Article

# Phantom PAINS: Problems with the Utility of Alerts for Pan-Assay INterference CompoundS

Stephen J. Capuzzi, Eugene N. Muratov, and Alexander Tropsha*

Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599, United States

**S** *Supporting Information*

**ABSTRACT:** The use of substructural alerts to identify **P**an-**A**ssay **IN**terference compound**S** (PAINS) has become a common component of the triage process in biological screening campaigns. These alerts, however, were originally derived from a proprietary library tested in just six assays measuring protein−protein interaction (PPI) inhibition using the AlphaScreen detection technology only; moreover, 68% (328 out of the 480 alerts) were derived from four or fewer compounds. In an effort to assess the reliability of these alerts as indicators of pan-assay interference, we performed a large-scale analysis of the impact of PAINS alerts on compound promiscuity in bioassays using publicly available data in PubChem. We found that the majority (97%) of all compounds containing PAINS alerts were actually infrequent hitters in AlphaScreen assays measuring PPI inhibition. We also found that the presence of PAINS alerts, contrary to expectations, did not reflect any heightened assay activity trends across all assays in PubChem including AlphaScreen, luciferase, beta-lactamase, or fluorescence-based assays. In addition, 109 PAINS alerts were present in 3570 extensively assayed, but consistently inactive compounds called Dark Chemical Matter. Finally, we observed that 87 small molecule FDA-approved drugs contained PAINS alerts and profiled their bioassay activity. Based on this detailed analysis of PAINS alerts in nonproprietary compound libraries, we caution against the blind use of PAINS filters to detect and triage compounds with possible PAINS liabilities and recommend that such conclusions should be drawn only by conducting orthogonal experiments.

CID: 9550911
ASSAYS TESTED: 686
ASSAY ACTIVE: 0
PAINS ALERTS: 2
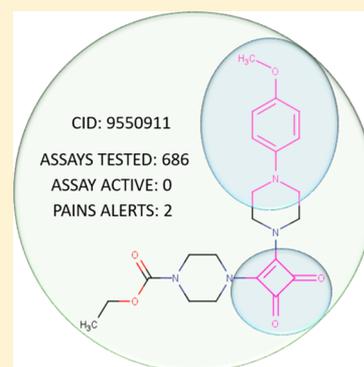
## 1. INTRODUCTION

The scientific community is in the grips of the data reproducibility crisis, highlighted by *Nature's* "Challenges in Irreproducible Research" initiative.[1,2] Oftentimes in drug discovery, compounds active in primary biological screens show no activity in follow-up studies.[3−5] The measured effect of false positives may be due to various mechanisms including those that interfere with the assay detection technology such as autofluorescence, hydrogen peroxide production, metal chelation, chemical aggregation, etc.[2−6]

In a highly cited study,[6] Baell and Holloway analyzed compounds that showed activity in multiple assays and suggested that these compounds may interfere with the bioactivity detection technology. Such compounds were cleverly dubbed PAINS, or **P**an-**A**ssay **IN**terference compound**S**. In an effort to provide a tool to enhance reproducibility and reliability of true hit identification in drug discovery, the authors then identified 480 "PAINS alerts", i.e., substructural features frequently found in PAINS, and suggested that these alerts could be used to flag false screening hits and annotate suspect compounds in screening libraries.[6]

Following the original publication,[6] which has garnered more than 800 citations according to Google Scholar at the time of this study, the concept of PAINS alerts (filters) has gained much attention, many supporters, and prompted many follow-up publications.[7−9] Several web-based applications relying on the original work by Baell and Holloway[6] have been developed to flag and filter compounds with PAINS alerts;[10,11] chemical databases, such as ZINC (http://zinc15.docking.org/) and ChEMBL (https://www.ebi.ac.uk/chembl/), also flag compounds containing PAINS alerts. On the scientific blogosphere, publications reporting compounds flagged with PAINS alerts as viable hits have been publicly ridiculed in a practice known as "PAINS-Shaming."[12]

The wide acceptance of the PAINS concept by the scientific community and the availability of PAINS filters have made it common for researchers to triage virtual screening hits flagged with these alerts prior to experimental validation.[13] Similarly, lead compounds resulting from experimental screening campaigns have typically been deprioritized for follow-up studies if they contained PAINS alerts.[14] Furthermore, scientific journals have begun to recommend that all hit compounds, virtual or otherwise, should be passed through one of the publicly available PAINS filters before the manuscript is considered for publication. For instance, the *Journal of Medicinal Chemistry* requires that "active compounds from any source must be examined for known classes of assay interference compounds".[15] The authors are asked to "provide firm experimental evidence in at least two different assays that
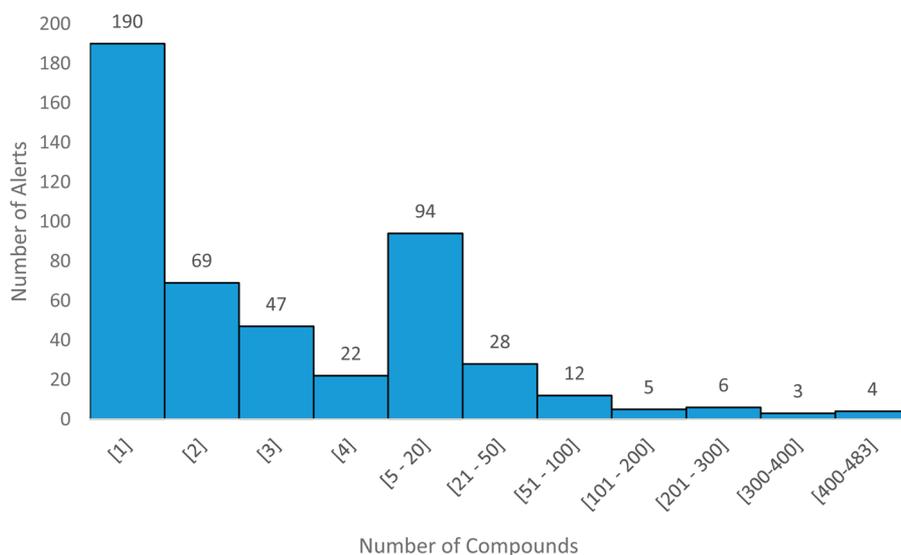
**Figure 1.** Probing the extrapolative power of PAINS alerts. A histogram showing the distribution of the number of PAINS alerts (amounting to 480 total) as the function of the number of compounds used to derive each alert. Note that 190 PAINS alerts were derived from one representative compound only whereas only 18 PAINS alerts were derived from samples including more than 100 compounds per alert.
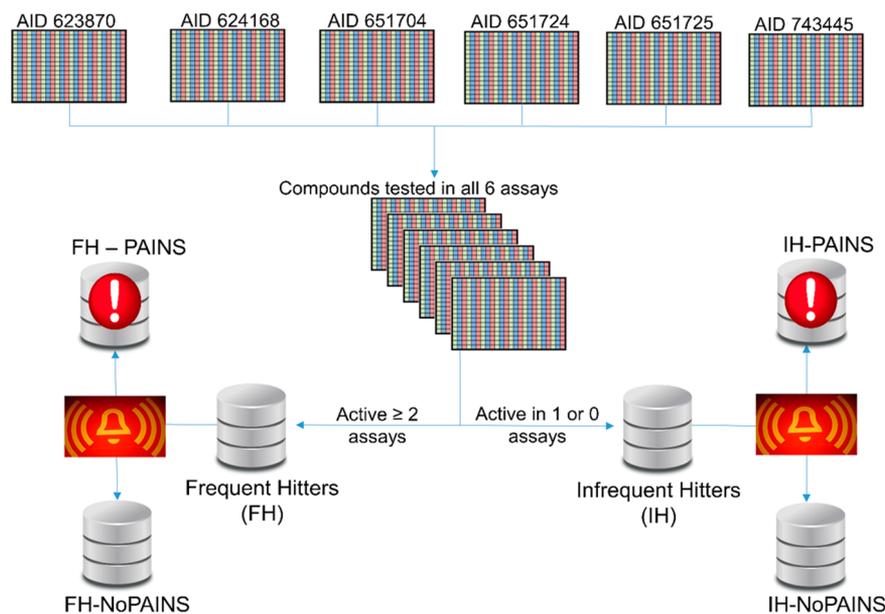


**Figure 2.** Study design for examining compounds in PubChem tested with AlphaScreen assay technology. Six assays targeting PPIs and using AlphaScreen were identified in PubChem. Only those compounds that were tested in all six assays were considered. Compounds were binned into two categories according to the number of active calls. Compounds in each category were then queried for PAINS alerts. The PubChem-wide bioassay activity of all compounds was then investigated.

reported compounds with potential PAINS liability are specifically active and their apparent activity is not an artifact".[15] Thus, compounds with potential PAINS liability are those flagged with PAINS alerts.

Amidst the generally wide acceptance of PAINS, there have been a few voices cautioning about the overarching utility of the alerts. Several authors have noted that the application of these alerts could discard viable drug candidates because such alerts have actually been found in approved drugs.[16,17] More substantial criticism of PAINS alerts has emerged as well on Internet forums (but not in peer-reviewed publications). Aware of these concerns, in the course of our own recent virtual screening investigations, we re-examined the original study[6] from which the 480 PAINS alerts were derived. We noticed

that the study[6] employed a relatively small (93 000 compounds) and proprietary library (complete chemical structures were not released) tested for one type of activity (protein–protein interaction inhibition) in just six HTS campaigns (three out of six targets were kept confidential) using a single detection technology (AlphaScreen).

Though considerable effort was made to divulge as much information as possible, due to the proprietary nature of the original study and unavailability of the chemical library explored therein, the detection of PAINS and the derivation alerts could not be fully and independently reproduced. That being said, upon further inspection of the 92 pages of Supporting Information,[6] we observed that more than half of the PAINS alerts were derived from one or two compounds only (Figure

**Table 1. Lack of Pan-Assay Activity for Compounds with PAINS Alerts in PubChem**[a]

| compound categories | $N_{compounds}$ | luciferase | β-lactamase | fluorescence | all assays |
|---|---|---|---|---|---|
| FH-PAINS[b] | 208 | 12% (93) | 4% (9) | 7% (312) | 10% (546) |
| FH-NoPAINS[b] | 694 | 6% (95) | 2% (9) | 3% (320) | 5% (550) |
| IH-PAINS[b] | 6413 | 3% (93) | 1% (10) | 2% (323) | 2% (550) |
| IH-NoPAINS[b] | 21500 | 1.5% (95) | 0.5% (9) | 1% (326) | 1% (555) |
| Random-PAINS[c] | 14611 | 3% (95) | 1% (12) | 2% (329) | 3% (562) |
| Random-NoPAINS[c] | 58722 | 2% (93) | 0.6% (13) | 0.8% (321) | 1% (550) |
| Drugs-PAINS[c] | 87 | 9% (71) | 7% (40) | 6% (223) | 24% (602) |
| Drug-NoPAINS[c] | 1373 | 5% (59) | 5% (33) | 3% (183) | 15% (458) |

[a]The average fraction of activity calls for PAINS and non-PAINS (defined as containing or lacking PAINS alerts, respectively) across both detection technology-specific assays and all assays in PubChem. The average number of assays in which the compounds were tested are shown in parentheses. [b]Defined by the compound profile in PPI assays utilizing AlphaScreen. [c]Defined by presence or absence of PAINS alerts.

1), with 68% (328 out of the 480 alerts) found in four or fewer compounds only, and more than 30% (190 PAINS alerts) found in one compound only showing "pan-assay" activity (Figure 1; Table S1). This preliminary analysis lead us to hypothesize that the majority of these alerts may have limited extrapolative power.

Given the aforementioned limited sources of PAINS alerts, we decided to probe into the pan-assay activity of PAINS and the reliability of PAINS alerts by analyzing publicly available data on extensively assayed compounds. To this end, we have (i) assessed the robustness of PAINS alerts at flagging frequent hitters among compounds assayed using the AlphaScreen technology as reported in PubChem; (ii) scanned PubChem to investigate the level of pan-assay activity of compounds with and without PAINS alerts; (iii) examined the frequency of PAINS alerts in extensively assayed, yet consistently inactive compounds known as "Dark Chemical Matter";[18] and (iv) profiled the PubChem-wide activity of FDA-approved drugs with and without PAINS alerts. Overall, using publicly available data, this study sought to evaluate the PAINS concept in general, with an additional focus on specific PAINS alerts established in the original investigation,[6] in order to provide both researchers and journal editors with insight into the utility of PAINS alerts as they currently stand.

## 2. RESULTS AND DISCUSSION

**Detection of PAINS in Chemical Libraries Tested with the AlphaScreen Technology.** We have identified six PubChem assays that measure protein−protein interaction (PPI) inhibition using AlphaScreen, i.e., the same type of activity and the same technology employed in the original study.[6] The study design is shown in Figure 2. The six originally studied assays were run at the relatively high compound concentration of 25−50 $\mu$M in primary screens, which may account for the high rate of interference, and two of the assays used hexa-his/Ni anchors.[6] We have chosen these six PubChem assays, similar to a study by Schorpp et al.,[19] in order to assess the robustness of PAINS alerts to flag frequent hitters across a similar, but not identical, series of assays. It should be noted that the anchorage and screening concentrations reported in PubChem were different from original study. However, current PAINS filters look solely for the presence of specific functional groups in assayed chemicals regardless of the assay conditions; thus, the difference in these conditions does not invalidate the use of PAINS alerts in this investigation.

As many as 153 339 unique compounds were found in PubChem to have been tested across all six assays (Table S2). Activity calls for each compound (Active, Inactive, and

Inconclusive) were recorded as defined by the assay depositor. Compounds were then binned into two categories: "Frequent Hitters" (active calls in at least two out of six assays) and "Infrequent Hitters" (active calls in one or zero assays), which is the same threshold as established in the original study.[6] Both categories were first queried for the presence of PAINS substructural alerts using the SMARTS implementation from PubChem Promiscuity,[20] then confirmed using SYBYL Line Notation (SLN) implementation from FAF-Drugs3.[10] Four categories arose: "Frequent Hitters-PAINS" (FH-PAINS), "Frequent Hitters-No PAINS" (FH-NoPAINS), "Infrequent Hitters-PAINS" (IH-PAINS), and "Infrequent Hitters-No PAINS" (IH-NoPAINS). There was a concordance of ∼99.9% between the SMARTS and SLN implementations for flagging compounds with PAINS alerts (Table S3). The enrichment value (EV), which was previously defined[6] as the percentage of compounds active in at least two of the six assays relative to the number of compounds that displayed no activity across all six assays, was calculated to compare FH-PAINS vs IH-PAINS (Table S4 and S5).

The results of our analysis are shown in Table 1. There were 902 compounds in the FH category, and only 208 (23%) of these contained PAINS substructural alerts (FH-PAINS). The remaining 694 FH lacked any PAINS alerts (FH-NoPAINS). For the IH, 146 224 (96%) compounds lacked PAINS alerts (IH-NoPAINS), but 6413 compounds (4%) still contained the alerts (IH-PAINS). Comparing the numbers of IH-PAINS and FH-PAINS leads to the apparent conclusion that, for this series of assays, the majority of compounds containing PAINS alerts (97%) were actually infrequent hitters.

The enrichment value calculated for PAINS-containing compounds (FH-PAINS and IH-PAINS) was only 3.5% (Table S4 and S5). Furthermore, if IH-PAINS that were active in one assay only were taken into consideration, the overall EV fell to 3.2% (Tables S4 and S5). The analysis of this series of assays indicates that PAINS alerts are found much more frequently in nonpromiscuous compounds.

To probe whether or not this observation is only limited to assays related to PPIs using AlphaScreen, we investigated the PubChem-wide bioassay activity of the same compounds. Given that IH-NoPAINS constituted the overwhelming majority of all compounds discussed above (146 224 compounds), the PubChem-wide activity of all IH-NoPAINS was not evaluated due to computational constraints. Instead, a random subset of 21 500 IH-NoPAINS (Table S7) were evaluated; this number was selected to preserve approximately the same ratio of frequent to infrequent NoPAINS (1:3.5) as was observed for the PAINS (cf. Table 1). PubChem
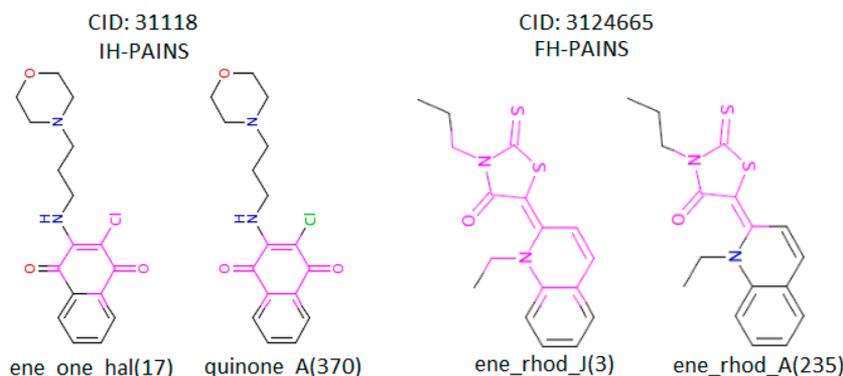
**Figure 3.** Compounds with multiple PAINS alerts. Two representative compounds from the IH-PAINS and FH-PAINS categories that contain multiple PAINS alerts.

Promiscuity[20,21] was used to retrieve the activity calls for all four aforementioned categories of compounds (FH-PAINS, IH-PAINS, FH-NoPAINS, and IH-NoPAINS) tested in luciferase-, $\beta$-lactamase-, and fluorescence-based assays (see Tables S4–S7). Lastly, we assessed activity calls across all bioassays in PubChem irrespective of the detection technology (Table 1).

We found that across all assays, including those that have been reported as particularly susceptible to interference,[5] FH-PAINS were active in more assays than FH-NoPAINS; however, IH-PAINS were active in fewer assays than FH-PAINS (Table 1). The reduced activity of IH-PAINS in the AlphaScreen assays, therefore, is not limited to this detection technology, as it can be observed over all reported assays in PubChem. Also, both categories of frequent hitters in AlphaScreen (FH-PAINS and FH-NoPAINS) showed greater PubChem-wide activity than infrequent hitters containing PAINS alerts (IH-PAINS). Therefore, the broader activity spectrum of these frequent hitters is independent of the presence or absence of any PAINS alerts, highlighting the importance of considering molecular entities as a whole rather than chemical fragments when trying to derive any structural rules governing assay promiscuity.

**Analysis of PAINS Alerts in Chemical Libraries Tested with the AlphaScreen Technology.** The specific alerts found in the above FH-PAINS and IH-PAINS categories were then investigated on an individual basis. In total, 163 individual PAINS alert types were observed in compounds among the two categories (Table S8). It should be noted that multiple PAINS alerts could be present within a single compound (Figure 3).

For the 208 FH-PAINS compounds, 41 individual PAINS alerts were detected (Table S8). Of these 41 alerts, only 7 alerts, i.e., quinone_A(370), mannich_A(296), ene_six_het_A(483), anil_di_alk_B(251), anil_di_alk_A(478), ene_one_hal(17), and imine_one_A(321), were found in more than 10 FH-PAINS compounds. The remaining 34 PAINS alerts were present in 10 or less FH-PAINS compounds.

For the 6314 IH-PAINS compounds, 162 individual PAINS alerts were detected (Table S8). Of these 162 alerts, 57 alerts were found in more than 10 IH-PAINS compounds. Moreover, 15 of these alerts were found in more than 100 IH-PAINS compounds. The anil_di_alk_A(478) alert, for example, appeared in 1083 IH-PAINS compounds.

Next, the PAINS alerts that were present in both the FH-PAINS and IH-PAINS were analyzed. Within these two categories, 40 individual PAINS alerts were shared, roughly ~25% of all observed alerts. Only one alert, i.e., anil_no_alk_A(1), was unique to FH-PAINS (Table S8); however, only 1 compound possessed this alert, which is consistent with the limited sample size (1 compound) used to derive this alert (cf. Figure 1). Similarly, 122 alerts were unique to IH-PAINS (Table S8). For this series of assays, ~75% PAINS alerts present in the PubChem library analyzed herein (122 out of 163) were found only in IH.

The enrichment value (EV), defined as percentage of compounds active in at least two of the six assays relative to the number of compounds active in 1 or 0 assays, was calculated for each of 40 shared PAINS alerts (Table 2). Only 6 alerts showed EVs greater or equal to 25%. However, 4 of these 6 alerts had less than 10 representative compounds. Therefore, 2 alerts, i.e., quinone_A(370) and quinone_D(2), were found in 10 or more compounds and had an EV greater or equal to 25%. The remaining 34 shared alerts had EVs less than 25%, and 32 of these 34 alerts had more than 10 representative compounds. Indeed, 6 shared alerts had EVs less than 1.0% despite being present in more than 100 compounds. For this series of assays, the vast majority of PAINS alerts were found among the IH-PAINS at much higher frequencies. The full analysis of all 40 shared alerts, including representative compound sizes and EVs, can be found in Table 2.

**Random PAINS in PubChem.** We also evaluated the PubChem-wide activity of compounds tested in at least 25 separate bioassays based only on the presence or absence of 480 originally established PAINS alerts, i.e., irrespective of any perceived promiscuity across a selected series of specific assays. Randomly selected compounds that were evaluated in the previous section were excluded. The resultant data set contained 73 333 individual compounds. The structures of these compounds were searched for PAINS alerts (described above) and binned into two categories: Random-PAINS (14 611 compounds) and Random-NoPAINS (58 722 compounds). We compared these two categories following the same protocol as described in the previous section (Table 1). The average pan-assay activity of Random-PAINS was just 3%, compared to an average of 1% for Random-NoPAINS (Table 1), i.e., Random-PAINS were marginally more active than Random-NoPAINS. Additionally, of the 14 611 Random-PAINS only 752 compounds (5% of the total) showed activity in at least 10% of all assays. Of the remaining 13 859 Random-PAINS (95%) that were active in less than 10% of all assays, 1146 had no activity at all, despite being tested in an average of 443 assays (Tables S9 and S10). These results indicate that the mere presence of a PAINS substructure does not give rise to any observed pan-assay activity, nor any marked interference trends in luciferase-, $\beta$-lactamase-, or fluorescence-based assays.

## Table 2. PAINS Enrichment in Six PubChem Assays Employing AlphaScreen[a]

| PAINS Alert | Substructure[b] | $N_{FH-PAINS}$ | $N_{IH-PAINS}$ | $N_{PAINS}$ | EV, % |
|---|---|---|---|---|---|
| quinone_B(5) | | 3 | 1 | 4 | 300.0 |
| imine_ene_A(5) | | 3 | 2 | 5 | 150.0 |
| het_65_Db(5) | | 4 | 3 | 7 | 133.3 |
| ene_rhod_J(3) | | 1 | 3 | 4 | 33.3 |
| quinone_A(370) | | 47 | 160 | 207 | 29.4 |
| quinone_D(2) | | 9 | 36 | 45 | 25.0 |
| dyes5A(27) | | 1 | 5 | 6 | 20.0 |
| anthranil_one_A(38) | | 3 | 16 | 19 | 18.8 |
| imine_one_sixes(27) | | 3 | 16 | 19 | 18.8 |
| het_pyridiniums_A(39) | | 5 | 29 | 34 | 17.2 |
| anil_alk_ene(51) | | 2 | 12 | 14 | 16.7 |
| thio_urea_D(8) | | 1 | 6 | 7 | 16.7 |
| ene_one_hal(17) | | 14 | 89 | 103 | 15.7 |
| anil_di_alk_B(251) | | 16 | 116 | 132 | 13.8 |
| ene_five_het_A(201) | | 7 | 51 | 58 | 13.7 |
| het_thio_5_imine_A(1) | | 1 | 9 | 10 | 11.1 |
| rhod_sat_A(33) | | 4 | 39 | 43 | 10.3 |
| dyes3A(19) | | 1 | 11 | 12 | 9.1 |
| sulfonamide_B(41) | | 2 | 25 | 27 | 8.0 |
| azo_A(324) | | 9 | 114 | 123 | 7.9 |
| mannich_A(296) | | 36 | 472 | 508 | 7.6 |
| anil_di_alk_F(14) | | 1 | 14 | 15 | 7.1 |
| imine_one_A(321) | | 12 | 215 | 227 | 5.6 |
| ene_one_ene_A(57) | | 2 | 40 | 42 | 5.0 |
| anil_no_alk(40) | | 2 | 48 | 50 | 4.2 |
| cyano_imine_A(37) | | 1 | 26 | 27 | 3.9 |
| ene_six_het_A(483) | | 18 | 556 | 574 | 3.2 |
| anil_di_alk_D(198) | | 5 | 155 | 160 | 3.2 |
| imine_one_fives(89) | | 1 | 38 | 39 | 2.6 |
| hzone_pipzn(79) | | 2 | 87 | 89 | 2.3 |
| anil_di_alk_E(186) | | 3 | 148 | 151 | 2.0 |
| thiophene_amino_Ab(40) | | 1 | 71 | 72 | 1.4 |
| catechol_A(92) | | 1 | 75 | 76 | 1.3 |
| anil_di_alk_A(478) | | 14 | 1083 | 1097 | 1.3 |
| imine_one_isatin(189) | | 1 | 111 | 112 | 0.90 |
| pyrrole_A(118) | | 2 | 269 | 271 | 0.74 |
| ene_rhod_A(235) | | 4 | 593 | 597 | 0.67 |
| anil_di_alk_C(246) | | 4 | 641 | 645 | 0.62 |
| ene_five_het_B(90) | | 1 | 161 | 162 | 0.62 |
| indol_3yl_alk(461) | | 1 | 354 | 355 | 0.28 |

[a]Forty alerts were present in both FH-PAINS and IH-PAINS. Two alerts showed EVs greater or equal to 25% and were found in 10 or more total compounds. Six alerts had EVs below 1.0%.

In fact, only two PAINS alert containing compounds, tanespimycin and dihydrexidine, were active in more than 50% of the assays (Figure 4). In total 202 PAINS alerts were
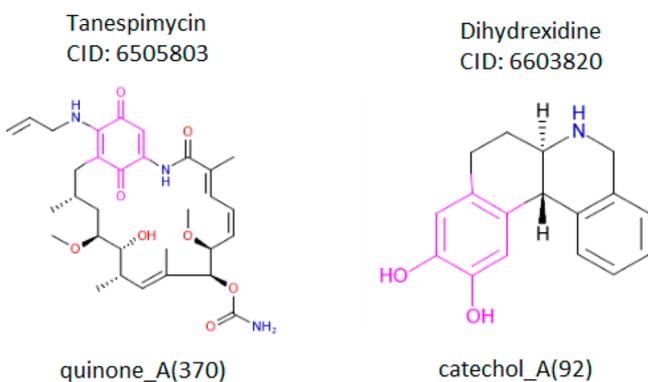


**Figure 4.** Random-PAINS displaying pan-assay activity. Tanespimycin and dihydrexidine are active in 85% and 50% of all assays in PubChem, respectively.

found among the Random-PAINS category. A PubChem-wide analysis of alerts in random PAINS is described in the Global Analysis of PAINS Alerts section.

**Analysis of PAINS Alerts in Dark Chemical Matter.** Following the observation that Random-PAINS can be consistently inactive across a large number of assays, we probed the so-called Dark Chemical Matter (DCM)[18] for the presence of PAINS alerts. DCM was defined by Wassermann et al.[18] as compounds that have not yet shown any activity when tested in a minimum of 100 assays. The complete data set of 139 352 DCM compounds, i.e. 128 997 PubChem and 10 355 Novartis DCM compounds, was downloaded from the Supporting Information of the respective study.[18]

The data set was examined with FAF3-Drugs,[10] and 3570 DCM compounds containing PAINS substructures were found, encompassing 109 of the 480 original PAINS alerts[6] (Table S11). Of these 109 PAINS alerts, 30 alerts were found in more than 10 compounds and 10 alerts were present in 100 or more compounds (Table 3). This analysis shows that even extensively assayed compounds containing PAINS alerts may be consistently inactive.

**Global Analysis of PAINS Alerts.** In this study, a total of 24 802 PAINS compounds were analyzed (208 FH-PAINS, 6413 IH-PAINS, 14 611 Random-PAINS, and 3570 DCM-PAINS) covering 220 specific PAINS alert types, which is ~46% of the original PAINS alerts.[6] In order to determine if specific PAINS alerts correspond to compounds with elevated assay promiscuity, we performed a global analysis of the PubChem-wide activity of all PAINS compounds investigated herein (Tables S12–S15). Since there is no agreed upon threshold of pan-assay activity, we selected assay activity of at least 10% as an arbitrary classifier. The global assay activity associated with all alerts can be found in the SI.

Of these 220 PAINS alert types, 32 alerts had greater than 10% assay activity in either all assays, luciferase-, $\beta$-lactamase-, or fluorescence-based assays (Table S12). However, only 12 of these alerts were present in more than 10 compounds (Table 4). It should be noted, however, that six of these alerts can also be found in DCM.

On the other hand, 176 (~80%) of the total PAINS alerts analyzed (220) were active in less than 10% of all investigated assays and technologies and 88 alerts were present in DCM

**Table 3. PAINS alerts enriched in Dark Chemical Matter[a]**

| PAINS Alert | Substructure[6] | $N_{DCM}$ |
|---|---|---|
| anil_di_alk_A(478) | | 902 |
| anil_di_alk_C(246) | | 492 |
| indol_3yl_alk(461) | | 343 |
| ene_six_het_A(483) | | 256 |
| mannich_A(296) | | 212 |
| imine_one_A(321) | | 193 |
| anil_di_alk_D(198) | | 184 |
| anil_di_alk_E(186) | | 164 |
| ene_rhod_A(235) | | 116 |
| pyrrole_A(118) | | 100 |

[a]Ten alerts are present in 100 or more in Dark Chemical Matter compounds ($N_{DCM}$).

(Tables S13 and S14). Eighty-four of 176 alerts were present in more than 10 compounds (Table S13). Interestingly, 6 of these alerts were found in more than 1000 compounds (Table 5). Finally, 12 alerts were found exclusively in DCM-PAINS (Table S15). Eleven of these alerts were found in less than 10 compounds, while 1 alert, i.e., hzone_phenol_B(215), was present in exactly 10 compounds.

There are 16 PAINS alerts that were derived from more than 150 compounds (cf. Figure 1). Of all 480 alerts, these 16 alerts were created from the most underlying data in the original study.[6] Given the prevalence and heightened promiscuity of compounds possessing these alerts in the original study, we specifically investigated whether any compounds in our collection flagged by these 16 alerts display suspect assay trends (Table 6). Aside from hzone_phenol_A(479) and

**Table 4. PAINS Alerts Displaying Elevated Assay Activity in PubChem[a]**

| PAINS Alert | Substructure[6] | $N_{PubChem}$ | $N_{DCM}$ | Lucifer. | β-lactamase | Fluoresc. | All Assays |
|---|---|---|---|---|---|---|---|
| quinone_A(370) |  | 653 | 15 | **10.1%** (91) | 4.9% (13) | 6.5% (312) | 8.4% (543) |
| quinone_D(2) |  | 170 | 5 | **11.8%** (82) | 2.0% (12) | 6.5% (277) | 8.4% (487) |
| ene_one_ene_A (57) |  | 140 | 3 | **16.3%** (93) | 9.2% (11) | 5.1% (308) | 8.2% (539) |
| het_pyridiniums_A (39) |  | 121 | 35 | **11.0%** (96) | 6.3% (11) | 5.4% (322) | 7.0% (557) |
| ene_five_het_E(44) |  | 55 | N/A | **12.5%** (107) | 1.1% (15) | 3.6% (372) | 4.9% (629) |
| dyes3A(19) |  | 45 | N/A | **21.0%** (107) | 5.3% (10) | 8.7% (333) | **12.3%** (582) |
| dyes5A(27) |  | 44 | 5 | **10.1%** (75) | 7.5% (11) | 5.9% (245) | 7.7% (430) |
| styrene_B(8) |  | 23 | N/A | 5.9% (73) | **10.2%** (11) | 5.4% (276) | 6.4% (471) |
| quinone_B(5) |  | 14 | N/A | 5.0% (93) | **10.1%** (13) | **15.5%** (343) | **12.6%** (571) |
| hzone_thiophene_A (11) |  | 14 | N/A | **10.8%** (99) | 4.8% (10) | 2.9% (331) | 4.1% (562) |
| keto_phenone_A (11) |  | 13 | 2 | **19.1%** (112) | 1.9% (12) | 2.1% (356) | 5.9% (630) |
| het_thio_676_A (10) |  | 11 | N/A | 7.1% (53) | **16.1%** (8) | 4.5% (204) | 7.0% (354) |
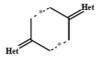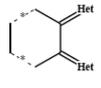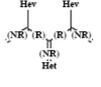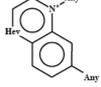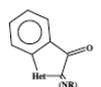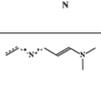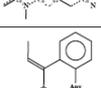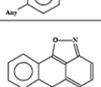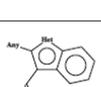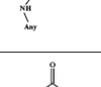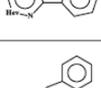
[a]Twelve alerts have more than 10 compounds ($N_{PubChem}$) and greater than 10% assay activity in either all assays, luciferase-, β-lactamase-, or fluorescence-based assays (bold). Six alerts were also present in DCM ($N_{DCM}$). The average numbers of assays in which the compounds were tested are shown in parentheses.

hzone_phenol_B(215), which were found exclusively in DCM, 14 of these 16 alerts were frequently assayed and abundantly present in the public collection. All 14 alerts displayed less than 10% activity in all PubChem assays despite being tested on average in more than 500 assays each. Among these specific alerts, the quinone_A(370) alert demonstrated the highest activity in all assays (8.4%).

Our findings using data in the public domain can be corroborated in part by other inquiries into the nature of promiscuous compounds. For instance, while attempting to use PAINS alerts to fill gaps in Eli Lilly's promiscuity filters, Bruns and Watson observed that "PAINS queries matched 286 promiscuous compounds that passed the Lilly rules, compared to 3986 in the non-promiscuous set, for an enrichment factor of 4.0".[22] Furthermore, they noted that "although 67 PAINS queries matched at least one promiscuous compound, only nine queries matched at least five promiscuous compounds and had an enrichment of at least 5."[22] These findings are consistent with our observations that PAINS alerts in public data frequently flag nonpromiscuous compounds or are manifested in only a small number of promiscuous compounds.

On the other hand, another study on frequent hitter behavior by researchers at AstraZeneca showed elevated "Frequent-hitter Incidence %" for 10 out of 15 PAINS alerts.[23] Although the

**Table 5. PAINS Alerts Not Displaying Elevated Assay Activity in PubChem[a]**

| PAINS Alert | Substruct.[6] | $N_{PubChem}$ | $N_{DCM}$ | Lucifer. | β-lactamase | Fluoresc. | All Assays |
|---|---|---|---|---|---|---|---|
| anil_di_alk_A (478) |  | 3695 | 902 | 2.2% (92) | 1.0% (10) | 1.0% (318) | 1.4% (545) |
| anil_di_alk_C (246) |  | 2014 | 492 | 0.8% (88) | 0.6% (10) | 0.6% (301) | 0.9% (516) |
| ene_six_het_A (483) |  | 1889 | 256 | 1.5% (99) | 0.8% (12) | 2.2% (351) | 2.1% (594) |
| ene_rhod_A (235) |  | 1706 | 116 | 3.0% (92) | 1.4% (9) | 3.0% (318) | 3.0% (542) |
| mannich_A (296) |  | 1435 | 212 | 4.3% (96) | 1.2% (11) | 1.9% (329) | 2.8% (563) |
| indol_3yl_alk (461) |  | 1097 | 343 | 1.9% (95) | 0.7% (10) | 0.7% (326) | 1.2% (557) |
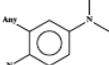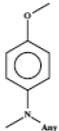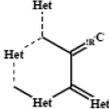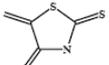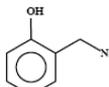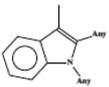
[a]These 6 alerts are present in more than 1000 compounds ($N_{PubChem}$) and have less than 10% assay activity in either all assays, luciferase-, β-lactamase-, or fluorescence-based assays. All six alerts were also present in DCM ($N_{DCM}$). The average numbers of assays in which the compounds were tested are shown in parentheses.

authors state that their "corporate data largely confirm previous observations of the PAINS classes", this study only investigated part of the first tier of the 480 PAINS alerts, i.e., the 15 out of 16 alerts derived from more than 150 compounds (cf. Figure 1), or ~3% of all alerts.[23] As can be seen, in that study, one-third of the profiled alerts did not show elevated frequent-hitter behavior, which is, in part, aligned with our general observations (Tables 5 and 6).

**PAINS Alerts in Drugs.** Other groups have noted that many drugs contain PAINS alerts,[6,16,17] and several careful and keen analyses have centered around this phenomenon.[24] It has also been observed that many of these PAINS alerts in drugs (but not all) map to poor ADMETox properties, such as quinone-containing drugs.[6] While this is an interesting observation, we view interference propensity and poor ADMETox properties as separate phenomena. Our group[25] as well as others[19] have also shown that a great majority of toxicity structural alerts, much akin to PAINS alerts, are overly sensitive and not predictive of actual in vitro or in vivo toxicity. Given that drug repurposing is currently widely used as a boon to traditional drug discovery,[26,27] we profiled the PubChem-wide bioassay activity of drugs with and without PAINS alerts (Table 1).

A list of 1460 approved small-molecule drugs was compiled from Drugs@FDA (https://www.accessdata.fda.gov/scripts/cder/drugsatfda/). Structures for these drugs were searched for PAINS alerts.[10,20,21] We identified 87 small-molecule approved drugs possessing 25 individual PAINS alerts (Table S16). As observed in the preceding sections, Drugs-PAINS are more active than Drugs-NoPAINS, having activity in 24% and 15% of all bioassays in PubChem, respectively (Table 1). According to current filters,[10,20] 16 of these drugs possess quinone PAINS alerts. The promiscuity of quinone-containing drugs have been extensively discussed in the PAINS literature[6,24] and is supported by our analysis (cf. Tables 2 and 4). For instance, the chemotherapeutic doxorubicin, which contains the quinone_A(370) alert, has been tested in more than 4000 assays with active calls ~85% of the time.

At the same time, however, the relationship between polypharmacology and PAINS has not yet been adequately explored. Many drugs show polypharmacological behavior and possibly derive their efficacy from interacting with multiple targets.[28] Indeed, a similar study on promiscuity in extensively assayed compounds found that drugs are more promiscuous than bioactive compounds,[28] which is evidenced in our analysis as well (Table 1). Polypharmacology may well account for the increased activity of both Drugs-PAINS and Drugs-NoPAINS relative to the other categories (cf. Table 1). While the phenomena of assay interference and polypharmacology have rightfully been contrasted,[28] there is very real possibility that a compound may both possess PAINS alerts and display polypharmacological behavior. Given that PAINS-containing drugs are now frequently used in drug repurposing screens, a larger discussion about the utility of PAINS alerts and polypharmacology should take place.

**Table 6. Global Assay Activity of Compounds in PubChem Possessing the Top 16 PAINS Alerts[a]**

| PAINS Alert | Substruct.[6] | $N_{PubChem}$ | $N_{DCM}$ | Lucifer | β-lactamase | Fluoresc. | All Assays |
|---|---|---|---|---|---|---|---|
| ene_six_het_A (483) |  | 1889 | 256 | 1.5% (99) | 0.8% (12) | 2.2% (351) | 2.1% (594) |
| hzone_phenol_A (479) |  | N/A | 2 | N/A | N/A | N/A | N/A |
| anil_di_alk_A (478) |  | 3695 | 902 | 2.2% (92) | 1.0% (10) | 1.0% (318) | 1.4% (545) |
| indol_3yl_alk (461) |  | 1097 | 343 | 1.9% (95) | 0.7% (10) | 0.7% (326) | 1.2% (557) |
| quinone_A (370) |  | 653 | 15 | 10.1% (91) | 4.9% (13) | 6.5% (312) | 8.4% (543) |
| azo_A (324) |  | 365 | 51 | 6.7% (92) | 1.4% (13) | 1.8% (310) | 2.9% (534) |
| imine_one_A (321) |  | 838 | 193 | 3.2% (97) | 1.0% (13) | 1.7% (343) | 2.3% (583) |
| mannich_A (296) |  | 1435 | 212 | 4.3% (96) | 1.2% (11) | 1.9% (329) | 2.8% (563) |
| anil_di_alk_B (251) |  | 450 | 25 | 6.5% (91) | 8.2% (11) | 5.3%(318) | 5.3% (541) |
| anil_di_alk_C (246) |  | 2014 | 492 | 0.8% (88) | 0.6% (10) | 0.6% (301) | 0.9% (516) |
| ene_rhod_A (235) |  | 1706 | 116 | 3.0% (92) | 1.4% (9) | 3.0% (318) | 3.0% (542) |
| hzone_phenol_B (215) |  | N/A | 10 | N/A | N/A | N/A | N/A |
| ene_five_het_A(201) |  | 193 | 16 | 3.2% (99) | 0.7% (11) | 2.6% (347) | 3.0% (590) |
| anil_di_alk_D (198) |  | 631 | 184 | 1.7% (86) | 0.4% (9) | 0.7% (294) | 1.2% (503) |
| imine_one_isatin(189) |  | 324 | 41 | 3.4% (99) | 0.3% (11) | 1.4% (346) | 1.9% (588) |
| anil_di_alk_E (186) |  | 589 | 164 | 1.6% (85) | 0.2% (9) | 0.8% (295) | 1.3% (504) |

[a]The bioassay activities of 14 out of 16 alerts have been profiled with more compounds than were used to derive the alerts originally.[6] Two alerts were found only in DCM ($N_{DCM}$). The average numbers of assays in which the compounds were tested are shown in parentheses.

**Beyond PAINS Substructures.** PAINS concept has been widely accepted by many experienced medicinal chemists both in academia and the pharmaceutical industry. Indeed, the original study from which the PAINS alerts were derived and the impetus behind it are an important step toward reproducibility and the appropriate use of resources in drug

discovery. However, our findings based on the analysis of public data suggest that many compounds containing PAINS alerts do not actually show high assay promiscuity, leading to the conclusion that these alerts should not be blindly used, in the absence of orthogonal experimental assays, to deprioritize a compound.

At the same time, it is undeniable that pan-assay interference *compounds* exist and care must be taken to avoid these compounds. Moreover, we recognize that true "PAINS" may be present in the data analyzed herein but have not been classified as such because the current alerts do not cover these compounds.[29] The issue of what constitutes a pan-assay interference compound thus remains unclear. For example, in *How to Triage PAINS-Full Research*, Dahlin and Walters define PAINS as "compounds that are recognized by the substructure filters reported by the Baell and Holloway article."[29] By this definition, all alerts (filters) are treated equally, regardless of the underlying data used to derive the alert or the actual promiscuity of flagged compounds. Yet our analysis indicates that the identification of such compounds should not be restricted to substructures alone. Substructural alerts, PAINS or otherwise, do not take into consideration the whole molecular environment,[25] as illustrated by PAINS alerts manifesting in both promiscuous and frequently inactive compounds (DCM). Attempts should then be made to move beyond substructural or fragment-based alerts. For instance, Yang and co-workers in their "BadApple" algorithm have extended the identification of promiscuous compound to larger scaffolds.[30]

In recent publications by Alves et al.,[31−33] quantitative structure−activity relationship (QSAR) models were used in conjunction with structural alerts for toxicity to dramatically improve the accuracy of prediction of multiple toxicity end points over alerts alone. The authors of the present study advocate the development of a similar approach for PAINS alerts. Such publicly accessible models, if successful, could be employed even for proprietary compounds insofar as chemical descriptors of PAINS alert-containing compounds could be shared without divulging actual molecular structures (given the proprietary nature of most compounds used to derive and evaluate PAINS so far[6,22,23]). The challenge is to build externally predictive QSAR models capable of classifying PAINS versus non-PAINS compounds. Using such models, predictions of suspect compounds could be made, giving higher confidence in the utility of the alert and the nefarious nature of the compound.

Meanwhile, the concept of PAINS alerts, at the very least, needs a redefined set of "best practices"[29] that covers the appropriate use of alerts, which may include cross-referencing the promiscuity profiles of structurally similar compounds or alert types in the public domain, annotation of particularly susceptible assays, targets, and conditions, pointers to the appropriate controls, and a generally agreed upon definition of pan-assay activity. It would be of great value if a community-wide effort to screen and analyze a large set of commercially available compounds representing all current PAINS alerts against multiple targets in various assays was performed by several independent groups.

## 3. CONCLUSIONS

It is imperative to establish target selectivity for any compound considered a viable chemical probe or drug candidate through rigorously acquired experimental data and meaningful SAR. Future studies may well establish some generalized approach

for detecting frequent hitters engendered by assay interference. However, until such approaches are developed and rigorously validated across a large number of molecules, researchers should be cautioned about using the current PAINS alerts as reliable indicators of nonspecific pan assay interference. Though it has been stated elsewhere that compounds flagged with PAINS alerts are not active in all assays or against all targets,[6,29] our analysis provides systematic and data-driven support of this claim across a large series of compounds, alerts, and assays. Our findings do demonstrate, with publicly available data at hand, that majority of the original PAINS alerts are not indicative of pan-assay compound promiscuity, that many compounds without PAINS alerts are as, if not more, promiscuous as those with the alerts, and that many compounds flagged by PAINS alerts show no activity. It is of great importance that reviewers and journal editors request experimental proofs of selectivity, such as orthogonal experimental assays, for hit and lead compounds reported in scientific manuscripts. However, the results of this study strongly suggest that such requests should not be based solely on the results of PAINS filters.[29]

## ■ ASSOCIATED CONTENT

**Ⓢ Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.6b00465.

Data are provided in Tables S1−S16 (ZIP)

## ■ AUTHOR INFORMATION

**Corresponding Author**

*E-mail: alex_tropsha@unc.edu.

**ORCID** Ⓞ

Alexander Tropsha: 0000-0003-3802-8896

**Notes**

The authors declare no competing financial interest.

## ■ REFERENCES

(1) Collins, F. S.; Tabak, L. A. Policy: NIH Plans to Enhance Reproducibility. *Nature* **2014**, *505* (7485), 612−613.

(2) Frye, S. V.; Arkin, M. R.; Arrowsmith, C. H.; Conn, P. J.; Glicksman, M. A.; Hull-Ryde, E. A.; Slusher, B. S. Tackling Reproducibility in Academic Preclinical Drug Discovery. *Nat. Rev. Drug Discovery* **2015**, *14*, 733−734.

(3) Dahlin, J. L.; Baell, J.; Walters, M. A. Assay Interference by Chemical Reactivity. *Assay Guidance Manual*; Eli Lilly & Company and the National Center for Advancing Translational Sciences: Bethesda, MD, 2015.

(4) Sassano, M. F.; Doak, A. K.; Roth, B. L.; Shoichet, B. K. Colloidal Aggregation Causes Inhibition of G Protein-Coupled Receptors. *J. Med. Chem.* **2013**, *56* (6), 2406−2414.

(5) Thorne, N.; Auld, D. S.; Inglese, J. Apparent Activity in High-Throughput Screening: Origins of Compound-Dependent Assay Interference. *Curr. Opin. Chem. Biol.* **2010**, *14* (3), 315−324.

(6) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from

Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719−2740.

(7) Dahlin, J. L.; Nissink, J. W. M.; Strasser, J. M.; Francis, S.; Higgins, L.; Zhou, H.; Zhang, Z.; Walters, M. A. PAINS in the Assay: Chemical Mechanisms of Assay Interference and Promiscuous Enzymatic Inhibition Observed during a Sulfhydryl-Scavenging HTS. *J. Med. Chem.* **2015**, *58* (5), 2091−2113.

(8) Erlanson, D. A. Learning from PAINful Lessons. *J. Med. Chem.* **2015**, *58* (5), 2088−2090.

(9) Mok, N. Y.; Maxe, S.; Brenk, R. Locating Sweet Spots for Screening Hits and Evaluating Pan-Assay Interference Filters from the Performance Analysis of Two Lead-like Libraries. *J. Chem. Inf. Model.* **2013**, *53* (3), 534−544.

(10) Lagorce, D.; Sperandio, O.; Baell, J. B.; Miteva, M. A.; Villoutreix, B. O. FAF-Drugs3: A Web Server for Compound Property Calculation and Chemical Library Design. *Nucleic Acids Res.* **2015**, *43* (W1), W200−7.

(11) Sterling, T.; Irwin, J. J. ZINC 15 − Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55* (11), 2324−2337.

(12) Baell, J. B. Screening-Based Translation of Public Research Encounters Painful Problems. *ACS Med. Chem. Lett.* **2015**, *6* (3), 229−234.

(13) Neves, B. J.; Dantas, R. F.; Senger, M. R.; Melo-Filho, C. C.; Valente, W. C. G.; de Almeida, A. C. M.; Rezende-Neto, J. M.; Lima, E. F. C.; Paveley, R.; Furnham, N.; Muratov, E.; Kamentsky, L.; Carpenter, A. E.; Braga, R. C.; Silva-Junior, F. P.; Andrade, C. H. Discovery of New Anti-Schistosomal Hits by Integration of QSAR-Based Virtual Screening and High Content Screening. *J. Med. Chem.* **2016**, *59* (15), 7075−7088.

(14) Williamson, A. E.; Ylioja, P. M.; Robertson, M. N.; Antonova-Koch, Y.; Avery, V.; Baell, J. B.; Batchu, H.; Batra, S.; Burrows, J. N.; Bhattacharyya, S.; Calderon, F.; Charman, S. A.; Clark, J.; Crespo, B.; Dean, M.; Debbert, S. L.; Delves, M.; Dennis, A. S. M.; Deroose, F.; Duffy, S.; Fletcher, S.; Giaever, G.; Hallyburton, I.; Gamo, F.-J.; Gebbia, M.; Guy, R. K.; Hungerford, Z.; Kirk, K.; Lafuente-Monasterio, M. J.; Lee, A.; Meister, S.; Nislow, C.; Overington, J. P.; Papadatos, G.; Patiny, L.; Pham, J.; Ralph, S. A.; Ruecker, A.; Ryan, E.; Southan, C.; Srivastava, K.; Swain, C.; Tarnowski, M. J.; Thomson, P.; Turner, P.; Wallace, I. M.; Wells, T. N. C.; White, K.; White, L.; Willis, P.; Winzeler, E. A.; Wittlin, S.; Todd, M. H. Open Source Drug Discovery: Highly Potent Antimalarial Compounds Derived from the Tres Cantos Arylpyrroles. *ACS Cent. Sci.* **2016**, *2* (10), 687−701.

(15) ACS Publications. Guidelines for Authors. *J. Med. Chem.* **2016**, No. 1, 5−6.

(16) Chai, C. L.; Mátyus, P. One Size Does Not Fit All: Challenging Some Dogmas and Taboos in Drug Discovery. *Future Med. Chem.* **2016**, *8* (1), 29−38.

(17) Senger, M. R.; Fraga, C. A. M.; Dantas, R. F.; Silva, F. P. Filtering Promiscuous Compounds in Early Drug Discovery: Is It a Good Idea? *Drug Discovery Today* **2016**, *21* (6), 868−872.

(18) Wassermann, A. M.; Lounkine, E.; Hoepfner, D.; Le Goff, G.; King, F. J.; Studer, C.; Peltier, J. M.; Grippo, M. L.; Prindle, V.; Tao, J.; Schuffenhauer, A.; Wallace, I. M.; Chen, S.; Krastel, P.; Cobos-Correa, A.; Parker, C. N.; Davies, J. W.; Glick, M. Dark Chemical Matter as a Promising Starting Point for Drug Lead Discovery. *Nat. Chem. Biol.* **2015**, *11* (12), 958−966.

(19) Schorpp, K.; Rothenaigner, I.; Salmina, E.; Reinshagen, J.; Low, T.; Brenke, J. K.; Gopalakrishnan, J.; Tetko, I. V.; Gul, S.; Hadian, K. Identification of Small-Molecule Frequent Hitters from AlphaScreen High-Throughput Screens. *J. Biomol. Screening* **2014**, *19* (5), 715−726.

(20) Canny, S. A.; Cruz, Y.; Southern, M. R.; Griffin, P. R. PubChem Promiscuity: A Web Resource for Gathering Compound Promiscuity Data from PubChem. *Bioinformatics* **2012**, *28* (1), 140−141.

(21) Schurer, S.; Vempati, U.; Smith, R.; Southern, M.; Lemmon, V. BioAssay Ontology Annotations Facilitate Cross-Analysis of Diverse High-Throughput Screening Data Sets. *J. Biomol. Screening* **2011**, *16* (4), 415−426.

(22) Bruns, R. F.; Watson, I. A. Rules for Identifying Potentially Reactive or Promiscuous Compounds. *J. Med. Chem.* **2012**, *55* (22), 9763−9772.

(23) M Nissink, J. W.; Blackburn, S. Quantification of Frequent-Hitter Behavior Based on Historical High-Throughput Screening Data. *Future Med. Chem.* **2014**, *6* (10), 1113−1126.

(24) Baell, J. B. Feeling Nature's PAINS: Natural Products, Natural Product Drugs, and Pan Assay Interference Compounds (PAINS). *J. Nat. Prod.* **2015**, *79*, 616.

(25) Alves, V. M.; Muratov, E. N.; Capuzzi, S. J.; Politi, R.; Low, Y.; Braga, R. C.; Zakharov, A. V.; Sedykh, A.; Mokshyna, E.; Farag, S.; Andrade, C. H.; Kuz'min, V. E.; Fourches, D.; Tropsha, A. Alarms about Structural Alerts. *Green Chem.* **2016**, *18* (16), 4348−4360.

(26) Oprea, T. I.; Mestres, J. Drug Repurposing: Far beyond New Targets for Old Drugs. *AAPS J.* **2012**, *14* (4), 759−763.

(27) Blatt, J.; Farag, S.; Corey, S. J.; Sarrimanolis, Z.; Muratov, E.; Fourches, D.; Tropsha, A.; Janzen, W. P. Expanding the Scopre of Drug Repurposing in Pediatrics: The Children's Pharmacy Collaborative. *Drug Discovery Today* **2014**, *19* (11), 1696−1698.

(28) Jasial, S.; Hu, Y.; Bajorath, J. Determining the Degree of Promiscuity of Extensively Assayed Compounds. *PLoS One* **2016**, *11* (4), e0153873.

(29) Dahlin, J. L.; Walters, M. A. How to Triage PAINS-Full Research. *Assay Drug Dev. Technol.* **2016**, *14* (3), 168−174.

(30) Yang, J. J.; Ursu, O.; Lipinski, C. A.; Sklar, L. A.; Oprea, T. I.; Bologa, C. G. Badapple: Promiscuity Patterns from Noisy Evidence. *J. Cheminf.* **2016**, *8* (1), 29.

(31) Alves, V. M.; Muratov, E. N.; Fourches, D.; Strickland, J.; Kleinstreuer, N.; Andrade, C. H.; Tropsha, A. Predicting Chemically-Induced Skin Reactions. Part I: QSAR Models of Skin Sensitization and Their Application to Identify Potentially Hazardous Compounds. *Toxicol. Appl. Pharmacol.* **2015**, *284* (2), 262−272.

(32) Alves, V. M.; Muratov, E. N.; Fourches, D.; Strickland, J.; Kleinstreuer, N.; Andrade, C. H.; Tropsha, A. Predicting Chemically-Induced Skin Reactions. Part II: QSAR Models of Skin Permeability and the Relationships between Skin Permeability and Skin Sensitization. *Toxicol. Appl. Pharmacol.* **2015**, *284* (2), 273−280.

(33) Alves, V. M.; Capuzzi, S. J.; Muratov, E. N.; Tropsha, A.; et al. QSAR Models Can Provide an Alternative to LLNA Testing for Assessing Human Skin Sensitization. *Green Chem.* **2016**, *18*, 6501.