administered materials in toxicity studies is fundamental, and characterizing delivered nanomaterials after administration in a test system or model provides the best quality data on dose and material properties that are related to observed responses, but this is limited by current methodological capabilities [2]. Further studies, especially *in vivo*, using different types of characterized materials, relevant routes of administration, and doses closely reflecting expected levels of exposure are needed to adequately evaluate the reproductive and developmental toxicity of nanomaterials.

Conflict of interest

None.

Acknowledgements

This study was conducted under the "Evaluating Risks Associated with Manufactured Nanomaterials" Project funded by the New Energy and Industrial Technology Development Organization (NEDO), Japan.

References

[1] Oberdörster G, Maynard A, Donaldson K, Castranova V, Fizpatric J, Ausman K, et al. Principles for characterizing the potential human health effects from exposure to nanoparticles: elements of a screening strategy. Part Fibre Toxicol 2005;2:8, doi:10.1186/1743-8977-2-8.
[2] Oberdörster G, Oberdörster E, Oberdörster J. Nanotoxicology: an emerging discipline evolution from studies of ultrafine particles. Environ Health Perspect 2005;113:823–39.
[3] Penn SG, He L, Natan MJ. Nanoparticles for bioanalysis. Curr Opin Chem Biol 2003;7:609–15.
[4] Liu WT. Nanoparticles and their biological and environmental applications. J Biosci Bioeng 2006;102:1–7.
[5] Colborn T, vom Saal F, Soto A. Developmental effects of endocrine-disrupting chemicals in wildlife and humans. Environ Health Perspect 1993;101:378–84.
[6] Brooking J, Davis SS, Illum L. Transport of nanoparticles across the rat mucosa. J Drug Target 2001;9:267–79.
[7] Wang J, Liu Y, Jiao F, Lao F, Li W, Gu Y, et al. Time-dependent translocation and potential impairment on central nervous system by intranasally instilled TiO_2 nanoparticles. Toxicology 2008;254:82–90.
[8] Lee KP, Trochimowicz HJ, Reinhardt CF. Pulmonary response of rats exposed to titanium dioxide (TiO_2) by inhalation for two years. Toxicol Appl Pharmacol 1985;79:179–92.
[9] Lindenschmidt RC, Driscoll KE, Perkins MA, Higgins JM, Maurer JK, Belfiore KA. The comparison of a fibrogenic and two nonfibrogenic dusts by bronchoalveolar lavage. Toxicol Appl Pharmacol 1990;102:268–81.
[10] Muhle H, Bellmann B, Creutzenberg O, Dasenbrock C, Ernst H, Kilpper R, et al. Pulmonary response to toner upon chronic inhalation exposure in rats. Fundam Appl Toxicol 1991;17:280–99.
[11] Warheit B, Sayes CM, Reed KL, Swain KA. Health effects related to nanoparticle eposures: environmental, health and safety considerations for assessing hazards and risks. Pharmacol Ther 2008;120:35–42.
[12] Heinrich U, Fuhst R, Rittinghausen S, Cretzenberg O, Bellmann B, Koch W, et al. Chronic inhalation exposure of Wistar rats and two different strains of mice to diesel engine exhaust, carbon black, and titanium dioxide. Inhal Toxicol 1995;7:533–56.
[13] Baan R, Staif K, Grosse Y, Secretan B, Ghissassi F, Cogliano V, on behalf of the WHO International Agency for Research on Cancer Monograph Working Group. Carcinogenicity of carbon black, titanium dioxide, and talk; 2006 [cited June 22, 2009], available from: http://oncology.thelancet.com.
[14] Landsiedel R, Kapp MD, Schulz M, Wiench K, Oesch F. Genotoxicity investigations on nanomaterials: methods, preparation and characterization of test material, potential artifacts and limitations-many questions, some answers. Mutat Res 2009;681:241–58.
[15] Ema M, Kobayashi N, Naya M, Hanai S, Nakanishi J. Genotoxicity evaluation of titanium dioxide nanoparticles. Jpn J Environ Toxicol 2009;12:71–84 [in Japanese].
[16] Fedulov AV, Leme A, Yang Z, Dahl M, Lim R, Mariani TJ, Kobzik L. Pulmonary exposure to particles during pregnancy causes increased neonatal asthma susceptibility. Am J Respir Mol Biol 2008;38:57–67.
[17] CEPA/OEHHA (California Environmental Protection Agency/Office of Environmental Health Hazard Assessment). Safe drinking water and toxic enforcement ACT of 1986 (Position 65). Notice to interested parties, February 21, 2003. Chemical listed effective February 21, 2003 as known to the State of California to cause cancer; 2003 [cited October 13, 2009], available from: http://oehha.ca.gov/prop65/prop65_list/files/22103not.pdf.
[18] Shimizu M, Tainaka H, Oba T, Mizuo K, Umezawa M, Takeda K. Maternal exposure to nanoparticle titanium dioxide during the prenatal period alters gene expression related to brain development in the mouse. Part Fibre Toxicol 2009;6:20, doi:10.1186/1743-8977-6-20.
[19] Takeda K, Suzuki K, Ishihara A, Kubo-Irie M, Fujimoto R, Tabata M, et al. Nanoparticles transferred from pregnant mice to their offspring can damage the genital and cranial nerve system. J Health Sci 2009;55:95–102.
[20] Komatsu T, Tabata M, Kubo-Irie M, Shimizu T, Suzuki K, Nihei Y, et al. The effects of nanoparticles on mouse testis Leydig cells *in vitro*. Toxicol in Vitro 2008;22:1825–31.
[21] Shelley SA. Nanotechnology: turning basic science into reality. In: Theodore L, Kunz RG, editors. Nanotechnology-environmental implications and solution. Hoboken: John Wiley and Sons, Inc. Publication; 2005. p. 61–107.
[22] Challier JC, Panigel M, Meyer E. Uptake of colloidal ^{198}Au by fetal liver in rats, after direct intrafetal administration. Int J Nucl Med Biol 1973;1:103–6.
[23] Takahashi S, Matsuoka O. Cross placental transfer of ^{198}Au-colloid in near term rats. J Radiat Res 1981;22:242–9.
[24] Sadauskas E, Wallin H, Stoltenberg M, Vogel U, Doering P, Larsen A, et al. Kupper cells are central in the removal of nanoparticles from the organism. Part Fibre Toxicol 2007;4:10, doi:10.1186/1743-8977r-r4-10.
[25] Myllynen P, Loughran MJ, Howard CV, Sormunen R, Walsh A, Vähäkangas KH. Kinetics of gold nanoparticles in the human placenta. Reprod Toxicol 2008;26:130–7.
[26] Wiwantitkit V, Sereemaspum A, Rojanathanes R. Effect of gold nanoparticles on spermatozoa: the first world report. Fertil Steril 2009;91:e7–8.
[27] Kim WS, Kim HC, Hong SH. Gas sensing properties of MoO_3 nanoparticles synthesized by solvothermal method. J Nanopart Res 2009, doi:10.1007/s11051-009-9751-6.
[28] Braydich-Stolle L, Hussain S, Schlager JJ, Hofmann MC. *In vitro* cytotoxicity of nanoparticles in mammalian germline stem cells. Toxicol Sci 2005;88:412–9.
[29] Makhluf SBD, Qasem R, Rubinstein S, Gedanken A, Breibart H. Loading magnetic nanoparticles into sperm cells does not affect their functionality. Langmuir 2006;22:9480–2.
[30] Bhabra G, Sood A, Fisher B, Cartwright L, Saunders M, Evans WH, et al. Nanoparticles can cause DNA damage across a cellular barrier. Nat Nanotechnol 2009;4:876–83.
[31] Lin YS, Haynes CL. Impact of mesoporous silica nanoparticle size, pore ordering, and pore integrity on hemolytic activity. J Am Chem Soc 2010;132:4834–42.
[32] Park MVDZ, Annema W, Salvati A, Lesniak A, Elsaesser A, Barnes C, et al. In vitro developmental toxicity test detects inhibition of stem cell differentiation by silica nanoparticles. Toxicol Appl Pharmacol 2009;240:108–16.
[33] Loutfy RO, Lowe TP, Moravsky AP, Katagiri S. Commercial production of fullerenes and carbon nanoparticles. In: Perspectives of fullerene nanotechnology. Netherlands: Springer; 2002. p. 35–46.
[34] Tsuchiya T, Oguri I, Yamakoshi YN, Miyata N. Novel harmful effects of [60]fullerene on mouse embryos in vitro and in vivo. FEBS Lett 1996;393:139–45.
[35] Sumner SCJ, Fennell TR, Snyder RW, Taylor GF, Lewin AH. Distribution of carbon-14 labeled C_{60}($[^{14}C]C_{60}$) in the pregnant and in the lactating dam and the effect of C_{60} exposure on the biochemical profile of urine. J Appl Toxicol 2009, doi:10.1002/jat.1503.
[36] Donaldson K, Tran L, Jimenez LA, Duffin R, Newby DE, Mils N, et al. Combustion-derived nanoparticles: a review of their toxicology following inhalation exposure. Part Fibre Toxicol 2005;2:10, doi:10.1186/1743-89772-2-10.
[37] Yoshida S, Hiyoshi K, Ichinose T, Takano H, Oshio S, Sugawara I, et al. Effect of nanopaticles on the male reproductive system of mice. Int J Andol 2008;32:337–42.
[38] Chan WH, Shiao NH. Cytotoxic effect of CdSe quantum dots on mouse embryonic development. Acta Pharmacol Sin 2008;28:259–66.
[39] Derfus AM, Chan WCW, Bhatia SN. Probing the cytotoxicity of semiconductor quantum dots. Nano Lett 2004;4:11–8.
[40] Bangs LB. New developments in particle-based immunoassays: introduction. Pure Appl Chem 1996;68:1873–9.
[41] Bosman SJ, Niet SP, Patton WC, Jacobson JD, Corselli JU, Chan PJ. Development of mammalian embryos exposed to mixed-size nanoparticles. Clin Exp Obst Gyn 2005;32:222–4.
[42] Shepard TH, Lemire RJ. Catalog of teratogenic agents. 11th ed. Baltimore: The Johns Hopkins University Press; 2004.
[43] Nielsen E, Thorup I, Schnipper A, Hass U, Meyer O, Ladefoged J, et al. Children and the unborn child. Exposure and susceptibility to chemical substances—an evaluation. Environmental project No. 589, Miljøproject; 2001 [cited October 18, 2009], available from: http://www2.mst.dk/udgiv/Publications/2001/87-7909-574-7/pdf/87-7909-333-7.pdf.
[44] Klimisch HJ, Andreae M, Tillmann U. A systemic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regl Toxicol Pharmacol 1997;25:1–5.
[45] Zhu X, Zhu L, Li Y, Duan Z, Chen W, Alvarez PJJ. Developmental toxicity in zebrafish (*Danio retio*) embryos after exposure to manufactured nanomaterials: buckminsterfullerene aggregates (nC_{60}) and fullerol. Environ Toxicol Chem 2007;26:976–9.

ラットを用いた短期反復投与毒性試験の低用量群に統計学的有意差が検出される割合

小林 克己*, 櫻谷 祐企*, 阿部 武丸*, 西川 智*, 山田 隼*,
広瀬 明彦**, 鎌田 栄一**, 林 真***

*独立行政法人 製品評価技術基盤機構 化学物質管理センター 安全審査課
〒151-0066 東京都渋谷区西原二丁目49-10 Tel：03-3481-1735 Fax：03-3481-1950 E-mail：kobayashi-katsumi@nite.go.jp

**国立医薬品食品衛生研究所 安全性生物試験センター
〒158-8501 東京都世田谷区上用賀一丁目18-1

***財団法人 食品農医薬品安全性評価センター
〒437-1213 静岡県磐田市塩新田582-2

1 要約

反復投与毒性試験の試験責任者は，低用量を無影響量として設定する。しかし，低用量群に統計学的有意差が散見される毒性試験は，少なくない。化審法によるラットを用いた109の28日間反復投与毒性試験を調査し，低用量群に統計学的有意差（$P < 0.05$）が認められる数とその割合を調査した。その結果，低用量群には205/12167（1.5%）の有意差が認められた。次に検査項目別では，尿検査が3.3%と高く，その他の検査項目の血液学的検査，血液生化学的検査，器官重量および器官重量・体重比は，1.1～1.8%程度であった。以上の調査結果，我々は低用量群の有意差検出率が2%（最大＜5%）程度であれば無毒性量を評価できると判断する。

2 はじめに

げっ歯類などを用いた毒性試験は，医薬品，農薬，動物用医薬品および一般化学物質についてそれぞれの所轄の官庁が定めたガイドラインによって実施されている。これら毒性試験の目的は，無影響量（NOEL）または無毒性量（NOAEL）の把握である。試験責任者は，用量設定予備試験の結果から本試験の各用量を決定する。そして本試験では，低用量群を無影響量と推定し試験を開始する。本試験終了後，各測定項目について統計学的および毒性学的有意差から無毒性量を決定する。しかし，これら毒性試験では，低用量群に統計学的有意差が認められる測定値が散見される。この場合は，背景値との比較，用量依存性の有無および関連項目との整合性などを吟味して無毒性量などを決定する。本調査報告の目的は，既存化学物質に対するラットを用いた短期28日間反復投与毒性試験から低用量群にどの程度の統計学的有意差（$P < 0.05$）が検出されているかを調査し，若干の考察を加え，試験責任者の知見にしていただきたい。

3 調査材料および方法

化審法ガイドライン（NITE, 2007）に従ったラットを用いた109の28日間反復投与毒性試験をインターネット（厚生労働省，2009）から取得した。調査した化審法28日反復投与毒性試験の基本的群構成を表1に示した。主な解析データは，投与28日後の定量値を用いた。各群の構成は，対照群を含めて4および5群（37試験）であった。多くの試験の1群内動物数は，5匹程度，用量の公比は，3が最も多かった。被験物質の投与は，全て胃ゾンデによる強制経口投与であった。

表1　試験群の構成と供試動物数

群	供試動物数	
	28日間の投与期間	14日間の回復期間
対照	10	5
低用量	5	—
中用量	5	—
高用量	10	5

調査に用いた検査項目（定量値）は，表2に示した。体重，飼料摂取量および飲水量などのデータはグラフによる開示のみで，これらのデータは，もし統計学的有意差が認められても経時的変化によって考察ができることから除外した。また尿検査および病理解剖・組織学的検査成績の定性データは，例数が少ないことから統計学的分析に問題があることから除外した。

表2　調査に用いた主な検査項目（定量値）

検査項目	主な測定項目
行動機能観察（FOB）	握力，開眼幅および自発運動量
尿検査	尿量，比重，浸透圧，Na，K，Cl
血液学的検査	白血球数およびその分画比，赤血球数，ヘモグロビン量，ヘマトクリット値，平均赤血球容積，平均赤血球血色素量，平均赤血球血色素濃度，血小板数および白血球百分率，プロトロンビン時間，活性化部分トロンボプラスチン時間など
血液生化学的検査	総蛋白濃度，総コレステロール濃度，ブドウ糖濃度，尿素窒素濃度，クレアチニン濃度，アルカリフォスファターゼ活性，GOT活性，GPT活性，γ-GTP活性，トリグリセライド濃度，無機リン濃度，カルシウム濃度，A/G，ナトリウム濃度，カリウム濃度および塩素濃度など 一部の試験は骨髄検査および電気泳動による蛋白分画検査を実施している。
器官重量および体重に対する相対重量比	脳，甲状腺，胸腺，心臓，肺，肝臓，脾臓，腎臓，副腎，精巣，卵巣など

表3に統計解析に使用した主な解析手法（Kobayashi *et al.*, 2008）を示した。調査した報告書は，片側検定，両側検定および記載なしが混在していた。この解析によって有意差（$P < 0.05$）が認められた検査項目を用量毎に集計した。

表3　げっ歯類を用いた28日間の反復投与毒性試験に使用された統計解析ツール

Bartlettの等分散検定，分散分析，Dunnettの多重比較検定，Schefféの多重比較検定，Duncanの多重範囲検定，Kruskal-Wallisの検定，Dunnett型ノンパラメトリック検定，Scheffé型ノンパラメトリック検定，Steelの多重比較検定など

4　調査結果および考察

調査試験109報ごとの対照群に対する低用量，中用量，高用量および最高用量群の5%水準による有意差検出数とその割合を表4に示した。1試験の調査項目は，平均111であった。対照群に対する有意差検出数とその割合は，低用量群，中用量群，高用量群および最高用量群がそれぞれ1.6，3.4，10および17%と当たり前であるが用量依存性が顕著に認められた。無影響量として設定した低用量群は，最大8.4%，最小0.0%，最頻値0.0%で平均1.6%の有意差が認められた。試験によって大きなばらつきが認められた。低用量群に全く統計学的有意差が認められなかった試験は，30/109であった。

次に109試験の調査項目別に対照群に対する各用量群の統計学的有意差数およびその割合を表5に示した。表下段の合計は，表4の合計と一致する。検査項目中，尿検査は，他の検査項目に比較して群間を通して有意差が検出されやすいことが認められた。尿検査以外の検査項目の血液学的検査，血液生化学的検査，器官重量および器官重量・体重比は，各群内で大きな差が無かった。

定期検査・解剖などが投与開始後26，52，78および104週で設定されている長期の慢性毒性試験など大規模試験は，極めて多くの測定値が得られる。このため低用量群には，偶発的な統計学的有意差が検出されやすい。低用量群に統計学的有意差が認められた場合の無毒性量は，毒性学的有意差の有無を検討し設定したい。以上の調査結果から，我々は，全調査項目中，低用量群に統計学的有意差が1～2%（＜5%）程度認められても無毒性量を評価できると判断する。

表4　化審法28日間反復投与毒性試験から得られた定量値に対する有意差（$P < 0.05$）検出数およびその割合
（その1，試験別からの結果）

試験番号	測定項目数	有意差検出数および（%）			
		低用量群	中用量群	高用量群	最高用量群
1	120	0 (0.0)	2 (1.7)	4 (3.3)	10 (11)
2	104	6 (5.7)	5 (4.8)	4 (3.8)	—
3	108	1 (0.9)	2 (1.8)	3 (2.7)	—
4	128	2 (1.5)	10 (7.8)	23 (18)	—
5	98	4 (4.8)	10 (10)	17 (17)	—
6	120	0 (0.0)	4 (3.3)	10 (8.3)	24 (20)
7	146	2 (1.3)	6 (4.1)	42 (28)	—
8	118	1 (0.8)	1 (0.8)	0 (0.0)	1 (0.8)
9	118	1 (0.8)	2 (1.6)	27 (22)	—
10	116	1 (0.8)	10 (8.6)	26 (22)	—
11	116	2 (1.7)	10 (8.6)	26 (22)	—
12	156	3 (1.9)	11 (7.0)	36 (23)	—
13	106	2 (1.8)	5 (4.7)	30 (28)	—
14	100	7 (7.0)	11 (11)	23 (23)	—
15	100	2 (2.0)	2 (2.0)	2 (2.0)	—
16	128	3 (2.3)	10 (7.8)	44 (34)	—
17	96	5 (5.2)	0 (0.0)	4 (4.1)	—
18	118	2 (1.6)	0 (0.0)	1 (0.8)	4 (3.3)
19	114	0 (0.0)	1 (0.8)	1 (0.8)	—
20	102	4 (3.9)	1 (0.9)	6 (5.8)	—
21	94	3 (3.1)	6 (6.3)	2 (2.1)	—
22	128	0 (0.0)	2 (1.5)	23 (17)	—
23	136	3 (2.2)	7 (5.1)	11 (8.0)	—
24	76	1 (1.3)	2 (2.6)	5 (6.5)	21 (27)
25	110	2 (1.8)	5 (4.5)	17 (15)	—
26	118	0 (0.0)	0 (0.0)	0 (0.0)	4 (3.3)
27	116	0 (0.0)	2 (1.7)	9 (7.7)	23 (19)
28	100	2 (2.0)	4 (4.0)	37 (37)	—
29	158	5 (3.1)	12 (7.5)	31 (19)	—
30	120	0 (0.0)	2 (1.6)	10 (8.3)	—
31	134	0 (0.0)	0 (0.0)	4 (2.9)	16 (11)
32	126	1 (0.7)	6 (4.7)	12 (9.5)	30 (23)
33	106	3 (2.8)	1 (0.9)	23 (21)	—
34	128	0 (0.0)	3 (2.3)	3 (2.3)	38 (29)
35	118	0 (0.0)	9 (7.6)	24 (20)	40 (33)
36	54	0 (0.0)	2 (3.7)	9 (16)	—
37	106	0 (0.0)	2 (1.8)	6 (5.6)	—
38	122	1 (0.8)	5 (4.0)	44 (36)	—
39	100	6 (6.0)	17 (17)	27 (27)	—
40	104	1 (0.9)	3 (2.8)	7 (6.7)	—
41	104	2 (1.9)	4 (3.8)	9 (8.6)	—
42	106	3 (2.8)	1 (0.9)	25 (23)	—
43	151	2 (1.3)	9 (2.9)	31 (20)	—
44	106	2 (1.8)	9 (8.4)	3 (2.8)	—
45	112	3 (2.6)	9 (8.0)	18 (16)	—
46	86	0 (0.0)	5 (5.8)	14 (16)	30 (34)
47	104	2 (1.9)	0 (0.0)	7 (6.7)	—
48	70	0 (0.0)	1 (1.4)	0 (0.0)	—
49	50	0 (0.0)	2 (4.0)	9 (18)	30 (60)
50	74	0 (0.0)	1 (1.3)	0 (0.0)	—
51	132	3 (2.2)	5 (3.7)	9 (6.8)	—
52	142	12 (8.4)	11 (7.7)	7 (4.9)	—
53	94	1 (1.0)	2 (2.1)	2 (2.1)	—
54	116	5 (4.3)	7 (6.0)	3 (2.5)	—
55	98	7 (7.1)	3 (3.0)	23 (23)	—
56	104	6 (5.7)	12 (11)	21 (20)	—
57	88	2 (2.2)	1 (1.1)	6 (6.8)	—
58	98	0 (0.0)	0 (0.0)	8 (8.1)	—
59	100	0 (0.0)	0 (0.0)	1 (1.0)	8 (8.0)
60	104	2 (1.9)	8 (7.6)	37 (35)	58 (55)
61	96	1 (1.0)	7 (7.2)	16 (16)	—
62	128	1 (0.7)	3 (2.3)	25 (19)	—
63	138	2 (1.4)	6 (4.3)	49 (35)	—
64	128	1 (0.7)	1 (0.7)	12 (9.3)	—
65	130	0 (0.0)	3 (2.3)	9 (6.9)	33 (25)
66	104	0 (0.0)	6 (5.7)	42 (40	—
67	94	6 (6.3)	3 (3.1)	4 (4.2)	—
68	106	3 (1.8)	3 (1.8)	3 (1.8)	—
69	110	0 (0.0)	0 (0.0)	2 (1.8)	21 (19)
70	122	0 (0.0)	2 (1.6)	4 (3.2)	0 (0.0)
71	162	1 (0.6)	2 (1.2)	9 (5.5)	—
72	136	1 (0.7)	3 (2.2)	18 (13)	—
73	104	1 (0.9)	0 (0.0)	1 (0.8)	4 (3.8)
74	106	0 (0.0)	0 (0.0)	7 (6.6)	19 (25)
75	112	7 (6.2)	1 (0.8)	1 (0.8)	—
76	118	1 (0.8)	3 (2.5)	6 (5.0)	13 (11)
77	90	0 (0.0)	0 (0.0)	2 (2.2)	14 (15)
78	116	0 (0.0)	4 (3.4)	2 (1.7)	5 (4.3)
79	106	1 (0.9)	0 (0.0)	4 (3.7)	8 (7.5)
80	62	2 (3.2)	2 (3.2)	2 (3.2)	3 (4.8)
81	96	1 (1.0)	1 (1.0)	0 (0.0)	—
82	124	3 (2.4)	3 (2.4)	30 (24)	53 (42)
83	110	1 (0.9)	3 (2.7)	13 (11)	—
84	100	1 (1.0)	1 (1.0)	2 (2.0)	—
85	102	1 (0.9)	2 (1.9)	2 (1.9)	14 (13)
86	120	4 (3.3)	12 (10)	15 (12)	29 (24)
87	162	1 (0.6)	4 (2.4)	16 (9.8)	—
88	134	3 (2.2)	1 (0.7)	3 (2.2)	17 (12)
89	90	1 (1.1)	2 (2.2)	2 (2.2)	—
90	108	1 (0.9)	2 (1.8)	1 (0.9)	—
91	100	2 (2.0)	2 (2.0)	4 (4.0)	—
92	92	0 (0.0)	0 (0.0)	0 (0.0)	2 (2.1)
93	110	2 (1.8)	5 (4.5)	21 (19)	33 (30)
94	100	0 (0.0)	0 (0.0)	1 (1.0)	—
95	104	1 (0.9)	1 (0.9)	1 (0.9)	—
96	104	5 (4.8)	11 (10)	22 (21)	—
97	98	4 (4.0)	5 (5.1)	19 (19)	—
98	114	0 (0.0)	0 (0.0)	8 (7.0)	43 (37)
99	118	3 (2.5)	3 (2.5)	0 (0.0)	—
100	88	1 (1.1)	1 (1.1)	1 (1.1)	—
101	118	3 (2.5)	3 (2.5)	12 (10)	—
102	120	1 (0.8)	0 (0.0)	3 (2.5)	18 (15)
103	116	0 (0.0)	1 (0.8)	1 (0.8)	14 (12)
104	150	1 (0.6)	6 (4.0)	15 (10)	—
105	134	5 (3.7)	7 (5.2)	14 (10)	—
106	100	0 (0.0)	0 (0.0)	6 (6.0)	11 (11)
107	136	1 (0.7)	1 (0.7)	4 (2.9)	—
108	130	2 (1.5)	4 (3.0)	16 (12)	31 (23)
109	116	1 (0.8)	1 (0.8)	2 (1.7)	9 (7.7)
試験数	109	109	109	109	37
合　計	12167	205	414	1318	731/4074
平均値	111	1.60%	3.40%	10%	17%
最大値	162	8.40%	17%	36%	55%
最小値	54	0.00%	0.00%	0.00%	0.00%
最頻値	—	0.00%	—	—	—

表5　化審法28日間反復投与毒性試験から得られた定量値に対する有意差（$P < 0.05$）検出数およびその割合（その2，測定項目からの結果）

測定項目	測定項目数	有意差検出数および（%）			
		低用量群	中用量群	高用量群	最高用量群
FOB	68	1 (1.4)	3 (4.4)	8 (11)	1/10 (10)
尿検査	392	13 (3.3)	30 (7.6)	81 (20)	40/96 (40)
血液学的検査	3586	56 (1.5)	106 (2.9)	318 (8.8)	176/1198 (14)
血液生化学的検査	4285	79 (1.8)	163 (3.8)	455 (10)	267/1426 (18)
器官重量	1928	22 (1.1)	44 (2.2)	188 (9.7)	103/672 (15)
器官重量・体重比	1908	34 (1.7)	68 (3.5)	268 (14)	144/672 (21)
合計	12167	205 (1.6)	414 (3.4)	1318 (10)	731/4074 (17)

謝辞

我々は，本成果が独立行政法人新エネルギー・産業技術総合開発機構（NEDO）の委託費事業によって作成されたことに感謝する。

引用論文および資料

1) 厚生労働省（2009）：
http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPage.jsp

2) Katsumi Kobayashi, K. Sadasivan Pillai, Yuki Sakuratani, Takemaru Abe, Eiichi Kamata and Makoto Hayashi (2008) : "Evaluation of statistical tools used in short-term repeated dose administration toxicity studies with rodents" : *J. Toxicol. Sci.*, Vol. 33 : No. 1, 97-104

3) NITE (2007)：
http://www.safe.nite.go.jp/kasinn/pdf/28test.pdf

毒性試験から得られる定量値に対する有効数値の桁数の差違およびMann-Whitney の U 検定とWilcoxon の検定の有意差検出の違い

小林 克己 *, 櫻谷 祐企 *, 阿部 武丸 *, 西川 智 *, 山田 隼 *,
広瀬 明彦 **, 鎌田 栄一 **, 林 真 ***

* (独) 製品評価技術基盤機構 化学物質管理センター 安全審査課
〒 151-0066 東京都渋谷区西原二丁目 49-10 Tel：03-3481-1735 Fax：03-3481-1950
E-mail：kobayashi-katsumi@nite.go.jp

** 国立医薬品食品衛生研究所 安全性生物試験センター

*** (財) 食品農医薬品安全性評価センター

抄録：12 受託試験機関による 106 のラットを用いた 28 日反復投与毒性試験（化審法）から得られた定量値の有効桁数を検討するため心臓の実重量について調査した。定量データの有効値の桁数の設定は，各試験機関で異なることがわかった。多くの試験機関は，3 桁が最も多く次いで 5, 4 桁の順であった。次に，分散などを用いない順位和検定の Mann-Whitney の U 検定を用いて桁数の違いによる有意差検出パターンを検索した。桁数が 3 と小さい場合は，桁数 4 および 5 に比較して有意差が検出しにくいことがわかった。この理由は，桁数が小さいと同一順位が増加する傾向を示すことにある。t-検定系は，桁数に大きく影響されない。動物数が 30 ～ 40 程度の場合，統計数値表を利用する Mann-Whitney の U 検定と計算による Wilcoxon の検定は，ほぼ同様の検出力を示す。したがって，短期毒性試験の場合，もし順位和検定を採用する場合は，片側検定の Mann-Whitney の U 検定または Wilcoxon の検定を用い，平均値および標準偏差の表示は，有効桁数を 3 で表示し，計算値（生データ）は，4 または 5 桁を採用したい。

1 はじめに

げっ歯類を用いた反復投与毒性試験から得られる定量データは，体重および飼料摂取量などの飼育管理データ，赤血球数およびヘモグロビン量などの血液学的検査，AST・ALT 活性値および総蛋白濃度などの血液生化学的検査，尿比重および尿量などの尿検査および肝および腎などの器官重量検査などその全項目は，50 以上（Kobayashi, 1997）である。これらの解析法のほとんどは，決定樹（Kobayashi *et al.*, 2008）によって対照群と用量群間差を吟味している。解析途中で，全群間に等分散が認められない場合は，順位和検定（rank sum test）となる。また比率・グレード値およびはじめから分散などの分布を利用しない順位和検定を採用する場合も少なくない。

このノンパラメトリック検定の順位和検定は，群間差を全群の順位の違いで解析している。群間に同一値が存在すると検出力は，低下する。したがって，個体値の桁数の設定によって有意差検出が変化することが推測できる。有効数値の桁数が多いと同一順位は，極めて少なくなると推測できる。

そこで我々は，化審法による公開試験報告書 109 について定量値の有効数値の桁数を調査し，有意差検出

パターンを調査した。同時に検出力の差をStudentのt-検定，Mann-WhitneyのU検定およびWilcoxonの検定と比較検討した。

2 調査材料および結果

2.1 有効数値の桁数

化審法（NITE, 2007）対応による12試験機関で実施されたラットを用いた28日間反復投与毒性試験（MHLW, 2009）149試験をインターネットから検索し，その内，試験機関によって測定値の桁数が3～5と大きく異なる心実重量（雄）を取得し，その心実重量個体値を試験機関別に，その有効数値の桁数を調査した（表1）。

表1 公開化審法28日間反復投与毒性試験から得られた心重量の試験機関別有効数値の桁数

試験機関名	調査試験数	心重量測定試験数	有効数値の桁数
A	24	7	3
B	19	9	3
C	18	13	3
D	25	17	5および4（2試験）
E	15	15	3
F	13	10	3
G	14	14	3
H	10	10	3
I	4	4	3
J	4	4	4
K	2	2	4
L	1	1	4
合計	149	106	

表2 有効数値の桁数別公開試験数

有効数値の桁数	試験数（%）
3	82（77）
4	9（8.5）
5	15（14）
合計	106（100）

その結果，149の調査試験の中で心実重量を測定している試験数は106であった。各試験機関は，決まった有効数値の桁数が設定されていた。その桁数は，3が最も多く全試験の77%，次いで5桁が14%，4桁が8.5%であった（表1および表2）。

心実重量は，雄が1g以上で，この場合の有効数値の桁数の表示が3桁（例1.12g）であるが，雌は1g以下で，この場合の有効数値の桁数の表示が2桁（例0.91g）が殆どであった。すなわち，雄と雌で有効数値の桁数が異なる場合が多かった。

2.2 有効数値の桁数の違いによる統計学的有意差の差違

調査試験報告書の中で高用量群に有意差が認められた肝実重量（g）の1試験を抽出し，有効桁数の違いによる有意差検出パターンを調査した。桁数を5, 4および3に設定し，桁数の変化に影響がないパラメトリック検定のStudent t-検定と同一順位があると有意差が変化するMann-Whitney U検定と比較検討した。Mann-Whitney U検定を選択した理由は，各群内動物数が6であることによる。両者の解析は，Excel 2008およびAOKI(2010)を用いて片側検定を採用した。

その結果は，表3に示した。このデータは，対照群に対して高用量群は，約12%重量の増加を示している。Studentのt-検定は，全桁数とも5%水準で有意差を示した。しかし，Mann-Whitney U検定は，同一順位が各群に1つ以上存在すると有意差が認められない。したがって，有効桁数を多く設定した場合は，同一順位の発生が少ないことから検出力が高くなる。

表3　化審法の28日間反復投与毒性試験から得られた雄の肝重量桁数の違いによる有意差検出パターン

有効桁数	群		確率（P）	
	対照（$N=6$）	高用量（$N=6$）	Student t-test	Mann-Whitney U test
5	10.391, 11.442, 13.653, 10.224, 10.783, 10.414	13.194, 11.444, 13.701, 11.572, 12.683, 12.661	0.0279	<0.05 ($U=5$)
平均値±標準偏差	11.151 ± 1.301	12.543 ± 0.889		
平均順位	4.3	8.6		
4	10.39, 11.44, 13.65, 10.22, 10.78, 10.41	13.19, 11.44, 13.70, 11.57, 12.68, 12.66	0.0279	有意差ナシ ($U=5.5$)
平均値±標準偏差	11.14 ± 1.30	12.54 ± 0.88		
平均順位	4.4	8.5		
3	10.4, 11.4, 13.7, 10.2, 10.8, 10.4	13.2, 11.4, 13.7, 11.6, 12.7, 12.7	0.0286	有意差ナシ ($U=6$)
平均値±標準偏差	11.1 ± 1.3	12.5 ± 0.8		
平均順位	4.5	8.5		

2.3　t-検定，Mann-Whitney の U 検定および Wilcoxon の検定の検出力

一般的に2群間検定の場合，F-test の結果，不等分散の場合は，Welch の検定となる。または始めから Mann-Whitney の U 検定を採用する場合がある。Mann-Whitney の U 検定は，一般的に標本数が50程度までは統計数値表（AOKI, 2010）が用意されている。標本数がそれ以上の場合は，正規化検定を計算式からZ値を算出して有意差の判断を行う。この検定法は，Wilcoxon の検定と呼ばれる。したがって，標本数が35の場合は，Mann-Whitney の U 検定か Wilcoxon の検定かその使用に迷う。すなわち，動物数が小数例の場合は，統計数値表（簡易表）による Mann-Whitney の U 検定で，また大数例は，計算による Wilcoxon の検定で解析するといわれている。

表4に標本数38のデータを示し種々の解析法による検出力の比較を示した。データは，38試験の Fibrinogen 量の雄雌別変動係数（%）である。解析プログラムは，Excel 2008 および AOKI (2010) を用いた。その結果，片側検定による Student t-検定と Mann-Whitney の U 検定と Wilcoxon の検定は，同一の検出力を示した。

表4　Fibrinogen 量の変動係数に対する Student の t-検定，Mann-Whitney の U 検定および Wilcoxon の検定の検出力（P）

性	N	平均値± 標準偏差	F-test	Student t-test	Mann-Whitney U test		Wilcoxon test	
					Two-side	One-side	Two-side	One-side
雄	38	6.28 ± 2.42	$P=0.1030$	$P=0.0117$	Not sig.	$P<0.05$	$P=0.0608$	$P=0.0304$
雌	38	7.78 ± 3.16			$U=541.5$		$Z=1.8752$	

3 考察

受託試験機関12による106の反復投与毒性試験から得られた定量値の有効桁数を調査するため心実重量について調査した。定量データの有効値の桁数の設定は，各試験機関で異なることがわかった。多くの試験機関は，3桁が多かった。次に，分散などを用いない順位和検定のMann-Whitney の U 検定を用いて桁数の違いによる有意差検出パターンを検索した。桁数が3と小さい場合は，桁数4および5に比較して有意差が検出しにくいことがわかった。この理由は，桁数が小さくなると同一順位が増加する傾向することである。しかし，t-検定系は桁数に大きく影響されない。

動物数が30～40程度の場合，統計数値表を利用するMann-Whitney の U 検定と計算によるWilcoxonの検定は，ほぼ同様の検出力を示す。したがって，短期毒性試験の場合，もし順位和検定を採用する場合は，片側検定のMann-Whitney の U 検定またはWilcoxonの検定を用い，平均値の表示は，有効桁数を3で表示し，計算値（生データ）は，4または5桁を採用したい。

謝辞

本調査報告は，独立行政法人新エネルギー・産業技術総合開発機構（NEDO）の補助金にて作成されたことに感謝する。

引用論文および資料

・大阪大学大学院薬学研究科・医薬情報解析学分野および大阪大学遺伝情報実験センター，大阪大学微生物病研究所（MEPHAS），(2010)：
http://www.gen-info.osaka-u.ac.jp/testdocs/tomocom/wilc1.html
・AOKI (2010)：
http://aoki2.si.gunma-u.ac.jp/lecture/Average/u-tab.html
・Kobayashi, K. (1997)：A comparison of one- and two-sided tests for judging significant differences in quantitative data obtained in toxicological bioassay of laboratory animals, Journal of Occupational Health, 39, 29-35
・Kobayashi, K., Pillai, K. S., Sakuratani, Y., Abe, T., Kamata, E. and Hayashi, M. (2008)：Evaluation and statistical tools used in short-term repeated dose administration toxicity studies with rodent, Journal of Toxicological Science, 33, (1), 97-104
・NITE (2007)：
http://www.safe.nite.go.jp/kasinn/pdf/28test.pdf
・Ministry of Health, Labour and Welfare, (MHLW) (2009)：
http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPage.jsp

The Journal of Toxicological Sciences (J. Toxicol. Sci.)
Vol.35, No.1, 79-85, 2010

Original Article

Relation between statistics and treatment-related changes obtained from toxicity studies in rats: if detected a significant difference in low or middle dose for quantitative values, this change is considered as incidental change?

Katsumi Kobayashi[1], Yuki Sakuratani[1], Takemaru Abe[1], Satoshi Nishikawa[1], Jun Yamada[1], Akihiko Hirose[2], Eiichi Kamata[2] and Makoto Hayashi[1,3]

[1]*Chemical Management Center, National Institute of Technology and Evaluation, 49-10 Nishihara-nichome, Shibuya, Tokyo 151-0066, Japan*
[2]*National Institute of Health Sciences, 18-1 Kamiyoga-ichome, Setagaya-ku, Tokyo 158-8501, Japan*
[3]*An-Pyo Center, 582-2 Shioshinden, Iwata-shi, Shizuoka 437-1213, Japan*

(Received October 8, 2009; Accepted November 11, 2009)

ABSTRACT — The purpose of a toxicity test is to determine the no-observed-effect level (NOEL) of test substance through biological and pharmacological techniques. If the low dose not does show statistically significant and biologically relevant changes in the data evaluated in a study, the usual practice is to consider this dose as the NOEL. To overcome this, 6 types of techniques that seemed to be appropriate are presented in this paper by investigating the results of several domestic and foreign theses on toxicology. The most appropriate techniques appear to be the trend test, comparison between treatment group and historical control by *t*-test, and confirmation that all individual values lie within the 95% confidence interval (2 SD) of the historical control value, if a significant difference is admitted in the low dose.

Key words: Toxicity, Rodents, Statistics, Historical control data, Incidental change, Standard deviation and error

INTRODUCTION

Toxicity test is necessary for evaluating the safety of industrial chemicals according to the Chemical Substances Control Law (1986). Quantitative data obtained from toxicity studies with test substance in rodents are analyzed by using decision tree procedure (Hamada *et al.*, 1998; Kobayashi *et al.*, 2008). If statistical analysis of such data reveals that low or mid dose data are significantly different as compared to the control group used as a reference for that particular study, it may not be possible to determine the no-observed-effect level (NOEL) of that test substance. This significant difference obtained in the low and mid dose groups is usually not considered as incidental. This can be confirmed by comparing the data of these groups with the historical control data obtained from the testing facilities or by confirming that there was no dose-related pattern observed, statistically or visually. Usually the comparison with the historical control data is made by checking the data of the treatment groups (low and mid dose), whether they lie within the width of 2 standard deviations (S.D.) of the historical control data. If the data of the treatment groups lie within the width of 2 S.D. of the historical control, it is assumed that the changes shown by these groups are incidental. However, the S.D. plotted shows the distribution pattern of all the individual historical control data, whereas the data of the treatment group is obtained by calculating the mean of 5 to 35 animals used in an experiment. Hence, the comparison between these 2 data sets may be erroneous since the quality of data of both the groups is different. In order to resolve these disagreements and to confirm incidental finding, this paper presents six alternative techniques. We would have liked to have documented the uniformity of each testing facility in the form of a thesis; however, the documentation of this kind is problematic since there are restrictions in publishing the toxicity results of a non-government organization and since the investigation

Correspondence: Katsumi Kobayashi (E-mail: kobayashi-katsumi@nite.go.jp)

lie within the width of ± 2 S.D. of the historical control value (Fig. 1).

(3) Also the difference between the mean value of the historical control group and that of the treatment group is analyzed by the *t*-test (Table 1).

The NOEL judgments

The NOEL is judged by the study director on the basis of his/her knowledge and experience in addition to the data presented in the thesis. In this case, priority may be given to the clinical relevance of the data as compared to the statistical significant difference. For instance, a significant difference might be detected by using statistics when considering parameters that have a small variance, such as the electrolyte concentration or specific gravity of urine samples. In this case, statistical results are ignored, and the clinical relevance is considered.

Checking the reliability of the control group

It is sometimes necessary to check whether or not the control group used in a study is in the normal range. The comparison of the treatment group with double controls (usually used in basal diet experiments) is also made by statistical analysis. Alternatively, the control group can be compared with the historical control value.

Investigation using published studies

28-day repeated dosing study in accordance with the Chemical Substance Control Law

The examples of the 28-day repeated dose toxicity study in rats are available in the public domain (http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPage.jsp) and are presented in Table 1.

When the data of the treatment groups were compared with the historical data, it was found that the mean values

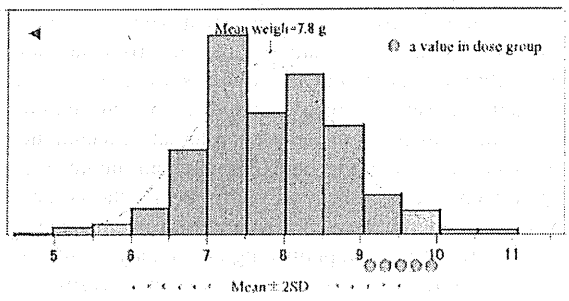


Fig. 1. Is there an individual of the examination group within mean ± 2 S.D. of the historical value?

of the treatment groups were within the range of ± 2 S.D. of the historical control data, but not within the range of ± 2 S.E. of it (except GPT of CAS No. 97-52-9) (Table 1). In the above studies, the authors judged whether the significant differences shown by the treatment groups were incidental or not by examining the data of the treatment groups with regard to the range of ± 2 S.E. of the respective historical control data. If the data of the treatment groups fall within the range of ± 2 S.E. of the historical control, they were considered as incidental. However, when the data were analyzed using *t*-test, adrenal weight (CAS No. 7756-94-7) and total protein (CAS No. 56-93-9) of the high dose were significant. Though the calculation procedure of *t*-test is based on the SEs of the treatment and historical control groups, it should be borne in mind that the number of animals used in these groups is different, being much larger in the latter, since the source of historical control data is several studies. But, while comparing the values within the range of the standard error, the number of animals used in each group is not taken into consideration.

In the repeated dose toxicity studies, when a significant difference between the treated and control values is detected at the low or medium doses, the determination of NOEL is difficult. An attempt was made to find out a solution to this problem by investigating 28-day repeat dose toxicity gavage studies in Sprague-Dawley (SD) rats wherein 126 test substances were analyzed in accordance with the Chemical Substance Control Law guidelines (http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPage.jsp) (Table 2). Among the studies investigated, only one testing facility in Japan described historical control mean value ± S.D. and the number of animals clearly in the report. The statements, such as "No dose-related pattern or dose dependency", made in the reports clearly show that there is no significant difference in the high and/or medium dose groups. It seems that the dose-response pattern and/or the dose dependency in most of the studies have been evaluated solely by a macroscopic decision. The following or similar statement from several report, supports our view, "this change was within the physiological range and/or the historical control range or is minimal change." In these reports, the authors have not given the range of values for the historical controls. Thus, the study director assumed that the change in this dosage group to be an incidental change without a statistics or solid scientific support.

Studies published in journals

On investigating the changes of parameters of dosage groups of several repeated dose toxicity studies in rodents

Table 3. Grounds considered as incidental changes to the significant differences in dosed group of repeated dose toxicity studies

Reason assumed to be incidental change	Reference
Presence of dose-related pattern, visually	Nishiguchi *et al.*, 1997; Nishiguchi *et al.*, 1994; Takagi *et al.*, 1992b; Nakano *et al.*, 1992; Tamura *et al.*, 1983; Yamazaki *et al.*, 2005; Chemical product Safety Center, 1994; Griffiths *et al.*, 2007; Topping *et al.*, 2007; Guijie *et al.*, 2006; Poon *et al.*, 1998; McClain *et al.*, 2006; Sato *et al.*, 2007; Hellwig *et al.*, 1993; Webb *et al.*, 1993; Mellert *et al.*, 2002; Bär *et al.*, 1995; Arterburn *et al.*, 2000; Goldsmith, 2000; Lee *et al.*, 2004; Janssen *et al.*, 2000; Kanki *et al.*, 2003; O'Hagan and Menzel, 2003; Nakamura *et al.*, 2001; Thomas *et al.*, 1991; Abdo *et al.*, 1986; Morgan *et al.*, 1989; Dunnick *et al.*, 1987; Okazaki *et al.*, 2002; Okazaki *et al.*, 1993; Kato *et al.*, 1993; Jeong *et al.*, 2006; Shim *et al.*, 2003
No change in a related parameter	Inui *et al.*, 1997; Takeuchi *et al.*, 1985; Jonker *et al.*, 1993; MacKenzie *et al.*, 1992a; Barber and Topping, 1995; Oshima *et al.*, 1999; Suzuki *et al.*, 1997; Graça *et al.*, 2007; Shimpo *et al.*, 1990
High or low value of control group	Inui *et al.*, 1997; Mellert *et al.*, 2002; Macrì *et al.*, 1987
Study director's judgment (no significant biological difference effect, negligible change, within physiological change or sporadic change)	Takagi *et al.*, 1994a; Nakano *et al.*, 1992; Kato *et al.*, 1991; Takahashi *et al.*, 1986; Omosu *et al.*, 2003; Griffiths *et al.*, 2007; Topping *et al.*, 2007; Inui *et al.*, 1997; Cho *et al.*, 2006; MacKenzie *et al.*, 1992a, 1992b; Juberg *et al.*, 1998; Horváth *et al.*, 2002; Kotkoskie *et al.*, 1998; Oshima *et al.*, 1999; Yi *et al.*, 2007
Within historical range (background data) or normal range (not described to macroscopic or statistic significant)	Takahashi *et al.*, 1986; Tamura *et al.*, 1983; Omosu *et al.*, 2003; Griffiths *et al.*, 2007; Guijie *et al.*, 2006; McClain *et al.*, 2006; Horváth *et al.*, 2002; Kitamura *et al.*, 2003; Hart, 1988; Suzuki *et al.*, 1994; Shiraishi *et al.*, 2006
Compared with another control other than vehicle control (using by double control)	Takahashi *et al.*, 1986; Webb *et al.*, 1993; Arterburn *et al.*, 2000
Statistics processing with normal values (historical data) range	Cerdá *et al.*, 2003

1. There is no significant difference if the mean value of if all the individual values of the dosage group fall within mean ± 2 S.E. for the historical control values.

2. There is no significant difference if all the individual values of the dosage group fall within mean ± 2 S.D. of the historical control values.

3. The statistical significant difference of the mean values between the dosage and historical control values group may be analyzed using the *t*-test.

4. The dose-response pattern/dose dependency may be analyzed using statistical techniques, for instance, Jonckheere trend test can be applied.

We recommend one in the above-mentioned four techniques to be used. However we suggest that the decision may not be made entirely on the basis of statistical analysis, but biological relevance of the statistical analysis may also be looked into.

ACKNOWLEDGMENTS

The research described in this paper was supported by a grant (Project name: Development of hazard assessment techniques using structure-activity relationship methods) from the New Energy and Industrial Technology Development Organization (NEDO). We gratefully thank Dr. K. Sadasivan Pillai from Frontier Tissueline Pvt. Ltd. (Chennai, INDIA) for his excellent advice concerning this article.

REFERENCES

Abdo, K.M., Rao, G., Montgomery, C.A., Dinowitz, M. and Kanagalingam, K. (1986): Thirteen-week toxicity study of d-α-tocopheryl acetate (Vitamin E) in Fischer 344 rats. Food Chem. Toxicol., 24, 1043-1050.

Arterburn, L.M., Boswell, K.D., Koskelo, E., Kassner, S.L., Kelly, C. and Kyle, D.J. (2000): A combined Subchronic (90-day) toxicity and neurotoxicity study of a single-cell source of docosahexaenoic acid triglyceride (DHASCO oil). Food Chem. Toxicol., 38, 35-49.

Bär, A., Til, H.P. and Timonen, M. (1995): Subchronic oral toxicity study with regular and enzymatically depolymerized sodium carboxymethylcellulose in rats. Food Chem. Toxicol., 33, 909-917.

Barber, E.D. and Topping, D.C. (1995): Subchronic 90-day oral tox-

icity study of suplatast tosilate (IPD-1151T) in rats. J. Toxicol. Sci., **17, Suppl. 2**, 11-38.

Nishiguchi, Y., Adachi, T., Nakazawa, M., Tawaratani, T., Uchimoto, H., Yoshida, M., Ishibashi, S., Kitayama, E., Iwakura, K. and Sumi, N. (1997): A 13-week oral repeated dose toxicity study of (+/-)-4-diethylamino-1,1-dimethylbut-2-yn-1-yl 2-cyclohexyl-2-hydroxy-2-phenylacetate monohydrochloride monohydrate (NS-21), a novel drug for urinary frequency and incontinence, in rats followed by a 5-week recovery test. J. Toxicol. Sci., **22, Suppl. 1**, 27-57.

Nishiguchi, Y., Ishibashi, S., Tarui, M., Fujimoto, M., Tawaratani, T., Yoshida, M., Uchimoto, H., Adachi, T., Hamasu, Y. and Sumi, N. (1994): 13-week oral toxicity study of lactitol (BS-4) in rats followed by 5-week recovery test. J. Toxicol. Sci., **19, Suppl. 3**, 305-326.

O'Hagan, S. and Menzel, A. (2003): A Subchronic 90-day oral rat toxicity study and *in vitro* genotoxicity studies with a conjugated linoleic acid product. Food Chem. Toxicol., **41**, 1749-1760.

Okazaki, K., Imazawa, T., Nakamura, H., Furukawa, F., Nishikawa, A. and Hirose, M. (2002): A repeated 28-day oral dose toxicity study of 17alpha-methyltestosterone in rats, based on the "enhanced OECD test guideline 407" for screening the endocrine-disrupting chemicals. Arch. Toxicol., **75**, 635-642.

Okazaki, S., Ueno, H., Anabuki, K., Tamura, K., Takeuchi, M., Hasegawa, H. and Kawase, Y. (1993): A 3-month oral toxicity study of MCP-1034 in rats with a recovery period of one month. Jpn. Pharmacol. Ther., **21**, 947-974.

Omosu, M., Eluard, B., Thien-A, H., Stepniewski, J.P., Vidal, J.M., Bode, G., Roeder, V., Sacre-Salem, B. and Guffroy, M. (2003): Oral repeated-dose toxicity studies and toxicokinetic studies of telithromycin in rats. Jpn. J. Chemother., **51**, 103-114.

Oshima, T., Sugiura, M., Takasu, K., Kimura, M. and Kudo, S. (1999): Repeated dose toxicity study of 3-amino-4-[[(Z)-4-[[4-(piperidiomethyl)-2-pyridyl]oxy]-2-butenyl]amino]-3-cyclobutene-1,2-dione monohydrochloride (Pibutidine hydrochloride) in rats (1): 13-week oral toxicity study with 4-week recovery test. Jpn. Pharmacol. Ther., **27**, 331-345.

Poon, R., Chu, I., Lecavalier, P., Valli, V.E., Foster, W., Gupta, S. and Thomas, B. (1998): Effects of antimony on rats following 90-day exposure via drinking water. Food Chem. Toxicol., **36**, 21-35.

Sakaki, H., Igarashi, S., Ikeda, T., Imamizo, K., Omichi, T., Kadota, M., Kawaguchi, T., Takizawa, T., Tsukamoto, O., Terai, K., Tozuka, K., Hirata, J., Handa, J., Mizuma, H., Murakami, M., Yamada, M. and Yokouchi, H. (2000): Statistical method appropriate for general toxicological studies in rats. J. Toxicol. Sci., **25**, 71-98.

Sato, S., Kitamura, H., Chino, M., Takei, Y., Hiruma, M. and Nomura, M. (2007): A 13-week oral dose subchronic toxicity study of gardenia yellow containing geniposide in rats. Food Chem. Toxicol., **45**, 1537-1544.

Shim, H.J., Kim, Y.C., Jang, J.M., Park, K.J., Kim, D.H., Kang, K. K., Ahn, B.O., Kwon, J.W., Kim, W.B. and Lee, M.G. (2003): Subacute toxicities and toxicokinetics of DA-8159, a new phosphodiesterase type V inhibitor, after single and 4-week repeated oral administration in rats. Biopharm. Drug Dispos., **24**, 409-418.

Shimpo, K., Takeuchi, M., Okazaki, S., Kiguchi, M., Hashimoto, Y., Aoki, Y., Kuwata, M. and Yamashita, K. (1990): Three-month oral subacute toxicity study of mofezolac (N-22) in rats. J. Toxicol. Sci., **15, Suppl. 2**, 11-41.

Shiraishi, K., Miyata, K., Houshuyama, S., Imatanaka, N., Umano, T., Minobe, Y. and Yamasaki, K. (2006): Subacute oral toxicity study of diethylphthalate based on the draft protocol for "enhanced OECD test guideline no. 407". Arch. Toxicol., **80**, 10-16.

Suzuki, S., Ogawa, Y., Kamata, E., Naito, K., Umemura, T., Saito, M., Kaneko, T. and Kurokawa, U. (1994): A 28-day repeated dose toxicity test of Phoxim in Wister rats. Eisei Shikenjo Houkoku, **112**, 40-47.

Suzuki, T., Koike, Y., Kitagaki, T. and Yamoto, T. (1997): A 90-day oral repeated dose toxicity study of azelnidipine in rats followed by a 28-day recovery period. Jpn. Pharmacol. Ther., **25**, 977-1004.

Takagi, A., Momma, J., Aida, Y., Takada, K., Suzuki, S., Naitoh, K., Tobe, M., Hasegawa, R. and Kurokawa, Y. (1992): Toxicity studies of a synthetic antioxidant, 2-2'-methylenebis (4-ethyl-6-tert-butylphenol) in rats. I. Acute and subchronic toxicity. J. Toxicol. Sci., **17**, 135-153.

Takagi, A., Takada, K., Sai, K., Ochiai, T., Matsumoto, K., Sekita, K., Momma, J., Aida, Y., Saitoh, M., Naitoh, K., Furuya, T., Hasegawa, R. and Kurokawa, Y. (1994): Acute, Subchronic and chronic toxicity studies of a synthetic antioxidant, 2,2'-methylenebis(4-methyl-6-tert-butylphenol) in rats. J. Toxicol. Sci., **19**, 77-88.

Takahashi, N., Kadota, T., Kawano, S., Ohta, K., Ishikawa, K., Kuroyanagi, K., Hamajima, Y., Ohta, S., Kai, S. and Kohmura, H. (1986): Toxicity studies of VP16-213(II)--Oral one-month subacute toxicity in rats. J. Toxicol. Sci., **11, Suppl. 1**, 17-49.

Takeuchi, M., Ichionohe, M., Okazaki, S., Sato, M., Kiguchi, M., Shimpo, K., Sudo, J. and Tanabe, T. (1985): Study on toxicity of halopredone acetate. (II). Subacute toxicity study in rats. J. Toxicol. Sci., **10, Suppl. 1**, 11-39.

Tamura, J., Sato, N., Ezaki, H., Miyamoto, H., Oda, S., Hirai, K., Takado, H., Matsumoto, M. and Shirai, T. (1983): Acute toxicity of ranitidine and its metabolite in mice, rats and rabbits, and subacute oral toxicity of ranitidine in rats. J. Toxicol. Sci., **8, Suppl. 1**, 1-24.

Thomas, W.C., McGrath, L.E., Baarson, K.A., Auletta, C.S., Daly, I.W. and McConnell, R.F. (1991): Subchronic oral toxicity of cellulose acetate in rats. Food Chem. Toxicol., **29**, 453-458.

Topping, D.C., Bernard, L.G., O'Donoghue, J.L. and English, J.C. (2007): Hydroquinone: acute and subchronic toxicity studies with emphasis on neurobehavioral and nephrotoxic effects. Food Chem. Toxicol., **45**, 70-78.

Webb, D.R., Wood, F.E., Bertram, T.A. and Fortier, N.E. (1993): A 91-day feeding study in rats with caprenin. Food Chem. Toxicol., **31**, 935-946.

Williams, D.A. (1971): A test for differences between treatment means when several dose levels are compared with a zero dose control. Biometrics., **27**, 103-117.

Williams, D.A. (1972): The comparison of several dose levels with zero dose control. Biometrics., **28**, 519-531.

Yamazaki, K., Aiso, S., Matsumoto, M., Arito, H., Nagano, K., Yamamoto, S. and Matsushima, T. (2005): Thirteen-week oral toxicity study of 1,4-dichloro-2-nitrobenzene in rats and mice. Ind. Health, **43**, 597-610.

Yi, Y., Yoon, H.J., Kim, B.O., Shim, M., Kin, S.O., Hwang, S.J. and Seo, M.H. (2007): A mixed polymeric micellar formulation of itraconazole: Characteristics, toxicity and pharmacokinetics. J. Control Release, **117**, 59-67.

Chemical Research in Toxicology

Article

pubs.acs.org/crt

Investigating the Relationship between in Vitro–in Vivo Genotoxicity: Derivation of Mechanistic QSAR Models for in Vivo Liver Genotoxicity and in Vivo Bone Marrow Micronucleus Formation Which Encompass Metabolism

Ovanes G. Mekenyan,[†] Petko I. Petkov,[†] Stefan V. Kotov,[†] Stoyanka Stoeva,[†] Verginia B. Kamenska,[†] Sabcho D. Dimitrov,[†] Masamitsu Honma,[‡] Makoto Hayashi,[‡] Romualdo Benigni,[§] E. Maria Donner,[∥] and Grace Patlewicz*[,∥]

[†]Laboratory of Mathematical Chemistry (LMC), As. Zlatarov University, Bourgas, Bulgaria

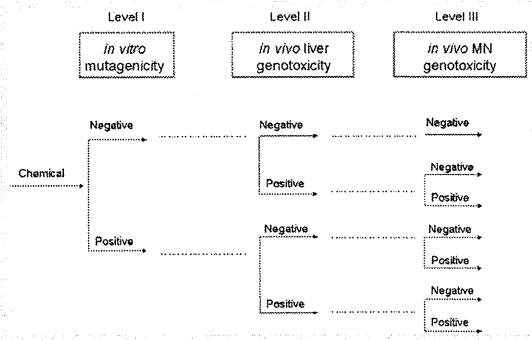
[‡]Division of Genetics and Mutagenesis, National Institute of Health Sciences, Tokyo, Japan

[§]Environment and Health Department, Istituto Superiore di Sanita', Rome, Italy

[∥]DuPont Haskell Global Centers for Health and Environmental Sciences, Newark, Delaware 19714-0050, United States

S *Supporting Information*

ABSTRACT: Strategic testing as part of an integrated testing strategy (ITS) to maximize information and avoid the use of animals where possible is fast becoming the norm with the advent of new legislation such as REACH. Genotoxicity is an area where regulatory testing is clearly defined as part of ITS schemes. Under REACH, the specific information requirements depend on the tonnage manufactured or imported. Two types of test systems exist to meet these information requirements, in vivo genotoxicity assays, which take into account the whole animal, and in vitro assays, which are conducted outside the living mammalian organism using microbial or mammalian cells under appropriate culturing conditions. Clearly, with these different broad experimental categories, results for a given chemical can often differ, which present challenges in the interpretation as well as in attempting to model the results in silico. This study attempted to compare the differences between in vitro and in vivo genotoxicity results, to rationalize these differences with plausible hypothesis in concert with available data. Two proof of concept (Q)SAR models were developed, one for in vivo genotoxicity effects in liver and a second for in vivo micronucleus formation in bone marrow. These "mechanistic models" will be of practical value in testing strategies, and both have been implemented into the TIMES software platform (http://oasis-lmc.org) to help predict the genotoxicity outcome of newly untested chemicals.



INTRODUCTION

Terms of Reference: Genotoxicity versus Mutagenicity.

Carcinogenicity and mutagenicity are among the toxicological end points that pose the highest concern for human health and are subject to regulatory testing for hazard and risk assessment. Much of the data that are currently available in the public domain have thus been derived from tests conducted to investigate potentially harmful effects on genetic material, that is, genotoxicity or mutagenicity. Since both terms, mutagenicity and genotoxicity, will be referenced in this paper, working definitions are given. According to academic definitions, genetic alterations that are fixed and can be inherited are termed mutations. These include different types of events such as base substitutions and deletions, structural chromosomal aberrations (CAs) (break and rearrangements), and numerical CAs (loss or gain of chromosomes, i.e., aneuploidy). The assays established to evaluate these events are described in brief. Genotoxicity is considered as a broader term—aside from mutations, it also encompasses other alterations of genetic material that are not fixed and are not inherited, such as DNA damage. Genotoxicity may or may not be transformed into mutations by the cell's machinery during cell replication, and it may be an indication of potential carcinogenesis associated with the exposure to a chemical agent. Appropriate in vivo experimental test systems used to evaluate genotoxicity include the bone marrow in vivo micronucleus test (MNT) assay, the unscheduled DNA synthesis (UDS) assay, and the alkaline single-cell gel electrophoresis assay (Comet assay). These tests are relevant to assess DNA-damaging and DNA-repair processes in specific organs of investigation in the whole animal such as liver. Therefore, the term *liver genotoxicity* was regarded as appropriate for the purposes of this study, although, overall, a wide array of other

Received: June 3, 2011

   dx.doi.org/10.1021/tx200547s | *Chem. Res. Toxicol.* XXXX, XXX, XXX–XXX

events aside from mutations are encompassed in these test systems.

Current Quantitative Structure–Activity Relationship (QSAR) Approaches. The importance of assessing genotoxicity coupled with the availability of experimental data has prompted many in silico studies. James and Elisabeth Millers's "electrophilic theory" introduced a chemical concept to help rationalize the mode of action of genotoxic carcinogens.[1] This prompted many evaluations to derive so-called structural alerts (SA), simple yet effective means of encoding qualitative mechanistic understanding for predicting potential mutagenicity/carcinogenicity. Seminal efforts include SA for carcinogenicity by John Ashby,[2] who subsequently extended his list with additional SA.[3] Bailey et al. compiled a set of 33 SAs for regulatory use within the U.S. Food and Drug Administration (FDA), which was predominantly based on the Ashby alerts.[4] Kazius et al. evaluated a mutagenicity database comprising 4337 mutagens and nonmutagens taken from the Toxnet database (http:/toxnet.nlm.nih.gov/) and derived 29 SAs for mutagenicity with associated detoxification fragments.[5] Some of these alerts exist in software platforms to enable routine use; for example, 17 SAs for mutagenicity are implemented into the OASIS tissue metabolism simulator (TIMES) software.[6] Benigni et al. combined the published information from Ashby, Bailey et al., and Kazius et al. with additional information from the OncoLogic (U.S. EPA) software (http://www.epa.gov/oppt/sf/pubs/oncologic.htm)[7] to arrive at a list of 33 SA for carcinogens and mutagens.[8]

Current quantitative strategies include (Q)SARs and expert systems. Two types of (Q)SAR models, local and global, exist to estimate the mutagenic potential of chemicals. Local (Q)SARs provide estimated results for closely related (congeneric) chemical structures. Such models are most predictive, but only if the essential features of the model domains are clearly represented. Models based on physicochemical descriptors with clear mechanistic meaning are particularly helpful in rationalizing genotoxic outcome as exemplified by Chung et al.[9] Other local models are based on mathematical representations of chemical structure, for example, topological indices, and thus are more difficult to interpret.[10]

Global (Q)SARs aim to provide mutagenicity estimations for a diverse (noncongeneric) set of chemicals. Such (Q)SARs may be additionally encoded into expert systems. For example, TOPKAT empirically makes predictions for a range of different end points including Ames mutagenicity and rodent carcinogenicity.[11] Other expert systems such as TIMES attempt to provide clear mechanistic meaning through the use of SAs, which address the reactivity toward DNA and/or proteins.[12,13] TIMES also includes 3D QSARs to underpin some of the available SAs. All of the aforementioned (Q)SARs have typically been derived on Ames (*Salmonella* mutagenicity data). TIMES includes a platform for in vitro CA data in addition to that for Ames.[13] There is a paucity of models for in vivo genotoxicity, but as highlighted in the survey by Benigni et al., there is only one publically available model for in vivo micronucleus.[14] The scarcity of such models may be due in part to experimental data being less readily available but also due to the complexicity of how to rationalize and interpret the outputs from the different test systems.

Our own investigation aims to fill in the above in vitro–in vivo genotoxicity gap by considering both the available test systems and how they are currently applied to formulate an approach for modeling in vivo genotoxicity. For convenience, we considered the REACH ITS[15] for mutagenicity since this described the typical assays used and how their outcomes should be interpreted for subsequent decision making. The actual experimental test systems are assumed to be reasonably familiar and are only briefly described in the next section.

Experimental Assays and Data for Rodent Mutagenicity and Genotoxicity. Integrated testing strategies, notably those described in the REACH Technical guidance,[15] outline the in vitro and in vivo systems that are most frequently used to evaluate the mutagenic potential of chemical substances. The in vitro systems include the bacterial reverse mutation test (Ames), an in vitro mammalian cell gene mutation test [such as the mouse lymphoma or hypoxanthine–guanine phosphoribosyltransferase (*hprt*) assay], the in vitro mammalian chromosome aberration (CA) test, and the in vitro MNT.[15] The Ames test uses amino acid-requiring strains of bacteria to detect (reverse) gene mutations (point and frameshift mutations). The in vitro mouse lymphoma assay (MLA), when correctly performed, detects structural chromosome aberrations, aneuploidy, and recombination events (e.g., such as gene conversion) that result in loss of heterozygosity. The *hprt* test identifies chemicals that induce gene mutations in the *hprt* gene of established cell lines. The in vitro mammalian CA test detects structural chromosome aberrations and increases in polyploidy. The in vitro MNT has the potential to detect both clastogenic (chromosome aberrations) and aneugenic (chromosome lagging due to dysfunction of mitotic apparatus) chemicals.

The scheme under REACH can be summarized as follows. As a first tier, three in vitro tests are recommended, which includes an Ames test, a mouse micronucleus/CA, and a mouse lymphoma/HRPT assay. If the results from all three tests are negative, then no more testing is merited, and a conclusion of nongenotoxicity can be made for the substance under study. If one or more tests are positive, then in vivo testing may be instigated. Obviously metabolism, pharmacokinetics, and toxicokinetics factors [absorption, distribution, metabolism, excretion (ADME)] are all inherent features in the in vivo genotoxicity tests, although the genetic end points for the tests address different genetic mechanisms. The UDS in vivo assay is used to evaluate the role of DNA repair. The in vivo Comet assay is a sensitive technique for the detection of DNA strand breaks; thus, it can be used for measuring DNA strand breaks in any tissue of an animal. Site-specific effects at contact tissues or the target tissue where the test compound accumulates or induces toxicity can be readily assessed. The specificity of the contact tissue under investigation is also feasible for the transgenic rodent gene mutation test (TGR), which measures gene mutations in vivo. However, the in vivo MNT is probably the most widely used test.[16] When performed appropriately, it detects both clastogenicity and aneugenicity.[17] The frequency of micronucleated polychromatic erythrocytes is traditionally determined from bone marrow samples, but with the emerging automated scoring methods, the emphasis is moving to assessing the induction of micronuclei in immature erythrocytes in peripheral blood samples.[18]

Most of the established in vitro mutagenicity tests, which are used for regulatory purposes, exhibit relatively high sensitivity for detection of genotoxic carcinogens.[19] However, particularly those based on cultured mammalian cells are thought to produce a remarkably high occurrence of irrelevant positive results (i.e., exhibit low specificity), when compared with rodent carcinogenicity.[19,20] To increase the specificity of predictions, regulators tend to interpret in vitro positive results in an in vivo perspective, that is, in vivo confirmation of in vitro mutagens.

dx.doi.org/10.1021/tx200547s | *Chem. Res. Toxicol.* XXXX, XXX, XXX–XXX

In addition, in vivo tests can also be utilized to identify chemicals producing in vivo only positive results (i.e., chemicals for which mutagenicity is not or poorly detected in vitro). Only a very limited number of chemicals have been found to be genotoxic in vivo and not in the standard in vitro tests. Most of these are pharmaceuticals such as atovaquone (95233-18-4), which is designed to affect pathways of cellular regulation, including cell cycle regulation. One of the most preferred in vivo assays, complementing genotoxicity test batteries, is the in vivo bone marrow MNT. The preference of this assay is attributed to both its wide mutagenicity range assessment (clastogenicity and aneugenicity) and its remarkably high specificity in concordance with the genotoxic carcinogenicity model, although it shows low sensitivity.[14,21] Therefore, it may be appropriate to include a second in vivo test if a positive in vitro result has not been adequately confirmed by the in vivo bone marrow MNT test. The UDS test is one complement to the bone marrow MNT since it is a surrogate in vivo gene mutation assay[21] measuring DNA excision repair of induced DNA damage. The utility of the Comet and the TGR assays to detect genotoxic damage in specific tissues, specifically DNA strand breaks and gene mutations has also been recognized.[15] Thus, an evaluation of in vivo genotoxicity potential could involve integrating outcomes from MNT and either UDS, Comet, and TGR tests depending on the outcomes that have been observed in vitro. UDS, Comet, and TGR can also be undertaken to address in vivo liver genotoxicity. Such tissue-specific assays are useful in in vivo follow-up tests especially since the liver is an organ of high metabolic capacity and therefore is frequently subjected to significant toxic overload.

Aims of the Study. Bearing in mind the way in which these different assays are integrated together, our goal was to investigate the in vitro and in vivo relationship, the so-termed in vitro–in vivo "gap" to inform the development of mechanistic (Q)SAR model(s). A large body of data covering in vitro mutagenicity, in vivo (liver) genotoxicity, and in vivo bone marrow MNT test results was collected for the same set of substances. The scope of the investigation can be summarized in the following three questions: (a) To what extent are in vitro mutagenic chemicals in vivo (liver) genotoxic, that is, what in vivo detoxification pathways exist? (b) To what extent are in vivo (liver) genotoxic chemicals in vivo bone marrow MNT positive? (c) Are there in vitro nonmutagenic chemicals that are in vivo liver or bone marrow genotoxic; that is, what in vivo bioactivation pathways exist? These questions were structured into a workflow (Figure 1) and enabled a stepwise evaluation of the in vitro–in vivo gap.

MATERIALS AND METHODS

Compilation of Data Set. Our training set comprised 557 chemicals ("557 list") with in vivo MNT data (Appendix I of the Supporting Information lists the substances and their overall calls). In vitro mutagenicity and in vivo (liver) data were collected for the same set of substances to the extent possible. This helped maximize the overlap between chemicals with various genotoxicity effects and the in vivo MNT data set. Documented in vitro mutagenicity data from multiple literature sources were identified for 397 noncongeneric chemicals within the training set (Appendix II of the Supporting Information). Positive calls were categorized by the digit 1, negative calls by 0, and N/A signified "no data available", based on the literature searches that were performed. Our in vitro data comprised that from the Ames assay, the CA assay, and the MLA, since these are the typical assays considered under REACH. Out of necessity and as typically the case for modeling efforts, reported study results were accepted as

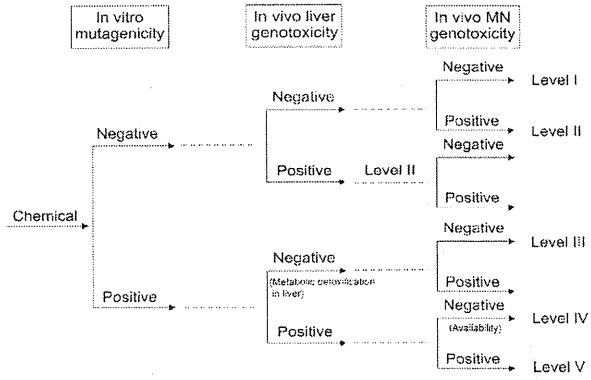


Figure 1. Workflow outlining the in vivo–in vitro gap.

reported, although an extensive effort was made in expert judgment and evaluation of the data quality and correctness of the calls.

Ames results with the rat liver S9 metabolic activation system were available for 283 noncongeneric chemicals. Of these chemicals, 109 (38%) were associated with positive calls and 174 (62%) with negative calls. Documented in vitro CA test data were identified for 296 chemicals, of which 186 (63%) were positive and 110 (37%) were considered negative. Data from 194 chemicals had been assessed in the in vitro MLA. The majority of the chemicals tested positive (148 chemicals, i.e., 76%) and 46 chemicals (24%) tested negative. For the 397 in vitro mutagenicity data, these comprised 267 positive calls (68%) and 124 negative calls (32%), and six calls were inconclusive. These substances were ethylene dichloride (107-06-2), sulfan blue (129-17-9), thiabendazole (148-79-8), methyl parathion (298-00-0), dibutylnitrosamine (924-16-3), C.I. direct black 38 (1937-37-7). In these six cases, only Ames and in vitro CA test outcomes were available with positive calls in Ames and negative calls in in vitro CA tests.

Results from in vivo Comet, UDS, and TGR assays were also collected to help evaluate in vivo liver genotoxic potential. Data were available for 185 diverse chemicals, which are listed in Appendix III of the Supporting Information. The Comet assay provided liver genotoxicity assignments for 127 (69%) of the 185 chemicals. Of the 127 chemicals, 78 (61%) were positive, and 49 (39%) were negative. The TGR comprised rodent liver genotoxicity data for 34 (18%) of the 185 chemicals; 27 (80%) of these were reported as positive, and 7 (20%) were negative. The in vivo UDS assay was associated with the least amount of liver genotoxicity data, only 24 (13%) of the 185 chemicals had overall calls, and five of them were observed to be positive in this assay (21%), and 19 were (79%) negative in this assay. Overall, of the 185 substances with liver assignments, 109 were associated with positive calls (59%) and 76 with negative calls (41%). The "557 list" included almost equal numbers of positive (267 chemicals, i.e., 48%) and negative (290 chemicals, i.e., 52%) MNT assignations performed in either bone marrow or peripheral blood. Figure 2 summarizes the distribution of assignments in each of the test systems.

The evaluation of this investigation was often hampered by conflicting in vivo MNT data available in the public domain. The compromised quality of these MNT data was attributed to the fact that many chemicals had been evaluated in the early 1980s; when species (rat vs mouse) and gender (male vs female) differences may not always have been considered, etc. To date, the validity of the in vivo MNT data has only been verified for chemicals where the in vitro mutagenicity outcome appeared to be negative, relative to the in vivo case (in either liver or bone marrow), where the genotoxicity result was positive. Expert judgment was relied upon to consider whether there were factors resulting in inconsistent in vitro results as compared with the in vivo situation, for example, rodent species differences, nonphysiological culture conditions, etc.

To illustrate the structural diversity of the training set, the 557 list was profiled against the set of DNA and protein binding alerts available within the OECD Toolbox v2.1. The distribution chart is

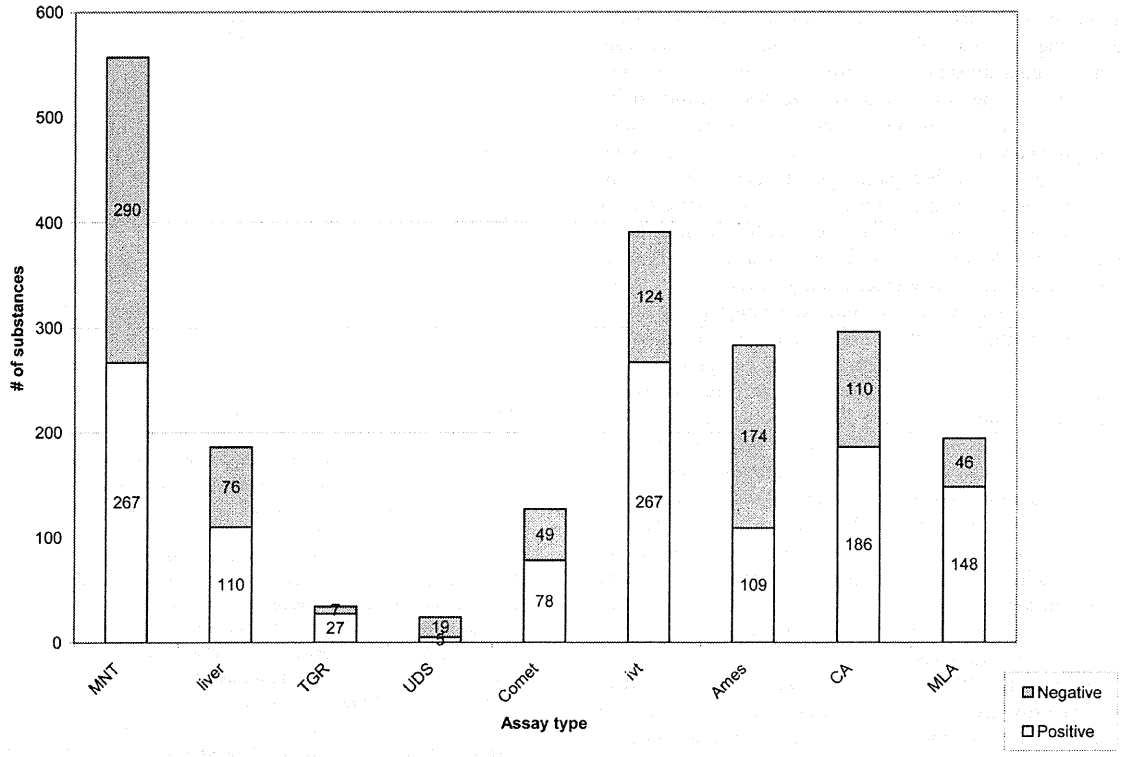


Figure 2. Distribution of the overall calls for each of the test assays under study.

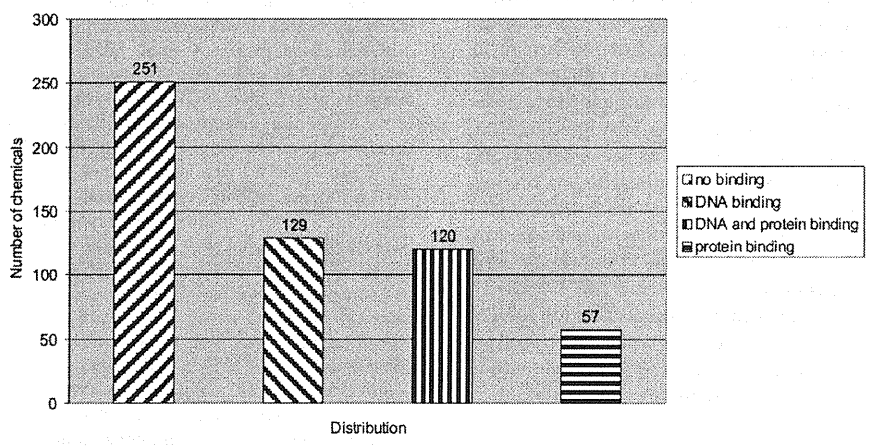


Figure 3. Distribution of training set chemicals across DNA and protein binding alerts.

shown in Figure 3. The results reveal that 251 (45%) of the 557 chemicals possess no DNA and/or protein binding alerts. One hundred twenty-nine of the remaining 306 (55%) chemicals have one or more DNA binding alerts, 57 chemicals have a protein binding alert, and 120 chemicals have both DNA and protein binding alerts. This distribution shows a broad spread of chemical mechanisms as depicted by the SAs triggered.

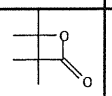
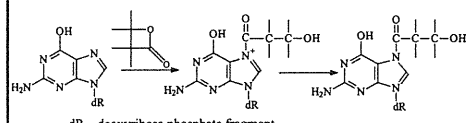
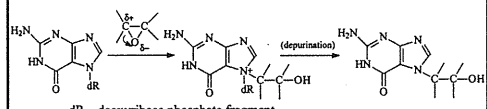
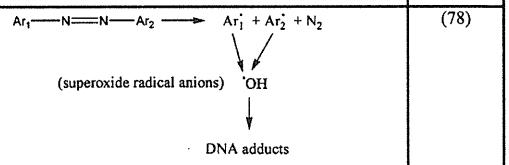
Our modeling approach sought to use the existing TIMES formalism and refine the components that had been originally developed to estimate Ames and in vitro CA. Here, we provide a brief overview of these components.

Modeling Reactivity to DNA and Proteins. According to the working hypothesis, interaction of chemicals with DNA and/or with specific proteins (such as histone, topoisomerase, spindle protein tubulus, and DNA repair enzymes) encompasses a diversity of genotoxic events, which can damage mammalian cells. For example, the formation of micronuclei arises as a result of the covalent interaction between chemicals with DNA and/or specific proteins. Accordingly, a reactivity component for an in vivo model, which predicts genotoxic effects such as formation of micronuclei or liver damages, should be based on the assessment of the potential of that chemical to interact with DNA and/or proteins.

TIMES models predicting the outcomes in Ames and the CA test have previously been published.[12,13] It has been established that the Ames test primarily accounts for the direct interaction of chemicals with DNA, whereas the in vitro CA test assesses both DNA and protein (e.g., histone, topoisomerase, spindle protein tubulus, and DNA repair enzymes) binding. This implies that Ames mutagenic chemicals should be CA positive, but the converse is not necessarily true. A recent comparative analysis of in vitro mutagenic data for a large

dx.doi.org/10.1021/tx200547s | Chem. Res. Toxicol. XXXX, XXX, XXX–XXX

Table 1. Alerting Groups and Descriptors Used in COREPA Models for Estimating Their Reactivity Associated with Supporting Mechanistic Information[a]

#	Alerting group	Chemical class	Descriptors in the COREPA model*	Interaction mechanism	Reference
1		Lactones	-	 dR – deoxyribose phosphate fragment Ring opening S_N2 reaction	(77)
2		Epoxides	*MW* E_{HOMO}	 dR – deoxyribose phosphate fragment Ring opening S_N2 reaction	(78)
3	—N=N—	Azo compounds	log K_{OW} Van der Waals surface	 (superoxide radical anions) $^{\cdot}OH$ DNA adducts Radical mechanism by reactive oxygen species (ROS) formation	(78)

[a]*E_{HOMO}, the energy of the highest occupied molecular orbital (eV); *MW*, molecular weight (Da); log K_{OW}, octanol–water partitioning coefficient (mol L_0^{-1} $mol^{-1}L_w$); and van der Waals surface area (Å^2).

number of chemicals confirmed this assumption. Eighty percent of chemicals that elicited bacterial mutagenicity (based on Ames test results) also induced CA, whereas only 60% of chemicals that induced CA were found to be active in the Ames test.[22,23] To distinguish these two mechanisms, the reactivity component of the newly derived models for MNT and liver genotoxicity was structured into two parts. The first part accounted for the interaction of chemicals with DNA. More than 60 alerting groups (being considered as a part of a future publication) were used to simulate covalent interaction with DNA. The use of each alert had been justified by the mechanistic interpretation of that interaction. Some alerts were additionally underpinned by mechanistically based COmmon REactivity PAttern (COREPA) 3D QSAR models.[24,25] Examples of these DNA binding alerts are presented in Table 1. The SAs are described together with physicochemical property/molecular parameter exclusion/inclusion rules. Supporting reaction mechanism information is also provided.

As seen from Table 1, the SAs can be categorized into two types: (1) those eliciting mutagenicity without the need for modulating factors (#1 in Table 1) and (2) those for which specific molecular parameter(s) define the degree of activation (#2 and #3 in Table 1).

The second part of the reactivity component accounts for the interaction of chemicals with specific proteins. More than 50 SAs were proposed that were associated with protein interaction (http://www.oasis-lmc.org/). Examples of protein binding alerts associated with parameters for reactivity and their supporting reaction mechanism information are presented in Table 2. These are characterized similarly—either requiring modulating factors (#1, #2, and #3 in Table 2) or not (#4 in Table 2).

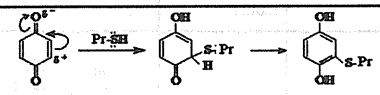
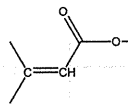
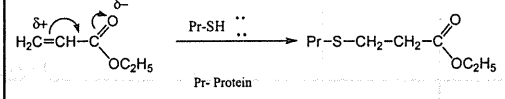
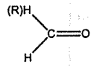
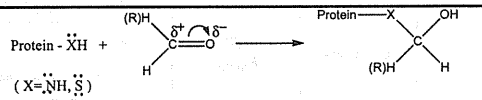
Most of the DNA binding alerts are also able to bind proteins. An example to demonstrate the mechanism by which a DNA binding alert interacts with proteins is presented for quinones in Figure 4.

Quinones are well-known mutagens, and they are included in the list of DNA-causing alerts. Topoisomerases are enzymes that participate in all stages of replication, functional activity, and structural maintenance of DNA. The inhibition of these enzymes by quinones is considered to elicit CA26. This is an example of how the same alert can elicit different outcomes depending on the interaction target. The structure of the reactivity component used in the in vivo genotoxicity models is provided in Figure 5.

A new chemical is first submitted to the reactivity component that encompasses the alerts associated with DNA interactions. A positive prediction for mutagenicity is assigned if the requirements for interaction with DNA are met, indicating that the ultimate mutagenic effect is due to this interaction mechanism. Regardless of whether the chemical meets the requirements for direct interaction with DNA, it is then forwarded to the second part of the reactivity component, which investigates the ability of the chemical to interact with proteins. This is to flag those cases where mutagenicity may arise by both mechanisms (direct interaction with DNA and interaction with protein) simultaneously. If the chemical passes through both parts of the reactivity component without being flagged for activity, a prediction of "unable to produce mutagenicity" is noted.

Conformational Analysis by Genetic Algorithm. To derive 3D QSARs, the flexibility of chemicals needs to be taken into account since this will give rise to the formation of many different conformers, and their reactivity profiles would accordingly differ. Common practice is to calculate molecular parameters for the lowest energy conformation, even though this necessarily may not be the form that drives the response and therefore not the most relevant one to study.[27] Given a systematic conformational analysis search would be computationally intensive (since the number of conformers would increase exponentially with the number of degrees of freedom), LMC derived a procedure to address the issue of conformation space using a genetic algorithm, which minimizes 3D similarity among generated conformers.[28] This made addressing the conformation space practical, even for large and very flexible chemicals. A procedure was also developed to saturate the conformation space, that is, to ensure consistency in the reproducibility of generated conformers and their distribution in the structural space.[28] This allowed the conformational space of chemicals to be populated with an optimal number of conformers.

Table 2. Alerting Groups for Protein Binding, Parameters for Reactivity, and Supporting Interaction Mechanisms[a]

#	Alerting group	Chemical class	Descriptors* in the model.	Interaction mechanism	Reference
1		Quinones	*MW*		(79)
3		Acrylates	log K_{OW}	 Pr- Protein Michael addition	(80)
4	 R=alkyl	Aldehydes	-	 (X=NH, S) (R= alkyl) Schiff base formation	(81)

[a] *MW*, molecular weight (Da); logK_{OW}, octanol–water partitioning coefficient (mol L_0^{-1} $mol^{-1}L_w$).

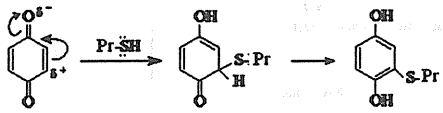


Figure 4. Interaction mechanism of quinones with proteins (Pr).

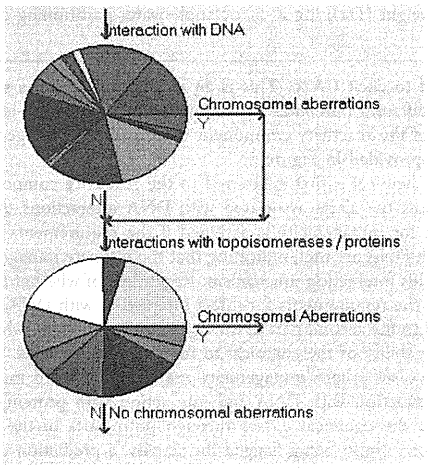


Figure 5. Structure of the reactivity component of the in vivo genotoxicity models.

TIMES. The TIMES platform comprises SA, 3D QSARs, and a metabolism simulator. This simulator comprises a list of hierarchically ordered transformations and a substructure matching engine for their implementation. The modeling is based on a probabilistic approach[29] whereby a hierarchy of transformations is defined by the probabilities of transformations determined in such a way as to reproduce a database of documented metabolic transformations or data for their rate of disappearance. The transformation probabilities are related to the feasibility of occurrence of various metabolic reactions. It is assumed that the transformations are independent and performed sequentially. Each molecular transformation consists of parent submolecular fragments, transformation products, and inhibiting masks. The latter play the role of reaction inhibitors. If a functional group assigned as a mask is attached to the target fragment, the execution of the transformation on the parent chemical is prevented. The presence of groups that can promote or inhibit metabolic reactions significantly increases the number of principal transformations. Currently, 343 principal transformations are used to model rat liver metabolism in vitro. The simulator starts by matching the parent molecule with the reaction fragment associated with the transformation having highest probability of occurrence. When a match is identified, the molecule is metabolized, and transformation products are treated as parent molecules for the next degradation step. The procedure is repeated for the newly formed chemicals until the product of probabilities of consecutively performed transformations reaches a user-defined threshold. The mathematical formalism defining the amount of metabolite, formation, and metabolism probabilities is described elsewhere.[6,29–31] The intent with this study was to refine the existing structure–activity and structure–metabolism rules within TIMES to account for the differences observed between the in vitro and the in vivo results. Where a realistic and feasible hypothesis could be generated and substantiated with data, these would inform the refinement of existing rules or introduction of new transformation rules.

RESULTS AND DISCUSSION

Workflow for Genotoxicity at Different Levels of Biological Organization. While the full set of data comprised 557 chemicals, a set of data where results from all assays were available were required to develop the mechanistic (Q)SAR models. Overall, calls for in vitro, liver genotoxicity, and in vivo MNT were available for 162 chemicals. Table 3 shows the list of 162 chemicals. A hierarchical workflow (Figure 6) outlines the results.

The first tier of in vitro tests comprises 162 chemicals that were either positive or negative in Ames, CA, and MLA. Four chemicals were assigned as inconclusive since Ames and CA data were found to be conflicting. All four were Ames positive but CA negative. The four chemicals were ethylene dichloride (107-06-2), thiabendazole (148-79-8), dibutylnitrosamine (924-16-3), and C.I. direct black 38 (1937-37-7). These were excluded from further study. Thirty-two (20%) of the 158 chemicals remaining were found to be in vitro negative, and 126 (80%) were found to elicit in vitro positive responses. Substances were categorized as negative if two or more results were negative and positive if they were positive in at least one of the three tests.

The 32 (20%) nonmutagenic chemicals in vitro were investigated in both liver and MNT in vivo tests. Thirty of the 32 in vitro nonmutagenic chemicals were confirmed negative in vivo

dx.doi.org/10.1021/tx200547s | Chem. Res. Toxicol. XXXX, XXX, XXX–XXX

Table 3. List of the 162 Chemicals and Their Summary Calls Both in Vitro and in Vivo Test Systems

CAS	name	ivt	liver	MNT
50-06-6	phenobarbital	1	1	1
50-32-8	benzo(*a*)pyrene	1	1	1
50-55-5	reserpine	0	0	0
51-03-6	piperonyl butoxide	1	0	0
51-79-6	urethane	1	1	1
52-24-4	thio-TEPA	1	1	1
56-04-2	methylthiouracil	0	0	0
56-23-5	carbon tetrachloride	0	0	0
56-57-5	4-nitroquinoline 1-oxide	1	1	1
56-75-7	chloramphenicol	0	0	0
57-14-7	dimazine	1	1	1
57-22-7	vincristine	1	0	1
57-30-7	phenobarbital, sodium	0	0	0
57-50-1	sucrose	0	0	0
57-57-8	propiolactone	1	1	0
57-97-6	7,12-dimethylbenz(A)anthracene	1	1	1
58-08-2	caffeine	1	0	0
58-89-9	lindane	0	0	0
59-05-2	methotrexate	1	1	1
59-89	*N*-nitrosomorpholine	1	1	1
60-09-2-3	*p*-aminoazobenzene	1	1	1
60-11-7	4-dimethylaminoazobenzene	1	1	1
60-35-5	acetamide	0	0	0
60-57-1	dieldrin	1	1	1
62-44-2	acetophenetidin	1	0	1
62-53-3	aniline	1	1	1
62-55-5	thioacetamide	1	0	1
64-86-8	colchicine	1	0	1
66-27-3	methyl methanesulfonate	1	1	1
67-20-9	nitrofurantion	1	1	0
67-66-3	chloroform	1	0	0
67-68-5	dimethyl sulfoxide	0	0	0
68-12-2	dimethylformamide	0	0	0
70-25-7	*N*-methyl-*N*′-nitro-*N*-nitrosoguanidine	1	1	1
71-43-2	benzene	1	1	1
75-07-0	acetaldehyde	1	1	1
75-09-2	methylene chloride	1	1	0
75-25-2	bromoform	1	0	0
75-56-9	propylene oxide	1	0	1
79-06-1	acrylamide	1	1	1
79-34-5	1,1,2,2-tetrachloroethane	1	1	1
81-07-2	saccharin	0	0	0
84-16-2	hexestrol	1	0	0
89-65-6	erythorbic acid	0	0	0
90-43-7	2-phenylphenol	1	1	0
91-20-3	naphthalene	1	0	0
91-59-8	2-naphthalenamine	1	1	1
91-64-5	coumarin	1	0	0
91-94-1	3,3′-dichlorobenzidine	1	1	1
92-52-4	biphenyl	1	1	0
92-67-1	4-biphenylamine	1	1	1
92-87-5	benzidine	1	1	1
95-50-1	1,2-dichlorobenzene	1	0	0
95-53-4	*o*-toluidine	1	1	0
95-80-7	2,4-diaminotoluene	1	1	0
95-83-0	4-chloro-1,2-diaminobenzene	1	1	1
96-09-3	styrene oxide	1	1	0
96-12-8	1,2-dibromo-3-chloropropane	1	1	1
96-45-7	ethylenethiourea	1	1	0
97-53-0	eugenol	1	0	0
97-56-3	*o*-aminoazotoluene	1	1	0
99-56-9	1,2-diamino-4-nitrobenzene	1	0	0
100-41-4	ethylbenzene	1	0	0
100-42-5	styrene	1	1	0
100-51-6	benzyl alcohol	1	0	0
100-75-4	1-nitrosopiperidine	1	1	0
101-14-4	4,4′-methylenebis(2-chlorobenzenamine)	1	1	1
101-77-9	4,4′-methylenebis(aniline)	1	1	1
103-33-3	aminoazobenzene	1	1	1
103-90-2	acetaminophen	1	1	1
104-55-2	cinnamaldehyde	1	0	0
105-11-3	*p*-quinone dioxime	1	0	0
105-60-2	hexahydro-2 h-azepin-2-one	0	0	0
106-46-7	1,4-dichlorobenzene	0	1	1
106-93-4	ethylene dibromide	1	1	0
106-99-0	butadiene	1	0	0
107-06-2	ethylene dichloride	no conclusion	1	0
107-13-1	acrylonitrile	1	0	0
108-88-3	toluene	0	0	0
108-95-2	phenol	1	1	0
110-00-9	furan	1	1	0
110-44-1	sorbic acid	0	0	0
110-86-1	pyridine	0	0	0
117-39-5	quercetin	1	0	0
117-81-7	bis(2-ethylhexyl)phthalate	0	0	0
118-96-7	2,4,6-trinitrotoluene	1	0	0
119-53-9	benzoin	1	0	0
119-93-7	tolidine	1	1	1
120-47-8	ethylparaben	1	0	0
120-71-8	*p*-cresidine	1	0	0
121-79-9	propyl gallate	1	0	0
123-91-1	1,4-dioxane	0	0	0
124-48-1	chlorodibromomethane	1	1	0
126-72-7	tris(2,3-dibromopropyl) Phosphate	1	1	1
128-37-0	butylated hydroxytoluene	1	0	0
128-44-9	saccharin, sodium	0	0	0
134-32-7	1-naphthylamine	1	1	1
136-40-3	phenazopyridine hydrochloride [USAN]	1	1	1
139-13-9	triglycollamic acid	1	1	0
140-11-4	benzyl acetate	0	0	0
140-88-5	ethyl acrylate	1	1	0
142-04-1	aniline HCl	1	1	1
147-94-4	cytosine arabinoside	1	0	1
148-79-8	thiabendazole	no conclusion	1	1
148-82-3	melphalan	1	1	1
301-04-2	lead acetate	1	0	0
305-03-3	chlorambucil	1	1	1
309-00-2	aldrin	1	0	0
366-70-1	procarbazine hydrochloride	1	1	1
427-51-0	cyproterone acetate	0	1	0
446-86-6	azathioprine	1	1	1
492-80-8	auramine	1	1	0
501-30-4	kojic acid	1	0	0
532-32-1	sodium benzoate	1	0	0
542-75-6	1,3-dichloropropene [BSI:ISO]	1	1	0
602-87-9	5-nitroacenaphthene	1	1	1
604-75-1	oxazepam	1	1	0
609-20-1	2,6-dichloro-*para*-phenylenediamine	1	1	1

dx.doi.org/10.1021/tx200547s | *Chem. Res. Toxicol.* XXXX, XXX, XXX–XXX