Available online at www.sciencedirect.com
ScienceDirect
Toxicology in Vitro 21 (2007) 1545–1551
Toxicology in Vitro
www.elsevier.com/locate/toxinvit

Struggles for equivalence: *In vitro* developmental toxicity model evolution in pharmaceuticals in 2006

Robert Chapin [a,*], Donald Stedman [a], Jennifer Paquette [a], Randal Streck [a], Steven Kumpf [a], Shibing Deng [b]

[a] *Investigative Developmental Toxicity Laboratory, Pfizer Global Research and Development, Pfizer, Inc., MS 8274-1336, Eastern Point Road, Groton, CT 06340, USA*
[b] *Non-Clinical Statistics Group, Pfizer Global Research and Development, Pfizer, Inc., Eastern Point Road, Groton, CT 06340, USA*

Received 19 September 2006; accepted 9 October 2006
Available online 17 October 2006

Abstract

Our group has been using the ECVAM Embryonic Stem Cell assay to predict developmental toxicity. In order to improve the separation of non-teratogens from weak teratogens, we have employed measures of gene expression, and different statistical methods from those originally used to develop the test. These approaches have fundamentally not improved the discrimination of 'weaks' from 'nons'. A realization that a very low value for cytotoxicity IC50 would drive a final result for the test in ways that were inappropriate for pharmaceuticals has led us to re-examine the cytotoxicity component. Our current efforts are focused on other, perhaps more sensitive, measures of cytotoxicity, combined with gene expression changes in mouse stem cells in an attempt to correctly identify weak teratogens and non-teratogens.

Keywords: Embryonic stem cell test; Development toxicity; *In vitro* assay

1. Introduction

The fields of toxicology and public health both have immense needs for tests that can determine the hazard of a compound exposure. These could be applied not only to new compounds as they are developed, but also to the thousands of existing compounds for which few or no health effects data exist. In an attempt to minimize the number of animals that must be used in such studies, and to speed the process of identifying tests that will provide trustworthy data, the European Centre for the Validation of Alternative Methods (ECVAM) was formed and is one of a cluster of groups actively developing new methods and working towards their validation. The intent would be to use these relatively short *in vitro* assays to identify those chemicals which appear the most toxic, and then to test those compounds further in more definitive assays.

ECVAM created a process for validating potential assays (Balls et al., 1995), assuming that the assays would be used at a number of different laboratories, and for varying purposes. It appears that the main goals for this validation process are: (a) to create an assay protocol that is portable (can be used at many different labs and give substantially the same result), and (b) to define how well each assay performs using chemicals whose in vivo activity is fairly well described.

ECVAM created an initial training set of chemicals and then a larger validation set. The chemicals in these groups were selected on the strength of the in vivo data that were available for them, and for the degree of effects they produced (Brown, 2002). Notable was the specific exclusion of receptor-directed compounds. This is appropriate, given the intended use of these assays as screens mostly for compounds whose human exposure would be environmental.

* Corresponding author. Tel.: +1 860 441 0571; fax: +1 860 501 3379.
E-mail address: Robert.e.chapin@pfizer.com (R. Chapin).

0887-2333/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tiv.2006.10.006

The assay that we will consider for this paper (the mouse embryonic stem cell test, [EST]) generates three concentration–response curves: one for general cytotoxicity using 3T3 fibroblasts, one for the cytotoxicity of the D3 mouse stem cells, and one for the inhibition of differentiation of those stem cells into beating cardiomyocytes. The concentration that inhibits each of these measures by 50% is put into three linear discriminant functions, developed by the ECVAM biostatistical task force (Holzhutter et al., 1996), and the products of the equations are examined. When the product of the first equation is larger than the other two, the compound is considered a non-teratogen. If the product of the second equation is larger, the compound is considered a weak teratogen, while the third equation predicts for strong teratogens (Genschow et al., 2002). The overall predictivity of this assay, using the chemicals selected by the ECVAM committee, was an admirable 78% (Genschow et al., 2004).

2. The pharmaceutical environment

The pharmaceutical industry is certainly a customer of these validation efforts. Pharma companies have many new compounds whose health effects are entirely unknown. After a new drug candidate has been shown to posses the desired properties (i.e., effectiveness against a disease model), the question of safety arises. While regulatory agencies require whole animal drug safety studies, preliminary *in vitro* assays can be very useful to help sort among different compounds.

The areas where the most safety-related attrition of new drug candidates occurs are cardiovascular and hepatic. Closely following these are genetic tox and embryofetal development. Most small-molecule pharma companies seem to experience these issues in approximately this order, and it is logical to address the issues of greatest concern first. *In vitro* assays can certainly help with both cardiovascular and hepatic safety issues. Sooner or later, attention turns to fetal development and avoiding compounds that produce fetal toxicity (growth reduction) or, of greater concern, fetal malformations. This is the main focus of our group at Pfizer. We are using the EST and whole embryo culture test (WEC) to give project teams some early idea of the likelihood that their compounds will produce embryofetal toxicity, and we will focus on the EST in this paper.

Pharmaceuticals represent a somewhat special case in the world of chemicals. New drugs are designed with high specificity; they are intended to hit only a single target molecule in the body. On the other hand, it is rare to find a medicine without side-effects (i.e., activity at some molecule other than the intended target). If the off-target proteins are closely related to the intended target (for example, a family of kinases vs. the intended target kinase), then such off-target effects will occur at very low concentrations (low nM range).

These two characteristics (highly selective to a certain receptor protein, and effectiveness (and toxicity) at very low concentrations) differ significantly from some of the characteristics of the ECVAM training set, which were explicitly non-receptor-directed. When evaluating assay performance for pharmaceutical candidates, it is useful to keep the genesis of the validation set in mind.

In pharmaceutical development, the project teams responsible for overseeing the development of each compound wish to advance only those compounds which have the greatest efficacy and the fewest liabilities. These liabilities could include, among others, an unacceptable pharmacokinetic profile or catabolic pathway or uneconomic synthesis or safety profile. Typically, the research teams want answers to these questions while using the least amount of drug possible, so being able to address safety issues while using the minimum amount of drug candidate can be an important consideration. *In vitro* assays can meet this need.

Additionally, project teams most often have several compounds from which they hope to select their lead molecule. It would be very useful for us to be able to rank compounds for toxicity, or to compare across candidate compounds for the likelihood of producing toxicity in vivo.

Another desired performance characteristic would be to separate weak developmental toxicants from non-toxicants with great confidence. While a few therapeutic areas can tolerate some embryofetal toxicity (cancer, for example), and this is not a concern for other patient populations (Alzheimer's, or bone loss), more therapeutic areas are emerging which can tolerate no developmental toxicity, and project teams demand high specificity in being able to separate compounds at this end of the activity spectrum.

One possible solution to the problem of separating active (i.e., embryotoxic) from non-active compounds would lie in the ability to create a category of compounds which were toxic to the fetus, but whose maternal toxicity would occur at lower doses. In vivo, such dose-limiting maternal toxicity would prevent the expression of embryofetal effects, and the compound could still be embryo-toxic, but it would not matter because the dose could never rise high enough to affect the fetus.

3. Our approach at Pfizer

In implementing the EST at Pfizer, our first step was to verify that our group could get the same answer as those obtained by the ECVAM validation labs. We were able to do that, looking at about 2/3rds of the compounds that ECVAM used. We focused mostly on those compounds which gave the most trouble to the ECVAM validating laboratories. We obtained essentially the same answers as the ECVAM labs (Table 1).

To address the issue of ranking compounds, we have tried function subtraction; that is, subtracting the results of Function 2 from Function 1. Theoretically, the compound that is a weaker teratogen will have a subtraction