Journal of Environmental Science and Health, Part C, 28:272–286, 2010

ISSN: 1059-0501 print / 1532-4095 online
DOI: 10.1080/10590501.2010.525781

Taylor & Francis
Taylor & Francis Group

Exploring *In Vitro/In Vivo* Correlation: Lessons Learned from Analyzing Phase I Results of the US EPA's ToxCast Project

Romualdo Benigni, Cecilia Bossa, Alessandro Giuliani, and Olga Tcheremenskaia

Environmental and Healt Department, Istituto Superiore di Sanita', Rome, Italy

The research on alternative toxicological methods provides, among other things, a privileged viewpoint on one of the central issues of modern biomedical research—the relationship between (a) biological phenomena observed at the level of tissues and organisms and (b) their cellular and molecular bases as studied in isolated systems in vitro. The newly released ToxCast Phase 1 results, subject to initial analysis, converge with evidence from other fields (e.g., research on drug design with intensive use of omics technologies, traditional research on alternative tests) in indicating a low degree of the in vitro/in vivo correlation overall. In addition, this and other approaches point to the need for combining biological and chemical information in exploring the in vitro to in vivo connection.

Keywords: alternative testing; three Rs; toxicity; reductionism; systems biology; ToxCast

INTRODUCTION

Toxicology today faces challenges that are both difficult and exciting. It is estimated that in the United States [1] and Europe [2] several thousand chemicals are in use without adequate toxicological information. Since the task of testing them systematically with classical animal assays is huge, there is a strong societal pressure to investigate and identify suitable alternative testing methods. For example, the new European legislation on chemicals, called REACH, explicitly mentions the possibility of using both experimental (in vitro) and

Address correspondence to Romualdo Benigni, Environment and Health Department, Istituto Superiore di Sanita', Viale Regina Elena 299, 00161 Rome, Italy. E-mail: romualdo.benigni@iss.it

non-testing (Structure-Activity Relationships, Read-Across, Categories) alternative methods [3]. As a matter of fact, research on the so-called Three Rs (Replacement, Reduction, and Refinement of animal testing) has been going on for years with the aims of shortening times of toxicity testing, protecting animal health and welfare and saving money [4].

Together with its paramount importance for defending human health, the research on alternative toxicological methods also provides a privileged viewpoint on one of the central issues of modern biomedical research—the relationship between (a) biological phenomena observed at the level of tissues and organisms and (b) their cellular and molecular bases as studied in dissected, isolated systems in vitro. Whereas much of the conceptual and technical tools developed by biologists in recent decades are aimed at finding the "molecular" keys of phenomena observed in the whole organism, there is growing recognition that this approach has serious limitations, and that a huge gap exists between in vitro observations and in vivo phenomena [5–8].

Recently, new impetus to the research on alternative methods has been given by the US Environmental Protection Agency's ToxCast project [1, 9, 10]. The pathway-based screening paradigm adopted by ToxCast is radically different from traditional in vitro testing. Basically, the perturbations provoked by chemicals to biochemical and biological pathways supposed to be critical to toxicity are analyzed. Such perturbations are studied in isolated systems in vitro with the use of modern High-Throughput Screening (HTS) techniques. The strategy of ToxcCast is to generate profiles of chemical and biological effects for large numbers of chemicals across a wide array of HTS assays, both cell-based and cell-free, and to search for patterns and mechanistically based associations (i.e., drawing from assay correspondence to a gene target and pathway) that correlate with animal toxicity measures and eventually with human adverse effects. Results would be used to prioritize chemicals for targeted testing, and the results of such testing would be used to refine mechanistic hypotheses and signature profiles used for prediction.

In the words of the authors, the goal of the program is " to identify mechanisms of chemically induced biological activity, prioritize chemicals for more extensive toxicological evaluation, and develop more predictive models of in vivo biological response. ... As a consequence, a reduction or replacement of animals in regulatory testing is anticipated to occur in parallel with an increased ability to evaluate the large numbers of chemicals that currently lack adequate toxicological evaluation. Ultimately, Tox21 is expected to deliver biological activity profiles that are predictive of in vivo toxicities" [10].

Among other things, a promising avenue is provided by the fact that, for the first time in this type of exercise, the chemicals are systematically characterized by chemical structure indexing and chemical descriptors. The comparison (correlation) of the three edges of ToxCast information (in vivo toxicity, in vitro HTS patterns, chemical structure) for identified subsets of chemical

sharing biological response characteristics may provide the ground to establish local predictive models based on the integration of in vitro and structure-activity relationships (SAR) approaches [9]. The diversity of chemicals in ToxCast (present and planned) and the richness of in vitro and in vivo endpoints data are promising enough to warrant interesting findings in the future.

Recently, a first round of ToxCast (Phase 1) has been concluded. More than 500 different HTS assays (both cell-free and cell-based) were applied to 320 compounds (309 unique chemicals): these are mostly agrochemicals for which animal toxicity data already existed (rodent carcinogenicity, developmental, and multi-generation toxicity). Just to give a few examples, the in vitro assays include effects of chemicals on proteins such as GPCR, kinase, phosphatase, protease, ion channel, cytotoxicity in transformed cell lines, modulation of TF activity in human hepatoma HepG2 cells, effects of chemicals on signaling pathways in primary human cells, and so on. The selection of the in vitro assays was chosen to cover a wide range of potential biological responses and also to be representative of known or hypothesized toxicity mechanisms.

The in vitro and in vivo data were made available through early release to EPA analysis partners and are being publicly released in association with ToxRef publications [11] and manuscripts by EPA authors detailing the assay systems and results. Taking advantage of the early data release to EPA analysis partners, a number of studies by external partners were presented and discussed at a ToxCast Data Analysis Summit (http://www.epa.gov/NCCT/toxcast/summit.html). After the Summit, in vitro data have been reviewed by authors, and a considerable number of them have been reassigned. The definitive results have been made available recently.

We have exploited the availability of the recently released ToxCast Phase 1 data to investigate the overall relationship between in vitro and in vivo measures, with an eye both to the basic scientific perspective and their relevance to the practical long-term goal of replacing animal testing. The results are then put in a wider perspective by comparing them with various types of existing evidence.

DATA AND METHODS

Besides a number of publications [1, 9, 10], a detailed description of the ToxCast program and of the recent ToxCast Data Analysis Summit is available at http://www.epa.gov/ncct/toxcast/summit.html.

In a first analysis performed in connection with the ToxCast Data Analysis Summit, the data were downloaded through the ToxCast Environmental Science Connector and were contained in the package: ToxCast_20090406.zip. The in vivo data were extracted from the ToxRefDB resource (see http://www.epa.gov/ncct/toxrefdb/ for current references). The results of our preliminary

analyses are in the poster: http://www.epa.gov/NCCT/toxcast/files/summit/ToxcastDataSummit_Poster_Benigni%20May2009.ppt.

The revised data, on which this analysis is based, are downloaded from: http://www.epa.gov/ncct/toxcast/data_sets.html, and are contained in the Data Set called: "ToxCast Phase I Data (AC50/LEC)."

Briefly, the data refer to 320 compounds (309 unique chemicals), mostly agrochemicals. Each chemical has the following toxicological data:

(a) 76 in vivo bioassays (Target organs (chronic toxicity); Reproductive, Developmental, Carcinogenicity);

(b) 524 in vitro assays (9 in vitro assay providers, for a total of 285 cell-based and 239 cell-free toxicity measures). The *in vitro* assays were run specifically within the ToxCast Project.

The names of the chemicals and assays can be found at the ToxCast website.

For the present large scale analysis, activity (positive/negative) data were used. The data were analyzed with multivariate data analysis methods [12], with the SAS/STAT statistical software (http://www.sas.com/technologies/analytics/statistics/stat/index.html).

In particular, Cluster and Discriminant analyses were used. Cluster Analysis is an unsupervised technique aimed at identifying classes of observations in such a way to maximize the ratio: between clusters variance/within cluster variance. In this way, Cluster Analysis points to groups of similar observations. Discriminant Analysis is a supervised method whose aim is to generate a (linear) function of descriptors of a data set, able to properly classify the observations in the set into predefined classes, e.g., toxic and nontoxic chemicals.

RESULTS AND DISCUSSION

Carcinogenicity is one of the major concerns for human health, and the contribution of chemicals to such burden is well documented [13]. Thus, it is of particular interest to assess how the panel of ToxCast in vitro assays correlates with rodent carcinogenicity data. As a matter of fact, a number of ToxCast assays are believed to provide information that is potentially informative of such mechanisms.

Characterization of ToxCast Carcinogenicity Data

Prior to a detailed analysis of the correlation between in vitro assays results and the carcinogenicity data, the ToxCast carcinogenicity results were characterized in comparison to the "traditional" carcinogenicity database. For this aim, we used the ISSCAN v3a database on chemical carcinogens. ISSCAN is a curated database [14] freely available from the website of the Istituto

Superiore di Sanita': http://www.iss.it/ampp/dati/cont.php?id = 233&lang = 1&tipo=7.

It appears that out of the 1153 substances contained in ISSCAN v3a, the overlap with ToxCast is only 46 chemicals. Thus ToxCast provides a considerable amount of new information to the public.

The proportion of carcinogens to noncarcinogens is quite different too. The ratio is 1:1.5 in ToxCast and is 2:1 in ISSCAN v3a. The "traditional" database has many more carcinogens, because it derives largely from studies aimed at ascertaining the carcinogenicity of "suspected" chemicals. On the contrary, because of their intended use the agrochemicals in ToxCast were developed in such a way as to minimize toxicity. Despite this, a considerable number of them are toxic anyway.

An important parameter is to what extent the rat and mouse carcinogenicity data are correlated. Both in ISSCAN v3a and ToxCast, rat and mouse carcinogenicity calls (yes/no) agree 65% of time. On the other hand, the correlation of the carcinogenic potency between rat and mouse is (a) in ToxCast $r = 0.21$, where the potency is defined as log (MW/LEL) (LEL = Lowest Effective Level); and (b) in ISSCAN v3a $r = 0.84$, where the potency is defined as log (MW/TD_{50}) (TD50 is the dose mg/kg body weight that reduces 50% of the animals without tumor) (our calculations for this paper). Whereas the 65% agreement between rat and mouse yes/no calls in the two databases is reassuring, the low correlation of carcinogenic potencies in ToxCast may point to the fact that LEL is not an accurate way of parameterizing potency. For this reason, the following analyses were performed by using the robust part of the toxicity information, i.e., the yes/no calls.

More important information is the underlying carcinogenicity mechanisms. Whereas detailed information on the individual carcinogens is not available, the prevalence of genotoxic and nongenotoxic carcinogens can be roughly estimated through structural considerations. By applying the updated list of Structural Alerts (SA) for carcinogenicity contained in Toxtree 1.6 [15, 16], it appears that (a) in ISSCAN v3a, among the carcinogens, around 70% have "genotoxic" SAs, and around 5% have "nongenotoxic" SAs; whereas (b) in ToxCast, the distribution among the carcinogens is around 35% with "genotoxic" SAs and 40% with "nongenotoxic" SAs. Even though the number of "nongenotoxic" SAs is still limited in Toxtree 1.6, it is evident that the types of carcinogenicity mechanisms are distributed quite differently in two databases. This should be taken into account when considering the ToxCast data.

Characterization of In Vitro Assays

In principle, the availability of large numbers of in vitro (n = 524) and in vivo (n = 76) measures in the ToxCast Phase 1 database permits the

exploration of an extremely vast range of in vitro/in vivo comparisons. However, the statistical theory indicates that in such cases the observation of correlations purely due to chance is extremely likely [17, 18]. Thus, averting chance correlations is a crucial pre-requisite.

A preliminary step of our analysis was the identification of the robust (correlated) part of the information provided by the ToxCast in vitro data. The 524 in vitro measures were subjected to k-means Cluster Analysis that pointed to an optimal partition into 6 clusters. The total variance explained by the clusters was 0.25, thus indicating the presence of a large proportion of uncorrelated (singular, or erratic) results. The 6 clusters partition was confirmed by Principal Component Analysis (results not shown).

Table 1 reports representative assays in each cluster. Inspection of the composition of the various clusters indicates that the assays that characterize each cluster have a similar mechanistic meaning: thus, it is possible to assign a biological significance to each of the clusters and give them a name. The names assigned by us to the clusters are the following: Cluster 1: "Cell growth, cell adhesion, and inflammation response"; Cluster 2: "Signaling and regulation (by post-translational modification); Cluster 3: "Nervous system (neurotransmitter receptors)"; Cluster 4: "Metabolism (cytochromes)"; Cluster 5: "Signaling and regulation (by transcription factors)"; Cluster 6: "Regulation (mixed)". Next, we provide as an example the reasoning behind the attribution of the name "Cell growth, cell adhesion and inflammation response" to Cluster 1. The same type of reasoning was applied to the other clusters.

As a matter of fact, Cluster 1 includes as its most representative assays (i.e., most correlated with the center of the cluster), single protein measurements together with global proliferation and cell number counts (Table 1).

The single protein measurements are Thrombomodulin (BSK_3C-SRB_down), Matrix metallopeptidase (BSK_KFCT_MMP9), and P-selectin (BSK_4H_Pselectin_down). The three proteins have to do with cell adhesion and cell-cell communication, and thus with the growth and proliferation features of cell populations. It is worth noting the presence in the same cluster of two of the major mediators of the inflammatory response: CD38 and CD69. Like Thrombomodulin, these are glycoproteins exposed on the surface of cells (in this case of the lymphocites) that permit to sense proliferation messages and to mediate cell adhesion and cell motion properties. On the other side, analogously to P-selectin, CD38 and CD69 are related to shape modifications of cells in response to specific stimuli and are strictly related to the cell movement toward a target.

Shifting to global parameters, Cluster 1 includes four tests of cell proliferation (BSK-hDFCGF-Proliferation_down, BSK-Sag-Proliferation_down, BSK_SM3C_Proliferation_down, BSK_3C_Proliferation down). Particularly interesting is the first one, which is obtained when in the presence of IL1-beta, TNFalfa, IFNgamma, BFGF, EGF, PDGF: these, consistently with CD38, CD69

Table 1: In Vitro Assay Clusters, Composition, and Own-Cluster Correlation

Cluster 1: Cell growth, cell adhesion, and inflammation response	
BSK_SAg_Proliferation_down	0.5535
BSK_SM3C_Proliferation_down	0.5050
BSK_4H_Pselectin_down	0.4725
BSK_3C_SRB_down	0.4663
CLM_CellLoss_72hr	0.4620
CLM_CellLoss_24hr	0.4571
ACEA_IC50	0.4519
BSK_hDFCGF_Proliferation_down	0.4415
ACEA_LOCdec	0.4335
BSK_KF3CT_MMP9_down	0.4310
BSK_3C_Proliferation_down	0.4209
BSK_SAg_CD38_down	0.4040
BSK_SAg_CD69_down	0.4011
Cluster 2: Signaling and regulation (by post-translational modification)	
NVS_ENZ_hPTPSHP1	0.8551
NVS_ENZ_hPTPb	0.8489
NVS_ENZ_hMAPK3	0.8245
NVS_ENZ_hPKBa	0.8245
NVS_ENZ_hPKBb	0.8245
NVS_ENZ_hSRC_Activator	0.8245
NVS_ENZ_hAurA	0.8196
NVS_ENZ_hMAPKAPK5	0.8196
NVS_ENZ_hPTP1b	0.8196
NVS_ENZ_hCK1D	0.8073
NVS_ENZ_hMet	0.8073
NVS_ENZ_hMsk1	0.8073
NVS_ENZ_hSGK1	0.8073
NVS_ENZ_hCASP5	0.7935
NVS_ENZ_hPTPBAS	0.7278
Cluster 3: Nervous system (neurotransmitter receptors)	
NVS_GPCR_hDRD4_4	0.7444
NVS_GPCR_hAdrb1	0.7297
NVS_GPCR_bNPYNon_Selective	0.5850
NVS_GPCR_hDRD2s	0.5666
NVS_GPCR_hM5	0.5080
NVS_GPCR_hAdra2C	0.5078
NVS_GPCR_hM1	0.4930
NVS_GPCR_g5HT4	0.4733
NVS_GPCR_mCCKAPeripheral	0.4653
NVS_GPCR_hM3	0.4400
NVS_GPCR_hOpiate_D2	0.4377
CLM_p53Act_1hr	0.4163
NVS_IC_rCaBTZCHL	0.4087
Cluster 4: Metabolism (cytochromes)	
NVS_ADME_rCYP2C11	0.5743
NVS_ADME_rCYP2B1	0.5581
NVS_ADME_hCYP2C18	0.5537
NVS_ADME_hCYP2C9	0.5529
NVS_ADME_hCYP2B6	0.5365
NVS_ADME_rCYP3A1	0.5051
NVS_ADME_hCYP3A5	0.4832
NVS_ADME_rCYP2A1	0.4590
NVS_ADME_rCYP2C6	0.4590

(Continued on next page)

Table 1: In Vitro Assay Clusters, Composition, and Own-Cluster Correlation (*Continued*)

NVS_ADME_rCYP3A2	0.4463
NVS_ADME_rCYP2A2	0.4448
NVS_ADME_hCYP1A1	0.4208
NVS_ADME_rCYP2D1	0.4207
NVS_ADME_rCYP2D2	0.4012
Cluster 5: Signaling and regulation (by transcription factors)	
ATG_MRE_CIS	0.6014
ATG_BRE_CIS	0.5641
ATG_Oct_MLP_CIS	0.5527
ATG_VDRE_CIS	0.4721
ATG_EGR_CIS	0.4717
ATG_NRF2_ARE_CIS	0.4011
Cluster 6: Regulation (mixed)	
NVS_ENZ_hIKKa	0.5770
NVS_ENZ_hPP2A	0.5583
NVS_ENZ_hPP1a	0.3173
NVS_NR_rAR	0.3023

and Thrombomodulin tests, refer to inflammatory and immune-mediated responses. In addition, CLM-cell loss-72 and 24H are cell toxicity tests linked to variation in cell growth, such as ACEA_IC50 and ACEA_LOCdec.

Overall, Cluster 1 can be assigned a general biological meaning in terms of transduction of primary stimuli with the onset of the following inflammatory response including both cell shape modification and proliferation control. As a matter of fact, the close link between cell adhesion and cell proliferation properties is at present at the center of intense research efforts [19].

Thus, the analysis of in vitro assays points to the presence of statistical correlations that also have biological significance, which is quite an important result. In addition it should be emphasized that the clusters identified in the data preliminarily released on the occasion of the ToxCast Analysis Summit were largely the same, and with the same biological meaning, as those identified in this paper with the definitive results. Thus the robust part of the information from in vitro data (around 25% of total) resists even large "perturbations" of the data.

In order to more formally quantify the in vitro/in vivo relationship, we compared selected in vivo measures with a panel of in vitro measures recognized as belonging to the robust part of the information.

In Vitro Assays as Predictors of Rodent Carcinogenicity

For the purposes of this exercise, aggregated in vivo tumorigenicity results for single rodent species (rat and mouse) were used as target endpoints for prediction. The correlation between rodent carcinogenicity and the robust part

Table 2a: Correlation of Clusters of In Vitro Assays with Rodent Carcinogenicity

	Squared Canonical Correlation (SQCC)	
Assays	Mouse	Rat
Cluster 1	0.03	0.03
Cluster 2	—	—
Cluster 3	—	—
Cluster 4	0.04	0.01
Cluster 5	—	—
Cluster 6	0.01	—

of the information carried by the in vitro assays was assessed as follows. Five central (i.e., mostly correlated within their own cluster) in vitro assays from each cluster were selected, and their correlation with rat (CHR_rat_tumorigen) and mouse (CHR_mouse_tumorigen) carcinogenicity was assessed by Linear Discriminant Analysis. The operation was repeated separately with each cluster. Table 2a shows that the maximum correlation was 0.04 (squared canonical correlation, corresponding to the proportion of variance explained). Using all 30 representative assays from the 6 clusters improved the squared canonical correlation, but only up to 0.04–0.06 for mouse and rat, respectively (Table 2b). Overall, the information carried by the ToxCast in vitro assays is a poor predictor of rodent carcinogenicity for the 309 ToxCast Phase 1 chemicals.

Other In Vitro/In Vivo Correlations

The availability of a rich data set in ToxCast provides the opportunity to explore the in vitro/in vivo relationship based on other in vivo toxicity endpoints besides rodent carcinogenicity. We considered *Rat Cholinesterase Inhibition* and *Mouse Liver Necrosis* endpoints, since the preliminary analysis of in vitro results pointed to the existence of clusters of in vitro assays that appeared directly related—in mechanistic terms—to the above in vivo endpoints.

Rat Cholinesterase Inhibition was modeled with assays in Cluster 3 (Nervous system) and *Mouse Liver Necrosis* was modeled with assays in Cluster 2 (Signaling and regulation) (for the variables in the clusters, see Table 1).

Table 2b: Global Correlation of In Vitro Assays with Rodent Carcinogenicity

	SQCC	
Assays	Mouse	Rat
30 representatives	0.04	0.06

Table 3: Correlation of In Vivo Toxicity Measures with Mouse Carcinogenicity

Variables Entered into the Model	SQCC (cumulative)
DEV_Rat_Urogenital_Ureteric	0.06
DEV_Rat_Skeletal_Appendicular	0.12
DEV_Rat_PregnancyRelated_Maternal	0.13
MGR_Rat_Ovary	0.15
MGR_Rat_Uterus	0.17
DEV_Rabbit_Cardiovascular_MajorV	0.18
DEV_Rabbit_Urogenital_Renal	0.20

Application of Discriminant Analysis showed that in neither case was a correlation between in vitro and in vivo measures (even though putatively related in mechanistic terms) apparent (maximum squared correlation coefficient 0.008; details not shown).

In Vivo versus In Vivo

Since the ToxCast chemicals are characterized also in terms of developmental and multigeneration toxicity measures, it was possible to study the degree of correlation between rodent carcinogenicity (variables CHR_rat_tumorigen and CHR_mouse_tumorigen) and the above in vivo toxicity data. This was performed by applying Discriminant Analysis.

Table 3 shows that the combination of a number of in vivo toxicity assays correlates with mouse carcinogenicity, with an overall squared canonical correlation of 0.20. Similarly, a combination of in vivo toxicity measures correlates with Rat carcinogenicity, with an overall squared canonical correlation of 0.17 (Table 4).

It should be emphasized that (a) these overall in vivo/in vivo correlations are more than double in magnitude than in vitro/in vivo correlations (Table 2); (b) no obvious mechanistic link between carcinogenicity endpoint and the toxicity measures in Tables 3 and 4 can be hypothesized.

Table 4: Correlation of In Vivo Toxicity Measures with Rat Carcinogenicity

Variables Entered into the Model	SQCC (cumulative)
DEV_Rabbit_Orofacial_JawHyoid	0.05
DEV_Rat_General_GeneralFetalPath	0.07
MGR_Rat_Mating	0.09
MGR_Rat_GestationalInterval	0.11
MGR_Rat_Spleen	0.13
DEV_Rabbit_Skeletal_Appendicular	0.14
DEV_Rat_Urogenital_Genital	0.15
MGR_Rat_Kidney	0.17

DISCUSSION

The analysis of ToxCast Phase 1 data provides initial large-scale indications regarding the relationship between in vitro assays and in vivo toxicological endpoints: the overall correlation is extremely weak with rodent carcinogenicity, which is a direct toxicological endpoint, or completely absent in the case of two indirect toxicological endpoints (i.e., *Rat Cholinesterase Inhibition* and *Mouse Liver Necrosis*). This evidence is even more striking in the light of the assumed mechanistic links between the above in vitro and in vivo endpoints. On the contrary, a proportion of the rodent carcinogenicity effects (around 20%) can be modeled by combinations of other in vivo toxicity effects (Tables 3 and 4) for which no obvious mechanistic link can be hypothesized with carcinogenicity. An hypothesis is that the Absorption Distribution Metabolism Excretion (ADME) properties that are typical of the whole animal and that play an important role in determining a compound's bioavailability and ultimate form in relation to various toxicity endpoints are the same for different in vivo toxicological endpoints. In this sense, we could hypothesize that the 20% figure be considered as a rough estimate of the contribution of ADME effects to the final carcinogenicity phenotype.

However interesting, the ToxCast Phase 1 results are limited in both biological and chemical scope, strictly related to the limited sample of chemicals tested. This dataset includes a large proportion of pesticidally active compounds designed to have some biological activity to quickly biodegrade and to be largely non-genotoxic. As shown previously, only about 35% of ToxCast carcinogens contain substructures or functional groups known to be linked to genotoxic carcinogenicity mechanisms, whereas around 75% of the carcinogens in the "traditional" chemical carcinogenicity database contain such functional groups. Hence, this chemical set is representative only of a particular area of the universe of chemicals. As a result, it is crucial to put the present results into a wider perspective and to check if they are supported—or contradicted—by other existing evidence.

Similarly to ToxCast, in recent years an intensive use of genomics and proteomics technologies has characterized the field of drug design. Wide panels of omics and HTS tools are used to identify in vitro promising compounds to be studied to a deeper extent in further steps of the drug development process as well as to predict undesirable toxic effects early in the design process. Here, contradictory evidence is coming to attention: it appears that in recent years the number of new drugs entering the US market has declined sharply, while spending by the pharmaceutical industry on research and development has steadily increased [20]. The two single most important reasons for attrition in clinical development are (a) lack of efficacy and (b) clinical safety or toxicology, which each are estimated to account for 30% of failures [7]. Failures have been largely ascribed to the lack of correlation between effects

Table 5: Replacement and Reduction of In Vivo Toxicological Assays through In Vitro Assays: State-of-the-Art

In Vivo Toxicity	
• Toxicokinetic	
• **Acute toxicity**	***Reduction***
• **Skin irritation and corrosion**	***Replacement***
• Skin sensitization	
• **Eye irritation**	***Reduction***
• Acute systemic and local toxicity	
• **Genotoxicity**	***Reduction***
• **Carcinogenicity**	***Reduction***
• Repeated dose toxicity	
• Reproduction	
• Developmental toxicity	
• Ecotoxicity	

observed in isolated receptors in vitro and those observed in whole animals and in humans [7, 8].

Further evidence on the in vitro/in vivo relationship comes from a recent review paper written by a committee of experts under the aegis of the European Food Safety Agency (EFSA) [4]. The paper provides a very detailed review on the state-of-the-art approaches incorporating replacement, reduction, and refinement of animal testing and examines a range of toxicological or toxicologically related endpoints, from toxicokinetic studies to genotoxicity, reproduction, and developmental toxicity. Table 5 shows a summary of the conclusions of the EFSA review paper, indicating which types of animal toxicity assays can be either replaced or substantially reduced with validated in vitro approaches.

The EFSA review paper points to skin irritation and corrosion testing in animals as the only assays that can confidently be replaced by in vitro alternatives. It should be emphasized that these are very local effects, with no systemic response involved. On the contrary, skin sensitization testing, where a systemic response is elicited, cannot at present be replaced or reduced with in vitro assays. For other endpoints (acute toxicity, eye irritation, genotoxicity, carcinogenicity), in vitro assays can help to direct the animal testing, thus prioritizing use of limited testing resources and reducing the number of animals required and the associated animal suffering. On the other hand, the development of alternative methods has proven to be more difficult for a range of other measures and endpoints, including toxicokinetics, skin sensitization, acute systemic and local toxicity, repeated dose toxicity, reproductive and developmental toxicity studies, and ecotoxicity (e.g., acute and chronic toxicity in fish and birds, bioconcentration in fish) [4].

All the above evidence, ranging from research on drug design with intensive use of omics technologies to more traditional research on alternative tests for regulatory purposes, converge with the initial evidence provided by ToxCast

data: isolated systems in vitro—when perturbed by chemicals—may respond in a way largely different from how they respond when they are integrated into whole organisms. The patterns of interactions among the different systems and organs in the whole organism appear to be major determinants of the response of the organisms, including toxicity, and, obviously, these networks cannot be fully accounted for by studies relying on in vitro isolated systems and pathways [6, 7]. For example, cancer is a condition that arises at the level of tissues and organs, and that calls into action different general responses, such as the immune system. The observed weakness of the in vitro/in vivo relationship suggests that the systemic component is an important force for mediating and transforming initiating events at the molecular level.

The existence of a substantial in vitro/in vivo gap implies that expectations regarding the systematic replacement of the classical animal toxicity assays with stand-alone in vitro alternatives are still quite premature and will not be fulfilled in the near future.

In this regard, an interesting success story is represented by the reduction of animal experimentation in carcinogenicity testing operated by in vitro short-term mutagenicity tests and chemical mechanistic knowledge. After decades of research and the generation of more than 100 short-term tests, it appears that the relationship between mutagenicity and carcinogenicity is valid only within the domain of DNA-reactive chemicals and that only the *Salmonella typhimurium*, or Ames test, has a clear-cut predictivity for chemical carcinogens [21, 22]. In addition, predictivity for chemical carcinogens is shown by the Syrian Hamster Embryo Cell Transformation assay ([23] and our unpublished results). The cell transformation assay detects phenotypic alterations, which are characteristic of tumorigenic cells. It should be emphasized that cell transformation can be produced via a plethora of different molecular mechanisms. On the other hand, *Salmonella* is sensitive to a very large family of carcinogens that are able to interact with DNA according to various molecular mechanisms (e.g., direct or indirect alkylation, acylation, intercalation, formation of aminoaryl DNA-adducts). Thus efficient stand-alone assays together with being based on biological hypotheses confirmed by large databases of results seem to be characterized by a remarkable degree of "aspecificity."

In conclusion, whereas bridging the gap between in vitro and in vivo is still a matter of further research, on the medium term it can be envisaged that progress in reduction of animal testing can be achieved by exploiting the whole range of tools available, such as Structure-Activity Relationship (SAR) concepts [3, 24], in vitro methods, and the combination of all available information into integrated testing strategies. Programs such as ToxCast can play a role by attempting to identify relationships between in vitro response patterns and toxicological effects for subgroups of chemicals/chemical classes that are beyond the reach of existing methods, such as *Salmonella* and Cell Transformation.

ACKNOWLEDGEMENTS

The continued interest and open, cooperative attitude of Ann M. Richard is gratefully acknowledged. This work was partially supported by the following projects: Project EU FP6 no. 037017 **OSIRIS** "Optimized Strategies for Risk Assessment of Industrial Chemicals through Integration of Non-Test and Test Information"; and Project EU FP7 no. 200787 **OpenTox** "An Open Source Predictive Toxicology Framework."

REFERENCES

1. Schmidt C. Testing for carcinogens: Shift from animals to automation gathers steam—slowly. *J Natl Cancer Inst*. 2009;101:910–912.

2. Pedersen F, de Brujin J, Munn SJ, Van Leeuwen K. Assessment of additional testing needs under REACH. Effects of (Q)SARs, risk based testing and voluntary industry initiatives. *JRC report EUR 20863 EN*. 2003, Ispra, EUR.

3. Worth AP, Bassan A, de Brujin J, Gallegos Saliner A, Netzeva TI, Pavan M et al. The role of the European Chemicals Bureau in promoting the regulatory use of (Q)SAR methods. *SAR QSAR Environ Res*. 2007;18:111–125.

4. European Food Safety Authority. Existing approaches incorporating replacement, reduction and refinement of animal testing: Applicability in food and feed risk assessment. *EFSA J*. 2009;1052:1–57.

5. van Regenmortel MHV. The rational design of biological complexity: A deceptive metaphor. *Proteomics*. 2007;7:965–975.

6. van Regenmortel MHV. Reductionism and complexity in molecular biology. *EMBO Rep*. 2004;5:1016–1020.

7. Hopkins AL. Network pharmacology: The next paradigm in drug discovery. *Nature Chem Biol*. 2008;4:682–690.

8. MacDonald JS, Robertson RT. Toxicity testing in the 21st century: A view from the pharmaceutical industry. *Toxicol Sci*. 2009;110:40–46.

9. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer MW, Kavlock RJ. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci*. 2007;95:5–12.

10. Kavlock R, Austin CP, Tice RR. Toxicity testing in the 21st century: Implications for human health risk assessment. *Risk Anal*. 2009;29:485–487.

11. Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM et al. In vitro screening of environmental chemicals for targeted testing prioritization—The ToxCast Project. *Environ Health Perspect*. 2010;218:485–492.

12. Benigni R, Giuliani A. Quantitative modeling and biology: The multivariate approach. *Am J Physiol*. 1994;266:R1697–R1704.

13. Tomatis L, Huff J, Hertz-Picciotto I, Sandler DP, Bucher J, Boffetta P et al. Avoided and avoidable risks of cancer. *Carcinogenesis*. 1997;18:97–105.

14. Benigni R, Bossa C, Richard AM, Yang C. A novel approach: Chemical relational databases, and the role of the ISSCAN database on assessing chemical carcinogenicity. *Ann Ist Super Sanità*. 2008;44:48–56.

15. Benigni R, Bossa C. Structure alerts for carcinogenicity, and the Salmonella assay system: A novel insight through the chemical relational databases technology. *Mutat Res Revs*. 2008;659:248–261.

16. Benigni R, Bossa C, Jeliazkova NG, Netzeva TI, Worth AP. The Benigni/Bossa rulebase for mutagenicity and carcinogenicity—A module of Toxtree. *EUR 23241 EN*. 2008, Luxenbourg, Office for the Official Publications of the European Communities. EUR—Scientific and Technical Report Series.

17. Hawkins DM. The problem of overfitting. *J Chem Inf Comput Sci*. 2004;44:1–12.

18. Topliss JG, Edwards RP. Chance factors in studies of quantitative structure-activity relationships. *J Med Chem*. 1979;22:1238–1244.

19. Evans EA, Calderwood DA. Forces and bond dynamics in cell adhesion. *Science*. 2007;316:1143–1153.

20. Young D. Drug innovation on the decline. *Am J Health Syst Pharm*. 2009;64:228–231.

21. Zeiger E. Identification of rodent carcinogens and noncarcinogens using genetic toxicity tests: Premises, promises, and performance. *Regulat Pharmacol Toxicol*. 1998;28:85–95.

22. Benigni R, Bossa C, Tcheremenskaia O, Giuliani A. Alternatives to the carcinogenicity bioassay: *In silico* methods, and the *in vitro* and *in vivo* mutagenicity assays. *Exp Opin Drug Metab Toxicol*. 2010;6:1–11.

23. Organization for Economic Co-operation and Development. Detailed review paper on cell transformation assays for detection of chemical carcinogens. OECD Series on Testing and Assessment. 2007, Paris, OECD.

24. Benigni R, Netzeva TI, Benfenati E, Bossa C, Franke R, Helma C et al. The expanding role of predictive toxicology: An update on the (Q)SAR models for mutagens and carcinogens. *J Environ Sci Health C Environ Carcinog Ecotoxicol Revs*. 2007;25:53–97.