



# Problems that Can Occur when Assaying Extracts to Pure Compounds in Biological Systems

David J. Newman, DPhil<sup>1,\*</sup>

NIH Special Volunteer, Wayne, PA, 19087



## ARTICLE INFO

### Article history:

Received 11 August 2021

Accepted 11 September 2021

### Key words:

Bioassay Pro

frequent hitters

invalid metabolic panacea compounds

PAINS compounds

## ABSTRACT

For a significant number of years, scientists of many persuasions have assayed natural product materials ranging from crude extracts to pure compounds, in a multitude of assays causally related to some biological processes. However, in a very significant number of submitted papers and published articles, what may be considered as canned biological assays were used, and if a positive effect was observed, then the authors would claim that the material assayed was a potential drug lead. This also occurred with pure synthetic compounds and compounds derived from natural products by simple chemical modifications.

However, what has now become quite obvious—with all such classes of materials—is that there are many promiscuous players with multiple bioactivities. These can range from relatively crude extracts, pure compounds from natural products, synthetic processes that produce natural product derivatives, and even compounds that are truly synthetic in origin. There is also a potential problem with the data from crude to purified extracts being used to claim some form of beneficial activities for such materials, to sell that particular mixture to the lay public, by very careful descriptions of its possible uses due to legal hurdles.

With the advent of artificial intelligence and very large compound databases, some of which may well contain impure materials, scientists from a variety of backgrounds have begun to utilize such listings to obtain compounds for their low to high throughput biological screens, without realizing that there are very significant numbers of active compounds (eg, pan assay interference compounds and invalid metabolic panaceas), that will hit in many different screens for a variety of reasons, thus leading to significant wasted efforts and published scientific articles that have incorrect results.

This commentary gives some of the history of such materials but is designed to be used as a warning to both researchers and in particular, journal editors, and reviewers, that reports of biological results that are claimed to be the result of the compounds used, need to be very carefully screened for results due to such promiscuous compounds, irrespective of their nominal source(s).

All literature searches were made by the author and the background knowledge has come from more than 55 years of research in industry and governmental laboratories in both the United Kingdom and the United States, for enzyme inhibitors/activators as well as antimicrobial and antitumor lead compounds mainly from natural product sources.

The conclusion that I came up with as a result is this: Caveat emptor. (*Curr Ther Res Clin Exp.* 2021; 82:XXX–XXX)

© 2021 Elsevier HS Journals, Inc.

© 2021 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Introduction

I was asked by the Editor in Chief to write a Commentary on the types of problems that can and do frequently occur when every type of compound—from simple extracts from biological materials to pure compounds—is assayed in some biological system(s) ranging from an isolated enzyme to a cell-based system. In the past 40

\* Address correspondence to: 664 Crestwood Rd, Wayne, PA 19087.

E-mail address: [djnewman664@verizon.net](mailto:djnewman664@verizon.net)

<sup>1</sup> NIH Special Volunteer, Wayne, PA, 19087.

or so years, I have been asked by editors of various journals to referee manuscripts in which the authors have reported the following (frequently as part of a title) “interesting biological results from assays of extract and/or natural product X,” that they consider worth having published in the (bio)chemical literature via that journal. This request is usually due to them considering that the results are indicative of significant drug capabilities of the material(s) used, but these data frequently consist of the results from what I will call canned biological assays.

For example, there are many such assays that claim to denote a “redox-active” potential from an extract/compound whether crude, partially purified, purified or even (semi)-synthetic, but in a fair amount of these cases, when investigated, the presence of promiscuous compounds that are nonspecific is well documented.<sup>1</sup> Although Feng et al<sup>1</sup> were referring to problem compounds in medium to high throughput drug screens, the same phenomena have been seen repeatedly, and nowadays, with the advent of machine learning/artificial intelligence (AI) used to select compounds from multiple external databases for subsequent assay(s). What can be implied from the data offered by Feng et al<sup>1</sup> (and others) is that many of the datasets/compounds that the AI system could utilize are not suitable for such systems and should be very carefully inspected before wasting time and effort on their use.

For example, the latest ZINC20 system has more than 1.3 billion entries.<sup>2</sup> The major unrecognized problem that can occur if an AI system is used to choose a series of compounds for subsequent testing is not due to the structures selected, but to the linked assay results in/from such databases. If the assay results are not reliable, the exercise is not going to produce reliable data for subsequent analysis.

Other potential sources of natural product libraries (some commercial, others open) are given in the 2020 review by Wilson et al,<sup>3</sup> although which are available for AI-based searching is not noted. The recent article by Rodrigues<sup>4</sup> demonstrates the good, the bad, and the ugly aspects of such data, and how scientists who wish to utilize the bioactivity databases that are available, need to thoroughly investigate the sources of the compounds whose subsequent data they wish to utilize.

### A Recent Example of Problems Using Natural Product Extracts

An example of what can happen with crude extracts from parts of plants is the recent article by Alhawarri et al<sup>5</sup> in the journal *Molecules*, (an open access journal). Page 3 of that article has a listing of the “assays” utilized, followed by a discussion of many of the classes of molecules that were “considered to be present.” However, many of the compounds listed were identified using what in basic analytical chemistry are known as spot tests.

To demonstrate such problems, the assay that they considered as relating to “antioxidant properties” of the crude extracts was used for more than 70 years as a basic method for determining phenolic components and was originally designed to show the presence of tyrosine, although as shown below it has many more activities associated with it.

What is not usually recognized or even mentioned by users is that there are many compound classes that react with this reagent, including proteins and free thiol groups. A partial listing of other compound types was given by Everette et al<sup>6</sup> in an excellent study in 2010 that should be required reading for any investigator before using this assay system.

This reagent system was the basis for the Lowry et al<sup>7</sup> protein assay from 1951, which was based on the 1922 article by Wu<sup>8</sup> that extended the original 1919 article by Folin and Wu.<sup>9</sup> Another point related to quantification is that assays based on the Folin reagent do not give linear results in protein determinations because it does

not obey Beer's Law,<sup>10</sup> and the assay is interfered with by multiple detergents.

### Specific Comments Related to Isolation and Assay Problems

#### *Ill-defined results*

Feng et al<sup>1</sup> and Rodrigues<sup>4</sup> cover purified compounds, but in the cases of natural products, one starts with a plant, microbe, or marine invertebrate and then proceeds basically by 1 of 2 routes.

The first is simply to macerate/extract the dried material using a number of solvents, frequently an aqueous alcohol mixture and/or a number of aliphatic solvents such as hexanes, although if mimicking an ethnobotanical process, the treatments may be different. However, once extracts have been obtained, they are screened using whatever assays the investigators have access to. At times, a low purity compound might be isolated, and it is tested in addition to the extracts in the assay system(s) available to the investigator. Although this may at first glance be similar to the bioactivity-driven method described below, it is not because the investigators are simply using the data to claim a specific activity.

The second method is known as a bioactivity-driven isolation process. In this case, which in general is only used in laboratories that have access to relatively sophisticated biological and chemical isolation/assay systems, the source material is treated in a series of isolation steps usually utilizing a variety of sequential solvent and/or chromatographic isolation. The difference is that at each step from the first crude product, its biological activity in the assay(s) desired is the driving force for the next purification step in the process. If the activity is lost, then one reevaluates the processes leading up to that result.

#### *Compounds that give spurious results*

Over the past few years, there have been a number of excellent articles describing pure compounds that in some cases might have come originally from natural sources, or be semisynthetic variants, and other compounds that consistently give what can be considered spurious results in multiple assays.<sup>11,12</sup> Why these reports have not altered the content of submissions to journals when authors repeatedly ignore this information may be due in a number of cases to a lack of access to scientific libraries (particularly in lesser-developed countries), or perhaps is coupled to the publish or perish systems that occur in certain countries, where an advanced degree is not awarded until an article reporting the results is published or presented in an international journal or conference. This latter requirement has led to the proliferation of very significant numbers of journals that will effectively publish any article if the necessary publication fees are paid, irrespective of the validity of the research.

### Data from Pure Materials with Unrecognized Interfering Assay Components

Feng et al<sup>1</sup> also described the problems with pure compounds irrespective of source, when they reported the effects in 2005 of detergent addition on 2 simple enzyme processes, a  $\beta$ -lactamase assay and a chymotrypsin assay. This report built upon data from 1997, 1999, and 2003 (references in Feng et al<sup>1</sup>) and came to the conclusion that some organic molecules form large colloidal aggregates that can sequester, and thus inhibit enzymes when assayed at 30  $\mu$ M. This is a common level used even with approximations for semipurified materials.

### Sub-Comment on Colloids

Although apparently out of order in this discussion, the finding of aggregation with or without colloid formation is relevant to natural product studies. In many labs, a crude extract is assayed at concentrations much in excess of 20 mg/mL, which can give nominal concentrations well over the 30  $\mu$ M level for individual components in a mixture. Thus, one should consider such problems before any such experiment because extracts and their subsequent fractions can well aggregate and/or form colloids due to other cellular components in the extracts.

### Pure Compounds from All Sources Identified as Pan Assay Interference Compounds

Baell et al<sup>11-13</sup> have reported extensively on pure compounds that are frequent hitters in biological assays. They use the term pan assay interference compounds (PAINS) to describe these molecules. Although they mainly utilized pure compounds in their reports, the substructures that enabled the multiple responses in different biological assays are found in many natural products, as they demonstrated in an article in the *Journal of Natural Products* in 2016.<sup>13</sup>

In 2015, another academic group<sup>14</sup> reported what they called "PAINS in the assay" when looking at the results from a high throughput screen (HTS) targeting the histone acetyltransferase Rtt109, where the actives fell into the PAINS category. They also listed the assays that had similar problems, with extensive citations, also commenting that cell-based assays are prone to such effects.

Very recently in 2021, Dahlin et al<sup>15</sup> published a review in *Cell Chemical Biology* covering nuisance compounds in cell-based assays. Although this group of coauthors from academia, the National Institutes of Health, and the pharma industry chose not to publish in an open access journal, it is well worth investigators obtaining a copy if only for the figures because they visually demonstrate the problems seen in a variety of cell-based assays, and the reference list is both extensive and up to date.

In addition, a recent report from Yang et al<sup>16</sup> indicated that in addition to PAINS compounds that are already identifiable, there are significant numbers of other agents with varying chemical properties, including confirming a number of the compound types and their chemical substructures mentioned in the Wilson et al<sup>17</sup> review that can interfere with assay components.

### Pure Compounds from All-Natural Product Sources Identified as Invalid Metabolic Panaceas

From the aspect of natural product-based interfering compounds, the analytical review published by Bisson et al,<sup>18</sup> identified what is effectively the natural products equivalent of PAINS. They called these agents invalid metabolic panaceas, and make the very valid point that the active principles in many crude to partially purified natural product-based therapies are polyfactorial agents. The group used data from the contents of NAPRALERT ([www.napralert.org](http://www.napralert.org)), a database covering more than 80 years of information that is hosted in the Pharmacy College at University of Illinois at Chicago. Table 1 in the review by Bisson et al<sup>18</sup> is an eye-opener as to the promiscuity of many known natural product compounds, including as numbers 1 through 5 in rank order: 1. quercetin, 2. gossypol, 3. alpha-pinene, 4. rutin, and 5. berberine. This demonstrates the incipient problems of simply assaying even pure compounds, let alone a crude to partially purified extract, and claiming a valid bioactivity.

### Current Methodologies in Extract Pretreatment

What is now becoming the major emphasis with natural product extracts before putting them in any assay, particularly an HTS assay, is the production of enriched fractions that are organized by polarity and molecular weight(s).<sup>19</sup> A process utilizing adsorption and polarity-based separations was designed at the National Cancer Institute (NCI) to generate 10 fractions per initial extract in a format amenable to high throughput screening (384 well plates) covering the processing of more than 200,000 extracts. This process built on many years of working with extracts from microbes, marine organisms, and plant collections by the NCI. Later work from the same NCI group demonstrated the further work process with these separated fractions. Consideration was also given to the concomitant production and storage of spectral data, usually MS<sup>n</sup> and at times Fourier transform IR, so that if a particular fraction produces an initial hit, the isolation of any active component(s) is relatively simple, although identification might take somewhat longer. All data are machine-readable and multiple copies of each plate will be produced.<sup>20</sup>

A later review from basically the same NCI group<sup>21</sup> discussed the problems, and some potential solutions, to the utilization of natural products (extracts to pure compounds) in HTS systems. That review included a discussion of methods of assessing the probability of identifying PAINS compounds in the extracts produced. The techniques proposed require significant preliminary determinations as to the intrinsic problems associated with specific assays. Such determinations will usually require that they be performed in sophisticated and well-funded laboratories to tag extracts and/or their fractions as containing potential PAINS-associated compounds, well before any significant further work is proposed on such extract fractions.

Part of the reasons for subsequently producing the polarity separated fractions referred to above and published in the last 3 references, in particular, Wilson et al,<sup>21</sup> was the experience gained by the NCI group on the problems associated with screening a large number of crude (unfractionated) natural product extracts during the 2012-2014 time frame, against a variety of cancer-cell related Bcl2 protein targets in conjunction with a California group and their subsequent follow-up of a few cases.<sup>22</sup>

### Comment on the Natural Product Curcumin

As the best relatively current example of how much time and treasure has been wasted due to a multiplicity of problems as a result of publications that claim, usually without proper controls in the assays used, that "Compound X" will be the next lead against major diseases, an article by Nelson et al<sup>23</sup> discussing curcumin should be required reading for any investigator studying natural products at any stage of purity.

The structures of the 3 major components of turmeric extract from which curcumin(oids) are extracted are shown in the Figure 1. The 3 compounds range from 1% to 6% by weight of turmeric on extraction from the host plant *Curcuma longa*, with the majority of the weight of the turmeric extract being due to carbohydrates. Of the 3 bioactive compounds, the major component curcumin is present at between 60% and 70% by weight followed by demethoxycurcumin at 20% to 27% by weight and then bisdemethoxycurcumin at 10% to 15% by weight depending upon the growth conditions and the collection area. Note that these proportions of the bioactive principles are "of the 1% to 6% by weight of the original extract."

From a chemical perspective, the ketone and hydroxyl groups in the middle of all the compounds permit keto-enol tautomerism to occur, and this leads to chemical instability. In addition, only

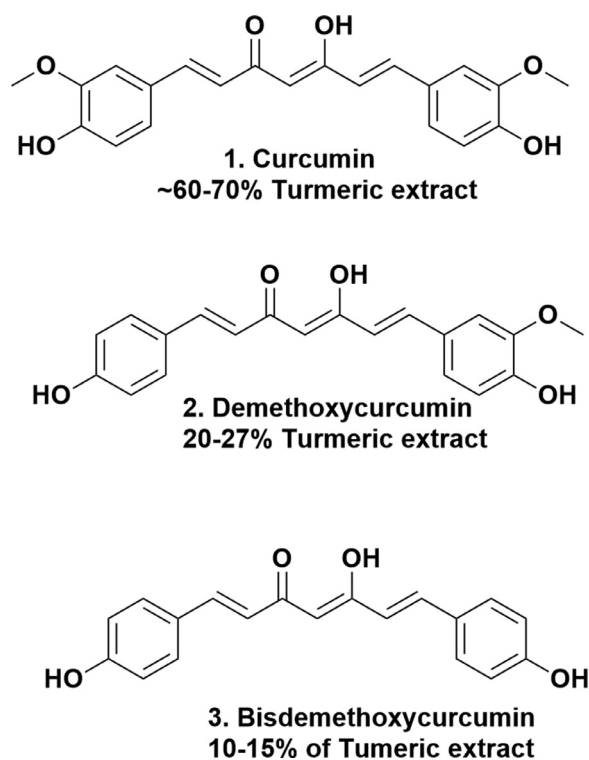


Figure 1. Curcumin components of turmeric extracts.

the enol tautomer is soluble between pH 3 to 9. Thus, biological availability is not what users might think.

Since the late 1990s, curcumin has been reported as a result of assays of various etiologies to have activity in at least the following disease states: anti-inflammatory, anti-HIV, antibacterial, anti-fungal, nematocidal, antiparasitic, antimutagenic, antidiabetic, and antifibrinogenic, in addition to being radioprotective, anticarcinogenic, and even useful in Alzheimer's disease—plus the old standby, antioxidant activity.

As of August 2021 there have been more than 250 clinical trials of curcuminoids (meaning regimens that included some version of curcumin) against diseases ranging from mucositis to various cancers. Of that number, only 13 studies were listed as “having results” in the National Institutes of Health Clinical Trials Database ([www.ClinicalTrials.gov](http://www.ClinicalTrials.gov)) as of the end of August 2021, with 1 being terminated at the Phase I or II level and only 1 going to Phase III as a potential treatment for radiation dermatitis. However, that trial, which finished in 2016, does not list any subsequent publication. Those 13 reports as noted came from more than 250 trials of curcuminoids in 1 form or another.

If one now uses the filters on the ClinicalTrials.gov database for “recruiting/not recruiting,” then there are 32 studies listed with 5 at the Phase III and 3 at Phase IV, 1 of which is not yet recruiting. It should be noted that Phase IV trials are supposedly for drugs that have been approved. However, no curcumin-based drugs have been approved as of the end of September 2019<sup>24</sup> and a check of subsequent data on drug approvals on the Food and Drug Administration website does not show any curcuminoid-based drug from October 2019 through the end of August 2021. Nor is there any published report of a successful clinical trial at Phase III using a double-blinded placebo controlled clinical study.

By using [ClinicalTrials.gov](http://ClinicalTrials.gov) as the resource and adjusting the filters to show diseases, drug treatment(s), methodologies, and so on, readers can see how many more trials of curcuminoids (and derivatives) at different clinical levels are underway. It should also

be noted that this database covers not only US-sponsored trials, but also frequently comparable studies from overseas. In addition, similar free databases are available in the European Union, Japan, China, and Australia, which can also be searched as readers desire.

## Conclusions

For the review of any article involving natural product sources irrespective of source, the experimental details should be the first section that is read in detail. If, for example, there is an aqueous ethanolic extraction of the crude mass, followed by fractionation of that extract using lower polarity solvents, then any subsequent results are at best debatable because the initial extraction will leave the less-polar fractions behind. If the initial work-up procedures have little or no relationship to using assays at each step, be very careful of the quoted results. If assays involve what can be considered to be analytical spot tests (see the earlier discussion on anti-oxidant assays above), be very aware of the promiscuity of such assays. If the assays only use purified compounds, be aware of the PAINS/invalid metabolic panacea compounds and their problems. Finally, the various problems mentioned above can be best expressed by the old Latin tag caveat emptor, which in English is “let the buyer beware.”

## Conflicts of Interest

The author has indicated that he has no conflicts of interest regarding the content of this article.

## References

- Feng BY, Shelat A, Doman TN, Guy RK, Shoichet BK. High-throughput assays for promiscuous inhibitors. *Nat. Chem. Biol.* 2005;1:146–148.
- Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, Moriz YS, Mayfield J, Sayle RA. Zinc20-A free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* 2020;60:6065–6073.
- Wilson BP, Thornburg CC, Henrich CJ, Grkovic T, O'Keefe BR. Creating and screening natural product libraries. *Nat. Prod. Rep.* 2020;37:893–918.
- Rodrigues T. The good, the bad, and the ugly in chemical and biological data for machine learning. *Drug Discov. Today Technol.* 2019;32:33–8.
- Alhawarri MB, Dianita R, Razak KNA, Mohamad S, Nogawa T, Wahab HA. Antioxidant, anti-inflammatory, and inhibition of acetylcholinesterase potentials of *Cassia timoriensis* DC flowers. *Molecules.* 2021;26:2594.
- Everette JD, Bryant QM, Green AM, Abbey YA, Wangila GW, Walker RB. A thorough study of reactivity of various compound classes towards the Folin-Ciocalteu reagent. *J. Agric. Food Chem.* 2016;58:8139–8144.
- Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* 1951;193:265–275.
- Wu H. A new colorimetric method for the determination of plasma proteins. *J. Biol. Chem.* 1922;51:33–39.
- Folin O, Wu H. A system of blood analysis. *J. Biol. Chem.* 1919;38:81–110.
- Buijss K, Maurice MJ. Some considerations on apparent deviations from Lambert-Beer's Law. *Anal. Chim. Acta.* 1969;47:469–474.
- Baell JB, Holloway GA. 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.
- Baell JB, Walters MA. Chemical con artists foil drug discovery. *Nature.* 2014;513:481–483.
- Baell JB. Feeling Nature's PAINS: Natural products, natural product drugs, and pan assay interference compounds (PAINS). *J. Nat. Prod.* 2016;79:616–628.
- Dahlin JL, Nissink WM, Strasser JM, Francis S, Liggins L, Zhou H, Zhang Z, Walters MA. PAINS in the Assay: Chemical mechanisms of assay interference and promiscuous enzymatic inhibition observed during a sulfhydryl-scavenging HTS. *J. Med. Chem.* 2015;58:2091–2113.
- Dahlin JL, Auld DS, Rothenaigner I, Haney S, Sexton JZ, Nissink JWM, Walsh J, Lee JA, Strelow JM, Willard FS, Ferrins L, Baell JB, Walters MA, Hua BK, Hadian K, Wagner BK. Nuisance compounds in cellular assays. *Cell Chem. Biol.* 2012;28:356–370.
- Yang Z-Y, Yang Z-J, He J-H, Lu A-P, Shao Liu S, Hou T-J, Cao D-S. Benchmarking the mechanisms of frequent hitters: limitation of PAINS alerts. *Drug Discov. Today.* 2021;26:1353–1358.
- Wilson BP, Thornburg CC, Henrich CJ, Grkovic T, O'Keefe BR. Creating and screening natural product libraries. *Nat. Prod. Rep.* 2020;37:893–918.
- Bisson J, McAlpine JB, Friesen, Chen S-N, Graham J, Pauli GF. Can invalid bioactives undermine natural product-based drug discovery? *J. Med. Chem.* 2016;59:1671–1690.

19. Thornburg C, Britt J, Evans J, Akee R, Whitt J, Trinh S, Harris M, Thompson J, Ewing T, Shipley S, Grothaus P, Newman D, Grkovic T, O'Keefe B. NCI program for natural product discovery: Generation of the largest, publicly accessible library of natural product fractions for high-throughput screening. *ACS Chem. Biol.* 2018;13:2484–2497.
20. Grkovic T, Akee RK, Thornburg CC, Trinh SK, Britt JR, Harris MJ, Evans JR, Kang U, Ensel S, Henrich CJ, Gustafson KR, Schneider JP, O'Keefe BR. National Cancer Institute (NCI) program for natural products discovery: Rapid isolation and identification of biologically active natural products from the NCI prefractionated library. *ACS Chem. Biol.* 2020;15:1104–1111.
21. Wilson BP, Thornburg CC, Henrich CJ, Grkovic T, O'Keefe BR. Creating and screening natural product libraries. *Nat. Prod. Rep.* 2020;37:893–918.
22. Hassig CA, Zeng F-Y, Kung P, Kiankarimi M, Kim S, Diaz PW, Zhai D, Welsh K, Morshedian S, Su Y, O'Keefe B, Newman DJ, Rusman Y, Kaur H, Salomon CE, Brown SG, Baire B, Michel AR, Hoye TR, Francis S, Georg GI, Walters MA, Divlianska DB, Roth GP, Wright AE, Reed JC. Ultra High Throughput Screening of Natural Product Extracts to Identify Pro-apoptotic Inhibitors of Bcl-2 Family Proteins. *J. Biomol. Screening.* 2014;19:1201–1211.
23. Nelson KM, Dahlin JL, Bisson J, Graham J, Pauli GF, Walters MA. The essential medicinal chemistry of curcumin. *J. Med. Chem.* 2017;60:1620–1637.
24. Newman DJ, Cragg GM. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 2020;83:770–803.