	GraphPad? Are there specific adjustments of the software preferences necessary? In general the decision criteria for positive or negative results should be more clearly defined and more attention should be paid to cytotoxic effects. For example, ketoconazole is considered positive in the antagonist assay and an IC50 of 6,98x10-6M is provided in table 12. However, at 10-6 M no effect on luciferase activity was observed in the antagonist assay and at 10-4 M ketoconazole is obviously highly cytotoxic (Fig. 10-2). The IC50 value seems to be based on a single data point (10-5 M) that revealed a significant change in luciferase activity with a cell viability >80% and, thus, is not really convincing. Since cytotoxicity can interfere with the detection of androgenic or anti-androgenic responses it might be more appropriate to regard data that is likely to be affected by cytotoxicity (like ketoconazole, but also DDT and methoxychlor) as equivocal / not conclusive or outside of the applicability domain of these assays. **Reviewer #2:** The following clarifying improvements need to be made with regard to the approximation of the IC50, EC50 and linear and non linear IC's 30 and 50, especially when the dose response curve is not complete. It cannot be considered a robust measure, only a less robust approximation. Also with respect of performance criteria and acceptable test performance that will need to be developed for the draft TG: for the agonist assay, there may well be problems with R1881 as it is an anabolic steroid with potentially highly limiting import restrictions and therefore import restriction for laboratory use also in some jurisdictions. This is probably because It is essentially the same as trenbolone, but as it has undergone 17alpha alkylation it can remain active after oral administration. It is also anticatabolic. Would MDHT be an appropriate alternative? **Reviewer #3:** Most of the protocols are very clear and adequate; however, there are some minor edits/additions that I suggest be made for ease of use. A graph of the expected DHT dose response would be helpful. How was concentration of DHT used in antagonist assays determined? It is stated in the report that it was about a PC80-100 but that is not stated in the protocol. My reservation here is that the report implies that sometimes 0.5 nM may be PC80 and at others a PC100. If this varies from plate to plate, then the PC10 and PC50 of test compounds may vary as much as 20% also. Inclusion of criteria for the DHT positive control is suggested. Each lab should run DHT dose response to confirm that they are getting correct response and to verify concentration of DHT needed for antagonist assay. Appendix 1 – section 1.8.2 and 1.8.3: The text says 175 cm2 flask, but I believe it should be 75 cm2. This appears to be a typographical error as 75 cm2 is correctly used through other parts of the protocol. Appendix 3, Section 4.5: Need to state the volume of lysis buffer to be added to each well as this would affect assay performance. I could not find this information in the SOP or any of the supporting protocols. Support protocol (SP) 1 – How long should cells be cultured before they can be used in an assay after reconstituted from frozen? It is my experience that usually 1 or 2 weeks of standard culture are required to allow full cell recovery and best

	performance in similar assays. What is required for this assay? Support protocol 2: Information on expected doubling time would be helpful. Clarify when or at what density, e.g. only at confluence, should cells be sub-cultured. Support protocol 4: Similar to mentioned above (for SP1), how long should cells be cultured before being used in an assay? Should only cells from confluent flasks be used in the assay? How often can cells be trypsinized without affecting assay performance? Perhaps cells for an assay should plan to be plated at the same time as propagation? Appendix 4 – (page 95) What is the purpose of adding Zeocin and Hygromycin? I am assuming they are selection antibiotics but I did not see it specifically stated anywhere. Page 98 – mid-page: What is meant by 56 nM DHT 0.1%DMSO solution? Should this be 56 nM DHT in DMSO? It is assumed so as then the 2 subsequent dilutions would result in 0.5 nM DHT in assay. Page 99 section 5.2 The referenced figure should be Fig 3 (not Fig 2). **Reviewer #4:** Only the decision criteria for evaluation of data are not clearly stated. Parameters used in the proposed assay system is not commonly used but considered to be effective for the purpose of the assay (prioritizing or screening). However, its calculation may not be easy for laboratories where newly introduce the assay. Therefore, it would be helpful to provide the validated calculation program (e.g. spreadsheet) as in ERTA assay (TG455). **The Panel agreed that this criterion has been partly met.**
4. Demonstration of within- and between-laboratory reproducibility	**Reviewer #1:** Intra- and inter-laboratory reproducibility has been demonstrated. For the agonist assay there was some variation in the Hill slope, but the calculated PC10, EC50 for the standard chemical (DHT) and a comparable strong agonist (R1881) were highly reproducible with very low CV values. The other three chemicals were identified correctly as negative. There were single experiments that resulted in PC10 values for Hydroxyflutamide and Diethylhexylphatalate and were not reproducible and therefore discarded but also demonstrate a high sensitivity of the system. The use of Hydroxyflutamide in the inter-laboratory study might have not been the best choice because of the problems with this substance in the agonist assay as described in paragraph 98, although in figure 10-2 no significant cytotoxicity can be observed up to concentration of 10-5 M (not 10-7 M as stated in paragraph 98). For the antagonist assay the CV values are somewhat higher but potential positives and negatives were correctly identified and the calculated IC50 and IC30 values reproducible in each lab and between labs. Only the one laboratory consistently detected anti-AR activity at lower concentrations which might be due to the lower Spike-in/PC DHT ratio. However, the reason for the lack of cytotoxicity data for the R1881 testing in all participating labs should be explained.

	Reviewer #2: Only for 5 substances for between laboratory reproducibility. Although this is generally acceptable for in vivo validations, for in vitro validations more substances are usually expected. Therefore, further demonstration of reproducibility in the naïve labs would be preferred. It would be particularly helpful to have information for a replacement for R1881, and a weak agonist and antagonist for which the dose response curve is not complete at 10uM so, up to cytotoxicity/limit of solubility testing (consideration of the solubility properties of the chemical)/ 100uM, plus another negative for both agonist and antagonist. These five examples of additional test substances would be very helpful for the development of the performance criteria for the draft TG. Testing at a max. concentration of 10uM may be too low to detect weakly active compounds, thus testing to the limit of solubility, the onset of cytotoxicity or higher doses may be required for pre-screening and subsequent screening if the pre-screening indicates that this is more correct and necessary. **Reviewer #3:** Generally CVs were low and reproducibility was quite good. However, a limited range of affinities of compounds was tested between labs. In the agonist assay, CVs for DHT run by three labs on three different days ranged from 0 to 8.0%. Reported overall CVs for each lab with DHT were 3.6 to 4.5% which is very good. CVs for coded compounds were also low. My reservation for the agonist assays is the fact that only strong agonists and negatives were tested in the interlaboratory study. No weak agonists were tested. Historically, variability often increases when weaker agonists are tested in vitro as the weak agonists are often a greater challenge. Variability in the pre-validation in one lab was acceptable for weaker agonists, which is sufficient as a larger number were tested in the assay, but no data is available for weak agonists from the inter-lab study to assess inter-lab reproducibility. In antagonist assays, only 2 antagonist were assessed, HF and BPA. Reproducibility for these 2 coded compounds was demonstrated and variability was acceptably low. Again, however, these were 2 of the more potent antagonists based on the pre-validation study results. My preference would have been to have at least one or two additional very weak antagonists evaluated in the inter-lab study. This is an important point as most environmental compounds identified to date that impact the AR are antagonists and many of them are weak antagonists in vitro. **Reviewer #4:** Provided in the validation report. **The Panel agreed that this criterion has been partly met.**

5. Demonstration of the test method's performance based on testing of representative reference chemicals	**Reviewer #1:** In the multi-lab study five coded chemicals were tested in four labs. The results correctly identified the expected positives and negatives in the agonist and antagonist assays. However, for the agonist assay the positives were DHT itself and R1881 that have comparable / identical activities to the control DHT and are structurally very similar /identical. Thus one can hardly conclude from this study that the assay will detect different kinds of chemicals that activate the AR. In addition these two chemicals have obviously very low PC50/PC10 values and it would have been more informative if chemicals with lower activities (higher PC10) values would have been tested. Testing these two substances only recapitulates the reproducibility of the agonist assay with the positive control substance as shown in table 15. For the antagonist assay it looks more convincing since two structurally distinct chemicals with anti-AR activity were identified correctly with similar but distinct IC50 / IC30 values. As pointed out in GD 34 "For the assessment of inter-laboratory reproducibility a subset of test substances used to assess accuracy might be appropriate, provided that the subset adequately represents an appropriate range of responses and physical / chemical properties for which the test method is proposed to be appropriate". In this respect, the data provided in this report do not allow a clear judgement of the relevance and reliability of the test method. **Reviewer #2:** *see respective answer to charge question 4* **Reviewer #3:** For the data shown the test method performed well in the pre-validation and inter-lab studies. In the inter-lab validation, compounds evaluated were tested using coded compounds to avoid unintentional bias. In the agonist assays 5 compounds were evaluated and 2 in the antagonist assay. The problem is that only relatively strong agonists and antagonists were evaluated. No weak agonists and no relatively weak antagonists were represented. Since one might expect, based on previous reports, that environmental compounds would have relatively weak activity, this is a limitation in the demonstration of assay performance. While a large number of compounds were not evaluated in the inter-laboratory study, I believe that testing a large number of compounds would be less valuable than testing a small number of compounds over a broader range of affinities. **Reviewer #4:** The performance was well-demonstrated by testing 45 chemicals including coded chemicals. Also, the test chemicals were basically selected from the well-known source (ICCVAM). The number of chemicals tested is not enough but in the range that can be tested, and the selection is reasonable, under the limited money and time. **The Panel agreed that this criterion has been partly met.**

6. Test methods evaluation related to existing relevant toxicity data	**Reviewer #1:** An obvious problem is the lack of a gold standard to compare the results with and to evaluate the relevance of the assay. It might be worth mentioning that the biological relevance of reference data might be for some substances not necessarily clear cut. In any case, the comparison of the data with the ICCVAM list (summarizing published results from other assays) and the additional data obtained from human AR-binding studies is an acceptable and in particular feasible way to evaluate the assay. The AR-Binding data is not discussed sufficiently, in particular the absence of significant binding of chemicals that elicit a response in the AR EcoScreen assay. It is argued in paragraph that this might be due to lower concentrations tested in the binding assay. This argument is not convincing because for the inter-lab study the maximum concentration tested was also 0,1M and, as a consequence, one would have to go back to the testing concentrations of the pre-validation studies (up to the available solubility or 1 M). Here it would be worthwhile to evaluate the relevance of positive responses in the AR EcoScreen assay that are either not detectable in the in vitro binding assay for technical reasons (comparison with published binding data) or mediated by other mechanisms than direct AR-binding. For example, for vinclozolin it has been shown that the substance itself, in contrast to its metabolites that are generated by hydrolysis, does hardly bind the AR receptor and weak in vitro ligand displacement activity for nonylphenol had also already been reported. However, there are also inconsistencies in the presented data. In Paragraph 116 and Table 14 (AR binding data) it is stated that 4-Hydroxytamoxifen were not detected as an AR agonist or AR antagonist despite significant binding to the AR receptor. However in table 8, it is listed as an antagonist in the AR EcoScreen, in accordance with published data. **Reviewer #2:** It is important to note that the ICVVAM lists were published in 2003, so are 8 years old, and I understand for the AR are due to be updated and corrected. In addition to comparison with AR binding assay data given in the report, comparison with the Hershberger TG would be helpful, and discussion of problems with comparison with the Hershberger, especially with the possible selection of additional test substances for the development of performance criteria. **Reviewer #3:** There is lack of a good "standard" for AR-TA assay data which makes this evaluation challenging. Comparisons were made to the ICCVAM list and to AR competitive binding data. These may or may not be the best data for such comparison. For agonist assays, concordance, sensitivity and specificity comparisons were most favourable when PC10 values from the AR-Eco-screen assay were compared to the ICCVAM list. There are two issues here: the use of PC10 values as criteria for a positive is questionable and questions exist with respect to the proper classification of some compounds on the ICCVAM list as positive or negative. With respect to the first issue, using Fig 8B as an example, actinomycin D, a prototypical inhibitor of gene transcription which I would expect to be a negative, was classified as a positive based on PC10 response. Further, there is no dose response – just a relatively flat line at or near the PC10 value.

	This compound should inhibit transcription of luciferase and be a negative. Hence, a compound that produced a similar curve to actinomycin-D should likely be questioned. Mifepristone produces a similar lack of dose response. Hence the biological relevance of a PC10 is questioned, especially without additional increase in response as concentrations are increased. Correlation to PC50 values would be preferred but due to solubility and cytotoxicity issues a PC50 values could not always be attained in the AR-Eco-screen, thus the correlation to PC50 was reduced as compared to a PC10 value. I also question whether the PC10 value is a statistically significant response. With respect to the second issue, classification of compounds on the ICCVAM list were determined from the literature and were compiled for data from many types of assays (including cell proliferation assays and yeast-based assays) not just AR-TA assays. Classification is currently based primarily on a weight-of-evidence type approach (number of papers reporting a positive) with little evaluation of the thoroughness of the assessment, quality of the data or methods/performance of the assay. In addition, with respect to antagonists, the ICCVAM list combined data which included the use of various differing "spike-in" compounds and their use at differing concentrations. Logic suggests that these factors would impact the results obtained in an assay and hence the classification of positive or negative. It is suspected that lower correlation of PC50 values may be related to these issues among others. That said, based on experience of the interaction compounds with the AR, the following additional comments should be considered: The approach taken to calculate EC50 and IC50 was the best approach (i.e. top and bottom of the test compound curve used without regard to the whether a full dose response was attained or not) is questioned. Calculation based on comparison to the dose response of the positive control (i.e. DHT) is suggested as percent of control. For example PC50 would be preferred. Poor correlation to the ICCVAM list for antagonists might be expected and may not be the best comparison until the ICCVAM list is updated. AR-Eco-Screen (based on data in Table 14) correlated well with AR binding data - Compared to the AR binding data the relevance of the assay was well demonstrated. **Reviewer #4:** This was done by comparing "human" AR binding assay data showing good performance. (see respective comment to charge question 2) There are no other suitable data sources to compare with. **The Panel agreed that this criterion has been partly met.**

7. Availability of all relevant data for expert review	**Reviewer #1:** For most parts the report did allow review, although the lack of primary data in some parts and the inconsistent use of different calculated values i.p. PC50, PC10, IC50, IC30 (with and without taking cytotoxicity data into account) caused difficulties to evaluate the data. In particular, a list of the substances and the respective primary data used to generate Table 9, Figure 9, Table 11 and Figure 11 would have been helpful. **Reviewer #2:** See the respective responses to charge questions 3 and 4. In addition, the discussion on PC50, PC10, EC, IC and lin.IC values needs further refinement and improved accuracy. **Reviewer #3:** A detailed test protocol was included.. I suggested a few minor edits above (see question 3) but overall the protocols were clear and easy to follow. Given the success of the multi-lab validation exercise, the protocol was sufficiently detailed to allow other labs to generate equivalent data. Benchmark data was provided by the use of a positive control (DHT) by which other labs could evaluate their performance. I might also suggest a small set of additional proficiency compounds over a range of activities. **Reviewer #4:** Data: Available; Protocols: The detailed protocols used during the multi-lab validation are available so that an independent lab can perform the assay. Also the performance criteria are provided. **The Panel agreed that this criterion has been met.**
8. GLP (ideally)	**Reviewer #1:** The inter-laboratory study was conducted under GLP conditions whereas the pre-validations studies (comprising most of the data) were not. In general that might be acceptable. However in this case only very few chemicals were tested in the inter-laboratory study, different test protocols had had been used for the pre-validation and inter-laboratory studies Since the validity of the test method cannot be concluded from the inter-laboratory study that was performed in accordance to GLP this requirement is not met. *(Comment by the Panel chair: The relevance of the reply to this question was discussed with the reviewer. As a result, Reviewer #1 considered this criterion as partly met.)* **Reviewer #2:** The pre-validation was in the spirit of GLP, and the actual validation was to GLP. **Reviewer #3:** It was clearly identified that data collected in the pre-validation assays were not collected under GLP. While

	GLP is preferable in all cases, I see no reason to exclude the pre-validation data based solely on this criteria. The Inter-laboratory validation was conducted under GLP and I suggest that this is sufficient. **Reviewer #4:** The data collected during the pre-validation phase were not obtained under GLP. However, it is considered to have no impact on the validation status as long as the all raw data of the assay are available. In addition, the results in pre-validation study had been already published in the peer-reviewed scientific journal. **The Panel agreed that this criterion has been partly met.**

Remark by the panel chair: Note that the initial replies of the reviewers as provided by February 11, 2011, have been collated here (see paragraph 3). Furthermore, the numbering of the reviewers does not correspond to the alphabetical order of Annex 1.

ANNEX 4a

January 24, 2011

Response to the inquiry dated Jan. 17, 2011

Masahiro Takeyoshi, Ph.D. , CERI-Japan

1. Fluoranthene stimulated Renilla activity in the antagonist assay at concentrations that also seem to stimulate a positive response in the agonist assay (10-5M). At least that is how it looks like comparing figure 8 and figure 10-2. However in table 10, the EC50 value is 1,29x10-8M.

So far, the cause for the dose-dependent induction of Renilla luciferase is unknown. However this phenomenon would not be relevant with androgenic activity.

Fluoranthene showed clear antagonistic response, it means this chemical interact with androgen receptor. So we think Fluoranthene may be partial agonist.

We have rechecked the values reported in the validation report, article published in Toxicology in vitro and the raw data related. Consequently, there is a mistake in the IC50 values reported in the article, IC50 values for Mifepristone and Fluoranthene were exchanged by mistake. And also we have proofreading error.

We cited the IC50 values from the article to the validation report, and resulted in the mistakes in the validation report.

Correct IC50 values are provided in the following table.

	IC50 reported	Correct IC50 value
Mifepristone	3.23x10-5	1.29x10-8
Fluoranthene	1.29x10-8	3.23x10-5

As a consequence of reinspection, there was no other mistake of IC50 values in the validation report.

2. DTT and Methoxychlor have been classified as negatives in the antagonist assay (Table 8), but IC50 values are provided in Table 12

These chemical showed cytotoxicity (below 80% viability) at 10-6M. So these chemicals were decided as negative in the antagonist assay.

IC50 values can be calculated if the cytotoxic effects were observed. So we provided the IC50 values for these two chemicals in table 12, But cytotoxicity of these chemicals were mentioned in Paragraph 111.

Where we have another mistake in Para 111; the cytotoxic dose of DTT and Methoxychlor are shown as "10-9 or 10-8 M", but correct values are "10-6 or 10-5 M".

As a consequence of reinspection about this issue, there was no other mistake in the validation report.

3. Corticosterone was classified as a positive in the agonist assay and in paragraph 103 it is argued that this might be due to the high concentrations tested. However, in table 10 the EC50 value for Corticosterone is 4,5x10-7 M.

EC50 values can be calculated automatically by the calculation softwares, such as GraphPad Prism, if the chemical has not high induction level. Indeed, EC50 can be calculated for corticosterone, but the chemical failed to yield PC50 value, so the chemical was decided as negative.

We have been advocated that EC50 is not appropriate to express the potency of the hormonal effect. Because EC50 is defined as 50% point between top and bottom values, even if the top value is only 3% level of the maximum induction level of positive control substance. This is the critical issue to evaluate the hormonal potency of chemicals by EC50 values. So we emphasize that PC50 value is suitable for evaluating hormonal potency because the actual induction level compared with maximum induction level of the positive control substance is considered in the PC50.

ANNEX 4b

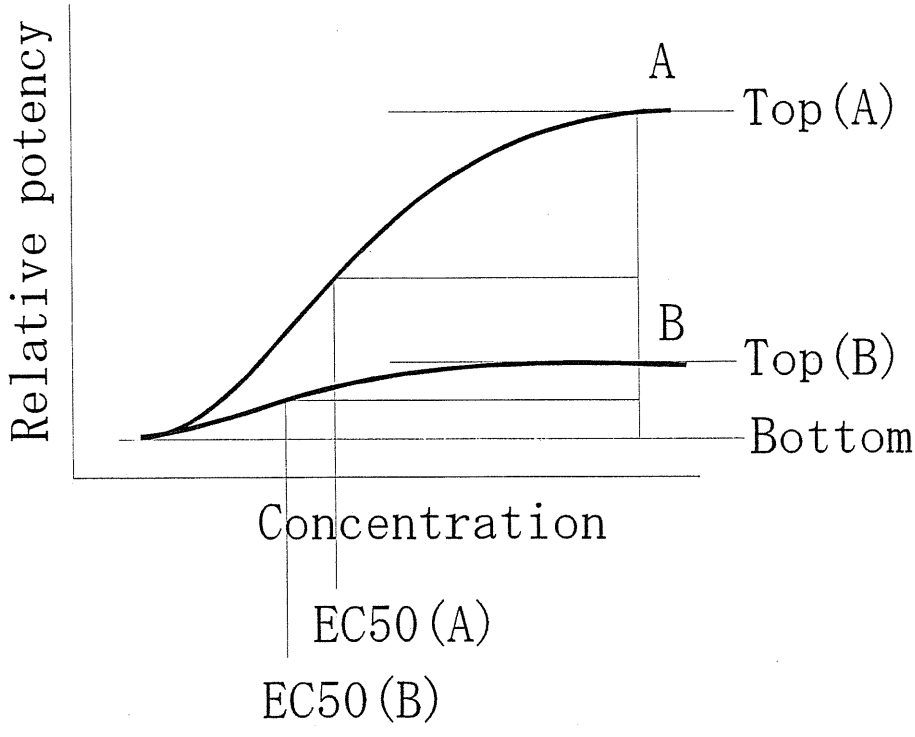
Questions (send by e-mail on January 28, 2011):

Two further issues in relation to the third point were brought up (see Annex 4a.3.):
1. The answer refers to the PC50. But should it not be the PC10?
2. How are the EC50 calculated? Do the 50% relate to the response of the positive control or to the maximum response achieved with the chemical?

Answers (received by e-mail on February 03, 2011):

To 1.:
Yes! In the agonist assay, positive response is decided if the chemical has PC10 response. Corticosterone has only PC10, and not has PC50. It means the chemical has low top response. Paragraph 103 refers only to the reason for getting positive response of this chemical.

To 2.:
EC50 is calculated with logistic equation by using Graphpad PRISM etc. This parameter is defined as 50% point between Top and Bottom responses. Please see the attached figure (below); Top values are varied from chemical to chemical. Corticosterone is corresponding the response curve (B). In this case, having low top response, EC50 value would be apparently lower value than that obtained regular response curve (A) having high top value.
Meanwhile, PC50 is defined as 50% point of maximum response induced by positive control substance.



厚生労働科学研究費補助金(化学物質リスク研究事業)
平成21年度～平成23年度分担研究報告書

遺伝毒性試験法コメットアッセイ(*in vivo*)のバリデーション

研究分担者　小島　肇　国立医薬品食品衛生研究所　薬理部新規試験法評価室

研究要旨
本研究班ではコメットアッセイの国際的なバリデーション実行委員会を組織し、本試験の専門家とコンセンサスを取りながら、コメットアッセイの標準プロトコルの合意を目指してバリデーション研究を実施してきた。
バリデーションPhaseⅣ-1において、参加13施設（うち、国内4施設）の協力を得て、4物質を用いたバリデーション研究を実施し、良好な施設間再現性が得られることを確認できた。バリデーションPhaseⅣ-2において、同施設の協力を得て、各施設が1～6物質を評価するバリデーションを実施し、作用機構に立脚した高い予測性を確認できた。

キーワード：遺伝毒性、コメットアッセイ、*in vivo*、バリデーション

協力研究者
宇野　芳文　田辺三菱製薬株式会社　研究本部　安全性研究所　第一部長
大森　崇　同志社大学文化情報学部　疫学・生物統計学研究室　准教授
林　真　食品医農薬安全評価センター　センター長（副理事長）

A．研究目的

遺伝毒性試験の中で、コメットアッセイとは単細胞ゲル電気泳動法とも呼ばれ、単離した細胞または核をアガロースに閉じ込めて融解した後、アルカリ処理で二本鎖DNAを単鎖にし、電気泳動による泳動パターンの変化によりDNA鎖切断などを検出する方法である。正常な細胞のDNAは非常に大きな分子であり、電気泳動してもほとんど移動せず、球形の核として観察される。一方、DNAで切断などが起こっている場合にはDNA断片の大きさに応じて移動し、球形の核を頭に尾を引いた彗星のような泳動パターンとなる[1)]。

本方法は、*in vivo*でも*in vitro*でも試験可能であること、細胞が得られるならばどのような臓器、器官でも試験可能であること、短期間で結果が得られること、初期のDNA損傷を検出できることから広く利用されている。しかし、研究室間の再現性を検証するバリデーションが実施されておらず、プロトコルの国際的な合意がなされてこなかった。このような課題を解決すべく、平成20年度（2008年度）までに実施された「厚生労働科学研究　リスク研究事業」において、日本環境変異原学会/哺乳類変異原性(MMS)研究会、欧米の研究機関と協力して国際的なプレバリデーション（phaseⅢまで）を実施し、コメットアッセイの標準プロトコルを確定した。これまでの経緯を図1に示した。平成21年度以降、さらにバリデーションを実施し、将来的にはOECDガイドラインへの掲載を目指すものである。

2009年度（平成21年度）には、Phase IV-1バリデーション研究として、平成20年度までに確定できたプロトコール(ver.14.1)を用いて、4被験物質（エチルメタンスルフォネート：EMS、2－アセチルアミノフルオレン:2-AAF、ニトロソジメチルウレア:NDU、マンニトール）をコード化し、参加13施設に1もしくは2物質を配布し、施設間再現性を検討した。2010年度（平成22年度）から2011年度（平成23年度）に掛けて、バリデーション PhaseⅣ-2 として参加14施設（うち、国内4施設）の協力を得て、各施設が 1～6 物質を評価するバリデーションを実施した。

また、その過程で明らかになったコメットとヘッジホッグの判別方法の不統一を解消し、国際的な基準の統一を図るため、図解集（カラーアトラス）を作成した。

B. 研究方法

B-1）組織

本バリデーション組織は2006年度（平成18年度）に設立した。2009年度は第6回国際バリデーシ

ョン実行委員会（以後、実行委員会と記す）を 9 月にフィレンツエ（イタリア）で、第 7 回実行委員会を 3 月にソルトレークシティ（米国）で開催した。2010 年度は、2 月にハンチントン（英国）で第 8 回実行委員会を、2011 年度は第 9 回実行委員会会議（以後、京都会議と記す）を 9 月に京都（同志社大）で開催した。

B-1-1）国際実行委員会

委員長　林　真（食品医農薬安全センター：以下、安評センターと記す）

In vivo 担当委員長

宇野芳文（田辺三菱製薬株式会社）

委員　L. Schectmann（前 Interagency Coordinating Committee on the Validation of Alternative Methods：ICCVAM）

R. Tice (The NTP Interagency Center for the Evaluation of Alternative Toxicological Methods: NICEATM)

R. Corvi（European Centre for the Validation of Alternative Methods：ECVAM）

事務局　小島　肇（国立衛研　安全性生物試験研究センター　薬理部：Japanese Centre for the Validation of Alternative Methods：JaCVAM）

B-1-2）国内実行委員会

委員長　林　真

委員　宇野芳文

浅野哲秀（元・日東電工株式会社）

中嶋　圓（安評センター）

森田　健（国立衛研　医薬安全科学部）

本間正充（国立衛研　変異遺伝部）

小島　肇

B-1-3）コンサルタント

P. Escobar (Boehringer-Ingelheim)

D. Lovell (Univ. of Surrey)

大森　崇（同志社大学文化情報学部）

大野泰雄（国立衛研）

B-1-4）バリデーション参加施設

表 1 に示す 14 施設が参加した。

B-2)試験方法

B-2-1）バリデーション Phase IV-1

Phase IV-1 バリデーション研究では、平成 20 年度までに確定できたプロトコール(ver.14.1)を用いて、4 被験物質（エチルメタンスルフォネート：EMS、2－アセチルアミノフルオレン：2-AAF、ニトロソジメチルウレア：NDU、マンニトール）をコード化し、参加 13 施設に 1 もしくは 2 物質を配布し、施設間再現性を検討した。

EMS　200mg/kg 濃度を陽性対照とし、被験物質の適用濃度及び溶媒は各施設に選択させた。

B-2-2）バリデーション Phase IV-2

PhaseIV-1 を経て微修正されたプロトコール(ver.14.2)を用い、合計40の被験物質を14施設に配布し（3物質／施設）、in vivoにおける遺伝毒性をどの程度評価できるかを目的に、次段階の本バリデーションを開始した。遺伝毒性発癌物質、非遺伝毒性発癌物質、遺伝毒性非発癌物質、非遺伝毒性非発癌物質で分類した40物質を各施設に配布した。被験物質の配布は、基本的に国立衛研　薬理部　新規試験法評価室で実施した。米国の参加施設については、NICEATMの支援を受けて配布した。被験物質コード番号は、A4201～A4240とした（表2参照）。

Ver.14.1から14.2におけるプロトコルの変更点を表3に示す。Ver.14と比較して大きな変更ではないと考えている。

プロトコルの概要を以下に示す。

①動物：Crl: CD（SD)ラット雄　7-9 週令を 5 匹／群使用

②投与方法：3 回の強制経口投与（初回投与 21 時間後に 2 回目、45 時間後に 3 回目を投与し、その 3 時間後に安楽死させ、臓器をサンプリング）

③適用臓器：胃及び肝臓

④サンプル：単一細胞を使用

⑤電気泳動：低温、20 分の泳動で実施

⑥指標：テールに含まれる DNA 量の細胞全体量に対する割合（%）の平均値、%DNA in tail

⑦標本観察：サイバーゴールドで染色して観察

なお、%DNA　in tailにおけるデータ採用基準は以下に示す通りである。

陰性対照　肝臓の平均値　1-8%

胃の平均値　1-20%

陽性対照：EMS(エチルメタン酸スルフォネート)

200mg/kg、経口2回投与、臓器を問わず、

溶媒との差　5%以上

溶媒との比　2倍以上

B-2-3)コメットアトラスの作成

今までのバリデーション研究の過程で結果が安定しない原因が種々明らかになり、その一つがコメット結果の判定方法の不統一によると判断とされた。そこで本問題を解消し、国際的な基準の統一を図るため、図解集（カラーアトラス）の作成を行った。

表１．Phase IV-1 及び２バリデーションの協力施設

施設名	国名	代表者
AstraZeneca	UK	Catherine Smith
Bayer HelthCare	Germany	Uta Wirnitzer
BioReliance*	USA	Buba Krsmanovic
Covance	UK	Lucinda Williams
Food and Drug Safety Center*	JPN	Kohji Yamakage
Health Canada	Canada	James P. McNamee
Huntingdon Life Sciences*	UK	Brian Burlinson
Johnson & Johnson	Belgium	Marlies De Boeck
Merck*	USA	Richard D. Storer
Mitsubishi Chemical Safety Institute	JPN	Hironao Takasawa
Novartis Pharma	Switzerland	Ulla Plappert-Helbig
Sumitomo Chemical	JPN	Sachiko Kitamoto
The Institute of Environmental Toxicology	JPN	Kunio Wada
ILS	USA	Cheryl A. Hobbs

*:Leading laboratory

表 2.　被験物質コード、物質名、カテゴリー、実施施設、溶媒、実験濃度

Test chemical code	Test chemical name (CASRN)	Category of genotoxicity and carcinogenicity	Lab tested (coded lab name)	Vehicle	Dose level (mg/kg/day)
A4114	2-Acetylaminofluorene (53-96-3)	Genotoxic carcinogen	Lab O	Corn oil	250, 500, 1000
A4201	1,3-Dichloropropene (542-75-6)	Genotoxic carcinogen	Lab C	Corn oil	50, 100, 200
A4202	Ethionamide (536-33-4)	Non-genotoxic non-carcinogen	Lab C	Corn oil	125, 250, 500
A4203	Buslfan (55-98-1)	Genotoxic carcinogen	Lab C	Corn oil	10, 20, 40
A4204	*N*-Nitrosodimethylamine (62-75-9)	Genotoxic carcinogen	Lab L	Saline	2.5, 5, 10
			Lab O *	Saline *	0.63, 1.25, 2.5 *
A4205	Ampicillin trihydrate (7177-48-2)	Non-genotoxic non-carcinogen	Lab L	Saline	25, 50, 100
				Corn oil	500, 1000, 2000
A4206	1,2-Dimethylhydrazine dihydrochloride (306-37-6)	Genotoxic carcinogen	Lab L	Saline	6.25, 12.5, 25
			Lab O *	Saline *	1.56, 3.13, 6.25 *
A4207	Isobutyraldehyde (78-84-2)	Non-genotoxic non-carcinogen	Lab B	Corn oil	500, 1000, 2000
A4208	Cisplatin (15663-27-1)	Genotoxic carcinogen	Lab B	0.5% CMC	6, 12.5, 25
A4209	Azidothymidine (30516-87-1)	Genotoxic carcinogen	Lab B	0.5% CMC	500, 1000, 2000

A4210	*p*-Dichloroaniline (106-47-8)	Genotoxic carcinogen	Lab D	Corn oil	37.5, 75, 150
A4211	*t*-Butylhydroquinone (1948-33-0)	Non-genotoxic non-carcinogen	Lab D	Corn oil	131.3, 262.5, 525
A4212	Methyl carbamate (598-55-0)	Non-genotoxic carcinogen	Lab D	Saline	500, 1000, 2000
A4213	Methyl methanesulfonate (66-27-3)	Genotoxic carcinogen	Lab G	Saline	20, 40, 80
A4214	2,6-Diaminotoluene (823-40-5)	Genotoxic non-carcinogen	Lab G	Corn oil	150, 300, 600
A4215	5-Fluorouracil (51-21-8)	Genotoxic non-carcinogen	Lab G	Saline	25, 50, 100
A4216	8-Hydroxyquinoline (148-24-3)	Genotoxic non-carcinogen	Lab N	Corn oil	125, 250, 500
A4217	Hydroquinone (123-31-9)	Genotoxic carcinogen	Lab N	Saline	125, 250, 500 [1)]
A4218	Saccharin (81-07-2)	Non-genotoxic carcinogen	Lab N	Corn oil	500, 1000, 2000
A4219	Sodium arsenite (7784-46-5)	Genotoxic carcinogen	Lab M	Saline	7.5, 15, 30
			Lab O *	Saline *	7.5, 15, 30 *
A4220	Thioacetamide (62-55-5)	Non-genotoxic carcinogen	Lab M	Saline	19, 38. 75
A4221	Diethanolamine (111-42-2)	Non-genotoxic carcinogen	Lab M	Saline	175, 350, 700
A4222	*p*-Phenylenediamine dihydrochloride (624-18-0)	Genotixic non-carcinogen	Lab K	Saline	25, 50, 100
A4223	*o*-Phenylphenol sodium salt (132-27-4)	Non-genotixic carcinogen	Lab K	Corn oil	250, 500, 1000
A4224	2,4-Diaminotoluene (95-80-7)	Genotoxic carcinogen	Lab K	Saline	100, 150, 200
					37.5, 75, 150
A4225	4,4'-Oxydianiline (101-80-4)	Genotoxic carcinogen	Lab H	0.5% CMC	50, 100, 200
A4226	*o*-Anisidine (90-04-0)	Genotoxic carcinogen	Lab O	Corn oil	150, 300, 600
A4227	Sodium chloride (7647-14-5)	Non-genotoxic non-carcinogen	Lab O	Water	500, 1000, 2000
A4228	Acrylonitrile (107-13-1)	Genotoxic carcinogen	Lab E	Corn oil	15.7, 31.3, 62.5
A4229	9-Aminoacridine hydrochloride monohydrate (52417-22-8)	Genotoxic non-carcinogen	Lab E	Corn oil	15.7, 31.3, 62.5
A4230	Ethanol (64-17-5)	Non-genotoxic carcinogen	Lab E	Saline	500, 1000, 2000
A4231	1,2-Dibromomethane (106-93-4)	Genotoxic carcinogen	Lab J	Corn oil	25, 50, 100
A4232	*p*-Anisidine (104-94-9)	Genotoxic non-carcinogen	Lab J	0.5% CMC	125, 250, 500
A4233	*o*-Anthranilic acid (118-92-3)	Non-genotoxic non-carcinogen	Lab J	0.5% CMC	500, 1000, 2000
A4234	Benzene (71-43-2)	Genotoxic carcinogen	Lab I	Corn oil	500, 1000, 2000
A4235	Di(2-ethylhexyl)phthalate (117-81-7)	Non-genotoxic carcinogen	Lab I	Corn oil	500, 1000, 2000
A4236	Trisodium EDTA monohydrate (10378-22-0)	Non-genotoxic non-carcinogen	Lab I	Saline	500, 1000, 2000

A4237	Cadmium chloride (10108-64-2)	Genotoxic carcinogen	Lab F	Saline	20, 40, 80
A4238	Chloroform (67-66-3)	Non-genotoxic carcinogen	Lab F	Corn oil	125, 250, 500
A4239	D,L-Menthol (15356-70-4)	Non-genotoxic non-carcinogen	Lab F	Corn oil	125, 250, 500

* 溶媒と実験濃度は国際実行委員会によって指示された。

表 3. Ver.14.1 から 14.2 におけるプロトコルの変更点

No.	項目	変更点
1	末梢血による小核試験のオプション	検出を意図するなら 4 回投与が必要かもしれない
2	サンプリングと溶解時間	記録
3	電気泳動時間	20 分間以上
4	計測可能な細胞数	90％以上の%DNA in tail の細胞は除外
5	標本枚数	2 スライド/動物で変更なし
6	陽性対照群の匹数	削減可能だがバリデーションでは 5 匹で変更なし
7	結果の評価	リニアトレンドテストでの検定結果も考慮

C. 研究結果

C-1) バリデーション Phase IV-1

陽性対照物質 EMS は、すべての施設で肝臓、胃とも溶媒(陰性)対照と比較して統計学的に有意であった。この結果から、すべての被験物質の結果は採用できると判断した。陰性対照のコメット値は、ほとんどの施設がデータ採用基準である肝臓 1-8％、胃 1-30％の中にあった。2 施設が肝臓で 1％未満であった。

それぞれの被験物質の肝臓及び胃における結果を図1～8に示した。図 1 及び2に示すように、EMS ではすべての施設において濃度依存的なコメットの増加を認めた。これは図5及び6に示すように、NDUでも同様の結果であった。一方、図3及び4に示すように、すべての施設でマンニトール、図7及び8に示すように、2/3施設で2－AAFにはコメット数の増加を認めなかった。2-AAFにおけるLab Kの結果は濃度依存性は明瞭ではないが、すべての濃度で有意な結果が得られた。媒体はコーン油を用いている。Lab Bは0.5% CMC溶液を媒体として実施したが、高用量（1000mg/kg）まで毒性が見られず、投与量が不十分であると考えられる。Lab Hではコードが開示されてしまい、2－AAFの毒性情報をもとに濃度を選択したため、毒性はでていないが、コメットも誘発されなかった。

C-2) バリデーションPhase IV-2

C-2-1) 被験物質の配布およびデータ収集

被験物質は、2010 年初めに 3 物質／施設に配布された。一部物質の決定が遅れたこと及び連絡の不手際もあり、メルクとバイオリライアンスに各 1 物質が届いていないことが 2010 年末に判明した。これらは 2011 年 2 月末までに送付した。結果的に、Lab H は 1 物質のみ実施したが、その他の施設は 3 物質を実施した。

2011 年 5 月までに、それまでに配布しや被験物質のデータはすべて収集され、その結果を 2011 年 9 月の京都での会議で議論した。結果として、A4219、A4204 及び A4206 の 3 物質について、再試験が必要と判断された。これらを施設に伝え、再試験を打診したが、実施不可能との返事を受けた。そこで、その実施を京都会議で受諾した Lab O により、3 物質の追加試験が実施されることになった。Lab O は 2012 年 2 月までに 3 物質の実験を終了し、結果を事務局に送付してきた。

C-2-2)データの採用

すべてのデータを統計学者がスクリーニングし、データを確定した。陰性対照（肝臓、胃）の結果を、図 9 及び 10 に示した。肝臓においては、1 施設の 1 回の結果を除いて、いずれも 1-8％のデータ受入れ基準を満たしていた。高い施設内再現性が確認できたが、施設間の値にはわずかに差が見ら

れた。図 9 に示すように、受入れ基準を満たさなかったデータは被験物質コード A4236 の陰性対照であり、%DNA in tail が基準の 1 %より低かった。その他のデータはすべて基準を満たしていた。この施設は全体的に陰性対照値が低めであることもあり、国際実行委員会は本データの採用を決めた。施設毎の溶媒対照としては、蒸留水、生理食塩水、0.5%CMC、コーン油が使われていたが、平均値は 2〜3.5%の間にあり、国際実行委員会は差がないと判断した（データは示していない）。

一方、胃においては、すべての結果が 1-20%のデータ受入れ基準を満たしており、高い施設内及び施設間再現性が確認できた。施設毎の溶媒対照としては、蒸留水、生理食塩水、0.5%CMC、コーン油が使われていたが、コーン油の平均値は 14.0(n=16)と他の平均値 9 前後よりやや高いと考察された（データは示していない）。

陽性対照（肝臓、胃）の結果を、図 11 及び 12 に示した。肝臓においても胃においても、すべての結果が統計学的に有意に増加していた。

以上の結果から、すべてのデータは以後の解析に使用できることを国際実行委員会が確認した。

その他の問題として、標本観察において Lab M がプロトコルに定められたサイバーゴールドではなく、エチジウムブロマイド（EB）を使用した。国際実行委員会では本逸脱について審議し、EB が一般的にコメットアッセイの DNA 染色に使用されている実績を考慮するとともに、陰性対照値がデータ採用基準内であったころから、この結果も採用となった。

C-2-3)被験物質結果の扱い

すべての被験物質コード、物質名、カテゴリー、実施施設、溶媒、実験濃度を表 3 に示した。なお、以下の議論はコード開示前になされた。判定結果はそれぞれの施設における判定と国際実行委員会の解釈が食い違う場合もあるが、参加施設も含めた京都会議で議論し、最終判定を決めた。

昨年度に報告したデータを吟味した結果、A4205 及び A4217 の結果が不採用と判断された問題については、最高投与量においても動物に毒性兆候がなく、陰性と判断できないとされたことによる。これらの物質では溶媒を変更して濃度を溶解度限界まで高める設定を行う等により再試験が実施され、A4205 は Lab L による溶媒変更して達成した最高適用濃度の再試験結果が採用された。一方、A4217 は Lab N により 750mg/kg で再試験がなされたが、結果として致死用量であったことが確認された。よって、第一回目の結果（500mg/kg）が最終結果として採用された。

京都会議では、A4219、A4204 及び A4206 に再試験が要求された。この理由として、A4219 は統計学的な評価が難しいにも関わらず、Lab M が陽性と判断したことによる。また、A4204、A4206 については、高用量において%DNA in tail の増加が認められるものの、病理学的検査で細胞毒性が認められ、より低用量での再試験が Lab L にて必要と判断されたことによる。

結果として、両施設とも再試験を受けられないと回答してきたことから、C-2-1)にも記載したように、Lab O にて再実験が実施されることになった。

最終的には、C-2-2)に記述したように、Lab O は 2012 年 2 月までにすべての実験を終了し、それらの結果が事務局に送付されてきた。

C-2-4)最終判定

表 4 に示す結果は、コード開示後に議論された結果であるが、すべての結果を含む最終判定は、まだ案の段階であり、国際実行委員会の最終確認がなされていない。添付資料 2 及び 3 に示す報告書案を示す。

これまでの結果及び判定を以下に示す。

1） 遺伝毒性発癌物質

19 物質中、12 物質で肝臓及び／または胃において、%DNA in tail の統計学的に有意な増加が認められた。12 物質中、A4220:Thiacetamide 及び A4228:Acrylonitritle の結果は、施設と最終結果が食い違っているが、いずれも陽性と国際実行委員会は判断している。残りの 7 物質の内、5 物質（A4114:2-Acetylaminofluorene、A4226: *o*-Anisidine、A4234:Benzene 、A4203:Buslfan 、A4217: Hydroquinone）は偽陰性と判断された。さらに、A4225:4,4'-Oxydianiline は胃において%DNA in tail が有意に減少した（肝臓は変化なし）。A4219: Sodium arsenite は LabM の報告において、肝臓で%DNA in tail が増加したとされ、再試験で LabO においては陰性と判定していることもあり、国際実行委員会は Equivocal と判断した。

2） 遺伝毒性非発癌物質

6 物質中、1 物質 A4214: 2,6-Diaminotoluene のみに肝臓で%DNA in tail が増加した。

3） 非遺伝毒性発癌物質

7 物質中、1 物質 A4238: Chloroform のみが肝臓で%DNA in tail が増加した。しかし、病理学的な解析も合わせ、最終的には陰性と判断されている。よって、本分類はすべての結果が一致したことになる。

4） 非遺伝毒性非発癌物質

8 物質中、1 物質 A4211: *t*-Butylhydroquinone のみが肝臓で%DNA in tail が増加した。国際実行委員会は陽性と判断したが、実施施設はヒストリカルコントロールの範囲内と判定し、陰性と判断している。

C-3)コメットアトラス

実験者の判定基準を明確にするため、現在、MMS 研究会が数百枚の写真から選んだ 150 例の画像を用意し、参加施設に配布して判定させている。得られたコンセンサスをもとに、判定基準を明示した。カラーアトラスを作成した。テキストとして発行するため出版社と打ち合わせた。

D. 考察

PhaseⅣ-1バリデーションの4被験物質の結果から、3-4施設／物質の施設間再現性は高いと判断された。また、陽性対照物質の結果から、施設内再現性の高さも確認できた。以上の点から、PhaseⅣ-1バリデーションはその目的である施設間再現性の確認を達成できたと考えている。

PhaseⅣ-2バリデーションにおいても、ほとんどの施設の結果が、陰性及び陽性対照の適合基準を満たしており、バリデーションを無事終了できた。被験物質の判定結果も、遺伝毒性発癌物質において、偽陽性物質が多く検出されたものの、コメットアッセイの作用機構を考慮すると妥当なものであり、作用機構が異なる小核試験と組み合わせることで予測性が確保できると現時点では国際実行委員会が判断している。

なお、以上の施設の判定結果と最終判定が食い違った物質、病理学的な指標を考慮して判定が変わった物質、ヘッジホッグの扱い等については、今後の国際実行委員会の議論により判定結果が変わりうることから明記を避けた。

E. 結論

バリデーションPhaseⅣ-1の13施設（海外9施設及び国内4施設）からなる多数施設のバリデーションにより、陽性対照物質の結果から、施設内再現性を確認でき、4被験物質のみではあるが、施設間の再現性を確認で　きた。

バリデーションPhaseⅣ-2において、バリデーションPhaseⅣ-2において、参加14施設（うち、国内4施設）の協力を得て、各施設が1～6物質を評価するバリデーションを実施した。結果として、平成24年初頭にすべての実験を終了した。

40物質の結果を解析中であるが、現時点では、判定結果もコメットアッセイの作用機構を考慮すると妥当なものであると国際実行委員会が判断した。

F. 参考文献

1) 日本トキシコロジー学会教育委員会編集、トキシコロジー、ｐ142、朝倉書店(2002)

2) FDA　Guidance、
http://www.fda.gov/cder/guidance/index.htm

3) Burlinson B, Tice RR, Speit G, Agurell E, Brendler-Schwaab SY, Collins AR, Escobar P, Honma M, Kumaravel TS, Nakajima M, Sasaki YF, Thybaud V, Uno Y, Vasquez M, Hartmann A; In Vivo Comet Assay Workgroup, part of the Fourth International Workgroup on Genotoxicity Testing Fourth International Workgroup on Genotoxicity testing: results of the in vivo Comet assay workgroup. Mutat Res. 627(1):31-5(2007)

4) Hartmann A, Agurell E, Beevers C, Brendler-Schwaab S, Burlinson B, Clay P, Collins A, Smith A, Speit G, Thybaud V, Tice RR; 4th International Comet Assay Workshop. Recommendations for conducting the in vivo alkaline Comet assay. 4th International Comet Assay Workshop, Mutagenesis. 18(1), 45-51(2003)

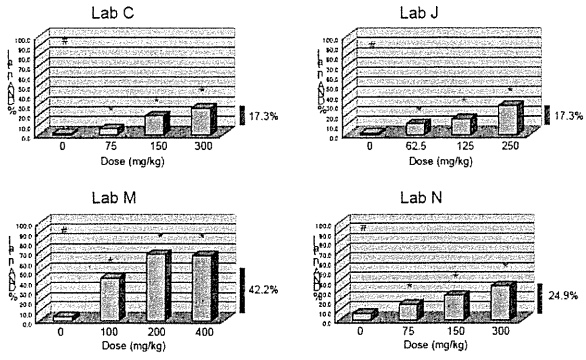


図1. EMS: 肝臓

* Dunnett test (P<0.05, two-sided); # Linear trend test (P<0.05, two-sided); - Positive control value

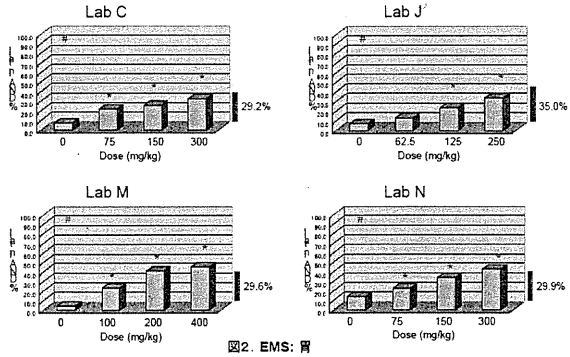


図2. EMS: 胃

* Dunnett test (P<0.05, two-sided); # Linear trend test (P<0.05, two-sided); - Positive control value

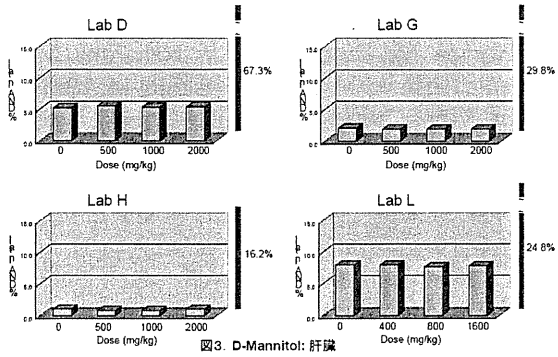


図3. D-Mannitol: 肝臓

No statistical significance with Dunnett test (P<0.05, two-sided) and linear trend test (p<0.05, two-sided), - Positive control value

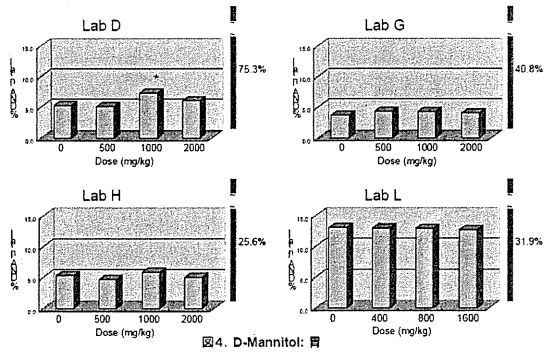


図4. D-Mannitol: 胃

* Dunnett test (P<0.05, two-sided); No statistical significance in linear trend test (p<0.05, two-sided); - Positive control value

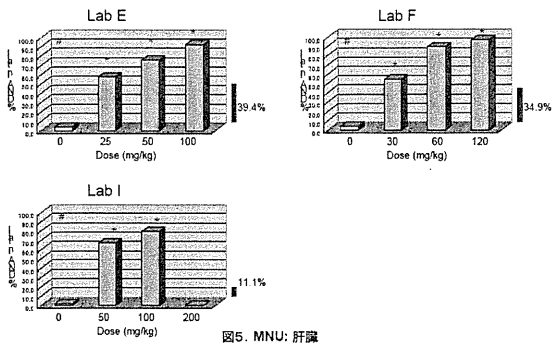


図5. MNU: 肝臓

* Dunnett test (P<0.05, two-sided); # Linear trend test (P<0.05, two-sided); - Positive control value

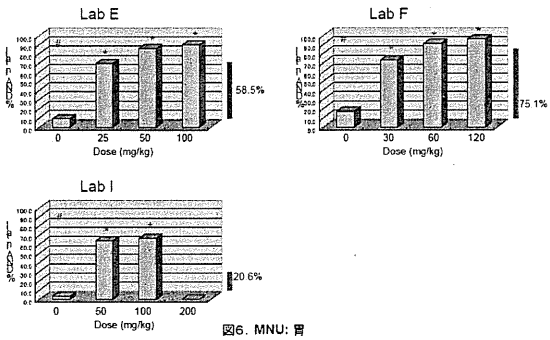


図6. MNU: 胃

* Dunnett test (P<0.05, two-sided); # Linear trend test (P<0.05, two-sided); - Positive control value

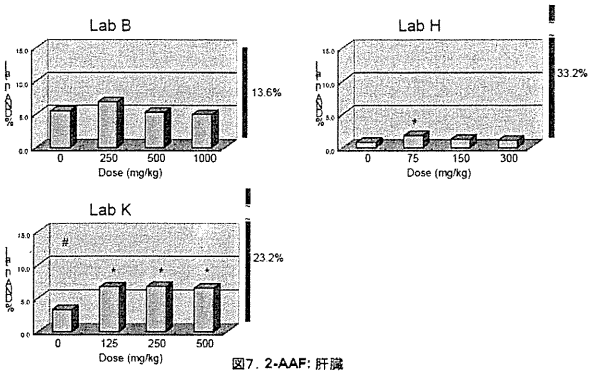


図7. 2-AAF: 肝臓

* Dunnett test (P<0.05, two-sided); # Linear trend test (P<0.05, two-sided); - Positive control value

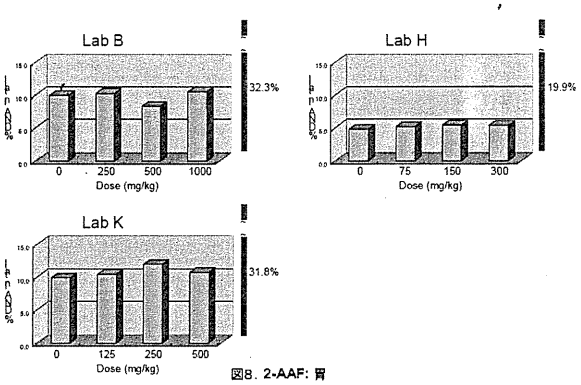


図8. 2-AAF: 胃

No statistical significance with Dunnett test (P<0.05, two-sided) and linear trend test (p<0.05, two-sided); - Positive control value

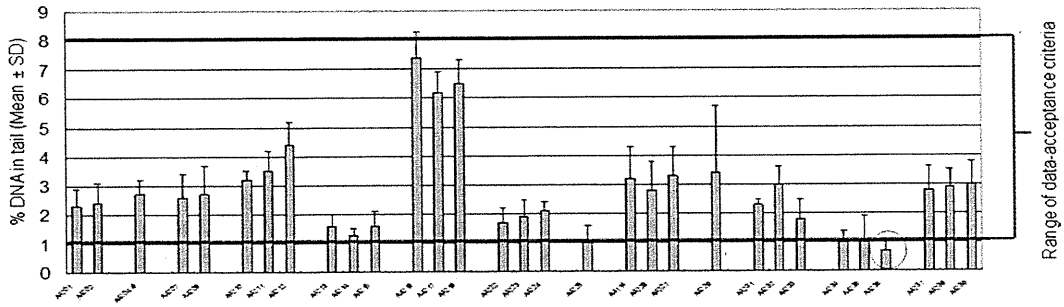


図 9．バリデーション Phase IV-2　　陰性対照値（肝臓）

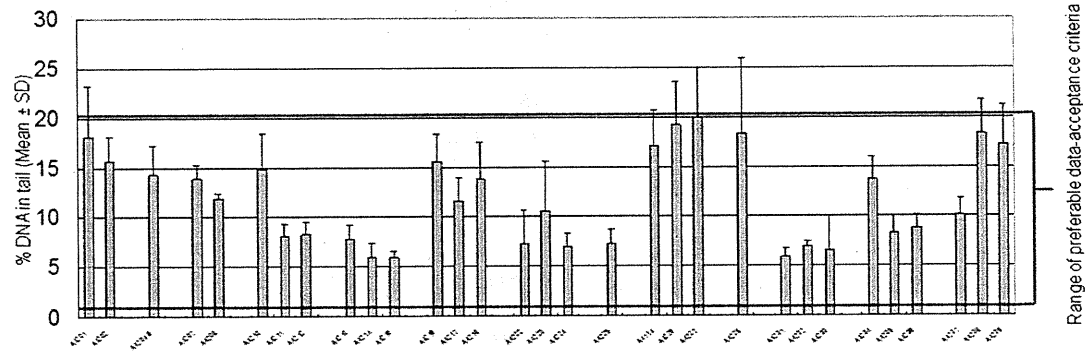


図 10．バリデーション Phase IV-2 陰性対照値（胃）

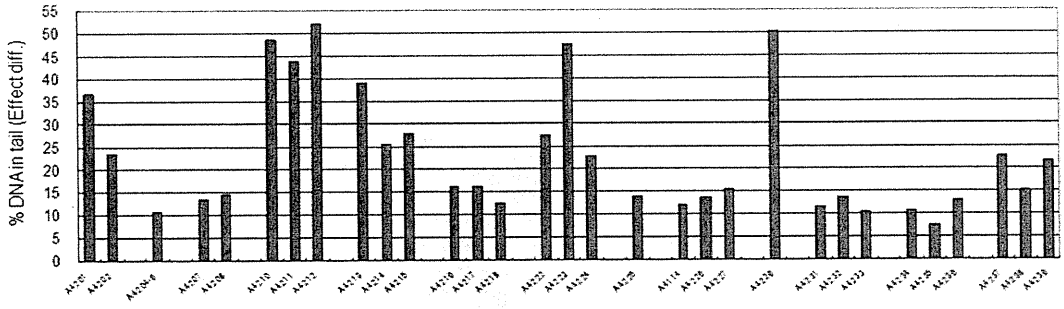


図 11．バリデーション Phase IV-2 陽性対照値（肝臓）