

# Post-High-Throughput Screening Analysis: An Empirical Compound Prioritization Scheme

TUDOR I. OPREA,<sup>1</sup> CRISTIAN G. BOLOGA,<sup>1</sup> BRUCE S. EDWARDS,<sup>2</sup>  
ERIC R. PROSSNITZ,<sup>2</sup> and LARRY A. SKLAR<sup>2</sup>

An empirical scheme to evaluate and prioritize screening hits from high-throughput screening (HTS) is proposed. Negative scores are given when chemotypes found in the HTS hits are present in annotated databases such as MDDR and WOMBAT or for testing positive in toxicity-related experiments reported in TOXNET. Positive scores were given for higher measured biological activities, for testing negative in toxicity-related literature, and for good overlap when profiled against drug-related properties. Particular emphasis is placed on estimating aqueous solubility to prioritize *in vivo* experiments. This empirical scheme is given as an illustration to assist the decision-making process in selecting chemotypes and individual compounds for further experimentation, when confronted with multiple hits from high-throughput experiments. The decision-making process is discussed for a set of G-protein coupled receptor antagonists and validated on a literature example for dihydrofolate reductase inhibition. (*Journal of Biomolecular Screening* 2005:419-426)

**Key words:** chemoinformatics, DHFR inhibition, GPCR antagonism, hit evaluation, lead discovery, post-HTS analysis, structure-activity relationships

## INTRODUCTION

THE VIRTUAL AND HIGH-THROUGHPUT screening (HTS) worlds have often witnessed a large amount of interesting molecules (often in the order of 100-1000) that warrant further attention.<sup>1</sup> The answer to the question “What molecule(s) to pursue next?” is often left to the medicinal chemist<sup>2,3</sup> or to the biologist<sup>4</sup>—who often have little else besides intuition to guide them in this analysis. We illustrate the decision-making process<sup>5</sup> based on a G-protein coupled receptor (GPCR) target, in which hits were evaluated via high-throughput flow cytometry.<sup>6</sup> In this particular workflow, we started with a 3-dimensional model of the GPCR (e.g., see Evers and Klebe<sup>7</sup>) and docked (see Muegge and Enyedy<sup>8</sup> for review) known weak (micromolar) and strong (nanomolar) antagonists from the literature. Given the high level of uncertainty of this type of model, we derived a pharmacophore<sup>9</sup> and allowed large tolerance levels for the search, which we applied to the Chemical Diversity Labs

(CDL)<sup>10</sup> library containing more than 450,000 compounds. The pharmacophore search produced 6500 compounds, of which a subset of 4324 compounds was available and screened. Based on the results from flow cytometry, 95 compounds (2.2%) were classified as HTS hits (i.e., some inhibition was observed), 76 compounds (1.8%) exhibited more than 50% inhibition, and 15 compounds (0.3%) were classified as antagonists, having  $K_i$  values less than 10  $\mu\text{M}$  on the GPCR target of interest. Because further *in vitro* and *in vivo* experiments are needed to improve the quality of these compounds for drug discovery, we discuss here the workflow for the decision-making process and the reasons for the empirical scoring scheme adopted here. This workflow is validated by applying the same empirical scheme on a set of 9 dihydrofolate reductase (DHFR) inhibitors, selected from 12 reported molecules.<sup>11</sup> These inhibitors were identified by screening a 50,000 small-molecule library via HTS then performing  $K_i$  determination for 62 compounds in secondary screening.<sup>11</sup>

## MATERIALS AND METHODS

### *Chemotype evaluation*

*Hits/Tested score.* Although only 33 compounds were tested in dose-response experiments ( $K_i$  values between 40 and 1  $\mu\text{M}$ ) for

<sup>1</sup>Division of Biocomputing and <sup>2</sup>Department of Pathology and Cancer Research and Treatment Center, University of New Mexico, Albuquerque.

Received Apr 30, 2004, and in revised form Sep 16, 2004. Accepted for publication Sep 26, 2004.

*Journal of Biomolecular Screening* 10(5); 2005  
DOI: 10.1177/1087057104272660

GPCR antagonism, our evaluation scheme takes into account the total number of compounds screened per individual chemotype, expressed as the hits/tested (HT): “tested” represents the total number of compounds evaluated on this target in this flow cytometry screen, whereas “hits” were considered all compounds that exhibited at least 68% inhibition. The HT score evaluates the hit rate of a particular chemotype and captures negative information (e.g., inactives) as well. Individual biological activities are considered separately (see below). We applied the following criterion, according to the principle that a higher H/T ratio should be reflected by a higher HT score value:  $H/T \geq 0.5$ ,  $HT = 50$ ;  $H/T \geq 0.35$ ,  $HT = 35$ ;  $H/T \geq 0.2$ ,  $HT = 20$ ;  $H/T \geq 0.1$ ,  $HT = 10$ ;  $H/T \geq 0.05$ ,  $HT = 5$ ; otherwise  $HT = 0$ . In the DHFR validation case, the chemotype evaluation was performed on 11 of the 12 reported molecules<sup>11</sup> using  $T = 20$  (tested molecules) for each of the 5 chemotypes. We chose  $T = 20$  as the default to minimize its influence on the rest of the analysis; the exact number of tested chemotype representatives was not reported. The H/T ratio scoring scheme for the DHFR example was the same as for the GPCR target. In this, and all other evaluation schemes presented in this work, scores are dimensionless. To distinguish between chemotypes, **G** and **D** are used to prefix GPCR antagonists and DHFR inhibitors, respectively.

*Intellectual property evaluation scheme.* We queried MDDR (MDL's drug data report<sup>12</sup>) and WOMBAT (World of Molecular Bioactivity<sup>13</sup>) for related substructures in an attempt to capture the value of various chemotypes with respect to drug discovery. We scored relevant chemotypes according to the principle that higher intellectual property (IP) scores should reflect a lower number of related substructures found in MDDR and WOMBAT. We awarded negative points for each hit in MDDR ( $-5$  for identical and  $-1$  for related [i.e., similar] substructures) and for hits in WOMBAT (we divided the total number of WOMBAT hits by  $-10$ ). Higher weight was given to MDDR hits because it implies that the IP value of that substructure is already considered by a 3rd party for commercial applications, whereas WOMBAT hits are related more to public domain (not necessarily patents). In summary, we penalized each chemotype that had hits in any of these 2 databases.

*Tox score and IPT score.* We ran multiple queries in the TOXNET<sup>14</sup> database system for information available on substructures specifically related to the compounds in question. Choosing the appropriate fragments to query is more art than science (e.g., cyanide is highly toxic by itself but safe(r) in cimetidine), but every attempt was made to choose the relevant structural family. Thus, TOXNET queries did not always include the smallest substructure but rather fragments that were deemed relevant to bioactivity in that class. We applied the following criteria, according to the principle that higher Tox scores should be reflected by higher Tox values: negative toxicity studies for the exact substructure,  $Tox = 100$ ; negative toxicity studies for related substructures,  $Tox = 50$ ; known (potential) toxicity studies for related substructures,  $Tox = -50$ ; known toxicity studies for the exact substructure,  $Tox = -100$ . In the absence of any other data,  $Tox = -10$  (default), because all

compounds are toxic (depending on the dose). Finally, we summed the HT, the IP, and Tox scores to evaluate the attractiveness of each chemotype in the IP and Tox (IPT) score.

### Individual molecule evaluation

*Physico-chemical property score.* Because the next stage of experiments is in vivo mouse studies, the estimated permeability for each confirmed HTS active was evaluated at the individual level, given the chemotype context. Given the relatively high margin of error for solubility prediction methods, the solubility/permeability (SP) score was computed in rather large intervals, deemed to match the property distribution of orally available drugs,<sup>15</sup> as follows:  $\text{LogP}/D > 6$ ,  $SP = -10$ ;  $\text{LogP}/D > 5$ ,  $SP = 10$ ;  $\text{LogP}/D > 3$ ,  $SP = 20$ ;  $\text{LogP}/D > 0$ ,  $SP = 50$ ;  $\text{LogP}/D > -2$ ,  $SP = 20$ ; otherwise  $SP = -10$ , where  $\text{LogP}/D$  reflects 2 separate descriptors, that is,  $\text{LogP}$ ,<sup>16</sup> or the octanol/water partition coefficient, and  $\text{LogD}_{7.4}$  (the  $\text{pH} = 7.4$  corrected  $\text{LogP}$ ). For the CDL database, these values are already computed with ChemoSoft.<sup>17</sup> For DHFR inhibitors, the corresponding values, precomputed with ACD/Labs software,<sup>18</sup> were extracted from SciFinder.<sup>19</sup>

*Minimum required aqueous solubility score.* To prepare for experiments in mice, we applied the minimum required aqueous solubility test ( $\text{min\_S}_{\text{aq}}$ ) as follows:

$$\text{Dose } (\mu\text{M}) = T_w \times K_i \times \text{Vol}, \quad (1)$$

where  $T_w$  is the therapeutic window (typically 10 to 20 times higher than the  $K_i$ ; in this study,  $T_w = 20$ );  $K_i$  is the measured inhibition constant, expressed in micromoles; and Vol is the total mouse blood volume (typically, and in this study, 5 mL).

$$\text{Dose (mg)} = \text{Dose } (\mu\text{M}) \times \text{MW}/1000, \quad (2)$$

where MW is the molecular weight (in atomic mass units) for each compound.

$$S_{\text{aq}} = \text{Dose } (\mu\text{M})/\text{SVol}, \quad (3)$$

where  $S_{\text{aq}}$  is the required solubility and SVol is the sample volume. In this study,  $\text{SVol} = 200 \mu\text{L}$  (this is the maximum fluid volume that can be given as bolus injection in mice).

$$\text{min\_S}_{\text{aq}} \text{ test: } \text{LogS}_{\text{aq}} > \text{LogS}_{\text{aq}} + \text{Log}10. \quad (4)$$

The  $\text{min\_S}_{\text{aq}}$  test forces the estimated  $\text{LogS}_{\text{aq}}$  (computed by either ChemoSoft<sup>17</sup> or ACD/Labs<sup>18</sup>) to exceed 10 times the required solubility dose (in equation 4,  $\text{Log}10$ ). The required solubility ( $S_{\text{aq}}$ ) is based on measured  $K_i$  values and, as such, reflects a “must-have” for the success of the in vivo experiments. The 10-times excess value reflects the possible errors in computing  $\text{LogS}_{\text{aq}}$ . The  $\text{min\_S}_{\text{aq}}$  test was applied to all compounds with available estimated  $\text{LogS}_{\text{aq}}$  and measured  $K_i$  values. Fifty points were awarded to compounds passing this test, and 50 points were subtracted from the score of compounds that failed this test.

*High activity and priority scores.* To take into account biological potency, we scored compounds according to their measured activity on the target of interest, as follows: For the GPCR antagonists, 100 points were awarded for  $10\ \mu\text{M} \geq K_i \geq 1\ \mu\text{M}$ ; 50 points were awarded for  $K_i$  between 10 and  $20\ \mu\text{M}$ , 20 points for  $K_i$  between 20 and  $40\ \mu\text{M}$ , and 10 points for  $K_i$  that had higher values. For the DHFR inhibitors, compounds were scored according to the  $K_i$  as follows: 80 points were awarded for  $5\ \text{nM} > K_i \geq 1\ \text{nM}$ , 60 points were awarded for  $10\ \text{nM} > K_i \geq 5\ \text{nM}$ , 40 points were awarded for  $50\ \text{nM} > K_i \geq 10\ \text{nM}$ , 20 points for  $0.1\ \mu\text{M} > K_i \geq 50\ \text{nM}$ , 10 points for  $0.5\ \mu\text{M} > K_i \geq 0.1\ \mu\text{M}$ , and 1 point for  $K_i$  above  $1\ \mu\text{M}$ . The score variation between the 2 targets reflects the limited distribution of the bioactivities for the DHFR inhibitors (only 9 structures scored) compared to the GPCR antagonists (33 structures ranked). The priority score was computed by summing the IPT (chemotype) score, the SP, and the  $\text{min}_S$  scores, as well as the  $K_i$  score.

## RESULTS AND DISCUSSION

### *Chemotype evaluation results for the GPCR set*

HT scores and bioactivities recorded for each chemotype in the MDDR and WOMBAT databases are presented in Figure 1. All chemical substructures depicted in Figure 1 follow the following (ISIS/base) convention: Dashed lines represent any bond type, A searches for “any atom except hydrogen,” and hydrogens are not explicit (unless specified). The HT evaluation scheme produces some interesting contrasts: 3 chemotypes (**G.A**, **G.B**, and **G.E**) are awarded the highest score, 50, whereas dihydropyridines and flavones (**G.C** and **G.H**) are awarded very low scores. The HT score, although not related to the potency of individual compounds, incorporates information about HTS negatives; these are usually ignored in structure-activity studies. Thus, low scores in families **G.C** and **G.H** reflect the fact that a relatively large number of analogs were found to be inactive on this GPCR target, compared to, for example, rhodanines (**G.E**), in which 3 out of 3 compounds are active.

Figure 1 highlights in bold the instances in which bioactivity information related to GPCR targets was recorded for this chemotype. Three out of 9 families have GPCR-related activities: The dihydropyridines (**G.C**) exhibit, besides calcium channel-blocking activity, antagonism on NPY, PAF, alpha adrenergic, and adenosine receptors. Chemotype **G.D** (in particular, 2,3-pyrrolidinediones) are CCR2 antagonists, besides exhibiting aldose reductase inhibition activity, whereas chemotype **G.H** (flavones) are dopamine-2 and sigma receptor antagonists, as well as kinase inhibitors. Figure 1 also underscores the potential promiscuity of some families (e.g., coumarins, **G.G**) that hit a variety of unrelated targets (including dopamine D4 receptors), whereas some other families (e.g., **G.A**) are much less explored according to patents and medicinal chemistry literature. Low scores in the IP arena (e.g., for dihydropyridines and coumarins, see Fig. 1) made these particular chemotypes less appealing: It is quite likely that

compounds of interest are already being investigated or patented for some related activity. Further optimization in this chemical space is therefore considered a more difficult task.

We did not limit ourselves to evaluating the IP potential of these chemotypes but proceeded to evaluate available toxicity information. Families **G.B** (pirazolo-pyridines) and **G.G** (coumarins) were penalized because of known contraceptive (**G.B**) and carcinogenic (**G.G**) activities, whereas families **G.E** (rhodanines) and **G.H** (flavones) were awarded high scores for being found relatively nontoxic. Given the composite IPT score, 3 chemotypes were highlighted for further investigation: rhodanines (**G.E**), flavones (**G.H**), and pyrrolidinediones (**G.D**).

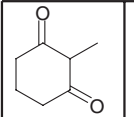
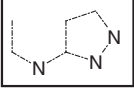
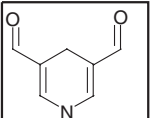
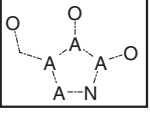
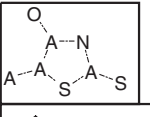
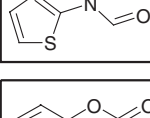
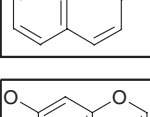
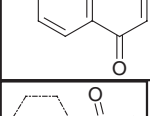
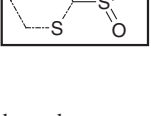
### *Individual molecule evaluation results for the GPCR set*

Individual molecules were evaluated in the context of chemotype scores, to which we added information related to the estimated physico-chemical properties (solubility/permeability), the minimum required aqueous solubility test ( $\text{min}_S$ ), and the measured biological activity, to derive the final priority score. The results for the top 12 ranked molecules are presented in Figure 2. Because solubility and high activity were given high scores when planning the mice experiments, it is not surprising that the rhodanine (**G.E**) chemotypes with  $K_i$  values of 4 and  $6\ \mu\text{M}$  are the top 2 molecules. It is reassuring to note that rhodanine itself is not carcinogenic and that rhodanine derivatives have not been heavily investigated in the GPCR arena (see Fig. 1).

Four other top-ranked molecules are from the **G.A** (cyclohexanedione) chemotype—with  $K_i$  values between 1 and  $21\ \mu\text{M}$ —and 1 of them passes the  $\text{min}_S$  test (Fig. 2). With as little as is known about the potential bioactivity and toxicity of these compounds, they are not considered the highest priority. The 3rd-ranked priority chemotype, pyrrolidinediones (**G.D**), is penalized due to its low solubility, even though the chemotype itself was considered quite interesting and even though the compound has promising activity ( $4\ \mu\text{M}$ ). Flavones (**G.H**) represent the 4th chemotype in the priority score, with  $K_i$  values between 2 and  $10\ \mu\text{M}$ ; 1 of them, with  $K_i = 6\ \mu\text{M}$ , almost passed the  $\text{min}_S$  test. This chemical family has a rather high number of tested analogs (182), of which only 11 show any activity (see Fig. 1). It was, therefore, difficult to assign them a high priority, even though related compounds pass the carcinogenicity tests.

### *Chemotype and individual molecule analyses for the DHFR inhibitors set*

The utility of the post-HTS prioritization scheme was further tested on a set of DHFR inhibitors, recently identified via HTS and confirmed by secondary screening.<sup>11</sup> The chemotype prioritization scheme, including HT scores, recorded bioactivities for each chemotype from MDDR and WOMBAT, as well as the TOXNET and IPT score (Fig. 3). Five chemotypes emerge as relevant from the original report: There are five 2,4-diamino-quinazolines (**D.A**), 2 methyl-substituted 5,6,7,8-tetrahydroquinazoline-2,4-diamines

Family	Smallest Substructure query	Tested	Hits	HT Score	MDDR Hits	MDDR Related	Bioactivity	Wombat Hits	Bioactivity	IP Score	TOXNET Information	Tox Score	IPT Score
G.A		15	11	50	0	0	None	0	None	50	Nothing recorded for this substructure	-10	40
G.B		8	5	50	0	2	Cognition enhancer, antidepressant	238	Potassium channel agonists, ACAT inhibitors	24.2	Contraceptive [3-amino-1H-pyrazolo [3,4-b] quinoline interferes with trophoblasts]	-50	-25.8
G.C		531	15	0	0	33	Calcium channel blockers, anti-hypertensive	458	Calcium channel blockers, [NPY Y1, PAF, alpha AR, adenosine] antagonists	-78.8	Nothing recorded for this substructure, other than Calcium channel blockade	-10	-88.8
G.D		11	4	35	3	0	CCR2 antagonists weak antimicrobial	25	Aldose Reductase Inhibitor	17.5	Enhances cytotoxic effect for doxorubicin and cisplatin via the extracellular signal-regulated kinase pathway	50	67.5
G.E		3	3	50	1	1	Aldose Reductase Inhibitor	20	[COX/PGF2 alpha production, Aldose reductase, PLA2] inhibitors	42	Rhodanine tested negative for mutagenicity; other results not available	100	142
G.F		34	12	35	0	12	PAF antagonists	116	[TNF alpha production, vitronectin R, GABA-A/BzR, ET-A] inhibitors	11.4	Nothing recorded for this substructure	-10	1.4
G.G		11	4	35	9	9	Anti-coagulant, anti-cholinergic, MAO inhibitors	295	[5-LO, HIV-RT, HIV-PR, D4-R, MAO, CYP2A5] inhibitors	-48.5	Coumarin tested positive in rat carcinogenicity	-100	-148.5
G.H		182	11	5	3	3	Calcium regulator, PKC inhibitor	307	[D2-R, sigma-R, 5-LO, aldose reductase, CDK-1] inhibitors	-43.7	Eflorate & Daidzein tested negative in Ames tests	100	56.3
G.I		2	1	35	0	6	Anti-coagulant, carbonic anhydrase inhibitors	0	None	29	nothing recorded for this substructure	-10	19

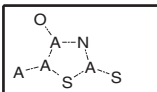
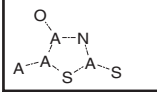
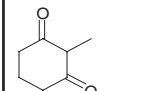
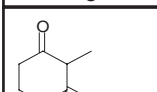
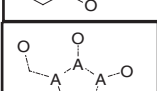
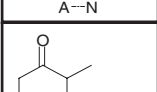
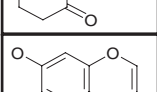
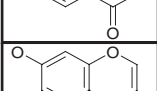
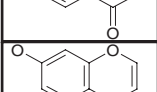
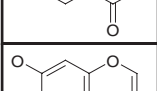
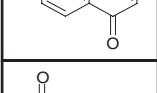
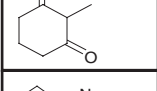
**FIG. 1.** High-throughput screening results, hits/tested (HT) score, documented bioactivity, intellectual property (IP) information, IP score, toxicity information, Tox score and IP and Tox (IPT) score for the G-protein coupled receptor set.

(**D.B**), two 2,4-diamino-pyrimidines (**D.C**), 1 benzyl-amidinothiourea (**D.D**), and 1 phenyl-biguanide (**D.E**). Another chemotype, reported for a single quinolinone, was excluded from the analysis due to its lack of DHFR activity ( $K_i > 10 \mu\text{M}$ ). Based on the MDDR and WOMBAT IP analysis, it becomes apparent that 3 of the 5 chemotypes (**D.A**, **D.B**, and **D.C**) have already been investigated for DHFR inhibition activity (highlighted in bold) and that 2 of these 3 chemotypes already include launched drugs, for example, trimetrexate (**D.A**) and trimethoprim (**D.C**). Because TOXNET provides pertinent information related to DHFR inhibitors based on the **D.A** and **D.C** chemotypes, they are awarded lower Tox scores. The 3rd drug captured by this analysis is the antimalarial proguanil from the **D.E** chemotype family, which is

awarded a higher Tox score, again due to relevant information available in TOXNET. Regardless of the Tox scores, the IP scores indicate the scenario that might occur in a drug discovery setting: Because chemotypes **D.A** and **D.C** are already well explored in the DHFR inhibitor IP arena, scientists working in drug discovery are likely to focus on the other 3 chemotypes. Because **D.B** was investigated in medicinal chemistry publications (as captured by WOMBAT), the drug discovery team would have to examine those reports and carefully evaluate the impact of prior art on any future patents.

It is not surprising that **D.E**, **D.B**, and **D.D** chemotype representatives are given the highest priority in Figure 4, once the measured biological activity, the estimated physico-chemical properties

## Empirical Compound Prioritization Scheme

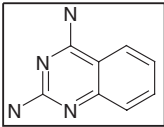
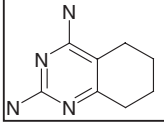
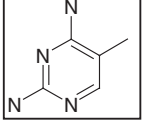
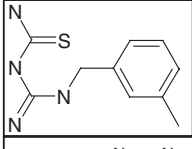
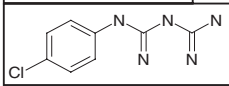
Chemotype	K <sub>i</sub> ( $\mu$ M)	MW	LogP	LogD <sub>74</sub>	LogS <sub>aq</sub>	Dose ( $\mu$ M)	Dose (mg)	S <sub>aq</sub>	min_S <sub>aq</sub>	Priority Score
	4	376.5	2.95	-1.38	0.03	0.4	0.15	-2.70	PASS	362
	6	441.3	3.49	-1.22	0.91	0.6	0.26	-2.52	PASS	332
	21	400.5	2.23	0.37	-0.38	2.1	0.84	-1.98	PASS	200
	10	413.5	2.26	1.50	-3.89	1	0.41	-2.30	FAIL	190
	4	409.5	4.01	2.55	-4.63	0.4	0.16	-2.70	FAIL	187.5
	10	447.4	3.11	2.29	-1.67	1	0.45	-2.30	FAIL	160
	2	392.3	4.54	3.45	-5.15	0.2	0.08	-3.00	FAIL	146.3
	6	456.5	4.36	3.12	-1.76	0.6	0.27	-2.52	FAIL	146.3
	10	357.4	3.60	3.31	-3.48	1	0.36	-2.30	FAIL	146.3
	10	406.4	4.92	3.95	-5.27	1	0.41	-2.30	FAIL	146.3
	1	474.6	3.93	3.95	-5.64	0.1	0.05	-3.30	FAIL	130
	4	562.7	4.92	3.70	-4.94	0.4	0.23	-2.70	FAIL	130

**FIG. 2.** Biological activity, estimated properties, the required minimum solubility test, and the final priority score for the top 12 ranking GPCR compounds.

(SP), and the minimum required aqueous solubility test (min\_S<sub>aq</sub>) are included in the final score. Trimethoprim (the 6th molecule in Fig. 4) is given a middle-tier priority, illustrating how good activity and physico-chemical properties can offset even the lowest IPT score for this group.

Our post-HTS evaluation scheme is validated by the fact that all top 5 molecules ranked in Figure 4 are previously unreported DHFR inhibitors (in the context of the initial publication<sup>11</sup>) and that

both the 1st and 4th molecule are potent and unique (in terms of IP position) DHFR inhibitors. That the top 5 molecules come from 4 different chemical classes, 3 of which belong to low-risk IP areas (the **D.E**, **D.B**, and **D.D** chemotypes), further illustrates the advantage of using such prioritization schemes, compared to simple bioactivity ranking. In contrast, if one were to use only K<sub>i</sub> values to rank the same 9 molecules discussed in Figure 4, the top candidate would be trimethoprim (**D.C**), followed by the benzyl-amidino-

Family	Smallest Substructure Query	Tested	Hits	HT Score	MDR Hits	MDR Related	Bioactivity	Wombat Hits	Bioactivity	IP Score	TOXNET Information	Tox Score	IPT Score
D.A		20	5	20	10	1	Trimetrexate is a DHFR inhibitor drug	308	123 DHFR inhibitors	-61.8	Well-described hematologic toxicity for Trimetrexate	-50	-111.8
D.B		20	2	5	0	0	None	41	41 DHFR inhibitors	0.9	Nothing recorded for this substructure	-10	-9.1
D.C		20	2	5	9	4	Trimethoprim, Brodimoprim and Epiroprim are DHFR inhibitor drugs	506	327 DHFR inhibitors	-94.6	Aneuploidy, chromosome aberrations, AMES test positive for Trimethoprim	-100	-194.6
D.D		20	1	0	0	0	None	0	None	0	Nothing recorded for this substructure	-10	-10
D.E		20	1	0	5	1	Proguanil-antimalarial drug	9	5HT3 antagonists	-26.9	Proguanil has well-studied, low toxicity	50	23.1

**FIG. 3.** High-throughput screening results, hits/tested (HT) score, documented bioactivity and intellectual property (IP) information, IP score, toxicity information, Tox score and IP and Tox (IPT) score for the dihydrofolate reductase (DHFR) set.

thiourea (**D.D**) and two 2,4-diamino-quinazolines (**D.A**). The most interesting molecule according to our own analysis, the phenyl-biguanide (**D.E**), would rank 5th, whereas the 5,6,7,8-tetrahydroquinazoline-2,4-diamines (**D.B**) chemotype would be at the bottom of the activity-ranked list. Albeit empirical, we believe this scheme is supported by the DHFR inhibitors analysis, further highlighting the need for more complex decision-making schemes when selecting candidates for additional investigation.

## CONCLUSIONS

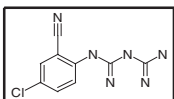
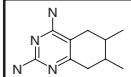
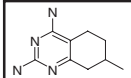
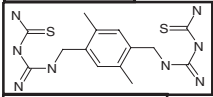
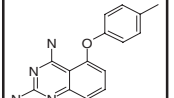
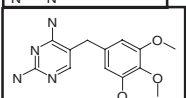
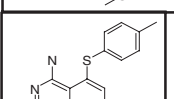
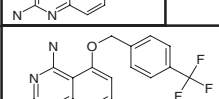
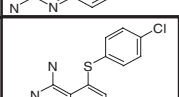
This article illustrates practical steps related to the evaluation of compounds for further experiments, once the HTS actives have been analyzed and confirmed and once their dose-response curves have been determined. The HT score captures information related not only to actives but also to inactives from the same chemical family, whereas the IP evaluation scheme and the Tox score are aimed at encoding information related to individual chemotypes, as they are available in various sources. At the individual molecule level, an example of how estimated physico-chemical properties can be included in the final prioritization score is given, starting with the property profile of known drugs.<sup>15</sup> For practical reasons, the minimum required aqueous solubility score is given a high weight, equal to the score awarded for high biological activity. Given enough resources, we encourage the use of measured

physico-chemical properties such as  $\text{LogS}_{\text{aq}}$  and  $\text{LogD}_{7.4}$  instead of computed ones to serve better the decision-making process.

Whenever the situation demands it, for example, a rather large number of distinct chemical families have emerged from primary and secondary HTS assays, the user can automatically derive Tox scores via expert systems, for example, using DEREK<sup>20</sup> and MULTICASE.<sup>21</sup> The output of these methods, when given in numeric form, can be directly used in the IPT score. An alternative is to sort the HTS hits according to chemotype and bioactivity values, then evaluate the top X structures (where X is 3, 5, etc.) from each chemotype by querying appropriate literature or rule-based systems, then incorporating these results at the compound (rather than chemotype) level. It is not surprising that the top 3 ranking GPCR chemotypes are among the top 4 recovered in the final priority score, but it is reassuring to note that, among these 3 chemotypes, rhodanine has high solubility, whereas the other 2 include analogs known to be active on GPCR targets. The fact that 2 DHFR inhibitors from previously unreported chemotypes are prioritized in the top 4 candidate molecules further validates this empirical scheme. We note that with the exception of a minor change related to biological activity scoring, the same empirical scoring scheme was applied in both the GPCR and DHFR data sets.

This article, aimed at showing how post-HTS analyses can be done, offers a practical method to encode information that is deemed relevant to preclinical drug discovery scientists and to doc-

## Empirical Compound Prioritization Scheme

Structure	Ki (nM)	MW	LogP	LogD <sub>74</sub>	LogS <sub>aq</sub>	Dose (nM)	Dose (µg)	S <sub>aq</sub>	min_S <sub>aq</sub>	IPT Score	Priority Score
	65	236.7	0.37	-1.60	-0.67	6.5	1.54	-4.49	PASS	23.1	183.1
	190	192.3	1.33	2.03	-2.09	19	3.65	-4.02	PASS	-9.1	160.9
	2300	178.2	1.54	0.85	-1.84	230	40.99	-2.94	PASS	-9.1	141.9
	26	366.5	-0.62	-0.62	-2.59	2.6	0.95	-4.89	PASS	-10	140
	41	266.3	3.38	1.25	-2.42	4.1	1.09	-4.69	PASS	-111.8	68.2
	3.8	290.3	0.79	0.52	-2.39	0.38	0.11	-5.72	PASS	-194.6	55.4
	73	282.4	3.62	2.05	-3.40	7.3	2.06	-4.44	PASS	-111.8	48.2
	130	334.3	3.07	0.88	-2.85	13	4.35	-4.19	PASS	-111.8	28.2
	61	302.8	3.99	2.62	-4.13	6.1	1.85	-4.52	FAIL	-111.8	-51.8

**FIG. 4.** Biological activity, estimated properties, the required minimum solubility test, and the final priority score for 9 dihydrofolate reductase (DHFR) inhibitors. MW = molecular weight; IPT = IP and Tox score.

ument the decision-making process<sup>5</sup> in the lead identification stage. By changing the weighting schemes to tailor particular projects or laboratories, one can monitor and understand the various factors that need to be considered in compound prioritization. For example, companies that have excellent IP portfolios in the dihydropyridine or 2,4-diamino-pyrimidine arenas might give these chemotypes (**G.C** and **D.E**, respectively) a much higher priority because it is less likely that **G.C** compounds will cause cardiac toxicity and because their toxicity profile is already well understood.

### ACKNOWLEDGMENTS

This work was supported by New Mexico Tobacco Settlement funds (TIO and CGB). The authors thank Nikolay Savchuk and Ilya Okun (Chemical Diversity Labs Inc.) for valuable comments and for providing the compound samples.

### REFERENCES

- Oprea TI, Li J, Muresan S, Mattes KC: High throughput and virtual screening: choosing the appropriate leads. In Ford M, Livingstone D, Dearden J, Van de Waterbeemd H (eds): *EuroQSAR 2002: Designing Drugs and Crop Protectants—Processes, Problems and Solutions*. New York: Blackwell, 2003:40-47.
- Rishton GM: Reactive compounds and in vitro false positives in HTS. *Drug Discov Today* 1997;2:382-384.
- Rishton GM: Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discov Today* 2003;8:86-96.
- Heyse S: Comprehensive analysis of high-throughput screening data. In Bornhop DJ, Dunn DA, Mariella RP, Murphy CJ, Nicolau DV, Nie S, et al (eds): *Proceedings of SPIE: The International Society for Optical Engineering, vol. 4626—Biomedical Nanotechnology Architectures and Applications*. Bellingham (WA): SPIE Press, 2002:535-547.
- Olsson T, Oprea TI: Cheminformatics: a tool for decision-makers in drug discovery. *Curr Opin Drug Discov Dev* 2001;4:308-313.

6. Edwards BS, Oprea TI, Prossnitz ER, Sklar LA: Flow cytometry for high throughput, high content screening. *Curr Opin Chem Biol* 2004;8:392-398.
7. Evers A, Klebe G: Ligand-supported homology modeling of G-protein-coupled receptor sites: models sufficient for successful virtual screening. *Angew Chem Int Ed* 2004;43:248-251.
8. Muegge I, Enyedy I: Docking and scoring. In Tollenaere J, De Winter H, Langenaeker W, Bultinck P (eds): *Computational Medicinal Chemistry and Drug Discovery*. New York: Marcel Dekker, 2004:405-436.
9. Güner O (ed): *Pharmacophore Perception, Development and Use in Drug Design*. La Jolla (CA): International University Line, 2000.
10. Chemical Diversity Labs, Inc. San Diego (CA). Retrieved from <http://www.chemdiv.com>
11. Zolli-Juran M, Cechetto JD, Hartlen R, Daigle DM, Brown ED: High throughput screening identifies novel inhibitors of *Escherichia coli* dihydrofolate reductase that are competitive with dihydrofolate. *Bioorg Med Chem Lett* 2003;13:2493-2496.
12. MDDR: Molecular Design, Ltd. San Leandro (CA). Retrieved from <http://www.mdli.com>
13. Olah M, Mracec M, Ostopovici L, Rad R, Bora A, Hadaruga N, et al: WOMBAT: World of Molecular Bioactivity. In Oprea TI (ed): *Cheminformatics in Drug Discovery*. New York: Wiley-VCH, 2005:223-239. Retrieved from <http://www.sunsetmolecular.com>
14. National Library of Medicine. TOXNET [Online]. Retrieved from <http://toxnet.nlm.nih.gov/>
15. Oprea TI: Property distribution of drug-related chemical databases. *J Comput Aided Mol Des* 2000;14:251-264.
16. Leo A: Calculating log P<sub>oct</sub> from structures. *Chem Rev* 1993;93:1281-1306.
17. Chemical Diversity Labs, Inc. CHEMOSOFT. San Diego (CA). Retrieved from <http://www.chemosoft.com>
18. Advanced Chemistry Development, Inc. Toronto, Ontario (Canada). Retrieved from <http://www.acdlabs.com/>
19. American Chemical Society. SciFinder Scholar 2004. Retrieved from <http://www.cas.org/SCIFINDER/SCHOLAR/index.html>
20. Sanderson DM, Earnshaw CG: Computer prediction of possible toxic action from chemical structure. The DEREK system. *Human Exp Toxicol* 1991;10:261-273.
21. Klopman G: MULTICASE: 1. A hierarchical computer automated structure evaluation program. *Quant Struct-Act Relat* 1992;11:176-184.

Address preprint requests to:

Tudor I. Oprea  
UNM Division of Biocomputing  
MSC08 4560  
One University of New Mexico  
Albuquerque, NM, 87131

E-mail: [toprea@salud.unm.edu](mailto:toprea@salud.unm.edu)