical properties investigated, LogD exhibits more pronounced variations as compounds move from preclinical stage to the market. Likewise, the difference in average values of LogD between launched and withdrawn or discontinued compounds appears to be greater, as compared to the corresponding trend in Fsp3, SXL, and ROG.

To establish whether or not the trends observed between various drug development stages are significantly different, we have used the KS-test, a statistical method often used for comparing distributions, and subsequently adjusted the p-values for multiple comparisons. Results summarized in Table IV indicate that out of the 15 comparisons made, SXL, ROG, Fsp3, and LogD exhibit statistically significant results 10, 8, 7, and 6 times, respectively. Thus, on the whole, 3D descriptors outperform their 2D counterparts. The differences in values of ROG, SXL, Fsp3, and LogD between phase 1 vs launched compounds, phase 2 vs preclinical, phase 2 vs launched, and preclinical vs launched are all statistically significant. In addition, the differences in Fsp3 and LogD fail to be statistical significant

dx.doi.org/10.1021/ci300445e | J. Chem. Inf. Model. 2013, 53, 327–342

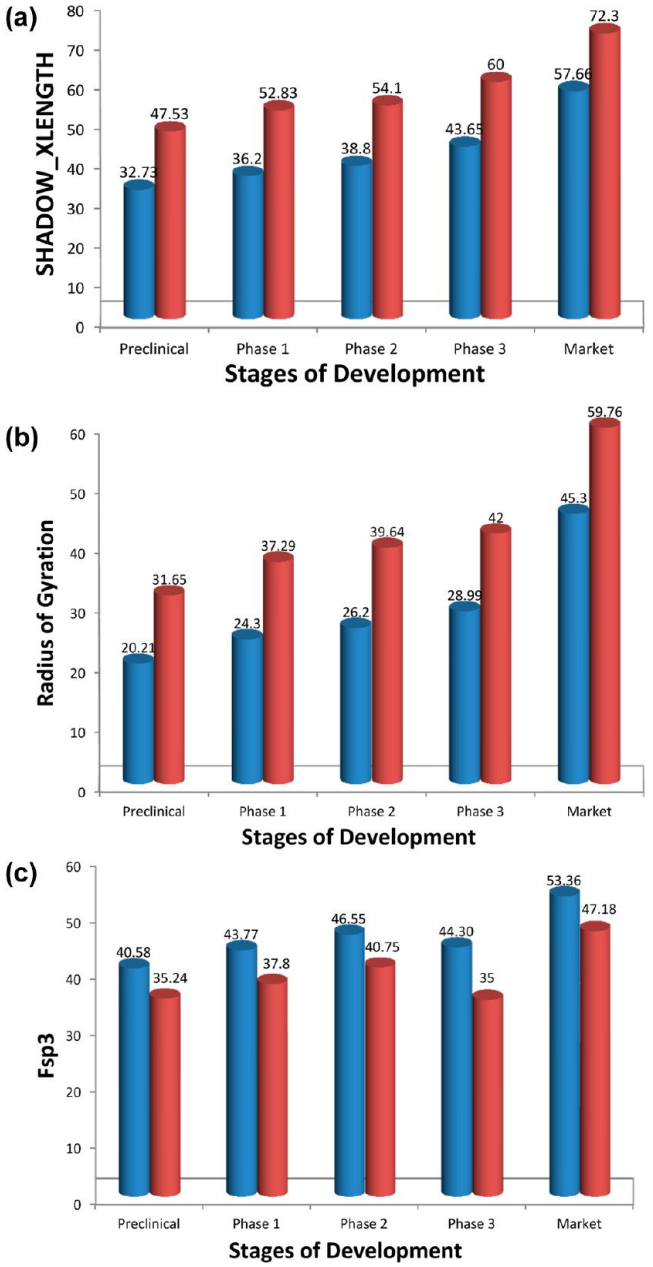


Figure 9. (a) Fraction (%) of MDDR compounds that meet the SXL cutoff (≤16.09). (b) Fraction (%) of MDDR compounds that meet the ROG cutoff (ROG ≤ 4.11). (c) Fraction (%) of MDDR compounds that meet the Fsp3 cutoff (Fsp3 ≥ 0.42).

between phase 1 vs withdrawn, phase 2 vs withdrawn, and preclinical phase vs withdrawn, while SXL succeed to be statistically different. However, Fsp3 outranks ROG, SXL, and LogD when it comes to comparing compounds in phase 2 vs those that have reached phase 3. Summary statistics and histograms for ROG, SXL, Fsp3, and LogD are shown in the Supporting Information.

Because more statistically significant results were obtained with SXL (67% of times), followed by ROG (53% of times), and slightly with Fsp3 (47% of times), whereas LogD appeared to score the poorest (40% of times), we decided to only focus on ROG, SXL, and Fsp3 in our subsequent studies. To investigate whether cutoffs derived for SXL, ROG, and Fsp3 with respect to off-target activity of nicotinic compounds can also be applied to other therapeutic areas, we computed the fraction of compounds that meet these criteria as compounds move from preclinical stage to the market. Results shown in Figure 9a–c suggest that as compounds migrate from early to later development stages, the fraction of nonfiltered compounds that satisfy the above-mentioned criteria increases from about 32 to 57%, 20 to 45%, and 40 to 53% with respect to SXL, ROG, and Fsp3, respectively. This corresponds to an overall difference of 25%, 25%, and 13%, for SXL, ROG, and Fsp3, respectively. Once filters are applied, the proportion of compounds satisfying the criteria changes from about 47 to 72%, 31 to 59%, and 35 to 47%, with respect to SXL, ROG, and Fsp3, respectively. This corresponds to an overall difference of 25%, 28%, and 12%, for SXL, ROG, and Fsp3, respectively. These results indicate that these criteria can indeed be useful for predicting clinical success and that 3D descriptors still appear to outperform their 2D counterpart.

Because we have found that the criteria based on these three descriptors as derived from nicotinic chemical landscape can still be extended to other chemical space and associated therapeutic areas, we then set out to focus on launched drugs in order to determine whether or not most of them would satisfy either SXL, ROG, or Fsp3 criteria, or their combination thereof.

Results shown in Table V indicate that about 84% of launched drugs satisfy either the criteria required for SXL (A ≡ SXL ≤ 16.09) or the criteria required for Fsp3 (C ≡ Fsp3 ≥ 0.42), as shown in the cell at the intersection of row #9 and the sixth column from the left. Adding criteria B to (AUC) gives (AUBUC) that results in the same proportion (84%), as shown on row #11. Interestingly, about 74% of launched drugs that fail the rule of five do indeed satisfy the SXL criteria (A) or the Fsp3 requirement (C). This result suggests that compounds that fail RO5 can further be checked for drug-likeness using SXL or Fsp3 criteria, thereby enhancing the accuracy of the screening process by reducing the number of potential false negatives. In addition, it is noteworthy to mention that Table V also shows that out of the 1415 launched drugs at the time of this study, about 358 of them fail Lipinski rule of five (25%). A total of 75% of these 358 drugs satisfy either SXL, Fsp3, or ROG, as shown in the rightmost cell of row #11.

To further investigate whether the above-mentioned criteria are indeed applicable to diverse therapeutic areas, we have partitioned the launched drugs into two categories, CNS vs non-CNS, and determined the fraction of launched drugs that satisfy either the SXL or Fsp3 criteria. Results shown in Figure 10 demonstrate that on the whole, this combination of criteria is equally met by both categories, with the CNS category exhibiting a slight edge. The trend is the same prior to and after filtering.

DISCUSSION AND CONCLUSION

Using a data set of nicotinic compounds, we have shown that stereochemical and shape complexity correlate with off-target activity. This finding is in agreement with the results obtained by Clemons et al.[15] in their studies using three much larger and more diverse library collections of different origins. We have also found that LogD directly correlates with promiscuity, in agreement with the finding reported by Hopkins et al.[57] observed between promiscuity and lipophilicity (ClogP) in a data set of 1098 diverse drugs compounds. In addition to these well-known 2D descriptors, we have shown that shape-based 3D descriptors such as shadow indices and radius of gyration strongly correlate with off-target activity, thereby affecting clinical success, as compounds move from preclinical stage to the market. In predicting clinical success, these shape-based 3D

dx.doi.org/10.1021/ci300445e | J. Chem. Inf. Model. 2013, 53, 327–342

Table V. Fractions of Launched Drugs Satisfying Either SXL, ROG, and/or Fsp3 Criteria[a]

Row number	Number of samples meeting criteria	Before filtering: N = 1415	%	After filtering: N=994	%	Not meeting RO5: N=358	%
1	A	816	58	719	72	92	26
2	B	641	45	594	60	46	13
3	C	755	53	439	44	240	67
4	A ∩ B	633	45	589	59	43	12
5	A ∪ B	824	58	724	73	95	27
6	B ∩ C	324	23	289	29	34	9.5
7	B ∪ C	1072	76	774	78	252	70
8	A ∩ C	424	30	354	36	66	18
9	A ∪ C	1147	81	834	84	266	74
10	A ∩ B ∩ C	319	23	286	29	32	9
11	A ∪ B ∪ C	1150	81	836	84	267	75

[a]A, B, and C designate the number of samples clearing the cutoff for SXL, ROG, and Fsp3, respectively. The Boolean symbols "∩" and "∪" represent "and" and "or", respectively. Best results are highlighted in green and second best in yellow.

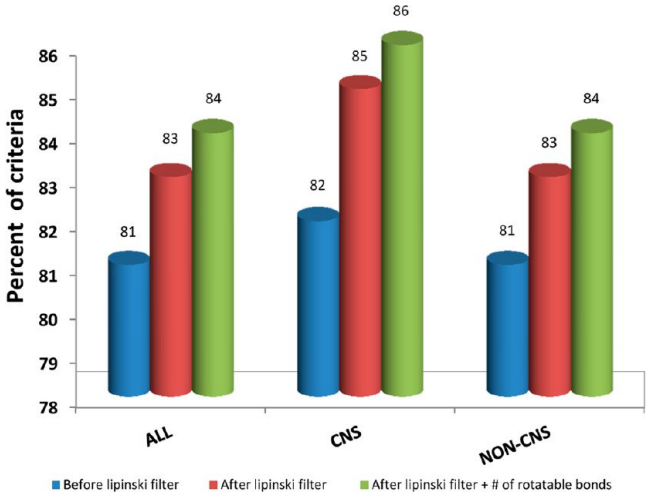


Figure 10. Fraction of launched drugs satisfying either SXL or Fsp3 criteria for CNS vs non-CNS drugs.

descriptors appear to outperform classical 2D descriptors such as Fsp3, the presence of stereo centers, and LogD.

We have also found that these shape-based 3D descriptors correlate well with experimentally observed solubility, a physicochemical property well-known to associate with clinical success. In addition, medium promiscuous compounds appear to exhibit higher values of these 3D descriptors than do their selective counterparts. Because sphere-like compounds exhibit lower values of radius of gyration and Shadow_Xlength than do their elongated counterparts, it appears that sphere-like compounds will tend to be more selective. This finding is in agreement with observations made by Clemons et al.[16] when they used principal moments of inertia as shape-based 3D descriptor to compare three diverse screening collections of known protein binding profile. One should note that shadow indices are shape-based descriptors calculated by projecting the molecular surface on three mutually perpendicular planes, XY, YZ, and XZ, as described by Rohrbaugh and Jurs.[58] Thus, not only do they depend on the conformation but also on the orientation of the molecule because the latter is first rotated to align the principal moment of inertia with the X, Y, and Z axis prior to calculating them. The radius of gyration is the mass weighted root-mean-square average distance of all atoms in the molecular system from their center of mass

$$\mathrm{ROG} = [\sum m(x^2 + y^2 + z^2)/\sum m]^{1/2} \quad (4)$$

where m is the mass of an atom and x, y, z are the atomic coordinates relative to the center of mass.

Given that both ROG and SXL are somewhat related to the molecular length, one would expect that in addition to describing the molecular shape, they could also correlate with the number of rotatable bonds, which is one of the well-known properties that characterize drug-likeness.[13] Indeed, Figure 6b indicates that the change in the trend for NROT of selective compounds vs

dx.doi.org/10.1021/ci300445e | J. Chem. Inf. Model. 2013, 53, 327−342

medium promiscuous in the nicotinic data set used herein goes in the same direction with SXL and ROG. As a matter of fact, we found that the Pearson correlation between ROG, SXL, and the number of rotatable bonds was equal to 0.83 and 0.81, respectively.

We have derived criteria for compound selectivity or lack of off-target promiscuity by using a minimal data set consisting of nicotinic compounds and have shown that these criteria can be applied to other therapeutic areas. Using the MDDR database, we have observed that SXL and ROG appear to be better at predicting entry into the clinic, as they successfully discriminate withdrawn compounds from those that are in preclinical stage and phase 1, respectively (Table IV). By contrast, Fsp3 and LogD fail to achieve such discrimination. Also, we have found that about 84% of launched drugs satisfy either Shadow_Xlength or Fsp3 criteria, i.e., SXL $\leq$ 16.09 or Fsp3 $\geq$ 0.42. Withdrawn and discontinued compounds do not meet the criteria of descriptors related to clinical success, as compared to launched drugs. Furthermore, we have found that about 75% of launched drugs that fail Lipinski's RO5 do indeed meet either the SXL or Fsp3 criteria. This is of particular interest because many contemporary drug discovery programs are seeking orally bioavailable compounds in non-Lipinski space.

Moreover, our finding that the accuracy of filtering for off-target activity can be improved by combining both 2D and 3D molecular descriptors is consistent with Yera et al. results obtained from their studies aimed at assessing predictions of on-target vs off-target activity.[59] For predicting the latter, they found that 2D descriptors resulted in a 37% true positive rate, whereas 3D descriptors yielded 61%, while the combination of 2D and 3D descriptors yielded 68%. Similarly, Nettles et al.[60] reported that they obtained a better performance when they used a systematic combination of both 2D and 3D descriptors to query annotated chemical databases. Moreover, our finding that shape-based descriptors are related to off-target promiscuity is consistent with the use of spherical harmonic representations in a shape-based approach to identify and characterize both promiscuous ligands and targets, which has recently been pioneered by Perez-Nueno et al.[61] On the basis of our findings described herein, we suggest that the following molecular descriptors be incorporated as filters for clinical success in virtual screening and compounds selection cascade: Shadow_Xlength $\leq$ 16.09, Fsp3 $\geq$ 0.42, ROG $\leq$ 4.11 and LogD at pH 7.4 $\leq$ 1.39. One should recall that compounds lacking aromatic rings or having a very limited number of aromatic rings will exhibit higher values of Fsp3 and lower LogD values as compared to their counterparts containing more aromatic rings.

We have shown that shape-based descriptors ROG and SXL correlate with off-target activity and solubility, which in turn influence clinical success. This finding led us to search for a potential correlation between these shape-based descriptors and clinical success. As stated in the Introduction section, side-effects, bioavailability, efficacy, and formulation are the main causes for drug attrition. Our studies have mainly focused on off-target activity, which is related to side effects. We have also investigated correlations with solubility, a property that can influence drug formulation. Because we have not taken into account the specific reasons why a particular drug was discontinued or withdrawn, our results do not lend support to a causative relationship between shape-based descriptors and drug attrition rate. Because of the complex nature of factors that impact clinical safety profiles, an extreme caution is required when one wants to use physico-chemical predictors as guidelines for clinical success.

It has been shown that well-tolerated protein kinase drugs bind to multiple kinases.[62−64] Examples include Gleevec (Imatinib) and Sutent.(SU11248). Likewise, some anti-psychotic compounds exhibit binding promiscuity for multiple G-protein-coupled receptors (GPCRs). Examples include Clozapine.[57] It has been experimentally demonstrated that better synergistic effects are obtained when a compound acts on more than one kinase as compared to the additive effect observed when each kinase is targeted individually.[65] Thus, promiscuity may also have a beneficial effect, provided that hitting additional targets does not cause side effects. Besides, hitting more than one target of interest may offer an advantage of providing a hint for compound repositioning. In the case of nicotinic ligands studied herein, results shown in Figure 3 indicate that the most likely additional targets to investigate for polypharmacology and repositioning may be the following targets that were most frequently hit: serotonin transporter, sigma receptors, serotonin 5HT3 receptor, and muscarinic receptor.

■ ASSOCIATED CONTENT

Supporting Information

Details on screening assays and targets, experimental procedure and details on binding affinity to nAChRs, Pearson and Spearman correlation matrices of a diverse set of 132 molecular descriptors, and summary statistics and histograms for ROG, SXL, Fsp3, and LogD. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 336-768-5532. E-mail: dckombo777@mail.com.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Phil Hammnond and Dr. Daniel Yohannes for helpful discussions. We are also grateful to Melanie Kiser and Kathy Dail for providing us with in vitro off-target binding data for nicotinic compounds. In addition, we thank JCIM reviewers for helpful discussions, suggestions, and comments.

■ ABBREVIATIONS

$OTHR_{10}$,Off-target hit rate at 10 μM ligand concentration; ROG,radius of gyration; SXL,shadow-Xlength; MP,medium promiscuous; S,selective; Fsp3,fraction of carbon Sp3; ROC,receiver operating characteristic; SD,standard deviation; MDDR,MDL drug data report; CNS,central nervous system; RO5,rule of five; NROT,number of rotatable bonds; HBD,number of hydrogen-bond donors; HBA,number of hydrogen-bond acceptors; MW,molecular weight; Sc,number of stereo centers; CNS,central nervous system

■ REFERENCES

(1) Lovering, F.; Bikker, J.; Humblet, C. Escape from flatland: Increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **2009**, *52*, 6752−6756.

(2) Mc Govern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K. A common mechansim underlying promiscuous inhibitors from virtual and high-throughut screening. *J. Med.Chem.* **2002**, *45*, 1712−1722.

(3) Mc Govern, S. L.; Shoichet, B. K. Kinase inhibitors: Not just for kinases anymore. *J. Med. Chem.* **2003**, *46*, 1478−1483.

(4) Feng, B. Y.; Shelat, A.; Doman, T. N.; Guy, R. K.; Shoichet, B. K. High-throughput assays for promiscuous inhibitors. *Nat. Chem. Biol.* **2005**, *1*, 146−148.
(5) Feng, B. Y.; Shoichet, B. K. Synergy and antagonism of promiscuous inhibition in multiple-compound librairies. *J. Med. Chem.* **2006**, *49*, 2151−2154.
(6) Metz, J. T.; Huth, J. R.; Hajduk, P. J. Enhancement of chemical rules for predicting compound reactivity towards protein thiol groups. *J. Comput.-Aided. Mol. Des.* **2007**, *21*, 139−144.
(7) Huth, J. R.; Song, D.; Mendoza, R. R.; Black-Schaefer, C. L.; Mack, J. C.; Dorwin, S. A.; Ladror, U. S.; Severin, J. M.; Walter, K. A.; Bartley, D. M.; Hajduk, P. Toxicological evaluation of thiol reactive compounds identified using a La assay to detect reactive molecules by nuclear magnetic resonance. *Chem. Res. Toxicol.* **2007**, *20*, 1752−1759.
(8) Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (Pains) from screening librairies and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53*, 2719−2740.
(9) Meanwell, N. A. Improving drug candidates by design: A focus on physicochemical properties as a means of improving compound disposition and safety. *Chem. Res. Toxicol.* **2011**, *24*, 1420−1456.
(10) Ehrlich, P. The theory and practice of chemotherapy. *Folia Serol.* **1911**, *7*, 697−714.
(11) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.
(12) Bhal, S. K.; Kassam, K.; Peirson, I. G.; Pearl, G. M. The rule of five revisited: Applying Log D in place of Log P in drug-likeness filters. *Mol. Pharm.* **2007**, *4*, 556−560.
(13) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615−2623.
(14) Ritchie, T. J.; Macdonald, S. J. F. The impact of aromatic rings count on compound developability—Are too many aromatic rings a liability in drug design? *Drug Discovery Today* **2009**, *14*, 1011−1020.
(15) Clemons, P. A.; Bodycombe, N. E.; Carrinski, H. A.; Wilson, J. A.; Shamji, A. F.; Wagner, B. K.; Koeler, A. N.; Schreiber, S. L. Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 18787−18792.
(16) Clemons, P. A.; Wilson, J. A.; Dančik, V.; Muller, S.; Carrinski, H. A.; Wagner, B. K.; Koehler, A. N.; Schreiber, S. L. Quantifying structure and performance diverisity for sets of small molecules comprising small-molecule screening collections. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 6817−6822.
(17) Philip, N. S.; Carpenter, L. L.; Tyrka, A. R.; Price, L. H. Nicotinic acetylcholine receptors and depression: A review of the preclinical and clinical literature. *Psychopharmacology (Berl.)* **2011**, *212*, 1−12.
(18) Taly, A.; Corringer, P. J.; Guedin, D.; Lestage, P.; Changeux, J. P. Nicotinic receptors: Allosteric transitions and therapeutic targets in the nervous system. *Nat. Rev. Drug Discovery* **2009**, *8*, 733−750.
(19) Lippiello, P. M.; Bencherif, M.; Hauser, T. A.; Jordan, K. G.; Letchworth, S. R.; Mazurov, A. A. Nicotinic receptors as targets for therapeutic discovery. *Expert Opin. Drug Discovery* **2007**, *2*, 1185−203.
(20) Romanelli, M. N.; Gratteri, L.; Guandalini, L.; Martini, E.; Bonaccini, C.; Gualtieri, F. Central nicotinic receptors: Structure, function, ligands, and therapeutic potential. *Chem. Med. Chem.* **2007**, *2*, 2−24.
(21) Breining, S. R. Recent developments in the synthesis of nicotinic acetylcholine receptor ligands. *Curr . Top. Med. Chem.* **2004**, *4*, 609−629.
(22) Mazurov, A. A.; Hauser, T. A.; Miller, C. H. Selective α7 nicotinic acetylcholine receptor ligands. *Curr. Med. Chem.* **2006**, *13*, 1567−1584.
(23) Bencherif, M.; Schmitt, J. D. Targeting neuronal nicotinic receptors: A path to new therapies. *Current Drug Targets CNS Neurol. Disord.* **2002**, *1*, 349−357.
(24) Peters, J.-U.; Schnider, P.; Mattei, P.; Kansy, M. Pharmacological promiscuity: Dependence on compound properties and target specificity in a set of recent Roche compounds. *ChemMedChem* **2009**, *4*, 680−686.
(25) Peters, J. U.; Hert, J.; Bissantz, C.; Hillebrecht, A.; Gerebtzoff, G.; Bendels, S.; Tillier, F.; Migeon, J.; Fischer, H.; Guba, W.; Kansy, M. Can we discover pharmacological promiscuity early in the drug discovery process? *Drug Discovery Today* **2012**, *17*, 325−335.
(26) Whitebread, S.; Hamon, J.; Bojanic, D.; Urban, L. In vitro safety pharmacology profiling: An essential tool for successful drug development. *Drug Discov. Today* **2005**, *10*, 1421−1433.
(27) Hamon, J.; Whitebread, S.; Techer-Etienne, V.; Le Coq, H.; Azzaoui, K.; Urban, L. In vitro safety pharmacology profiling: What else beyond hERG? *Future Med. Chem.* **2009**, *1*, 645−665.
(28) *MDDR*; Accelrys, Inc.: San Diego, CA, 2006.
(29) Mazurov, A.; Miao, L.; Klucik, J. Heteroaryl-Substituted Diazabicycloalkanes: Methods for Its Preparation and Use Thereof. U.S. Patent 7732607, 2010.
(30) Mazurov, A.; Klucik, J.; Miao, L.; Phillips, T. Y.; Seamans, A.; Schmitt, J. D.; Hauser, T. A.; Johnson, R. T.; Miller, C. H. 2-(Arylmethyl)-3-substituted quinuclidines as selective α7 nicotinic receptor ligands. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2073−2078.
(31) Bencherif, M.; Jordan, C.; Hauser, T.; Toler, S. M.; Letchworth, S. R.; Kombo, D. C. Treatment with alpha-7 Selective Ligands. WO Patent 2010/56622, 2010.
(32) Breining, S. R.; Mazurov, A.; Miller, C. H.; Phillips, T. Y.; Miao, L.; Bhatti, B. S.; Hawkins, G. D. Pharmaceutical Compositions and Methods for Relieving Pain and Treating Central Nervous System Disorders. U.S. Patent 7897611, 2011.
(33) Caldwell, W. S.; Dull, G. M.; Bhatti, B. S.; Hadimani, S. B.; Park, H.; Wagner, J. M.; Crooks, P. A.; Lippiello, P. M.; Bencherif, M. Hydroxybenzoate Salts of Metanicotine Compouinds. U.S. Patent 7459469, 2008.
(34) Breining, S. R.; Mazurov, A.; Miller, C. H.; Phillips, T. Y.; Miao, L.; Bhatti, B. S.; Hawkins, G. D. Pharmaceutical compositions and Methods for Relieving Pain and Treating Central Nervous System Disorders. U.S Patent 7402592, 2008.
(35) Dull, G. M.; Munoz, J. A.; Moore, J. R.; Genus, J. Hydroxybenzoate Salts of Metanicotine Compounds. U.S Patent 7459469, 2008.
(36) Bhatti, B. S.; Gatto, G. J.; Klucik, J. The Use of N-aryl Diazaspirocyclic Compounds in the Treatment of Addiction. European Patent 1784184, 2007.
(37) Strachan, J. P.; Cuthberson, T. J.; Wirth, D. D.; Dull, G. M.; Letchworth, S. R.; Jordan, K. G.; Kombo, D. C. Salt forms of 3 -cyclopropylcarbonyl-3,6-diazabicyclo[3.1.1.] Heptane WO Patent 2012/125518, 2012.
(38) Gopalakrishnan, M.; Honore, M., P.; Lee, C.; Malysz, J.; Ji, J.; Li, T.; Schrimpf, M., R.; Sippy, K. B.; Anderson, D. J. Pharmaceutical Compositions and Their Methods of Use. European Patent 2226074, 2011.
(39) Bhatti, B. S.; Schmitt, J. D.; Clark, T. J.; Miller, C. H. Pharmaceutical Compositions and Methods for Use. U.S. Patent 6579878, 2003.
(40) Crooks, P. A.; Deo, N. M. Pharmaceutical Compositions and Methods for Use. U.S. Patent 6211372, 2001.
(41) Coe, J. W.; Brooks, P. R.; Vetelino, M. G.; Wirtz, M. C.; Arnold, E. P.; Huang, J.; Sands, S. B.; Davis, T. I.; Lebel, L. A.; Fox, C. B.; Shrikhande, A.; et al. Varenicline: An $\alpha4\beta2$ nicotinic receptor partial agonist for smoking cessation. *J. Med. Chem.* **2005**, *48*, 3474−3477.
(42) Bhatti, B. S.; Strachan, J.-P.; Breining, S. R.; Miller, C. H.; Tahiri, P.; Crooks, P. A.; Deo, N.; Day, C. S.; Caldwell, W. S. Synthesis of 2-(pyridin-3-yl)-1-azabicyclo[3.2.2]nonane, 2-(pyridin-3-yl)-1-azabicyclo[2.2.2]octane, and 2-(pyridin-3-yl)-1-azabicyclo[3.2.1]-octane, a class of potent nicotinic acetylcholine receptor ligands. *J. Org. Chem.* **2008**, *73*, 3497−3507.
(43) Mazurov, A.; Miao, L.; Xiao, Y.; Yohannes, D.; Akireddy, S. R.; Breining, S. R.; Kombo, D.; Murthy, V. S. Derivatives of Oxabispidine as Neuronal Nicotinic Acetylcholine Receptor Ligands. European Patent 2356125, 2011.

(44) Mazurov, A. A.; Miao, L.; Bhatti, B. S.; Strachan, J. P.; Akireddy, S.; Murthy, S.; Kombo, D.; Xiao, Y. D.; Hammond, P.; Zhang, J.; Hauser, T. A.; Jordan, K. G.; Miller, C. H.; Speake, J. D.; Gatto, G. J.; Yohannes, D. Discovery of 3-(5-chloro-2-furoyl)-3,7-diazabicyclo[3.3.0]octane (TC-6683, AZD1446), a novel highly selective $\alpha4\beta2$ nicotinic acetylcholine receptor agonist for the treatment of cognitive disorders. *J. Med. Chem.* **2012**, *55*, 9181−9194.
(45) Mazurov, A. A.; Kombo, D. C.; Hauser, T. A.; Miao, L.; Dull, G.; Genus, G. F.; Fedorov, N. B.; Benson, L.; Sidach, S.; Xiao, Y.; Hammond, P. S.; James, J. W.; Miller, C. H.; Yohannes, D. Discovery of (2S,3R)-N-[2-(Pyridin-3-ylmethyl)-1-azabicyclo[2.2.2]oct-3-yl]benzo-[b]furan-2-carboxamide (TC-5619), a selective $\alpha7$ nicotinic acetylcholine receptor agonist, for the treatment of cognitive disorders. *J. Med. Chem.* **2012**, *55*, 9793−9809.
(46) Hahn, M. Receptor surface models. 1. Definition and construction. *J. Med. Chem.* **1995**, *38*, 2080−2090.
(47) 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* (Database issue), D1035−41.
(48) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36* (Database issue), D901−906.
(49) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *1* (Database issue), D668−672.
(50) Hou, T. J.; Xia, K.; Zhang, W.; Xu, X. J. ADME evaluation in drug discovery. 4. Prediction of aqueous solubility based on atom contribution approach. *J. Chem. Inf. Comp. Sci.* **2003**, *44*, 266−275.
(51) Field, A. *Discovering Statistics Using SPSS (Introducing Statistical Method)*, 3rd ed.; Sage Publications Ltd: Thousand Oaks, CA, 2009.
(52) Storey, J. D. A direct approach to false discovery rates. *J. R. Stat. Soc. B* **2002**, *64*, 479−498.
(53) Azzaoui, K.; Hamon, J.; Faller, B.; Whitebread, S.; Jacoby, E.; Bender, A.; Jenkins, J. L.; Urban, L. Modeling promiscuity based on in vitro safety pharmacology profiling data. *ChemMedChem* **2007**, *2*, 874−880.
(54) Medina-Franco, J. L.; Maggiora, G. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. A similarity-based data-fusion approach to the visual characterization and comparison of compound databases. *Chem. Biol. Drug Des.* **2007**, *70*, 393−412.
(55) Akella, L. B.; DeCaprio, D. Cheminformatics approaches to analyze diversity in compound screening librairies. *Curr Opin. Chem. Biol.* **2010**, *14*, 325−330.
(56) Halgren, T. A.; MMFF, V. I. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720−729.
(57) Hopkins, A. L.; Mason, J. S.; Overington, J. P. Can we rationally design promiscuous drugs? *Curr. Opin. Struct . Biol.* **2006**, *16*, 127−136.
(58) Rohrbaugh, R. H.; Jurs, P. C. Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. *Anal. Chim. Acta* **1987**, *199*, 99−109.
(59) Yera, E. R.; Cleves, A. E.; Jain, A. N. Chemical structural novelty: On-targets and off-targets. *J. Med. Chem.* **2011**, *54*, 6771−6785.
(60) Nettles, J. H.; Jenkins, J. L.; Bender, A.; Deng, Z.; Davies, J. W.; Glick, M. Bridging chemical and biological space: "Target fishing" using 2D and 3D molecular descriptors. *J. Med. Chem.* **2006**, *49*, 6802−6810.
(61) Perez-Nueno, V. I.; Venkatraman, V.; Mavridis, L.; Ritchie, D. W. Predicting drug polypharmacology using novel surface property similarity-based approach. *J. Med. Chem.* **2011**, 3 (Suppl. 1), O19.
(62) Hampton, T. Promiscuous anticancer drugs that hit multiple targets may thwart resistance. *J. Am. Med Soc.* **2004**, *292*, 419−422.
(63) Fabian, M. A.; Biggs, W. H. I., III; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lélias, J. M.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **2005**, *23*, 329−336.
(64) Atwell, S.; Adams, J.; Badger, J.; Buchanan, M. D.; Feil, I. K.; Froning, K. J.; Gao, X.; Hendle, J.; Keegan, K.; Leon, B. C.; Müller-Dieckmann, H. J.; Nienaber, V. L.; Noland, B. W.; Post, K.; Rajashankar, K. R.; Ramos, A.; Russel, M.; Burley, S. K.; Buchanan, S. G. A novel mode of Gleevec binding is revealed by the structure of spleen tyrosine kinase. *J. Biol. Chem.* **2004**, *279*, 55827−55832.
(65) Kung, C.; Kenski, D. M.; Dickerson, S. H.; Howson, R. W.; Kuyper, L. F.; Madhani, H. D.; Shokat, K. M. Chemical genomic profiling to identify intracellular targets to a multiplex kinase inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3587−3592.