COMMENT

Chemical con artists foil drug discovery

Naivety about promiscuous, assay-duping molecules is polluting the literature and wasting resources, warn **Jonathan Baell** and **Michael A. Walters**.

Academic researchers, drawn into drug discovery without appropriate guidance, are doing muddled science. When biologists identify a protein that contributes to disease, they hunt for chemical compounds that bind to the protein and affect its activity. A typical assay screens many thousands of chemicals. 'Hits' become tools for studying the disease, as well as starting points in the hunt for treatments.

But many hits are artefacts — their activity does not depend on a specific, drug-like interaction between molecule and protein. A true drug inhibits or activates a protein by fitting into a binding site on the protein. Artefacts have subversive reactivity that masquerades as drug-like binding and yields false signals across a variety of assays[1,2].
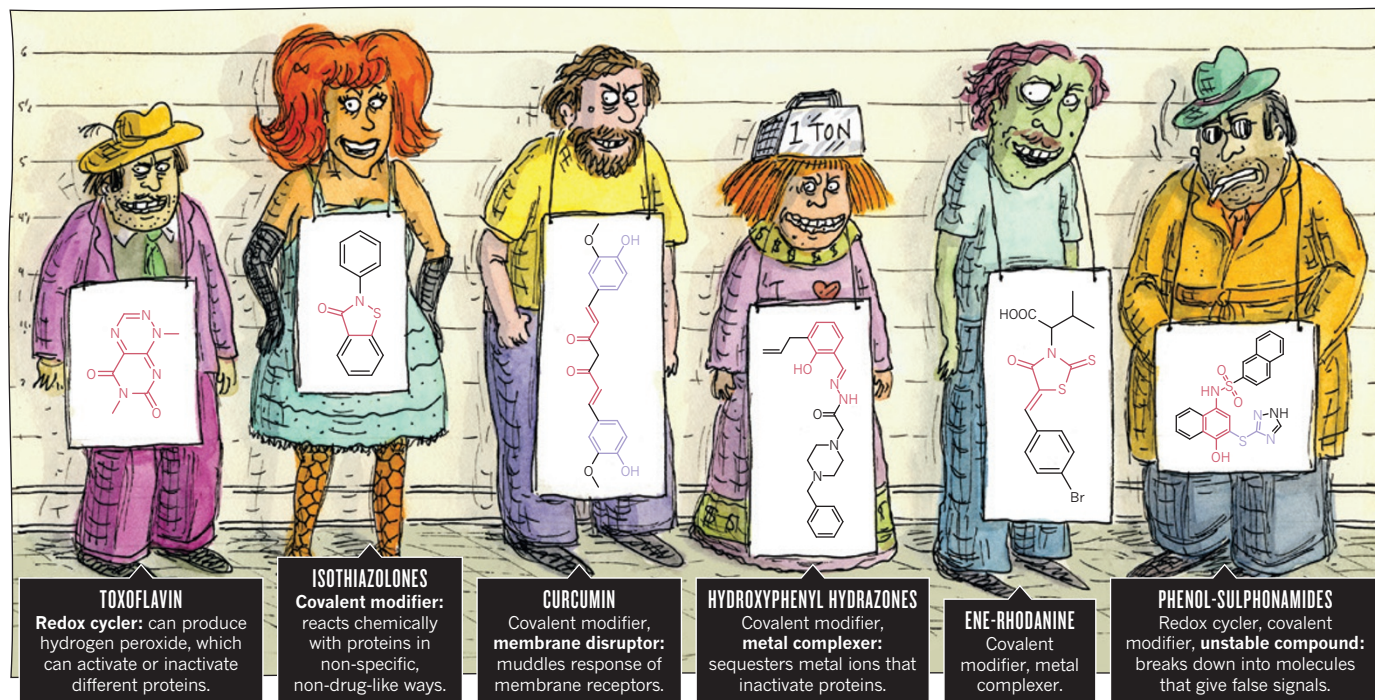
These molecules — pan-assay interference compounds, or PAINS — have defined structures, covering several classes of compound (see 'Worst offenders'). But biologists and inexperienced chemists rarely recognize them. Instead, such compounds are reported as having promising activity against a wide variety of proteins. Time and research money are consequently wasted in attempts to optimize the activity of these compounds. Chemists make multiple analogues of apparent hits hoping to improve the 'fit' between protein and compound. Meanwhile, true hits with real potential are neglected.

Until the past decade or so, screening ▶

WORST OFFENDERS

Pan-assay interference compounds (PAINS) fall into hundreds of chemical classes, but some groups occur much more frequently than others. Among the most insidious are the eight shown here (reactive portions shown in red and purple). These and related compounds should set off alarm bells if they show up as 'hits' in drug screens.

TOXOFLAVIN
Redox cycler: can produce hydrogen peroxide, which can activate or inactivate different proteins.

ISOTHIAZOLONES
Covalent modifier: reacts chemically with proteins in non-specific, non-drug-like ways.

CURCUMIN
Covalent modifier, **membrane disruptor:** muddles response of membrane receptors.

HYDROXYPHENYL HYDRAZONES
Covalent modifier, **metal complexer:** sequesters metal ions that inactivate proteins.

ENE-RHODANINE
Covalent modifier, metal complexer.

PHENOL-SULPHONAMIDES
Redox cycler, covalent modifier, **unstable compound:** breaks down into molecules that give false signals.

▶ work was mainly performed at pharmaceutical companies, and supported by experienced chemists. It is now increasingly common in academic environments, in which the same support may not exist. An example of this trend, the Academic Drug Discovery Consortium, founded in 2012, already boasts more than 115 centres in 9 countries. Unfortunately, growing numbers of academic drug researchers are typically not trained to separate hits into compounds good, bad and ugly[3].

The apparent activity of PAINS is so seductive that work continues despite published reports explaining that a compound interferes with assays. One under-appreciated study[4] identified half-a-dozen molecules that showed activity against one-third or more of the proteins that they were screened against.

Repeated identification of the same types of molecule as promising hits against different proteins is polluting the chemical literature. Publications falsely revalidate molecules as good drug leads and feed Sisyphean cycles of 'screen, publish, flounder'. Chemical companies include these artefacts in their sales catalogues as published protein inhibitors, and other biologists start using them in their own studies.

Often termed 'bad actors', PAINS are actually excellent actors: they impersonate promising leads only too well. One of us (J.B.) first published a guide[1] to identifying PAINS in 2010, after his lab spent the equivalent of two or three employee years on futile work. The other of us (M.A.W.) became an anti-PAINS advocate after preparing a manuscript and patent application for a potential fungicide, only to realize that many of the compounds identified as 'actives' were not. For instance, some compounds underwent non-specific chemical reactions with proteins. More than a year was spent working out what was really going on, using assays for confounding properties such as aggregation, chemical decomposition, protein reactivity and fluorescence.

> *"Most PAINS function as reactive chemicals rather than discriminating drugs."*

Better awareness of these molecules and a few precautionary practices could, we estimate, save years of biomedical-research work and millions of dollars in wasted experiments (see 'PAINS-proof drug discovery').
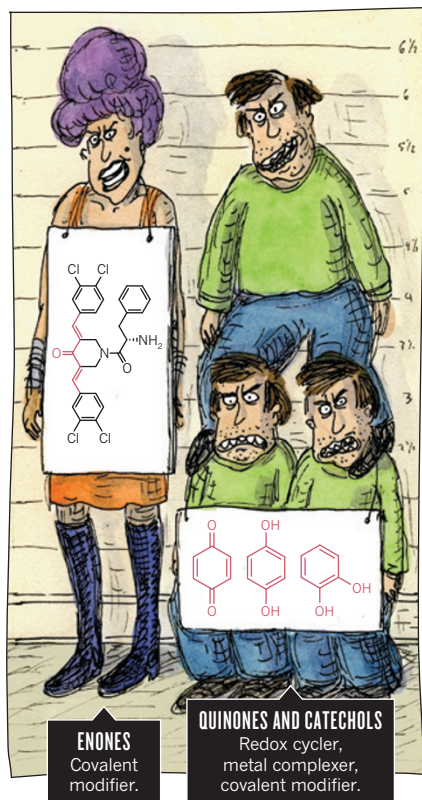
MISLEADING MIRAGE

In a typical academic screening library, some 5–12% of compounds are PAINS[1], which can be rediscovered in multiple assays. This reflects the proportions in the pre-assembled libraries that most academic laboratories buy or access, for example the Library of Pharmacologically Active Compounds (LOPAC) from chemical-supply company Sigma-Aldrich, and the Molecular Libraries Small Molecule Repository (MLSMR) from the US National Institutes of Health.

Most PAINS function as reactive chemicals rather than discriminating drugs. They give false readouts in a variety of ways. Some are fluorescent or strongly coloured. In certain assays, they give a positive signal even when no protein is present. Other compounds can trap the toxic or reactive metals used to synthesize molecules in a screening library or used as reagents in assays. These metals then give rise to signals that have nothing to do with a compound's interaction with a protein. Other PAINS coat a protein or sequester metal ions that are essential to a protein's function, or they may alter proteins chemically without fitting specifically into a binding site. All of these mechanisms thwart further attempts to improve a molecule's biological activity by tweaking its structure (the art known as medicinal chemistry).

PAINS often interfere with many other proteins as well as the one intended. For example, under certain assay conditions, some molecules — redox cyclers — produce hydrogen peroxide[5], an antiseptic that is also produced by certain immune cells. The hydrogen peroxide inactivates the target protein and makes the compound look like a good inhibitor. But the compound itself may not bind to the protein at all.

After a hit is identified, often the next step is to test it in cells. PAINS frequently have more than one interference mechanism. They might produce a desired cellular readout, such as growth inhibition, even without acting as expected. Researchers are too readily convinced that a compound has potent activity against a protein target. They purchase similar compounds from commercial

ENONES
Covalent modifier.

QUINONES AND CATECHOLS
Redox cycler, metal complexer, covalent modifier.

THREE TIPS
PAINS-proof drug discovery

Learn disreputable structures.
Pan-assay interference compounds (PAINS) encompass some 400 structural classes, but more than half of PAINS in a typical library fall into just 16 easily recognizable categories[1,2]. Software tools can filter PAINS from screening libraries, but they are no match for sharp-eyed scientists.

Researchers should familiarize themselves with the most common structures and consult with PAINS-savvy medicinal chemists when these structures show up in hits. Scan compounds for functional groups that could have reactions with, rather than affinity for, proteins. These may not be flagged as PAINS, but can be similarly misleading.

Check the literature. Search by both chemical similarity and substructure to see if a hit interacts with unrelated proteins or has been implicated in non-drug-like mechanisms. Online services such as

SciFinder, Reaxys, BadApple or PubChem can assist in the check for compounds (or classes of compound) that are notorious for interfering with assays.

Assess assays. For each hit, conduct at least one assay that detects activity with a different readout. Be wary of compounds that do not show activity in both assays. If possible, assess binding directly, with a technique such as surface plasmon resonance.

Drill into further details. Compounds that become more active over time are probably acting through non-drug-like mechanisms. When a compound is tested with a protein and then diluted away, its activity should decrease. If not, it might be a PAINS.

Verify the identity and purity of hits. Sometimes a positive readout is due to an unstable breakdown product of the chemical identified from the screening library[1,2]. Remake or repurify these molecules and test them again. **J.B. & M.A.W.**

suppliers and use expensive medicinal-chemistry resources to make and test more analogues. Sometimes a molecule fails only after going through patenting and assessment in animals[6]. It is regrettably easy, in our experience, to get a misleading readout in an animal model that is not related to the anticipated mechanism of action.

In fact, real hits — molecules that do interact specifically with the desired protein — often do not show activity in cells until structures are modified to bind more efficiently or to enter cells more readily[7].

ALL PAIN, NO GAIN
Some of the compounds that should ring the most warning bells are toxoflavin and polyhydroxylated natural phytochemicals such as curcumin, EGCG (epigallocatechin gallate), genistein and resveratrol. These, their analogues and similar natural products persist in being followed up as drug leads and used as 'positive' controls even though their promiscuous actions are well-documented[8,9].

Rhodanines exemplify the extent of the problem. A literature search reveals 2,132 rhodanines reported as having biological activity in 410 papers, from some 290 organizations of which only 24 are commercial companies. The academic publications generally paint rhodanines as promising for therapeutic development. In a rare example of good practice, one of these publications[10] (by the drug company Bristol-Myers Squibb) warns researchers that these

types of compound undergo light-induced reactions that irreversibly modify proteins. It is hard to imagine how such a mechanism could be optimized to produce a drug or tool. Yet this paper is almost never cited by publications that assume that rhodanines are behaving in a drug-like manner.

Very occasionally, a PAINS compound does interact with a protein in a specific drug-like way. If it does, its structure could be optimized through medicinal chemistry. However, this path is fraught — it can be difficult to distinguish when activity is caused by a drug-like mechanism or something more insidious. Rhodanines also occur in some 280 patents, a sign that they have been selected for further drug development. However, to our knowledge, no rhodanine plucked out of a screening campaign is in the clinic or even moving towards clinical development. We regard the effort to obtain and protect these patents (not to mention the work behind them) as a waste of money.

In a move partially implemented to help editors and manuscript reviewers to rid the literature of PAINS (among other things), the *Journal of Medicinal Chemistry* encourages the inclusion of computer-readable molecular structures in the supporting information of submitted manuscripts, easing the use of automated filters to identify compounds' liabilities. We encourage other journals to do the same. We also suggest that authors who have reported PAINS as potential tool compounds follow up their original reports with studies confirming the subversive action of

these molecules. Labelling these compounds clearly should decrease futile attempts to optimize them and discourage chemical vendors from selling them to biologists as valid tools.

Most of all, academic drug discoverers must be more vigilant. Molecules that show the strongest activity in screening might not be the best starting points for drugs. PAINS hits should almost always be ignored. Even trained medicinal chemists have to be careful until they become experienced in screening. Take it from us: do not even start down these treacherous routes. ∎

Jonathan Baell *is professor of medicinal chemistry at Monash University in Melbourne, Australia.* **Michael A. Walters** *is research associate professor of medicinal chemistry at the University of Minnesota in Minneapolis, USA.*
e-mails: jonathan.baell@monash.edu; mwalters@umn.edu

1. Baell, J. B. & Holloway, G. A. *J. Med. Chem.* **53,** 2719–2740 (2010).
2. Baell, J. B. *Future Med. Chem.* **2,** 1529–1546 (2010).
3. Dahlin, J. L. & Walters, M. A. *Future Med. Chem.* **6,** 1265–1290 (2014).
4. Hu, Y. & Bajorath, J. *AAPS J.* **15,** 808–815 (2013).
5. Baell, J. B. *Drug Discovery Today* **16,** 840–841 (2011).
6. Ramesh, C. et al. *J. Med. Chem.* **53,** 1004–1014 (2010).
7. Baell, J. B., Ferrins, L., Falk, H. & Nikolakopoulos, G. *Aust. J. Chem.* **66,** 1483–1494 (2013).
8. Priyadarsini, K. I. *Curr. Pharm. Des.* **19,** 2093–2100 (2013).
9. Ingólfsson, H. I. et al. *ACS Chem. Biol.* **9,** 1788–1798 (2014).
10. Voss, M. E. et al. *Bioorg. Med. Chem. Lett.* **13,** 533–538 (2003).