Bioorganic & Medicinal Chemistry Letters 20 (2010) 5308–5312

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Using an in vitro cytotoxicity assay to aid in compound selection for in vivo safety studies

Nigel Greene *, Michael D. Aleo, Shirley Louise-May, David A. Price *, Yvonne Will

Pfizer Global Research and Development, Eastern Point Road, Groton, CT 06340, USA

ARTICLE INFO

Article history:
Received 12 May 2010
Revised 19 June 2010
Accepted 24 June 2010
Available online 1 July 2010

Keywords:
c log *P*
Toxicology
TPSA
Attrition
Cytotoxicity
Physico chemistry

ABSTRACT

Recent publications have demonstrated that using calculated physiochemical properties can help in the design of compounds that have a decreased risk of significant findings in rodent toxicology studies. In this Letter, we extend this concept and incorporate results from a high throughput cytotoxicity assay to help the drug discovery community select compounds for progression into in vivo studies. The results are presented in an easily interpretable odds ratio so that teams can readily compare compounds and progress potential clinical candidates to the necessary rodent in vivo studies.

Numerous drugs have been withdrawn from the market place over the last 20 years for a variety of reasons, however, recent publications suggest that safety concerns are among one of the leading causes for attrition.[1–3] With the increased investment required for prosecution of projects once they obtain clinical status, pre-clinical attrition is preferred; however, increasing the likelihood of delivering compounds of high quality to the clinic is the real goal. In response to this challenge, research groups are increasing the use of both, in vitro and in vivo safety models earlier in the drug discovery process to help differentiate compounds and increase the likelihood of the compound entering the clinic. A recent review of safety screening approaches demonstrates that many in vitro safety assays are used in risk identification mode where specific relationships exist between well defined mechanisms of toxicity and discrete in vivo end points.[4] Examples include the Ames assay as an assessment of genetic toxicity and the potential risk of carcinogenicity in humans or hERG patch clamp assays and the potential for cardiovascular risk in man. Both Ames and hERG patch clamp assays have become mainstays of safety screening paradigms in the modern drug discovery process.[5,6]

Many research institutions are now driving the use of short term, repeat dose in vivo safety assessment studies as early as possible in a project's life.[6] These exploratory toxicology studies are then supplemented with the necessary required regulatory studies to enable the compound to progress safely to clinical studies. It is hoped that this aggressive use of early toxicology studies will reduce overall research costs and cycle times across the industry. Given the lag time between the discovery cycle and the market there has not been sufficient time to assess the impact of this new paradigm on the development of new chemical entities (NCEs).

A more difficult decision facing project teams is deciding which, and how many, compounds should be progressed to exploratory in vivo studies to investigate safety profiles of compounds in animals. These decisions are often poorly informed with little in vitro safety data to aid compound selection for initial in vivo safety studies. These studies often require relatively large quantities of high purity material to be prepared and so present a significant investment for project teams. The selection of compounds can be aided by assessments of off-target pharmacology and compounds that demonstrate promiscuity are often removed from consideration.[7] Other non-safety related parameters such as ease of synthesis and pharmacokinetics are also taken into consideration but the decisions made are still highly qualitative rather than data driven. In this Letter, we describe how an in vitro ATP depletion assay, that measures a compound's ability to cause cytotoxicity in transformed human liver epithelial (THLE) cells can be used effectively to prioritize compounds for in vivo studies.

To be able to compare a continuous concentration measurement in an in vitro cytotoxicity assay to an in vivo measurement of toxicity it is critical to have a measure of the exposure at which the toxicity was observed in the in vivo study. Often literature reports of toxicology studies do not include information on the drug levels in the circulating blood stream making these types of

* Corresponding author. Tel.: +1 860 686 9180.
E-mail address: david.a.price@pfizer.com (D.A. Price).

0960-894X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2010.06.129

comparisons virtually impossible. In this experiment, we used Pfizer proprietary compounds, studied in short term exploratory toxicology studies, where this detailed information was available. For this analysis 290 compounds were selected for screening in the THLE assay. The compound set represents a diverse spectrum of chemical structures, intended target pharmacologies and disease indications. The purity and correct compound mass for the samples were confirmed using NMR and LC–MS. Within the format of the cytotoxicity assay described (see Supplementary data), compounds were determined to have LC_{50} values ranging from <10 μM through to an artificial maximum of >300 μM.

The exploratory in vivo safety studies for this analysis typically involved dose escalation over three or more dose levels for a duration of four days or longer. No distinction was made regarding the species or the sex of the animals assessed in these studies and all compounds were administered orally. The rationale for combining species was that the in vitro assay was performed in a human cell line and measured cytotoxicity and was considered able to detect toxicity in all pre-clinical species. In an attempt to differentiate the structural determinants of toxicity from identified pharmacological causes, we eliminated those in vivo findings where the primary pharmacology of the compound had a well known link to the discrete adverse safety outcome. In addition, corresponding exposure data (C_{max}) and plasma protein binding data were also recorded

For the purpose of this exercise two systems for classifying compounds according to the observation seen in its exploratory toxicology study were used. The first system is the binary classification previously used to look at the influence of physicochemical properties on the incidence of adverse findings.[8] This classification scheme uses a 10 μM threshold for C_{max} and compounds that showed any form of significant adverse finding in the exploratory toxicology study below this threshold were classified as 'toxic' and any compound that had no significant findings at C_{max} levels above this threshold were classified as 'clean'. For some 78 compounds in the original data set, the doses in the in vivo study were such that the highest exposure with no significant findings was below the 10 μM C_{max} threshold but the lowest dose with significant findings was above the 10 μM C_{max} threshold. In these cases it was impossible to put these compounds into one or other of the classes and so these were removed from any analysis that used this binary classification leaving 212 compounds in the data set.

The second system is a crude way to look at the relative severity of findings in the in vivo study at a particular dose and C_{max} concentration based on a simple scoring system for severity.[9] The findings for each dose group within a study were broken out into the three main classes of observations made during the experiment organ function, pathological damage and systemic toleration.

Dose groups where significant perturbations in both serum chemistry and hematology parameters were observed were assigned an organ function score of '4'. Dose groups where only perturbations in serum chemistry parameters or hematology parameters were found were given an organ function score of '2' and those with no compound related changes in either of these categories were given an organ function score of '0'. If a dose group had a microscopically observable necrosis in one or more tissues, then the dose group was assigned a pathological damage score of "4". If the dose group was reported to have any microscopic observation other than necrosis, then it was assigned a pathological damage score of '2' and if there were no microscopic observations for the dose group then a score of '0' was assigned. The overall score for the dose group could then be calculated by adding the organ function score and the pathological damage score. However, dose groups where systemic toleration observations were unscheduled deaths were given an overall score of '12' since organ function and pathological damage could not be determined accurately in these cases.[8]

For the results of this analysis to be broadly applicable it is important that the set of compounds are of diverse chemical structure and representative of the current pharmacological space, (Fig. 1). The molecular property distribution of the cytotoxicity dataset overlaps well with a selection of compounds taken from a diverse subset of the Pfizer file chosen for its coverage of chemical space. From analysis of chemical structure alone there did not appear to be any learning around structural motifs that appeared to be higher risk in terms of the THLE assay.

A full listing of all the compound in vitro results and in vivo classifications are listed in the Supplementary data table. Within the dynamic range of the in vitro assay, the data was subdivided into four categories; compounds with THLE LC_{50} <10 μM, <25 μM, <50 μM and <100 μM. Of the compounds with THLE LC_{50} <10 μM none of them were defined as clean from the in vivo exploratory toxicology studies, and 8 compounds were classified as toxic from the in vivo exploratory toxicology studies. Therefore, the sensitivity of the THLE assay with the cutoff of <10 μM, where sensitivity is equal to the number of correctly identified toxic compounds divided by the total number of toxic compounds, was 5% (8:155) with a false positive rate of 0%, where false positive rate equals the number of compounds incorrectly called toxic divided by the total number of toxic calls, when compared to the gold standard assay (in vivo exploratory toxicology studies). Altering the threshold of the THLE assay, we then examined the ability to further discriminate between toxic and clean compounds. As the threshold increases in the THLE assay the sensitivity and false positive rates are also modified. The best compromise was to select LC_{50} <50 μM where 35 compounds from the 155 defined as toxic from the exploratory in vivo studies were identified. This gives a sensitivity of 23% (35:155) and a false positive rate of 8% (3:38). These results are summarized in Tables 1 and 2.

In summary, the relative odds of being toxic at 10 μM when THLE LC_{50} <50 μM = 35:3 = 11.7 whereas the odds of being toxic at 10 μM when THLE LC_{50} >50 μM = 120:54 = 2.2 which means compounds with a THLE LC_{50} <50 μM are five times (11.7 divided by 2.2) more likely to see findings in an exploratory toxicology study at a C_{max} <10 μM than if it's LC_{50} >50 μM.

Comparing the in vitro assay response to the measure of severity presents a greater challenge since the former is a continuous scale and the latter is a categorical measure at a defined C_{max}. Since the range of C_{max} values observed for this data was spread over multiple orders of magnitude, a log scale conversion was performed to give a normal distribution. In addition, we binned each compound-dose group combination according to the compound's LC_{50} value and the observed score for the dose group. The average log (C_{max}) was then calculated for each bin and the results are shown in Table 3.

The differences between these bins, either row to row or column to column are considered statistically significant with approximate p-values <0.001, however, precise calculation of p-values is hindered by the categorical nature of the severity scores at the dose group level. Nevertheless, average log (C_{max}) of compounds eliciting no adverse outcome in the exploratory toxicity studies (i.e., 0) are approximately 4.7-fold higher on average for compounds with

Table 1
THLE assay result versus in vivo outcome

THLE LC_{50} bucket	# Compd toxic at 10 μM (total: 155)	# Compd clean at 10 μM (total: 57)
<10 μM	8	0
10–25 μM	7	3
25–50 μM	20	0
50–100 μM	19	3
>100 μM	101	51