narcosis. Any of these discoveries must have made him an accomplished medical scientist. Because of his scientific principles, he strictly defined *observers* and *experimenters*, critically. He called observers as those who do not alter "nature", but statically observe the ostensible world; whereas experimenters are those who purposely alter "nature" to obtain a reaction, and seek natural responses behind the phenomenal world. He strongly recommended the use of living organisms to obtain responses, and seek natural reactions behind the phenomenal world. This is the reason why he emphasized the use of vivisection in science throughout his life. In his major discourse on scientific methods, "An Introduction to the Study of Experimental Medicine" (1865), Claude Bernard described what makes a scientific theory good and what makes a scientist important and a true discoverer. Unlike many scientific writers of his time, Bernard writes about his own experiments and thoughts, and uses the first person[2].

Although Bernard was the first scientist who initiated the use of animals in experiments, his original aims at that time were to criticize physicians and to rescue humans from iatrogenic accidents due to poor and insufficient surgical treatments. However, his wife and daughter initiated the first "animal rights campaign" immediately after his death[3], because of their intense aversion to Bernard's animal studies without using anesthesia, namely, vivisection, although this is ironically the best and appropriate method of determining the response of experimental animals.

It is about a century since Bernard started a systematic education on animal experiments. Experimental studies using animals changed last decades because of not only a greater awareness about animal welfare, but also greater decreases in the need for conventional experiments. Accordingly, in 1984, the International Guiding Principles for Biomedical Research Involving Animals was established. Then, in 1985, the European Convention also established the Protection of Vertebrate Animals Used for Experimental and other Scientific Purposes.

In Japan as well also, laws for animal care were successively passed in the 1970's. The Act for Animal Welfare and Proper Administration was passed in 1973, and the Guidance Documents for Experimental Animal Maintenance and Proper Administration in 1980. The Guideline for Experimental Animal Use was established in the same year in 1980 by the Japanese Academy of Science, the Guideline for Proper Use of Experimental Animals in 1987 by the College Union, and the Extension of Animal Life and Ethics by the Japanese Academy of Science in 1996. Recently, the establishment of the Act for Animal Welfare and Proper Use in Experiment was issued in 1999. Despite these guidelines, we could not eliminate all the animal experiments at the moment. However, we are now at the turning point in the history of experimental animal use.

2. Bruce N. Ames – accomplished the most prominent alternative study –

Bruce N. Ames

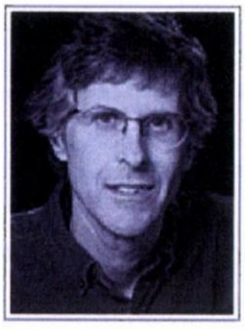
We now introduce a scientist who developed the revertant mutagenesis assay, Bruce Ames[4]. Ames is now a Professor of Biochemistry and Molecular Biology at the University of California, Berkeley. He is a member of the National Academy of Sciences and he was on their commission on life sciences. His publications of more than 450 led to his being among the most cited scientists.

The idea of mutation induced by chemical compounds was first described in 1944[5-6]; this was about 30 years prior to Ames' development of revertant mutagenicity assay. Chemical mutagenesis became the focus of considerable attention, because large amounts of industrial chemicals started to be used in various industries in the mid-twentieth century. Afterwards, because Ames' test enabled the detection of most mutagenic compounds, it has contributed greatly to a large reduction of the number of experimental animals used for in vivo mutagenicity bioassays. What Bruce Ames originally proposed was to use an induced bacterial gene mutation as an evaluation tool for mammalian mutagenesis. He attempted to develop a system for incorporating mammalian microsomal metabolism to the assay also by him, which is presently known as the S9-mixture[7]. It took a very long-time to establish the test system after considerable debate, because scientists at that time had to learn the difference between direct genotoxic carcinogenesis and indirect genotoxic carcinogenesis, namely, epigenetic carcinogenesis. However, after the establishment of the test system, innumerable experimental rats and mice were saved from carcinogenicity bioassay. Although Ames did not intend to save experimental animals by his invention, knowing such common rule of mutagenicity in genetics between Salmonella and mammalians made innumerable number of reduction in experimental animal use possible. Thus, from the study of Ames, our conclusion on alternative studies, is, that an

essential strategy for reaching an alternative goal may be the "Development of True Sciences."

The current major interests of Bruce N. Ames are reported to be the determination of optimum micronutrient intake for minimizing human DNA damage as a preventive measure against cancer, and the study of other degenerative diseases associated with aging.[8-9]

3. Patric O. Brown – gene chip technology



Patrick O. Brown

The third person who we introduced is Patrick Brown, who invented a new methodology, that is, gene chip technology[10]. The gene chip technology and the consequent toxicogenomics[11] that he developed were supposed to rapidly minimize experimental animal use to a large extent[12] (Meeting proceedings from ECVAM-ICCVAM/NICEATM, 2006).

The establishment of the genome sequencing program in 2000 was supposed to be a strong driving force for the progress of alternative studies, particularly via toxicogenomics. All the information derived from animal experiments is incorporated in the genome expression database, that is, "computer mouse", which may be virtually used in the near future even without actual animal experiments.

The method established by Patrick Brown is "molecular microscopy", which enables the differentiation of patterns of gene expression profiles[13]. We showed sample expression profiles of genotoxic compounds studied by the consortium of International Life Science Institute (ILSI), which showed a short-term differential prediction of chemicals with DNA-binding affinity, such as cisplatin, methotrexate, mitomycin C, and chemicals with indirect genotoxicity, such as, taxol, hydroxiurea, and etoposide. Such a rapid and easy prediction may greatly contribute to the realization of essential purposes leading to the development of 3Rs.

Concerning the gene expression profiles, linear increase in dose-response relationship obtained by a conventional testing protocol may not be always applicable each other. In the presented example of microarray data after radiation exposure, because the expression levels of some genes increase with radiation dose and those of some genes decrease with increasing radiation dose, the dose-response relationship obtained by a conventional toxicological testing protocol can be assumed as the only phenomenologic outcome on the basis of one aspect. Rather, we recommend that the dose-response relationship should be considered complex, and that these combination profiles per se, may be essential biomarkers. The authors showed other sample data obtained after whole-body radiation in which one can observe dose-related expression profiles, on one hand, and dose-specific expression profilings, on the other.

Another issue that the authors introduced was age-related stochastic and probabilistic gene expression profilings, which can also be visualized in nontreated senescent mice when one focuses on their individual gene expression. By linear configuration for gene expression, one can clearly recognize that the divergent expression profiles of each individual mouse were not due to an error, but biological diversity with aging. Moreover, representative responsible genes showed clear differences between 2-month- and 21-month-old profiles, which elucidated the age-related responsible gene ontology, represented by the senescence-specific genes[14].

In the cases of experimental myeloid leukemias, spontaneous leukemias are differentiated from those of radiation-induced myeloid leukemias by their different responsible gene intensities in the line configuration of the expression gene profilings. They are also differentiated by the analysis of principal components, which are observed from the three dimensional expression. These databases are also supposed to be essential information for developing 3Rs supported by basic science.

Toxicogenomics sometimes makes the categorical border between physiology and toxicology ambiguous. Similar genes, such as those encoding apoptosis-related genes, caspases, participate simultaneously as physiologic and toxicologic parameters. Toxicogenomics sometimes changes a toxicologic paradigm. Depending on such fluctuating changes in the cell cycle genes, for example, and many other cellular functions, which may be mild or severe, the degree of oscillatory ranges differs from one another, which may be new risk factors.

Conclusion

Lastly, as we mentioned above, the use of experimental animals has, unfortunately, not been completely eliminated to date. Thus, in this regard, we would like to emphasize that "science should progress further". Certainly, one may not accept any risky drugs that have not undergone preclinical animal testing for use in one's children. On the other hand, no one may believe that animal studies will be continued for more than 4-500 years from now. We believe that experimental animals may be eventually replaced by other technical systems developed in the future, although such systems are still technologically immature to replace everything at this moment.

Table 1. Surveillance of experimental animals used in Japan.

	1995	1998	2001
rats*	2.09 (100)	1.53 (73)	1.24 (59)
mice *	6.68 (100)	[redacted]	2.80 (42)

* Million / (%)

Surveyed by the Society of Experimental Animals in Japan[1].

The authors emphasized that animal testing may be eventually replaced by other new technologies, and animal testing would eventually disappear. Some people, however, believe that animal testing should be replaced immediately by other technologies; hopefully today, if not today, maybe tomorrow! These gaps may be filled by nominal driving forces such as humane animal welfare, industrial economy, and politics. However, the essential driving force for this matter may be the development of science itself, particularly by the development of "genome sciences". In other words, an elimination of animal experiments may be l'oiseau bleu (blue bird) of each scientist for the development of future science.

A recent survey by the Experimental Animal Society of Japan showed marked decreases in the number of experimental animals used[15]. As shown in the **Table 1**, the numbers of rats and mice used decreased to 59% and 42%, respectively, since 1995. The possible reason for these decreases is the obtainment of considerably clear-cut experimental results using a relatively small number of genetically modified animals, whereas unreliable experimental results are obtained with a relatively larger number of wild-type animals. These data strongly suggest the future possible reduction in the use of experimental animals.

References

1. Society of Experimental Animal: Annual surveillance of experimental animal use. Experimental Animal News 52 (5): 143-151, 2003.
2. Wikipedia; http://en.wikipedia.org/wiki/Claude_ Bernard
3. Society Against Animal Cruelty and Animal Protection established in 1854.
4. Ames BN, Lee FD, and Durston WE: An improved bacterial test system for the detection and classification of mutagens and carcinogens, Proc. Natl. Acad. Sci. USA., 70: 782-786, 1973.; Ames BN, McCann J, and Yamasaki E: Methods for detecting carcinogens and mutagens with the *Salmonella*/ Mammalian-microsome mutagenicity test. Mutation Research 31, 347-364, 1975.
5. Auerbach C. and Robson JM: Chemical production of mutations. Nature 154, 81, 1944.
6. Muller HJ: Artificial transmutation of the gene. Science 66: 84-87, 1927.
7. Ames BN, Durston WE, Yamasaki E and Lee FD: Carcinogens are mutagens: A simple test system combining liver homogenates for activation and bacteria for detection. Proc Nat'l Acad Sci USA. 70:2281, 1973.
8. Ames, B.N., DNA dames from micronutrient deficiencies is likely to be a major cause of cancer. Mutat Res 2001, 475: 7-20.
9. Liu, J., D. Killilea, and B.N. Ames, Age-associated mitochondrial oxidative decay: Improvement carnitine acetyltransferase substrate binding affinity and activity in brain by feeding old rats acetyl-L-carnitine and/or *R*-α-lipoic acid. Proc Nat'l Acad Sci U S A. 2002. 99: 1876-1881.
10. Brown PO and Botstein D: Exploring the new world of the genome with DNA microarrays. Nat Genet., 21: 33-37, 1999.
11. Inoue T, Pennie WD eds.: Toxicogenomics, Springer-Verlag Tokyo, pp. 1-11, 2003.
12. Corvi R, Ahr H-J, Albertini S, et al. Meeting report: Validation of toxicogenomics-based test systems: ECVAM-ICCVAM/NICEATM considerations for regulatory use. Environ Health Perspect 114 (2): 420-429, 2006.
13. Hirabayashi Y, Inoue T, Toxicogenomics applied to hematotoxicology. In: Joegen Borlak Ed. Handbook of Toxicogenomics-Strategies and Applications. Wiley-VCH, Verlag GmbH & Co. KGaA, Weinheim, 2005, pp. 583-607.
14. Authors: unpublished observation.
15. Society of Experimental Animal: Annual surveillance of experimental animal use. Experimental Animal News 52 (5): 143-151, 2003.

内分泌かく乱化学物質研究の世界的動向

国立医薬品食品衛生研究所
安全性生物試験研究センター
井上 達

1996年にロンドン郊外のWeybridgeで、内分泌かく乱化学物質に関する初めての国際ワークショップが開催されてから10年が過ぎ、昨2007年1年間は、さまざまな回顧と展望が語られた(2007年5月は、DDTの全面禁止につながる『沈黙の春』の著者レイチェル・カーソンの生誕100年でもあった)。この間、2002年には、WHOがグローバルアセスメントをまとめ、2005年11月には、Weybridge10周年を記念したワークショップがヘルシンキで開催され、研究の進捗状況が報告された。

本稿では、内分泌かく乱化学物質研究を、ヒトを含む野生生物が環境との間に営む生体異物相互作用と捉え、この10年間の研究のあゆみの中から、今後の研究に求められていることを述べてみたい。

Weybridgeワークショップからヘルシンキ会議まで

当時、ワニ、カメ、あるいはカモメなど種々の野生生物では、それぞれ血漿エストラジオールの減少、ヴィテロジェニン(卵黄タンパク成分)の増加、あるいは卵殻の菲薄化等、さまざまな異常生態が観察され、それがジコフォール、DDT、あるいはPCBsなどさまざまな農薬、化学物質によるものではないかと危惧され、話題になりはじめていた。これはさらに、ヒトで、尿道下裂のような小児の先天性奇形、あるいは俗にキレるといった表現で表される小児の精神神経学的な障害、さらに乳がんや前立腺がんなどの頻度の亢進などにも、関連があるものとされるに至った(図1)。

こうした中で開かれた欧州連合の主催によるWeybridge会議では、世界保健機構(WHO)と経済開発協力機構(OECD)がそれぞれ役割を決め、この問題に協力

井上 達
国立医薬品食品衛生研究所
安全性生物試験研究センター長

横浜市立大学医学部卒。東京都老人総合研究所、米国ブルックヘブン国立研究所、放射線医学総合研究所などを経て、2001年より現職。専門は、実験病理学、実験動物学、分子毒性学。現在は化学物質の生体影響発現機構の研究を進めている。著書に、「Toxicogenomics」(編著、Springer-Verlag社、2003年)など。

危惧された表徴所見	生物学的蓋然性	疫学的所見	研究課題
先天性奇形（尿道下裂など）	yes	???	発生障害の機構研究（含:胎生期ウィンドウ問題、核内受容体問題など）
生殖能の低下	yes	???	繁殖毒性障害の機構研究 化学物質とホルモン受容体の相互作用研究
精神神経学的障害	yes	???	高次生命系のかく乱の可能性（中枢神経系、内分泌系、免疫系の発生生物学）
発がん性の亢進（乳がん・前立腺がん.）	????	???	組織特異的遺伝子発現;エストロゲン・シグナル応答性遺伝子発現
野生生物所見（雌雄差、など）	yes & ???	——	感受性の種差を裏付ける機構研究（種差の特徴; 生殖腺の形態形成）
低用量反応	yes	——	受容体原性毒性反応機構
相乗／相加反応	yes	——	エストロジェンシグナル応答性遺伝子発現

（米国学士院）

図1　内分泌かく乱仮説と生物学的蓋然性

して取り組むことになった。その申し合わせの1つとして、WHOは内分泌かく乱問題に関する世界的なアセスメント、グローバルアセスメントの編纂を20人ほどの編集委員をもって、3年計画で開始した。編集委員会では個々の事象を1つ1つ検討し、2002年、それまでその存在の如何についてもはっきりしていなかった内分泌かく乱現象に対し、最終的に、内分泌かく乱物質問題がすでに存在する概知の事象であろうと結論づけ、出版物として刊行した。

この時のまとめは4点あり、1)身体が形成されていく過程での曝露が、刺激に対する応答機能を恒久的に変化させてしまう可能性をもつこと、2)成熟した動物への曝露は、ホメオスタシス(恒常性)に基づいた応答により、顕著な影響を示さないかもしれないこと、3)発育段階の違いや季節変動などで同じホルモン様の影響が異なった結果を呈する可能性のあること、そして、4)内分泌系の異なった要素間でのクロストークにより予想外の影響を生ずる可能性のあることなどを挙げた。そして何より、内分泌かく乱化学物質問題は仮説の問題ではなく、すでに存在する既知のことがらであること、従って何らかの対応が必要なことを結論したことがその後に与えた影響には顕著なものがある。

グローバルアセスメントでは、内分泌かく乱化学物質が、成獣には影響を起こしにくいようであること、しかし、胎生期と新生児期の性成熟過程にあたる形態形成期では不可逆的影響を及ぼす可能性がある、といった点を強調した。この指摘により欧州機構や各国の取り組みが本格化した。昨年末のヘルシンキ会議は、さらに5年経って欧州機構などのこの問題に関する研究支援が本格的に軌道に乗った中で行われ、多くの新しい有用な研究情報が紹介された。会議最終日、ダイオキシン受容体研究で知られるジューコ・ツオミストが挨拶に立ち、野生生物の深刻な実態に比較した時のヒトのリスクを次のようにまとめた。「多くの内分泌かく乱物質は、経口的に摂

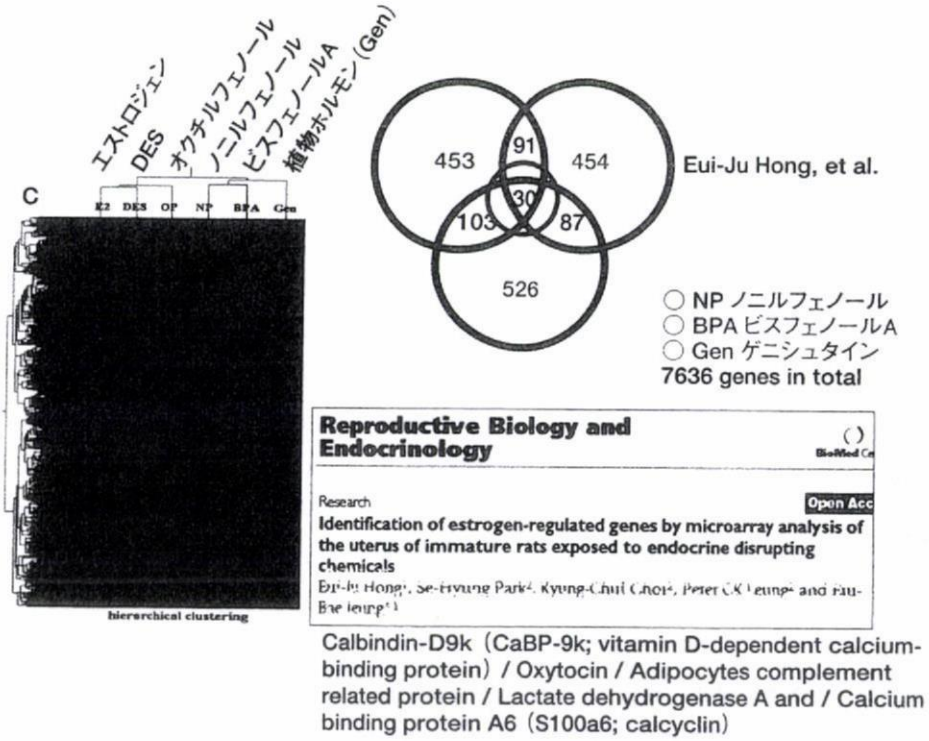


図2　遺伝子発現で見た内分泌かく乱化学物質の性質

左のキルト模様のようなグラフは、生体がどう反応しているのかを遺伝子の反応パターンで示している。左図のエストロジェン、DES、オクチルフェノールなどは、いずれも女性ホルモン様の作用をもつとされている。しかし、グラフに明らかなとおり、横一直線に共通した反応は見られない。右のバイグラフも同様であり、ノニルフェノール、ビスフェノールA、ゲニシュタインの3つの化学物質で共通に発現する遺伝子は30個検出されるのみである。つまり、内分泌かく乱化学物質といっても、ヒトに共通に影響を与える目印になる遺伝子は見られず、したがって、十把一絡げにその特徴を判断するわけには行かない、ということがわかる。

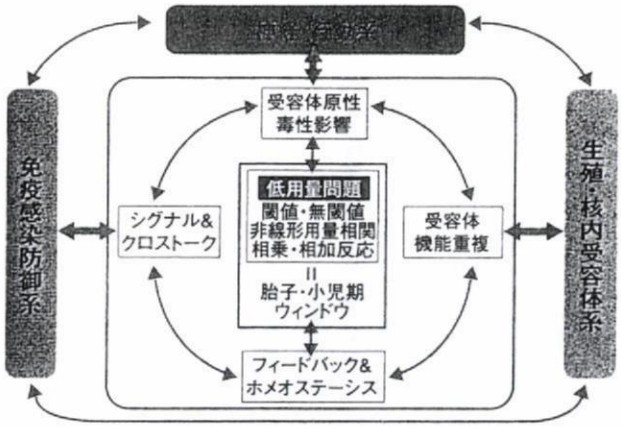


図3　分泌かく乱化学物質の高次生命系への影響（模式図）
神経・行動系、免疫感染防御系、そして生殖・核内受容体系、の3系統に重点を置いた研究が進められている。これらの臓器では、ホルモン受容体が普段から発現していて、共通のさまざまな因子の発現が観察される。それぞれが長期の記憶装置をもっていることも特徴である。

取されるであろう。しかし、ヒトの摂取する食物と水とは、その衛生管理が整っており、そこで、ヒトへの曝露は、事実上認められない」というものである。おそらく、そうした紙一重の状態が、ヒトへのリスクを押さえているというのが実態なのだろう。

新しくわかったこと、今後必要性な研究

以上がこれまでの概略であるが、初期、危惧されていた事柄などを通覧すると、人々の身のまわりにはホルモン様物質がたくさんあり、その中では生体ホルモンでさえ、本来、ある程度生体に有害な性質をもつ。雌雄対偶動物の身体の中ではこれが"漏れ出さないよう"緻密な自己防護システムが備わっていること、あるいは内分泌かく乱化学物質とよばれるものに、物質としての共通作用の乏しいこと（図2）、さらには、そうした化学物質の複合作用の有無など、多くの事柄で、多分に整理されないまま、机上の論議が行われてきた感が否めず、実験生物学の立場から見ると、充分に的を射た議論が行われてこなかったように思われる。

他方、筆者らは、内分泌かく乱化学物質の生体影響研究では、ホルモン受容体が普段から発現していて、いろいろな共通の補助因子の発現が観察され、それぞれが長期の記憶装置をもっていること、神経・行動系、免疫・感染防御系、そして生殖・核内受容体系などの諸系統に注目することの重要性を、これまでにも啓蒙書[1]を出版したりして強調してきた（図3）。幸いなことに、この考えはかなり妥当な判断だったようで、一連の研究からは、0.01～0.2mg／kg／dayという無作用量以下の低用量のビスフェノールAに、神経・行動異常を引き起こす作用が見出されており[2]、諸分野で、注目すべき新たな知見と今後の研究の展望が明らかに成りつつある。ここでは、特記すべき点のみ、かいつまんで紹介する。

1．用量－反応曲線と低用量への外挿性

毒性学による安全性の試験では、高用量の反応から直線回帰をして低い用量での反応性を予測することが多い。したがって、低用量での作用が明らかになりつつある内分泌かく乱化学物質の作用曲線について、無作用量以下の低用量域で外挿表徴型と異なった反応がないかどうか

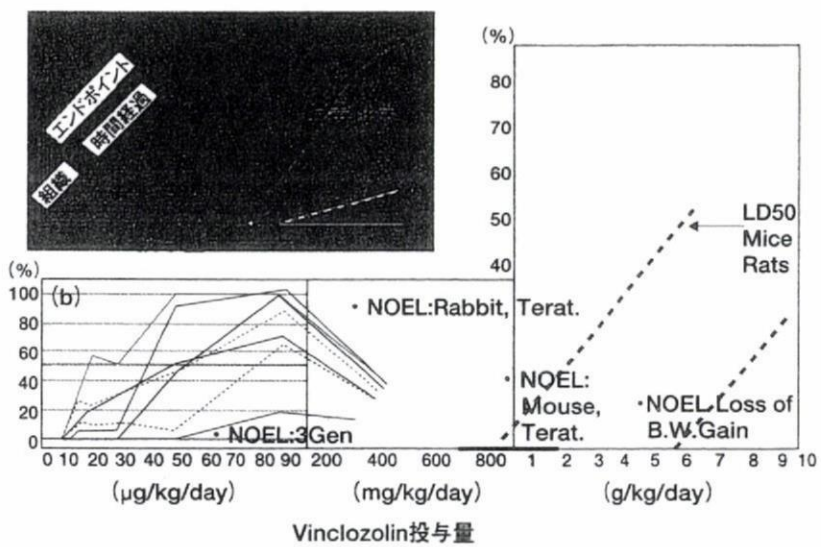


図4　低用量域の変化と稀少特性
既存のデータを整理すると、従来は、無作用量、無毒性量と定めていた用量より低い濃度で、さまざまな影響が観察されることが報告されている。
（b：米国の国立環境影響研究所NIEERLのEarl Gray博士による）

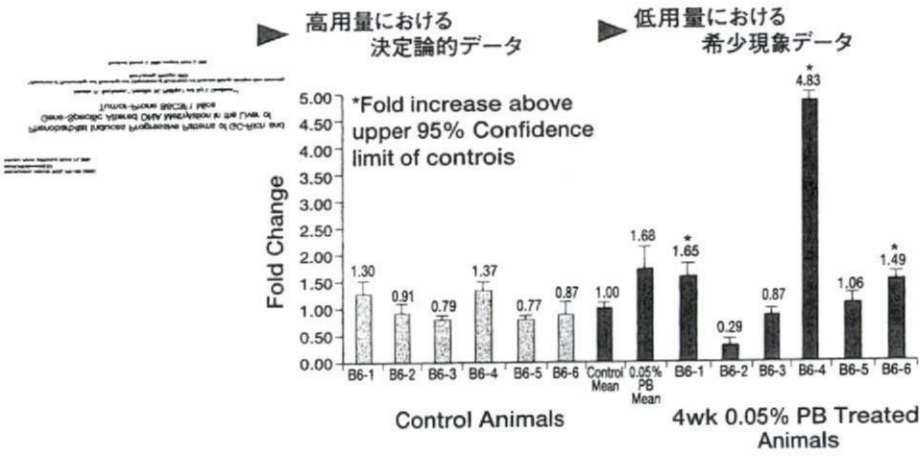


図5　低用量域の変化と稀少特性

低用量域で種々の試験法に沿って実験をすると、純系動物を用いた実験でも、変動幅の広い結果得られることがある。図は、フェノバルビタールによるメチル化部位の形成確率を、ネズミー匹毎に検出したもの。対照群と異なり、個体毎の値が広く分布している。内分泌かく乱化学物質影響にもこうしたエピジェネティック特性が想定される。(ミシガン州立大学のJay Goodman教授による)

の如何については、早くから議論があった。そして実際に既存のデータを整理すると、従来無作用量とか、無毒性量と定めた用量より低い濃度でさまざまなデータが認められた。米国環境防護庁所轄の国立環境影響研究センターでは多年度計画を立てて検討に入っている[3] (図4)。

2. 低用量反応の問題点

メチル化は、たとえ純系動物でも確率論的に形成され、等質の結果が出ないことがわかってきた (図5)。こうした結果に対して平均値を取ると、ばらつきは背景データに隠れてしまうので、"プロクラステスのベッド"*で知られる通り、禁忌とされる。

*プロクラテスは、一面に宝石を散りばめた黄金のベッドの寸法よりも客の寸法が短いと、ベッドの寸法に合うよう4人の力持ちの大男に引き延ばさせたり、ベッドよりも長いと頭や足を切り落とさせたという。

3. 発がん蓋然性の問題

発がん性との関係では、発がんの蓋然性、つまり、これらの物質によってがんが起こりやすそうな体内環境が形成される可能性があるのかどうかという点である。女性ホルモンとダイオキシン類の1つが、正常のmycという遺伝子と協同して、テロメラーゼという、細胞を無限増殖へ導く遺伝子の活性を引き上げた、という報告がある[4]。こうした変化は、がん化につながる可能性があるため、早急に検証することが必要である。

4. 思春期早発の蓋然性と加齢影響

内分泌かく乱化学物質の性質の1つとして、思春期を早く発来させ、早期の老化を引き起こすなどといった点に危惧があった。図6に示したように、ビスフェノールAの効果では、投与した動物の寿命曲線が、グラフの白○印の対象に較べて、死亡が早期化し、傾きも急峻になる傾向があることがわかる[5]。もしこの寿命曲線が正しいとすると、これまで見てきた生殖リスク評価センター(CERHR)の判断とはまったく違った結果になるわけであり、これについても、よく調べる必要がある。

5. 内分泌器官の拡張

さて、次に注目すべき点は、結論から述べると、これまで内分泌器官と一括して述べたが、よく調べてみると、肝臓や脂肪細胞など、従来必ずしも内分泌器官と考えてこなかった臓器が、内分泌器官の役割をしていると、い

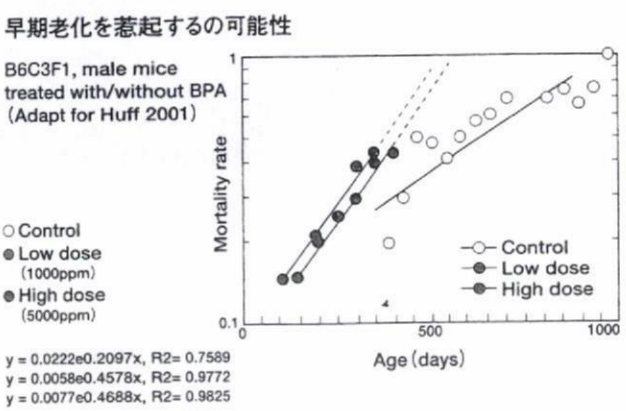


**図6　ビスフェノールAによる
エピジェネティック発がんと促進老化**

内分泌かく乱化学物質の性質の1つとして、思春期の早発や早期老化が危惧されてきた。ビスフェノールAを投与した動物の寿命曲線は、白○印の対象に較べて、死亡が早期化し、傾きも急峻になる傾向が見られる。ビスフェノールAは、エピジェネティック発がんの促進が見られるようである。(文献5より引用改変)

A critical period, called the window, is known during pre- and post-natal life at various time points at/during which organism seems to be particularly sensitive to exposure to various chemicals.

図7　低用量問題と胎生期ウィンドウ

内分泌かく乱化学物質は、形態形成期の臨界投与期に障害を引き起こすのでこれをウィンドウ効果とよんでいる。臨界投与量(0.003mg/kg)のDESという物質で尿道下裂が引き起こされる時期も、生後7日の新生児期に限局している(基生研、井口泰泉教授による)。世界保健機構が子どもの健康に力をいれる所以である。

えなくもないということである。たとえば、ノニルフェノールという物質は、通常の方法で見るとごく弱い女性ホルモン様の作用をもっているのであるが、肝臓に注目すると、女性ホルモンよりも、ずっと強い活性をもっていることが、基礎生物学研究所・井口泰泉教授らの研究でわかった。すると、肝臓は、内分泌器官なのだろうかという疑問がもたれる。

6. 内分泌の概念の拡張

続いて、内分泌器官の概念の拡張にかかわる事柄である。異物受容体とよんでいるダイオキシン受容体は、エストロジェンが存在しない状態では、P300と名づけているタンパク分子の助けで転写活性化を担い、なんと、女性ホルモン様の作用をもつことがわかった。しかも、エストロジェンがある時は、この分子は、これと反対にUbiquitin ligaseとよばれる複合体を形成し、エストロジェン受容体を壊して、抗女性ホルモン様の役割を発揮することもわかってきた。このように、ホルモン受容体でもない、こうした生体内分子としての異物受容体が、ホルモン様の作用を発揮するということは、内分泌かく乱問題の分子的基盤が、大きく拡大していくことを示している。これは非常に大きな驚くべき問題といわねばならない。しばらくして、日本発のこの東京大学分子細胞生物学研究所・加藤茂明教授のグループによる研究は、Natureに紹介された[6]。

おわりに

最後に一言、"子どもの問題"に触れて、本稿を終了したい(図7)。

低用量問題は、胎生期ウィンドウに密接に関連していることを前述した。これは神経・行動学的不可逆性にも関連している。したがって、内分泌かく乱化学物質研究は"子ども"の毒性学研究に重なり合うところが大きい[7]。近年、"子どもは小さな大人ではない"、ということが指摘されるようになった。WHOはこの点を重視して事業目標の重点に据えている。この概念が広く認識されることの意義には疑義がない。そうした前進にもかかわらず蛇足として指摘したい点は、この"子どもの特性"が量的な差異ではなく、質的な差異に基づいている、ということである。しばしば、"子どもの反応性"が脆弱で可塑的であるから安全係数を負荷するといった試みがなされている(たとえば米国EPAでは、一定の判断の下に5を除する)。筆者は、これはまさに"子どもを小さな大人"と扱ったもの以外の何物でもないことを指弾している。つまり"子ども"には、数値で大人と比較しきれない、本質的に大人と異なった事柄がある、という蓋然性は、ここでは考慮に入れられていないに等しい。この点を正しく認識するためには、成長過程の"子ども"についてのさらに深い研究が必要である。

[参考文献]

1) 井上達, 井口泰泉：生体統御システムと内分泌かく乱, シュプリンガー・フェアラーク東京, pp.324, 2007.
2) Kubo K, et al.：Neurosci. Res., 45: 345-356, 2003.
3) Office of Research and Development, US-EPA: Multi-year plan (FY2000-2012) for endocrine disruptors. http://www.epa.gov/nerl/goals/ecosystem/edc.html
4) Sarkar P, et al.：Intn'l. J. Oncol., 28: 43-51, 2006.
5) Huff J：Odontology, 89: 12-20, 2001.
6) Ohtake F, et al.：Nature, 423: 545-550, 2003.
7) http://cfpub.epa.gov/ncer_abstracts/index.cfm/fuseaction/display.researchCategory/rc_id/842

Meeting Report: Validation of Toxicogenomics-Based Test Systems: ECVAM–ICCVAM/NICEATM Considerations for Regulatory Use

Raffaella Corvi,[1] Hans-Jürgen Ahr,[2] Silvio Albertini,[3] David H. Blakey,[4] Libero Clerici,[5] Sandra Coecke,[1] George R. Douglas,[4] Laura Gribaldo,[1] John P. Groten,[6] Bernd Haase,[7] Karen Hamernik,[8] Thomas Hartung,[1] Tohru Inoue,[9] Ian Indans,[10] Daniela Maurici,[1] George Orphanides,[11] Diana Rembges,[5] Susanna-Assunta Sansone,[12] Jason R. Snape,[13] Eisaku Toda,[14] Weida Tong,[15] Joost H. van Delft,[16] Brenda Weis,[17] and Leonard M. Schechtman[18,19]

[1]European Centre for the Validation of Alternative Methods (ECVAM), Institute for Health and Consumer Protection (IHCP), Joint Research Centre of the European Commission (JRC), Ispra, Italy; [2]Bayer HealthCare AG, Wuppertal, Germany; [3]Hoffmann-La Roche, Basel, Switzerland; [4]Environmental Health Centre, Health Canada, Ottawa, Ontario, Canada; [5]Physico-Chemical Exposure, IHCP, JRC, Ispra, Italy; [6]TNO, Utrecht, the Netherlands; [7]QIAGEN, Hilden, Germany; [8]U.S. Environmental Protection Agency, Washington, DC, USA; [9]National Institute of Health Sciences, Tokyo, Japan; [10]Health Safety Executive, London, United Kingdom; [11]Syngenta, Macclesfield, United Kingdom; [12]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, United Kingdom; [13]AstraZeneca, Brixham, United Kingdom; [14]Organisation for Economic Co-operation and Development, Paris, France; [15]Food and Drug Administration, National Center for Toxicological Research, Jefferson, Arkansas, USA; [16]University of Maastricht, Maastricht, the Netherlands; [17]National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA; [18]U.S. Interagency Coordinating Committee on the Validation of Alternative Methods, Research Triangle Park, NC, USA; [19]U.S. Food and Drug Administration, National Center for Toxicological Research, Rockville, Maryland, USA

This is the report of the first workshop "Validation of Toxicogenomics-Based Test Systems" held 11–12 December 2003 in Ispra, Italy. The workshop was hosted by the European Centre for the Validation of Alternative Methods (ECVAM) and organized jointly by ECVAM, the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). The primary aim of the workshop was for participants to discuss and define principles applicable to the validation of toxicogenomics platforms as well as validation of specific toxicologic test methods that incorporate toxicogenomics technologies. The workshop was viewed as an opportunity for initiating a dialogue between technologic experts, regulators, and the principal validation bodies and for identifying those factors to which the validation process would be applicable. It was felt that to do so now, as the technology is evolving and associated challenges are identified, would be a basis for the future validation of the technology when it reaches the appropriate stage. Because of the complexity of the issue, different aspects of the validation of toxicogenomics-based test methods were covered. The three focus areas include *a)* biologic validation of toxicogenomics-based test methods for regulatory decision making, *b)* technical and bioinformatics aspects related to validation, and *c)* validation issues as they relate to regulatory acceptance and use of toxicogenomics-based test methods. In this report we summarize the discussions and describe in detail the recommendations for future direction and priorities. *Key words:* acceptance, alternatives, biomarker, predictive test, regulatory use, standardization, toxicogenomics, toxicology, validation. *Environ Health Perspect* 114:420–429 (2006). doi:10.1289/ehp.8247 available via *http://dx.doi.org/* [Online 17 August 2005]

Toxicogenomics, an emerging field in molecular toxicology, offers the promise of new approaches to identify and characterize such factors as the biologic activity of new and existing chemicals and drugs and could play an important role in hazard assessment for human health. This revolutionary field can potentially affect many scientific and medical areas, including the development of a new generation of alternative predictive testing and screening methods that could lend themselves to the reduction, refinement, and replacement of animals used for such purposes.

The European Centre for the Validation of Alternative Methods (ECVAM), the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), and the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) are currently investigating the specific considerations necessary for adequate validation of toxicogenomics-based test methods. The primary objective of ECVAM and ICCVAM/NICEATM is to facilitate development, validation, and regulatory acceptance of new, revised, and alternative test methods that reduce, refine, and replace the use of animals (referred to as the three Rs; Russell and Burch 1959) in testing while maintaining and promoting scientific quality and the protection of human health, animal health, and the environment. The efforts of such organizations as ICCVAM/NICEATM and ECVAM have helped foster the principles of the three R's and have contributed to progress in the use of alternative methods for regulatory, research, and educational purposes.

Experience in the validation of conventional alternative test methods has led to an understanding that new and innovative approaches likely will be necessary to standardize test methods based on toxicogenomics and to evaluate the scientific validity and regulatory applicability of such test methods. It is envisioned that the entire validation process will be more complex and challenging than that typically encountered thus far for other alternative test methods. This is because not only will the technology itself need to be standardized and validated, but the methods that are based upon the technology and their predictive aspects will also need to undergo validation if they are to be employed in regulatory decision-making processes. In addition the validation process must be able to accommodate the anticipated rapid changes in technology that could affect the performance of the test method and its reliability for a specific purpose.

Toxicogenomics-based methods are being widely applied in toxicology and biomedical research. Because data are already being generated using these technologies, it is both timely and important to address the subject of validation now with the aim of establishing a foundation that will facilitate future regulatory acceptance of scientifically validated

Address correspondence to R. Corvi, European Centre for the Validation of Alternative Methods, IHCP, Joint Research Centre of the European Commission, Via E. Fermi 1, 21120 Ispra, Italy. Telephone: 39-0332-785266. Fax: 39-0332-785845. E-mail: raffaella.corvi@jrc.it

Supplemental Material is available online (http://ehp.niehs.nih.gov/members/2005/8247/suppl.pdf).

We gratefully acknowledge W.S. Stokes, Director of the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, for contributions to the planning and organization of the workshop as well as insightful and constructive comments.

This document represents the consensus of the participants' views expressed as individual scientists and does not necessarily represent the policies and procedures of their respective institutions.

The authors declare they have no competing financial interests.

Received 22 April 2005; accepted 17 August 2005.

toxicogenomics-based test methods. By addressing the critical validation issues early, and in parallel with the evolutionary and maturation phases of the technologic development of toxicogenomics-based methods, it should be possible to preempt many potential pitfalls and data gaps encountered with retrospective method evaluations that could impede validation of this promising research and regulatory tool. Such a strategy will also facilitate early buy-in and confidence in the technologies by the regulatory arena in its quest for new, improved, and relevant methods by which to help ensure human health, protect the environment, and demonstrate responsiveness to animal welfare issues.

In consideration of all these related issues, ECVAM and ICCVAM/NICEATM held the first of a planned series of workshops to address the validation principles that lend themselves to toxicogenomics-based test methods, for example, gene expression technologies and associated bioinformatics. Given the complexity of the rapidly evolving toxicogenomics field, a variety of issues were addressed. These included but were not limited to *a*) differences in and evolution of technology platforms including changes in genome coverage for model species; *b*) quality assurance (QA) and Good Laboratory Practice (GLP) compliance; *c*) technology standardization, transferability, and reproducibility; *d*) relevance to *in vivo* biological responses; *e*) yardsticks against which toxicogenomics responses should be measured; *f*) data evaluation, statistical approaches, and databases; *g*) validation approaches; and *h*) regulatory acceptability.

To begin to examine these complex issues, three breakout groups were formed. Each group concentrated on different aspects of the validation of toxicogenomics-based test methods, and the discussions were shared with the other participants in plenary sessions. The three focus areas were *a*) biological validation of toxicogenomics-based test methods for regulatory decision making, *b*) technical and bioinformatics aspects related to validation, and *c*) validation issues as they relate to regulatory acceptance and use of toxicogenomics-based test methods.

Validation of Toxicogenomics: Focus on the Biological Systems

The biological issues related to the validation of toxicogenomics-based test methods involved two strategies proposed for developing and validating such methods so that they can be employed to support regulatory decision making. One strategy involves phenotypic anchoring of gene expression changes to identify molecular mechanisms and candidate biomarkers of toxicity (i.e., single genes, proteins, or biological pathways). A second strategy involves the identification and validation of predictive gene expression signatures of toxicity. Validation considerations specific to data quality and cross-platform and interlaboratory variability that are common to both strategies were identified. It is acknowledged that any new toxicogenomics-based methods will need to address established validation criteria for determination of reliability and relevance (Balls et al. 1995; ICCVAM 1997, 2003) as well as articulate the advantages and limitations of a given toxicogenomics-based test method. In addition biological validation of such a test method, that is, assessment of the concordance of gene changes with biological events, is essential but is contingent upon validation of the technology itself, which is addressed elsewhere in this article.

Strategy 1: use of toxicogenomics data to define mechanism and identify biomarkers. Toxicogenomics offers the opportunity to enhance existing toxicity prediction strategies through elucidation of biological mechanisms around critical events. This sentiment is captured in the recent U.S. Environment Protection Agency (EPA) and U.S. Food and Drug Administration (FDA) strategies regarding the inclusion of genomics data in submissions of regulated substances (U.S. EPA 2002; U.S. FDA 2005). Although these agencies currently preclude basing regulatory decision making on genomics data alone, they do encourage the voluntary submission of well-documented, quality genomics data. Both agencies are considering the use of submitted data on a case-by-case basis for assessment purposes (e.g., to help elucidate mechanism of action or contribute to a weight-of-evidence approach) or for populating relevant comparative databases by encouraging parallel submissions of genomics data and traditional toxicologic test results. This approach is appropriate given the state of scientific knowledge of toxicogenomics and the requisite need for a clear understanding of the toxicologic relevance of the gene expression signals detected by this technology. There is a small but rapidly increasing number of published reports demonstrating a linkage between gene expression changes and adverse phenotypic changes (Huang et al. 2003; Orphanides 2003). These reports provide qualitative evidence of the power of genomics to link phenotype with gene expression, thereby contributing to an understanding of mechanism of action. Some such reports demonstrate the predictive power of these data to classify compounds. However, they fail to address adequately quantitative dose- and time-dependent (e.g., threshold) responses that are the hallmark of toxicologic evaluation, making their immediate acceptance in regulatory arenas circumspect.

Nonetheless, toxicogenomics data may eventually be useful in hazard and risk assessment if data quality and validity can be adequately substantiated. Some regulators are finding that these data have the potential to add to the body of knowledge about compound mechanism of action. With appropriate dose- and time-dependent measurements, gene and protein changes can be used to mark the molecular events that occur as an organism moves through the continuum from exposure to response. The obvious benefit is the identification of early markers of response, including responses that mark the point of departure from adaptation to toxicity. In addition, it may be possible to detect unforeseen effects at very low doses or in unexpected tissues (Brown et al. 2002). This is important because changes in gene or protein expression alone are not sufficient to differentiate toxicity from biologic adaptation after exposure to an exogenous compound. The challenge for predictive toxicology is to link changes in gene and protein expression to sequential changes in phenotype, both adaptive and adverse, in a manner that is consistent with the underlying biologic mechanisms. For example, gene expression profiling has been used to classify hepatotoxins based on mechanism of action and to differentiate early, presumably adaptive, responses from later responses that are reflective of toxicity (Hamadeh et al. 2002a, 2002b; Waring et al. 2001, 2003). The gene expression changes correlated well with changes in histopathology and clinical chemistry, supporting the liver as target organ for the test compounds.

Although good technical progress has been made in recent years, additional proof-of-principle studies are needed for the regulatory community to become more accepting of the use of toxicogenomics data as part of the regulatory decision-making process. It would be important to demonstrate, for instance, that toxicogenomics not only can confirm what is already known about specific compounds and toxic end points (i.e., phenotypic anchoring) but also can accurately predict toxicity for unknown compounds. The task is to present regulatory scientists with new knowledge gained from toxicogenomics approaches in a familiar context. Ideally, at least in the short term, the focus will be the identification of single, or small sets of, genes or proteins that serve as biomarkers of response, as opposed to signatures of response that are the typical output of microarray experiments. Simple biomarkers of response are favored over complex expression signatures because they are familiar in toxicology assessment, are easy to maintain over time (e.g., are independent of the microarray platform), and can be readily validated. Validation strategies for toxicogenomics-based markers can be modeled after protocols for existing biomarkers. Thus, global gene expression technologies such as microarrays can be used to identify a specific gene marker,

or a suite of markers, that can then be validated by conventional methods such as Northern blot analysis, *in situ* hybridization, and quantitative polymerase chain reaction. This approach has advantages because regulatory agencies such as the U.S. FDA have proposed procedures to address gene and protein biomarkers, and other organizations, such as the Organisation for Economic Co-operation and Development (OECD 2005), are embarking on establishing similar guidance (Supplemental Material, Section 1; http://ehp.niehs.nih.gov/members/2005/8247/suppl.pdf).

Proof-of-principle studies could be conducted concurrently with existing regulatory test methods using similar samples of test compounds. In such situations, it may be appropriate to use *in vivo* systems, which are widely accepted by the regulatory community. Parallel *in vitro* studies could be conducted in situations where an appropriate test system is available. It may be wise to focus initial efforts on defining relationships between gene expression changes and toxicity for individual compounds or compound classes with well-defined end points. The experimental design should address conventional aspects of dose and time (dose response), species and strain susceptibility, group size and sex, and selection of end points for study (e.g., histopathology, clinical chemistry). Numerous commercial microarray platforms offer genomewide coverage for model systems such as rat, mouse, *Caenorhabditis elegans*, and humans. Commercial microarrays are also available for genes that are highly expressed in specific tissues (e.g., liver, breast) and during specific biological processes such as metabolism (e.g., P450 enzymes). Both genomewide and dedicated arrays can be used with RNA samples from *in vivo* and *in vitro* (tissue and cell culture) systems, enabling parallel studies to be conducted with a single microarray platform. This is important because the results of microarray experiments can vary depending on the array design and the selection and performance of gene probes on the array. Encouraging results on cross-platform comparisons and between-laboratory reproducibility are now emerging (Bammler et al. 2005; Chu et al. 2004; Irizarry et al. 2005; Larkin et al. 2005; Yauk et al. 2004). Toxicogenomics studies conducted in parallel and comparative systems can demonstrate the biologic relevance of *in vitro* models as surrogates for *in vivo* models without the need to address cross-platform (technologic) issues (Boess et al. 2003; Huang et al. 2003). Although initial efforts should focus on defining simple gene and protein biomarkers for specific compound classes, end points, and model systems, the end goal is to establish a compendium of compound-specific knowledge that transcends technology platform. Ideally, the markers should be robust enough to withstand technologic advances in toxicology that add to the existing knowledge about the compound. Once sufficient and adequately validated data are available, toxicogenomics can become part of a hierarchical approach to compound assessment.

The use of toxicogenomics to identify (screen) compounds with the potential to cause adverse effects may present opportunities to reduce the need for full animal tests, or perhaps refine animal use, and/or reduce the numbers of animals needed when *in vivo* tests are necessary. Of course, the statistical power of any test will influence the number of animals used in an *in vivo* test as well. Screening-type assessments may be appropriate for priority setting, dose setting, chemical ranking, and so forth. The extent of validation required for screening tests may be different than that required for full replacement tests because negative compounds might still undergo full animal testing. Establishing a compendium of compound-specific information will enable regulators and sponsors to access what is known about a compound across multiple test systems, species, and end points, thereby improving the biological relevance of regulatory decisions to safeguard human health and the environment.

Strategy 2: use of gene expression signatures to predict toxicity. Toxicogenomics holds great promise for improving predictive toxicologic assessments. Gene expression profiling has been used to classify compounds by chemical class and mechanism (Hughes et al. 2000; Scherf et al. 2000; Steiner et al 2004; Thomas et al. 2001), tumors by origin and type (Chung et al. 2002), and breast cancer patients for follow-up chemotherapy (van 't Veer et al. 2002). In all cases, classification was based on a set of discriminatory gene elements, between 10 and several hundred, identified from a larger pool of genes on a microarray. The pattern of gene expression, not the measurement of a single or a small set of genes, was the basis for classification. A variety of gene expression analysis algorithms were used to discriminate samples based on gene expression signature. In all cases, the compound class or tumor status was known *a priori*, and gene expression signatures for known samples were used to predict classification for other known but blinded samples (Blower et al. 2002; Brindle et al. 2002). Such models are currently being developed in the private sector (e.g., Gene Logic, Iconix) and are commercially available but cannot, as yet, be exploited by regulators and the scientific community because the underlying data sets and algorithms have not been made available outside the private sector.

Predictive model development will require an extensive "training" set of gene expression measurements for classes of model compounds in a variety of test systems, both *in vivo* and *in vitro*, at multiple doses and time points. Initial studies can be conducted concurrently with conventional testing systems as a way to confirm model predictions. In the short term, it is unlikely that sufficient data will be available for gene expression signatures to replace conventional approaches. Until then, such data can be used as part of a hierarchical approach to toxicity testing in conjunction with accepted methods routinely used for regulatory purposes. In the long-term, sufficient data should accumulate from well-designed validation studies such that gene expression signatures could be part of a battery of tests that reduce or replace animal procedures.

Model validation will necessitate multiple independent data sets and application of sophisticated statistical approaches. Acceptance of these models will require that research and regulatory communities have access to the data analysis tools used to build the models, and that they become familiar with the limitations and uncertainties of using these complex computational models. Confidence in and acceptance of these models will also require rigorous performance standards and appropriate controls to ensure reproducibility and stability over time (see below) and adequate sensitivity and specificity to discriminate toxic from nontoxic responses. Initial model development could easily be accelerated through coordinated sector-spanning efforts. Coordinated efforts across academia, government, and industry partnerships will accelerate progress in defining gene sets that are robust and discriminatory both within and across technology platforms. This is an ideal scenario given the rapidly advancing pace of technology development.

An important aspect of any toxicogenomics validation strategy is the need to measure the range of biological variability of gene responses for a given test system. Ideally, this should be accomplished by one species, tissue, and end point at a time, in order to adequately assess cross-species differences that often hamper risk assessments. Measurements of biologic variability under baseline and toxicant-challenged conditions will enable regulators to better discriminate biologically relevant responses from baseline homeostatic fluctuation. This is an important issue for toxicogenomics, as studies conducted on cell culture populations demonstrate a wide range of biological variability in gene expression measurements for individual cells under both baseline and challenged conditions (Kuang et al. 2004). Therefore, it is necessary to define criteria to adequately address biological variability in a data submission and to establish whether the burden of maintaining these data is that of the regulator or sponsor.

The recommendations related to the biological validation of toxicogenomics-based test methods are listed in Table 1.

Standardization and Validation of Toxicogenomics-Based Methods: Focus on the Technology

Considerations given to validation of the technology encompassed the technical and bioinformatics issues related to the validation of toxicogenomics-based test methods. The starting premise adopted was that with the availability of bioinformatics expertise, biological data generated from toxicogenomics studies could be interpreted with a high degree of confidence. The ultimate aim was to identify a strategic approach that would enable credible biological observations and consequential judicious regulatory decisions, and that this approach would be independent of the toxicogenomic platform used. Moreover, standardization and validation of toxicogenomic platforms were seen as essential for identifying and reducing technologic artifacts. Standardization would also be required to increase the certainty by which biological observations could be extrapolated across and between different microarray platforms. It is therefore important to build on the learning of previous and ongoing efforts in standardization of toxicogenomics (reviewed by Sansone et al. 2004).

Three distinct levels where validation is necessary were identified (see Figure 1 and discussion below). The first level of validation is the responsibility of the array manufacturer or provider and has to be performed only once. This can be seen as a "one-off validation" and relates to both the microarray quality and the instrumentation. The second level of validation is the responsibility of both the experimental toxicologist and the array manufacturer or provider. This can be seen as "routine validation" or best practice to allow data comparability. It encompasses quality control (QC) aspects of the critical experimental components and is a process that occurs on a regularly scheduled basis. The third level of validation, that is, determination of reliability and relevance, is needed every time a change is introduced into the test procedure. Performance standards developed based upon the original test method would serve as the criteria against which the revised method would be compared. Despite these multilevel validation needs, it was repeatedly emphasized that significant technologic development and progress in microarray platforms are still under way and that efforts to validate and standardize these technologic platforms must not be at the expense of innovation.

One-Off Validation

The one-off validation is the responsibility of the array manufacturer or array provider. This is required to ensure that the array platform being used is robust and that the inherent variability within the platform is transparent to the user and the regulator (Figure 1). The following were identified as being necessary for microarray-based toxicogenomics to be used in regulatory assessments:

- Microarrays should be fabricated in accordance with the principles of Good Manufacturing Practice (GMP).
- Specifications and performance criteria for all instrumentation and method components should be available.
- All quality assurance/quality control (QA/QC) procedures should be transparent, consistent, comparable, and reported.
- The array should have undergone sequence verification, and the sequences should be publicly available.
- All data should be exportable in a MAGE (MicroArray and Gene Expression)-compatible format.

Table 1. Recommendations: focus on biological systems.

- Encourage increased use of toxicogenomics-based approaches to define the mechanistic context of toxic responses to exogenous compounds
- Promote greater understanding of the relationships between gene expression responses and altered phenotype, considering the biological pathways affected, dose response, and the point of departure from adaptive to toxic response
- Favor the identification of biomarkers that are independent of technology platform but acknowledge the potential strengths of pathway analysis
- Characterize the range and extent of biological variability of responses for the test systems (e.g., diurnal effects, animal care and use, age-related context)
- Encourage the immediate use of toxicogenomics-based approaches in conjunction with conventional toxicity testing approaches
- Explore the extent to which toxicogenomics can address cross-species responses and specific disease states
- Promote the conduct of parallel and comparative *in vivo* and *in vitro* studies to identify *in vitro* systems that can serve as surrogates for *in vivo* systems
- Characterize predictive toxicology models with respect to parameters such as dose, time, study design, relevance; characterize the system to fulfill validation criteria
- Promote the identification of gene and protein biomarkers as early (prognostic) markers as a refinement to existing toxicity testing methods
- Establish a compendium of toxicant information based on gene expression responses for model compounds across multiple species, end points, and test systems
- Foster the development of effective partnerships between academic, government, and industry groups to promote collaborative efforts to validate toxicogenomics-based test methods and generate sufficient high-quality data to support regulatory decision making

Routine Validation

Routine validation is an ongoing process that is the responsibility of the experimental toxicologist and the array manufacturer or provider (Rockett and Hellmann 2004). Again, for microarray-based toxicogenomic assays to be used in regulatory decision making the following important factors were identified (Figure 1):

- Oligos, cDNAs, or clones that are arrayed should be randomly sequence-verified to ensure that no errors are introduced between batch syntheses. This verification process should be recorded and reported by the manufacturer
- All reagent components should be identified. Reagents should be prepared according to GMP and/or GLP as appropriate. Data regarding batch variability should also be recorded and reported
- Common reference RNA standards (housekeeping genes) should be adopted to facilitate comparison between array platforms. This may be achieved in collaboration with the international Microarray Gene Expression Data (MGED) Society and other related efforts (see below).

Biological standards. Performance standards, test component standards, and QC measures are key components of any validation strategy for a toxicologic test method. Establishing standards is particularly important for gene expression technologies due to the inherent technologic and biological "noise" in these systems. Commonly used biological standards are reference RNAs that are competitively hybridized with the sample of interest in two-channel array formats, and *in vitro* RNA transcripts that are "spiked into" RNA samples of interest in either one-channel or two-channel array formats. Establishing accepted RNA standards will address concerns of regulatory reviewers about data quality and variability within and between laboratories and across different technology platforms. The standards will also provide a common benchmark for regulators to assess platform performance over time. To achieve this goal, we must establish standards that maintain a defined level of accuracy, sensitivity, specificity, and reproducibility across platforms.

Reference RNAs can be derived from tissue extracts, cell lines, or both and serve a variety of purposes. Workshops sponsored by governments and industry have focused on defining the specifications for reference RNAs for clinical and regulatory applications (Joseph 2004). The consensus is a that multiple RNA standards are needed to measure the accuracy, dynamic range, sensitivity, and specificity of varied technology platforms under varied conditions. Important questions are whether regulatory agencies will define preferred sources of RNA standards, and, if so, who will generate and maintain baseline information about these

standards. Although the selection of a given RNA standard depends primarily on the purpose and application, all RNA standards should be tested for a clearly defined number of copies of a given sequence within an RNA preparation over some linear range (Cronin et al. 2004).

Some initiatives are raising awareness of the effects of variables that might hamper data comparability and are working toward developing best practice guidelines for microarray-based measurements (Hopkins et al. 2004). For example, recommendations for best practice in array normalization, together with performance characteristics in terms of sensitivity, accuracy, and comparability of different array platforms (cDNA and oligo, spotted and *in situ* synthesis), are beginning to emerge together with proposals for transparency and availability through publicly accessible databases (http://www.vam.org.uk). Other initiatives are considering the use of quality metrics for standardizing and validating array-based toxicogenomics measurements. The extent to which such efforts will be pursued and the impact they will have upon the standardization issues that are a necessary prerequisite to the validation exercises remain to be seen.

Quality assurance and Good Laboratory Practice. GLP is intended to promote proper documentation, quality, and authenticity of toxicity test data and is required for data acceptance by regulatory agencies (e.g., U.S. FDA, U.S. EPA). At the international level, GLP has been promulgated under the OECD guidelines program (OECD 1998). As part of the progression toward regulatory acceptance, toxicogenomics experiments should ideally be conducted in accordance with GLP. However, at present, most large-scale toxicogenomics efforts are not arising from GLP-compliant laboratories, and requiring compliance for data submission could greatly hamper the technical advancement of new technologies and retard their migration into the regulatory arena. To avoid discouraging technologic progress while maintaining a level of GLP conformity, it could be argued that for research and technical development and improvement purposes, it might be acceptable if array-based studies could at least measure up to the reporting standards required by GLP. However, with the adoption of the toxicogenomics-based technologies into regulatory decision-making practices, GLP compliance undoubtedly will be expected. Procedural aspects of GLP compliance not currently captured in MIAME-Tox (minimum information about a microarray experiment for toxicogenomics) will need to be identified but can be incorporated over time. Until then, it may be possible to allow for proof-of-principle and prevalidation studies to be conducted in accordance with the "intent" of GLP practices by requiring submitters to adequately document procedures and control measures and make experimental data open to regulatory review. "Best practices" for toxicogenomics can be established until formal procedures are adopted. This may be a more realistic solution that permits the advancement of science while addressing the need for QA and QC.

Validation as a Result of Procedural Changes

This third level of validation is necessary whenever a technical or methodologic change is introduced into the test. Such changes might, on one hand, be restricted to the microarray technology (e.g., modification or addition of sequences to a microarray, changes in data analysis procedures). Alternatively, they could involve the experimental design (e.g., dose, time, cell culture procedures). One consideration is that a distinction between minor and major procedural changes that might be incorporated into a test would help determine the extent of such validation necessary. Additionally, to facilitate the process, performance standards should be defined based upon the original validated test procedure. Minor changes would entail a demonstration of equivalence of results obtained with the modified test to that obtained from the validated test. Major changes would involve the need to define a new set of reference materials to be tested and a more extensive validation. Guidance on the use of performance standards and the elements comprising them have been published (ICCVAM 2003) and have been employed for *in vitro* dermal corrosion assessment methods (ICCVAM 2004). Such guidance can also help facilitate the establishment performance standards for toxicogenomics-based test methods in which procedural modifications have been introduced after an initial validation exercise, thereby providing a basis for the comparison of reliability and accuracy of the modified method relative to the validated and accepted reference test method.

The concept of performance standards was originally developed to evaluate the acceptability (accuracy and reliability) of proposed test methods that are based on similar scientific principles and that measure or predict the same biologic or toxic effect as an accepted (previously validated) test method. Because some regulatory authorities and international test guidelines programs (e.g., OECD) have restrictions regarding the use of proprietary test methods (methods that are copyrighted, trademarked, or patented), performance standards also allow for the development and validation of comparable nonproprietary methods based on performance standards derived from the corresponding proprietary antecedent method. Under these circumstances, performance standards allow the characteristics and functional attributes of a proprietary method or technique to be described and offer a procedure for evaluating the performance of methods claimed to be substantially similar. A method that meets the established performance standards is

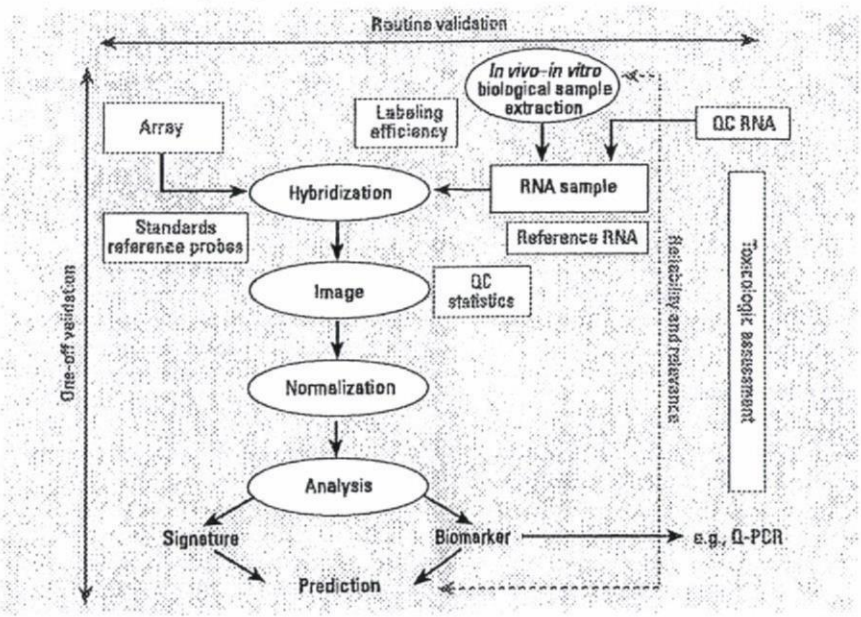


Figure 1. Scheme of the different steps in a toxicogenomics-based test. Three distinct levels were identified where validation is necessary: one-off validation (left), which should be performed once and is mainly related with the quality of the microarray and the instrumentation (blue); routine validation and QC (top), representing the ongoing requirements that are the responsibilities of the experimental toxicologist and the manufacturer (red); and the extent of validation necessary whenever a technical or methodologic change is introduced in the test (right): a method should meet the preestablished performance standards in order to be considered reliable and relevant as the original test method (green). Q-PCR, quantitative PCR.

considered sufficiently accurate and reliable for the specific testing purpose for which it is designed and is viewed as comparable with the original test method upon which it is based. If the correct performance standards have been developed, a method for which the results have the same accuracy and reliability as the original should by definition also be as relevant as the original method.

The conceptual framework and scope of performance standards could be expanded or adapted to include innovations or advancements in areas such as microarray or protein or metabolite separation and identification technology, where proposed improvements might or might not be generally or completely analogous to those in existing systems but would still enable similar applications. Performance standards could still provide a gauge for evaluating newer or revised technologies to ensure that their reliability and accuracy were at least comparable with that of existing acceptable techniques using similar chemicals even if essential test method components (i.e., structural, functional, and procedural elements of a validated test method to which a proposed, mechanistically and functionally similar test method should adhere) were not substantially similar.

This level of validation, which does not imply that a test needs to be completely revalidated, is of extreme importance for tests based on rapidly evolving technologies. It would be a mistake to immobilize these technologies by enforcement of a strict and inflexible validation approach that would hamper progress and test improvement. Finally, a periodic reassessment of a test method's performance (accuracy and reliability) employing established performance standards would help ensure adherence to essential test method components and the reliability and accuracy of the modified test method relative to the validated antecedent method (Hartung et al. 2004). Such assurance could be best established and reported by international validation bodies such as ECVAM and ICCVAM/NICEATM, which could track the history, performance, and validation status of a given test.

Data Management

The lack of robust QC procedures and capture of adequate metadata has caused problems with the analysis and reproducibility of array-based transcriptomics investigations. Consequently, the international MGED Society proposed standards for publication (Nature 2002) that were designed to clarify the MIAME guidelines (Brazma et al. 2001). As a result, a number of journals now require that articles containing microarray experiments must be compliant with the MIAME standard; some also require that the data integral to the article's conclusions be submitted to the ArrayExpress database at the EBI (European Bioinformatics Institute) (Brazma et al. 2003), GEO (Gene Expression Omnibus) at NCBI (National Center for Biotechnology Information) (Edgar et al. 2002), and CIBEX (Center for Information Biology Gene Expression database) at DDBJ (DNA Databank of Japan) (Ikeo et al. 2003)—the European, American, and Japanese database counterparts, respectively.

There is a critical need for public toxicogenomics databases because of the significant volume of data associated with these experiments, the complexity of comparing different gene annotations and splice variants across platforms, and the need for a resource for complex informatics analyses of the traditional toxicology and microarray data in parallel. However, to fully achieve the potential of this emerging interdisciplinary field, it is necessary that we move toward the establishment of a common public infrastructure for exchanging toxicogenomics data (Mattes et al. 2004). The infrastructure should address *a*) the technical problems involved in data upload, *b*) the demand for standardizing data models and exchange formats, *c*) the requirement for identifying minimal descriptors to represent the experiment, *d*) the necessity of defining parameters that assess and record data quality, and *e*) the challenge of creating standardized nomenclature and ontologies to describe biological data. The goal is also to create an internationally compatible informatics platform integrating toxicology/pathology data with transcriptomics, providing the scientific community with easy access to integrated data in a structured standard format, facilitating data analysis and data comparison, and enhancing the impact of the individual data sets and the comprehension of the molecular basis of actions of drugs or toxicants. Ultimately, such a knowledge-base could be maintained (respecting confidentiality as appropriate) as a reference for regulatory organizations to evaluate toxicogenomics and pharmacogenomics data submitted by registrants to those organizations.

The potential exists for the international development of this public infrastructure. As part of the collaborative undertaking with the International Life Sciences Institute Health and Environmental Sciences Institute (ILSI–HESI) Technical Committee on the Application of Genomics to Mechanism Based Risk Assessment (http://www.hesiglobal.org/committees), the European Molecular Biology Laboratory of the European Bioinformatics Institute (EMBL–EBI; Brazma et al. 2003; http://www.ebi.ac.uk/microarray/Projects/toxnutri/index.html), the National Institutes of Health/National Institutes of Health National Institute of Environmental Health Sciences National Center for Toxicogenomics (NCT; Waters et al. 2003; http://www.niehs.nih.gov/nct/), and the U.S. FDA NCT (Tong et al. 2003; http://www.fda.gov/nctr/science/centers/toxicoinformatics/index.htm) have worked closely together. The respective databases are based on the international standards developed by the MGED Society (Brazma et al. 2001; Spellman et al. 2002). After the very favorable response that the MIAME received from the microarray community and key scientific journals (Ball et al. 2002, 2004; Nature 2002), the MIAME checklist was extended to describe array-based toxicogenomics experiments. The MIAME-Tox checklist (MGED 2004) is an attempt to define the minimum information required to interpret unambiguously and potentially reproduce and verify array-based toxicogenomics experiments. MIAME-Tox also supports a number of other objectives, for example, linking data from different experimental domains within a study and linking several studies from one institution and exchanging toxicogenomics data sets among public databases. The major objective of MIAME-Tox is to guide development of toxicogenomics databases and data management software. Without a sufficient depth of data in these resources, the scientific community's opportunity to develop consensus on analysis and application of these data for risk assessment or screening may be limited. The availability of this level of information regarding platform specification, appropriate common reference standards, and the toxicologic study alone will facilitate the predictive value of toxicogenomics across different array-based platforms. This, in turn, will result in a greater appreciation of and confidence in the value of toxicogenomics within a regulatory context, such that testing strategies can be optimized, predictive alternative models can be identified, and animal use can be reduced (Supplemental Material, Section 2; http://ehp.niehs.nih.gov/members/2005/8247/suppl.pdf).

Moreover, the long-term provision of a MIAME-Tox–compliant database with a MAGE-ML (Microarray Gene Expression Markup Language) export is required for the long-term storage of toxicogenomics data. This would directly support the role of ECVAM, ICCVAM/NICEATM, and other validation bodies in the validation of toxicogenomics-based test methods.

The recommendations related to the technical and bioinformatics aspects of validation are listed in Table 2.

Regulatory Acceptance of Validated Toxicogenomics-Based Methods

Regulatory scientists are increasingly being called upon to consider incorporation of toxicogenomics data in regulatory assessment processes that involve evaluation of potential human health or environmental hazard and risk. Those scientists will need to be able to

judge the level of confidence to place in both *in vivo* and *in vitro* toxicogenomics-based test methods and the resulting data that might be submitted in support of regulatory decision making. Whether a method has been determined to be valid for a specific purpose will be an important factor for the consideration of its use for regulatory purposes. Furthermore, the level of confidence held by regulators will influence regulatory acceptance of methods and data, and will affect both the further pursuit of toxicogenomics technologies and technologic improvements and the extent of industry application of these technologies.

Potential uses of toxicogenomics data in the regulatory area. The potential of toxicogenomics-based methods in contributing to regulatory assessment processes is broad. Examples might include, but would not be limited to, obtaining microarray data from individual *in vivo* bioassays or *in vitro* cell or tissue-based assays or from batteries of assays, using conventional or high-throughput approaches. In accordance with the current developing state of the science, realistic possibilities for initial uses of toxicogenomics data in regulatory settings might be first in the realm of hazard assessment, such as to support chemical mechanism of action arguments. Other early uses might include aiding individual chemical/chemical mixture screening or ranking exercises to set priorities for toxicity testing or to sort chemicals into batches. These types of applications might involve identification of individual genes or gene patterns associated with particular toxic effects or pathways, adaptive responses, or metabolic pathways. However, global pattern recognition–type techniques are, as yet, not considered to be ready to fully replace traditional bioanalytical methods for predicting toxicity or elucidating information on mechanism of action or biochemical pathway component identification.

Using only human or animal *in vitro* or *in vivo* data derived from toxicogenomics technology to estimate such parameters as adverse/no adverse effect levels or to determine dose–response relationships for conducting risk assessments is regarded as a much longer term goal. However, for hazard assessment purposes, the possibility of considering toxicogenomics data along with other types of toxicologic information and data [e.g., from *in vivo* and *in vitro* studies, determinations of quantitative structure–activity relationships (QSAR) or SAR] in a weight-of-evidence approach on a case-by-case basis was not discounted. Regulatory bodies have begun to craft preliminary proposals, policies, and guidance for the submission and use of omics-type data in regulatory deliberations and to provide encouragement for the use and further development of the technology (U.S. EPA 2002; U.S. FDA 2005). Additionally, organizations such as the OECD are actively working with member countries on approaches that seek to harmonize the use of omics-derived information for hazard assessment related to health and environmental effects.

Harmonization of toxicogenomics-based test methods will first necessitate the standardization and validation of the specific test protocol(s) developed for a specific purpose(s), as conducted by international validation bodies such as ECVAM and ICCVAM/NICEATM. It will then be important for such organizations to interface with the OECD to ensure the appropriate crafting of harmonized OECD toxicogenomics-based test guidelines that are based upon standardized, adequately validated procedures, that are considered practical, and that permit consistent regulatory judgments.

Case for a modular approach to validation. Because of the extraordinary rate at which toxicogenomics technologies are evolving, current validation processes might need to adapt so as to accommodate the rapidly developing changes and advancements while still observing the basic tried-and-true validation principles. To meet this anticipated need, a modular approach to validation (Hartung et al. 2004) was considered, not to abridge the process but to allow for more flexibility in data collection and evaluation throughout the progressive changes that the technology will undergo. Typically, in the conventional validation procedures for an alternative test method, a sequential approach to the process is taken. The test protocol is first optimized and its transferability is determined. The resulting standardized method is then evaluated for within-lab and between-lab reproducibility and for its accuracy. Thus, an optimized, standardized protocol linked to specific test method elements and a prediction of outcome for given classes of chemicals are evaluated together for performance characteristics and applicability. Such a linear validation model, although effectively employed for other test methods, might not be optimal for dynamic test methods in which changes are rapidly introduced that improve or alter the protocol or the technology incorporated in the protocol in any substantive way. The linear validation model might result in unnecessary delays in incorporating innovations into toxicogenomics-type test methods. In contrast, with a modular approach to validation, which capitalizes on the fundamental classic concepts of validation as defined by ECVAM and ICCVAM (Balls et al. 1995; ICCVAM 1997, 2003), the different steps in the validation process are subdivided into independent modules, each of which can be assessed individually so that those components that have been completed need not undergo repeated validation. Further validation activities would instead be directed to only that part of the process flow where needed. The proposed model would accommodate validation of innovation affecting only a particular part of the sequence such that incorporation of advancements in a particular sector into testing strategies would less likely be impeded. At the same time, a modular approach to validation could efficiently handle information/data gaps that could be filled over time without derailing the validation stages already achieved. The modular approach, complemented with the use of performance standards (see "Validation as a Result of Procedural Changes" above), is expected to facilitate and help expedite the validation of the toxicogenomics technology and test methods that are based on toxicogenomics.

The modular approach follows the fundamental classic concepts of validation as defined by ECVAM and ICCVAM. Validation is defined as the process by which the relevance and reliability of a test method for a specific purpose are determined (Balls et al. 1995; ICCVAM 1997, 2003). Adequate validation

Table 2. Recommendations: focus on technology.

- Validation and QA/QC should be mandatory during the manufacturing of the arrays
- The array should undergo sequence verification and sequences should be available in the public domain
- MIAME guidelines should be adhered to
- Initially, develop "best practices" for toxicogenomics, including the interpretation of data and how to manage uncertainties and limitations
- Subsequently develop guidance for and adherence to GLPs for toxicogenomics experiments
- Common reference standards should be considered
- A workshop should be convened to address the development of standards for RNA sample preparation (and other biologic aspects of microarray analyses)
- Develop a "common" RNA standard including developing consensus about sources and maintenance of baseline data for regulatory and research purposes
- Studies should be MIAME-Tox compliant
- Performance standards should be developed and implemented to evaluate reliability and accuracy of test methods incorporating procedural modifications
- An ongoing dialogue should be maintained between scientists in the various relevant disciplines, including bioinformaticians, through meetings, published papers, and advisory/discussion panels (e.g., ILSI-HESI committee, NCT consortium, OECD panel)
- Ensure that validation efforts and QA/QC criteria are not restrictive to the technology or its advancement
- Explore whether toxicogenomics measurements can define toxicologic effects quantitatively
- Develop prediction models (e.g., algorithms) for toxicogenomics-based test methods
- Develop a data infrastructure for capturing, storing, and reporting toxicogenomics data
- Ensure continuation of financial support for long-term public database maintenance

involves development of a standardized test method protocol and assessment of the protocol's within- and between-laboratory variability, predictive capacity/accuracy, usefulness and limitations, and adherence to performance standards.

Standards for comparison. As technologic advancements are made and new, modified, or revised toxicogenomics-type test methods are put forward for consideration, it will be necessary to have a means by which the performance of proposed methodologies can be compared with that of existing (traditional and nontraditional) methods, especially those that employ animals. The lack of an approach rooted firmly in high-quality science could jeopardize attempts to seek or gain regulatory acceptance of toxicogenomics-based test methods and strategies. Evaluations of test method performance might be based on comparisons made between particular parameters, as dictated by the specific intent for which the assay was developed. Examples include the following:

- *In vivo–in vivo* study comparisons to examine concordance of gene changes with such factors as onset, duration, severity, dose, age, possible temporal changes of effects, and species differences
- *In vitro–in vivo* study comparisons to explore gene changes associated with a critical event or end point in an *in vitro* cell-based assay and an established *in vivo* biomarker of toxicity
- *In vitro–in vitro* study comparisons to analyze the responses of human and animal cell systems to xenobiotics
- Technologic comparisons to evaluate the effects of proposed technical improvements (e.g., comparing gene changes using different techniques of array/platform preparation)

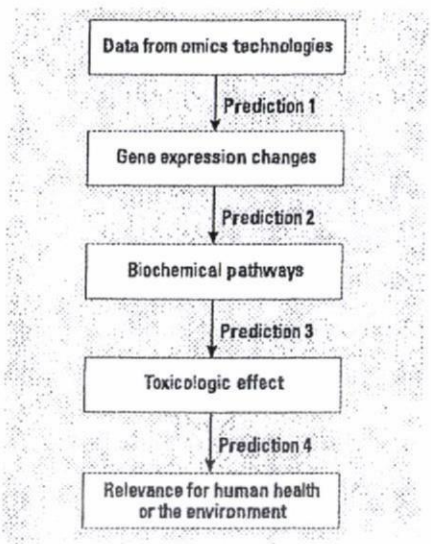


Figure 2. Process flow showing different independent prediction levels considered important in assessing validity of a toxicogenomics-based test method.

Accordingly, to determine the appropriate types of validation activity and comparison in a given situation, it is important that the specific purpose of the proposed methodology and a detailed description of all relevant procedures be clearly elaborated (Balls et al. 1995; Hartung et al. 2004; ICCVAM 1997, 2003).

Toxicogenomics data from **in vitro** *systems and data relevance.* At the present time, toxicogenomics data derived from *in vitro* systems have been considered to have limited utility in regulatory applications. However, a great deal of interest exists for the further development of *in vitro*–based toxicogenomics methods, for an examination of their potential applicability in the regulatory arena, and for an appraisal of their potential for contributing to improvements in animal welfare. It is anticipated that technologic advancements will ultimately facilitate the use of *in vitro*–based methods as adjuncts to or surrogates for *in vivo*–based methods. Possible areas where validated *in vitro*–based toxicogenomics test methods might play a future role include *a*) preliminary assessments (prescreens), *b*) complementary testing that might assist in obtaining additional (e.g., mechanistic) information, and *c*) surrogate tests that could help in the refinement, reduction, and replacement of animals used for omics-based or traditional testing methods. One exciting aspect of toxicogenomics technology is the prospect of being able to identify species differences and/or similarities in the response to a xenobiotic. Although this is not viewed as near-term prospect, it obviously has potential applications for hazard and risk assessment purposes and could also have an impact on previous regulatory decisions when the technology becomes sufficiently advanced to permit such uses for it.

Additional regulatory acceptance issues. In considering approaches to validation, achieving regulatory acceptance of toxicogenomics-based methods or acceptance of information/data derived from such methods is an important goal. Regulators will be asked to evaluate whether data submitted using omics technologies can be used in support of a particular or broader based toxicologic, pharmacologic, or physiologic premise. For example, experiments using microarrays demonstrated increased expression of a cluster of related genes that was associated with enhanced activity and production of a microsomal enzyme important in the metabolic activation of a chemical to a toxic entity, which in turn was associated with a histopathologic biomarker lesion in the liver with a known human cancer correlate. Each of the events in this example can be thought of as a sequence of separate critical steps or information levels (Figure 2) that progressively connect omics data (from microarrays) to gene expression changes (increased expression), to a biochemical pathway (liver enzyme induction leading to toxic metabolite formation), to a toxicologic effect *in vivo* (liver lesion) with human relevance (cancer). Moving between two levels involves a prediction of outcome linking both steps. At each of these prediction junctures, regulators would be looking for evidence to scientifically substantiate moving to the next step and whether the prediction linking the levels (e.g., in this example, prediction 1, 2, 3, or 4 in Figure 2) was adequately validated. Theoretically, with this type of system, validated links could be established between any two levels. Technologic advancements or new information could be independently incorporated into a given level and considered and evaluated for the specific relevant prediction juncture. In this way, each of the prediction levels can be assessed independently and the validity of the links determined.

In the future toxicogenomics-based test methods may be shown to have been adequately validated and technically suitable for certain specific purposes, but regulatory acceptability and implementation will depend partly on whether the methods validated can be used for a given regulatory agency or program, that is, they are applicable to the products that fall within their regulatory purview. Some regulatory bodies may have internal peer-review processes, specific regulatory mandates, and/or regulatory assessment procedures that also have a role in the determination of test method applicability in regulatory programs, even though a test method may have been appropriately validated.

The widespread use of omics technologies will also bring about increasing demands on the regulatory community in terms of training of regulatory personnel in areas such as potential applications; data QC, analysis, and interpretation; statistical analysis; limitations of the technology; and how the information might be incorporated into safety, hazard, and risk assessment processes. To satisfy these needs, regulatory agencies have been engaging in developing and implementing training procedures, hiring scientists with the necessary technical knowledge and experience, establishing centers of excellence and dedicated laboratories focused specifically on the various omics and related informatics areas [e.g., National Center for Toxicological Research (U.S. FDA), NCT (NIEHS), Minister of Health Labour and Welfare–National Institute of Health Sciences Project in Japan, Netherlands Genomics Initiative, and EMBL–EBI, where informatic scientists are working with experimental practictioners and the MGED Society to ensure that transcriptomic experiments can be mapped on to regulatory toxicology studies]. In addition the regulatory arena has found that maintenance of open lines of communication with appropriate external scientists facilitates cooperation and the sharing of technical aspects, skills,

and practical experiences that help to broaden the collective knowledge base. Regardless, as the technology evolves further and finds wider application and acceptance, it will be necessary to address such fundamental matters as *a*) the generation, management, and interpretation of massive amounts of data; *b*) the consequent complex questions that will undoubtedly arise (e.g., what constitutes an adverse effect as identified using the technology; how does a given gene pattern correlate with a particular toxic end point or relate to onset, duration, and severity of effects, and to age, dose, and species?); and *c*) the limitations to the technology. Addressing such issues efficiently will warrant an ongoing dialogue between regulators and practitioners and a willingness to share relevant experiential and theoretical knowledge. Standard submission and presentation formats compatible with electronic data submission likely would need to be developed. Programs and staff would need to learn how information from the new technologies might be incorporated in regulatory practices and decision-making processes and would also have to face possible incongruities between toxicogenomics-derived data and existing or future submissions of conventional toxicity data. A number of regulatory authorities have already begun to contemplate and make provisions for this enormous and challenging task, but others may not yet have committed the resources to do so.

The recommendations related to regulatory acceptance and use of toxicogenomics-based test methods are listed in Table 3.

Conclusions

This workshop was organized as a result of the rapid growth and technologic advancements in the field of toxicogenomics; the promise it offers for numerous scientific arenas, especially human health and the environment; and the interest demonstrated by regulatory agencies as well as by the industrial sector. Consequently, it has become apparent that a considerable effort needs to be invested in the appropriate validation of both the technology alone and those test methods that incorporate the technology. The workshop provided a platform for technical experts in the field to become cognizant of the validation principles and regulatory issues to be encountered and for regulators and principal validation bodies to gain a better sense of those technologic aspects that would lend themselves to standardization, harmonization, and validation. Thus, this workshop was an important initiative that fostered an exchange of information fundamental to the ultimate adoption of toxicogenomics-based test methods for regulatory decision-making purposes. It is envisioned that the conclusions and recommendations that resulted will be a basis for future validation considerations for test method applications of toxicogenomics technologies in the regulatory arena and evaluating their potential utility for hazard/safety/risk assessments.

Several aspects of the validation of toxicogenomics that were identified as needing further exploration to help facilitate regulatory acceptance of future toxicogenomics-based test methods are as follows:

- Conduct toxicogenomics-based tests and the associated conventional toxicologic tests in parallel to *a*) generate comparative data supportive of the use of the former in place of the latter or *b*) provide relevant mechanistic data to help define the biological relevance of such responses within a toxicologic context
- Determine and understand the range of biologic and technical variability between experiments and between laboratories and ways to bring about greater reproducibility
- In the short term, favor defined biomarkers that are independent from technology platforms, and therefore are easier to validate; in the longer term, focus on pathway analysis (i.e., system biology approach) rather than just on individual genes
- Harmonize reference materials, QC measures, and data standards and develop compatible databases and informatics platforms that are key components of any validation strategy for a toxicologic method; this can only be achieved by promoting partnerships and collaborations among ongoing initiatives in toxicogenomics, standardization, and validation
- Determine performance standards for toxicogenomics-based test methods that will serve as the yardsticks for comparable test methods that are based on similar operational properties
- Define further the modular validation scheme that would allow keeping up with methodologic improvements and innovations without having to repeat the entire validation process but would, however, integrate ECVAM and ICCVAM principles of validation and acceptance.

Table 3. Recommendations: focus on regulatory acceptance of toxicogenomics-based methods.

- Build on and/or learn from previous and ongoing efforts in toxicogenomics, standardization, validation, and harmonization efforts where possible (e.g., MIAME, ICCVAM, ECVAM, NCT, EMBL–EBI, ILSI–HESI, U.S. FDA, U.S. EPA, OECD)
- Fund pilot programs to test possible validation strategies and processes
- Identify training needs and assist in developing training vehicles and ways of presenting the state-of-the-science to regulators and the regulated community (including electronic means)
- Maintain transparency of validation processes
- Explore additions, amendments, and revisions to ICCVAM and ECVAM validation guidance that would accommodate new and rapidly changing technologies
- Implement the modular approach to validation to accommodate existing knowledge and future technical developments
- Establish performance standards for toxicogenomics-based test methods and have them accommodate rapid technologic advancements and procedural modifications
- Explore, develop, and support sector-spanning worldwide harmonization entities
- Create confidence among regulators by involving them early on in discussions and various scientific forums that would facilitate application of the technology for regulatory purposes
- Encourage industry and other parties to share data, in part, to support validation comparisons
- Promote high-quality science in supporting the use and development of the technology for regulatory purposes to further protection of human health and the environment
- Consider opportunities for synergy between QSAR, pharmacokinetic, and pharmacodynamic modeling, and other *in silico* efforts and the toxicogenomics communities

References

Ball C, Brazma A, Causton H, Chervitz S, Edgar R, Hingamp P, et al., on behalf of the MGED Society. 2004. Standards for microarray data: an open letter [Letter]. Environ Health Perspect 112:A666–A667.

Ball CA, Sherlock G, Parkinson H, Rocca-Sera P, Brooksbank C, Causton HC, et al. 2002. Standards for microarray data. Science 298:539.

Balls M, Blaauboer BJ, Fentem JH, Bruner L, Combes RD, Ekwall B, et al. 1995. Practical aspects of validation of toxicity test procedures. Altern Lab Anim 23:129–147.

Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, Bradford BU, et al. 2005. Standardizing global gene expression analysis between laboratories and across platforms. Nat Methods 2:351–356.

Blower PE, Yang C, Fligner MA, Verducci JS, Yu L, Richman S, et al. 2002. Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data. Pharmacogenomics J 2(4):259–271.

Boess F, Kamber M, Romers S, Gasser R, Muller D, Albertini S, et al. 2003. Gene expression in two hepatic cell lines, cultured primary hepatocytes and liver slices, compared to the *in vivo* liver gene expression in rats: possible implications for toxicogenomics use in *in vitro* systems. Toxicol Sci 73(2): 386–402.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29(4):365–371.

Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, et al. 2003. ArrayExpress—a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 31:68–71.

Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HW, et al. 2002. Rapid and non-invasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabolomics. Nat Med 12:1439–1445.

Brown HR, Ni H, Benavides G, Yoon L, Hyder K, Giridhar J, et al. 2002. Correlation of simultaneous differential gene expression in the blood and heart with known mechanisms of adriamycin-induced cardiomyopathy in the rat. Toxicol Pathol 30(4):452–469.

Chu TM, Deng S, Wolfinger R, Paules RS, Hamadeh HK. 2004. Cross-site comparison of gene expression data reveals high similarity. Environ Health Perspectives 112:449–455.

Chung CH, Bernard PS, Perou CM. 2002. Molecular portraits and the family tree of cancer. Nat Genet 32(suppl):533–540.

Cronin M, Ghosh K, Sistare F, Quackenbush J, Vilker V, O'Donnell C. 2004. Universal standards for gene expression. Clin Chem 50(8):1464–1471.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression

Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207–210.

Hamadeh HK, Bushel PR, Jayadev S, Martin K, DiSorbo O, Sieber S, et al. 2002a. Gene expression reveals chemical-specific profiles. Toxicol Sci 67:219–231.

Hamadeh HK, Bushel PR, Jayadev S, Martin K, DiSorbo O, Sieber S, et al. 2002b. Prediction of compound signature using high density gene expression profiling. Toxicol Sci 67:232–240.

Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, et al. 2004. A modular approach to the ECVAM principles on test validity. Altern Lab Anim 32:467–472.

Hopkins D, Gout S, Burns M, Nixon G, Andersen M, Foy C, et al. 2004. Comparability of Gene Expression Measurements on Microarrays. LGC/MFB Report. Available: http://www.mfbprog.org.uk/themes/theme_publications_item.asp?intThemeID=22&intPublicationID=961 [accessed 9 January 2006].

Huang E, Ishida S, Pittmann J, Dressman H, Bild A, Kloos M, et al. 2003. Gene expression phenotypic models that predict the activity of oncogenic pathways. Nat Genet 34(2):226–230.

Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, et al. 2000. Functional discovery via a compendium of expression profiles. Cell 102:109–126.

ICCVAM. 1997. ICCVAM Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publ no 97-3981. Research Triangle Park, NC:Interagency Coordinating Committee on the Validation of Alternative Methods, National Institute of Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov [accessed 8 July 2005].

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publ no 03-4508. Research Triangle Park, NC:Interagency Coordinating Committee on the Validation of Alternative Methods, National Institute of Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov [accessed 8 July 2005].

ICCVAM. 2004. ICCVAM Recommended Performance Standards for *In Vitro* Test Methods for Skin Corrosion. NIH Publ no 04-4509. Research Triangle Park, NC:Interagency Coordinating Committee on the Validation of Alternative Methods, National Institute of Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov [accessed 8 July 2005].

Ikeo K, Ishi-i J, Tamura T, Gojobori T, Tateno Y. 2003. CIBEX: Center for Information Biology Gene Expression database. C R Biol 326(10–11):1079–1082.

Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, et al. 2005. Multiple-laboratory comparison of microarray platforms. Nat Methods 2:345–349.

Joseph LJ. 2004. RNA reference materials for gene expression studies. RNA metrology: forecast calls for partial clearing. Clin Chem 50(8):1290–1292.

Kuang Y, Biran I, Walt DR. 2004. Simultaneously monitoring gene expression kinetics and genetic noise in single cells by optical well arrays. Anal Chem 76(21):6282–6286.

Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J. 2005. Independence and reproducibility across microarray platforms. Nat Methods 2:337–343.

Mattes WB, Pettit SD, Sansone SA, Bushel PR, Waters MD. 2004. Database development in toxicogenomics: issues and efforts. Environ Health Perspect 112:495–505.

MGED Society (Microarray Gene Expression Data Society). 2004. RSBI Tiered Checklist (RSBI TC). A Proposed Framework Structure for Reporting Biological Investigations. Available: http://www.mged.org/Workgroups/rsbi/rsbi.html [accessed 9 January 2006].

Nature. 2002. Microarray standards at last [Editorial]. Nature 419:323.

OECD. 1998. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. 1: OECD Principles on Good Laboratory Practice. Paris:Organisation for Economic Co-operation and Development. http://www.oecd.org/document/63/0,2340,en_2649_34381_2346175_1_1_1_1,00.html [accessed 8 July 2005].

OECD. 2005. OECD Activities to Explore and Evaluate Regulatory Application of Genomic Methods; Toxicogenomics. Paris:Organisation for Economic Co-operation and Development. Available: http://www.oecd.org/document/29/0,2340,en_2649_34373_34704669_1_1_1_1,00.html [accessed 2 February 2006].

Orphanides G. 2003. Toxicogenomics: challenges and opportunities. Toxicol Lett 140–141:145–148.

Rockett JC, Hellmann GM. 2004. Confirming microarray data—is it really necessary? Genomics 83:541–549.

Russell WMS, Burch RL. 1959. The Principles of Humane Experimental Technique. London:Methuen. Available: http://altweb.jhsph.edu/publications/humane_exp/het-toc.htm [accessed 8 July 2005].

Sansone SA, Morrison N, Rocca-Serra P, Fostel J. 2004. Standardization initiatives in the (eco)toxicogenomics domain: a review. Comp Funct Genom 5:633–64.

Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, et al. 2000. A gene expression database for the molecular pharmacology of cancer. Nat Gen 24:236–244.

Spellman PT, Miller M, Stewart J, Troup C, Sarkans U, Chervitz S, et al. 2002. Design and implementation of microarray gene expression markup language (MAGE-ML). Genome Biol. 3(9):RESEARCH0046. Available: http://genomebiology.com/2002/3/9/research/0046 [accessed 9 January 2006].

Steiner G, Suter L, Boess F, Gasser R, de Vera MC, Albertini S, et al. 2004. Discriminating different classes of toxicants by transcript profiling. Environ Health Perspect 112:1236–1248.

Thomas RS, Rank DR, Penn SG, Zastrow GM, Hayes KR, Pande K, et al. 2001. Identification of toxicologically predictive gene sets using cDNA microarrays. Mol Pharmacol 60(6):1189–1194.

Tong W, Cao X, Harris S, Sun H, Fang H, Fuscoe J, et al. 2003. ArrayTrack—supporting toxicogenomic research at the U.S. Food and Drug Administration National Center for Toxicological Research. Environ Health Perspect 111:1819–1826.

U.S. EPA. 2002. Interim Genomics Policy. U.S. Environmental Protection Agency Science Policy Council. Washington, DC: U.S. Environmental Protection Agency. Available: http://epa.gov/osa/spc/htm/genomics.htm [accessed 8 July 2005].

U.S. FDA. 2005. FDA Guidance for Industry—Pharmacogenomic (PG) Data Submissions (Final: March 2005). Washington, DC:U.S. Food and Drug Administration. Available: www.fda.gov/cder/guidance/6400fnl.htm [accessed 8 July 2005].

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415(6871):530–536.

Waring JF, Cavet G, Jolly RA, McDowell J, Dai H, Ciurlionis R, et al. 2003. Development of a DNA microarray for toxicology based on hepatotoxin-regulated sequences. Environ Health Perspect 111:863–870.

Waring JF, Jolly RA, Ciurlionis R, Lum PY, Praestgaard JT, Morfitt DC, et al. 2001. Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. Toxicol Appl Pharmacol 175(1):28–42.

Waters MD, Boorman G, Bushel P, Cunningham M, Irwin R, Merrick A, et al. 2003. Systems toxicology and the chemical effects in biological systems knowledge base. Environ Health Perspect 111:811–824.

Yauk CL, Berndt ML, Williams A, Douglas GR. 2004 Comprehensive comparison of six microarray technologies. Nucleic Acid Res 32(15):e124.