temperature, electrophoresis conditions (pH, V/cm, mA, and temperature at the start of unwinding and the start and the end of electrophoresis) and staining procedure; criteria for scoring comets and number of comets analyzed per slide, per tissue and per animal; evaluation criteria; criteria for considering studies as positive, negative or equivocal.

3.1.3.5. Results

Signs of toxicity, including histopathology in the appropriate tissue(s) if applicable; individual and mean values for DNA migration (and ranges) and % hedgehogs in individual tissue, animal, and group; concurrent positive and negative control data; and statistical evaluation.

3.1.3.6. Discussion of the results and/or conclusion, as appropriate.

4. ARCHIVES AND REVIEW

The study report and all raw data (including slide samples and image data) from this study will be retained according to the SOP in each testing facility. All raw data will be submitted to the management team for review if required.

5. NOTES

1) We evaluated the data of the 3rd phase validation studies as to whether or not fewer (two, three or four) animals were sufficient in the positive control group to show a statistically significant increase in the Effect (difference) with a one-tailed student's t-test ($P<0.025$). The analysis results were presented and discussed at the Florence meeting held on August 25-26, 2009, and the participants felt that the reduction of animal number would be possible but the slight decrease in the statistic power might require additional experiments and result in the increase in animal usage. Thus the VMT decided to continue using five animals as the positive control in this validation effort. We may need to further investigate the appropriate number of animals/group afterwards based upon power calculation.
2) We will likely need to specify shelf life for some solutions as we reconcile lab-specific protocols.
3) The VMT extensively discussed at the Osaka meeting held on Feb. 4-6, 2009 how a preliminary dose-finding study should be done to choose an appropriate high dose level, because selection of a suitable high dose would be closely related to the sensitivity/specificity of genotoxicity assays in general. The VMT decided to request each facility to submit its own protocol for dose-selection, and the VMT will review them and then direct each facility to use its own protocol as it is or to follow a

dose-finding study protocol recommended by the VMT.

4) When following the regimen for EMS as a positive control, micronucleus (MN) induction will be detected in bone marrow but not in peripheral blood. To also detect MN induction in peripheral blood, it would be necessary to administer EMS as well as the other test chemicals three times. It was also pointed out at the Florence meeting (August, 2009) that four times administration of test chemicals excluding the positive control, EMS, would be needed if we expect to detect micronuclei in the peripheral blood.
5) In this validation study, Comet analysis for the liver and the stomach will be conducted. Comet analysis along with MN for the bone marrow and/or the peripheral blood are optional in this validation study.
6) At the Florence meeting, it was pointed out that the duration of tissue sampling should be kept to a set time (e.g. within 10 min) and the duration for lysis should be controlled, in order to obtain more stable negative control values. The VMT considers that such action would be preferred and recommended but not required of participant laboratories because the feasibility would depend on the performance of each laboratory. To further address this issue, the duration of tissue sampling and the duration for lysis should be recorded in the study report of each facility.
7) The size of the liver portions will be at the discretion of the laboratory, because there is no recommendation for standardizing this step.
8) In each electrophoresis run, there should be the same number of slides from each animal in the study; see Attachment 1, an example of how to keep track of each slide during each electrophoresis run. Each laboratory will need to provide its own electrophoresis box chart, as different boxes can accommodate different numbers of slides.
9) Under those electrophoresis conditions, it is expected that an average DNA migration obtained in the negative control group will be 1-8% tail DNA for the liver, and 1-20% tail DNA for the stomach. These ranges were set based on the analysis with negative control data from the 2nd and 3rd phase validation studies, i.e. the average +/- 3XS.D. values were as follows in the 2nd and 3rd phase validation studies, respectively: 3.8+/-4.8 (n=15 from 5 labs) and 3.1+/-3.9 (n=12 from 4 labs) in the liver, and 12.5+/-6.9 (n=12 from 4 labs) and 8.8+/-9 (n=10 from 4 labs) in the stomach. The reason why the lowest value is set at 1 is to be able to detect a significant decrease in % DNA in the tail. The decrease in DNA migration is expected for cross-linkers, and if such agents are intended to be detected using the Comet assay then a decrease

in migration would be easier to detect when the negative control value is at the higher end of the acceptable range. If the negative control average deviates from the range, the duration of electrophoresis will be adjusted to achieve this range.

10) An investigation was conducted to compare with two slides/animal and three slides/animal about some data of the 3rd phase validation study, and the result was presented and discussed at the Florence meeting. As there was no difference between them as far as the present analysis method was used, the VMT decided to use two slides/animal.

11) In order to obtain suitable areas for observation, dilution of cell suspension may be required during the single cell preparation process.

12) This instruction indicates that if a comet is analyzable by the software program then it should be analyzed. However the following cases will be excluded from the analysis: a) analyzable but the recognition by software is considered incorrect (e.g. the automatic recognition of nucleus center is shifted); and b) the staining of nucleus and/or migration is considered poor. At the Florence meeting, more detailed analysis methods were discussed and agreed to, i.e. cells should be classified into three categories, scorable, non-scorable and hedgehog, and also scorable cells with a 90% or more DNA in the tail should not be adopted as part of the data for analysis. The VMT will prepare a color atlas to instruct how to distinguish comet and hedgehog.

13) 'Tail length' is defined as 'Tail migration' in some image analyzers such as Comet IV.

14) At the Atagawa meeting held on March 13-14, 2008, there was discussion about the need to collect data on tail length and Olive tail moment in this validation study. Again, there was brief discussion about this point at the Osaka meeting. The consensus was that % DNA in tail seems to be a sufficient endpoint for validation and therefore these parameters would no longer be analyzed statistically. However, data on tail length and tail moment will continue to be collected in this validation study in case there is a reason to analyze these data in the future.

15) Effect (difference) seems to be more suitable for revealing variation between labs than Effect (ratio), which was pointed out at Osaka meeting in the discussion of the data of 3rd phase validation study.

16) At the present moment, there is no evident data on the consistency between the percentage of "hedgehogs" and histopathology. In this validation study, histopathology will be used as a primary endpoint to evaluate cytotoxicity, although both of the data will be collected for further analysis on the consistency between the percentage of "hedgehogs" and histopathology.

6. REFERENCES

Burlinson B, et al., Fourth International Workgroup on Genotoxicity Testing: result of the in vivo comet assay workgroup. Mutation Res., 627, 31-35, 2007.

Collins AR, et al., Direct enzymatic detection of endogenous oxidative base damage in human lymphocyte DNA. Carcinogenesis, 14, 1733-1735, 1993.

Hartmann A, et al., Recommendation for conducting the *in vivo* alkaline Comet assay. Mutagenesis, 18(1), 45-51, 2003.

Lovell DP, G Thomas G, R Dubow., Issues related to the experimental design and subsequent statistical analysis of in vivo and in vitro comet studies. Teratog Carcinog Mutagen. 19(2), 109-119, 1999.

Olive PL, et al., Heterogeneity in radiation-induced DNA damage and repair in tumor and normal cell using the "comet" assay. Radiat. Res., 122, 86-94, 1990.
Tice RR et al., Single cell gel/Comet assay: guidelines for in vitro and in vivo genetic toxicology testing. Environ. Mol. Mutagen., 35, 206-221, 2000.

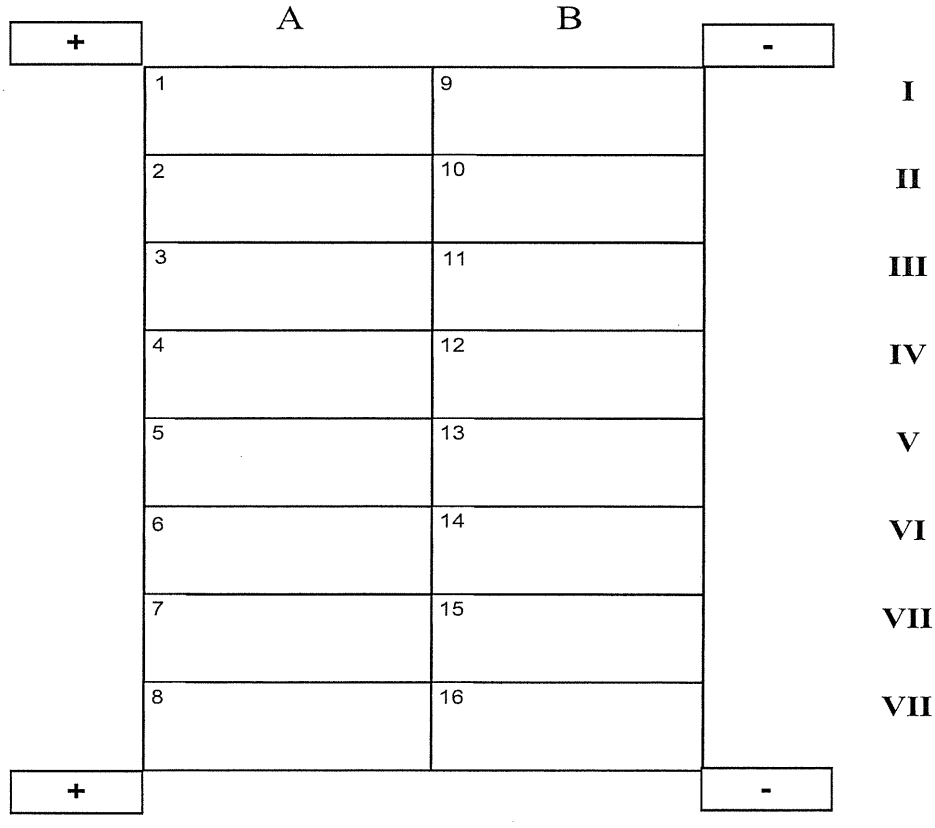
Wiklund SJ, E Agurell., Aspects of design and statistical analysis in the Comet assay. Mutagenesis 18(2):167-175, 2003.

Attachment 1:

SLIDES UNWINDING & ELECTROPHORESIS RECORDING SHEET

Electrophoresis Run #				**Initials & Date**
Approximate alkaline electrophoresis buffer volume in chamber				
Unwinding				
Time	Total	Start	End	
Buffer Temperature				
Electrophoresis				
Running time	Total	Start	End	
Volts				
Milliamperes				
Buffer Temperature				
Thermometer No.				
Electrophoresis chamber No.				
Power supply No.				

Diagram Electrophoresis Chamber



添付資料２

Title: Report of the JaCVAM initiative international validation study of the *in vivo* rodent alkaline Comet assay for the detection of genotoxic carcinogens: the 4th (definitive) phase-1st step

Issued: Yoshifumi Uno, D.V.M., Ph.D., a Validation Management Team (VMT) member
Date: March 21, 2012
Status: Draft version-1

Notes: this document is prepared to summarize the *in vivo* Comet assay validation process and results in the 4th (definitive) phase-1st step. The methods are mentioned minimally in this document, because the details are described in the study protocol and the study plan. An article for submission to a scientific journal will be provided separately based on this document, the study protocols and the study plans (and the other documents if necessary and available).

A Table of Contents

1. Introduction

An *in vivo* rodent alkaline Comet assay is a genotoxicity assay to detect DNA damage in almost all tissues theologically, and is practically used worldwide for detecting genotoxic chemicals. The assay is applied to some regulatory aspects for investigation of genotoxic potential of test chemicals, and currently expected as a promising second *in vivo* genotoxicity assay in the ICH-S2(R1) guidance behind *in vivo* micronucleus assay with bone marrow and/or peripheral blood. The methods of Comet assay were often discussed in the meetings of the International Workshop on Genotoxicity Testing (IWGT) and the International Comet Assay Workshop (ICAW), and the consensus articles have been published on the methodology (1-3).

The assay, however, has not been validated formally with a standardized study protocol. In addition, since the reports on predictive capability of *in vivo* rodent Comet assay for carcinogenicity are limited (4-6), the investigation of predictive capability in multiple laboratories using one study protocol validated would be more useful to understand the accurate performance of *in vivo* Comet assay. The Japanese Environmental Mutagen Society/the Mammalian Mutagenicity Study Group (JEMS/MMS) decided to have a (international) collaborative study of *in vivo* Comet assay in 2003, and conducted a preliminary collaborative study on Comet assay procedure, e.g., comparison of assay results between whole cells and isolated nuclei (7). Contemporaneously, the UK Environmental Mutagen Society/the Industrial Genotoxicity Group had a concern about the establishment of OECD guideline on Comet assay (Dr. Brian Burlinson, personal communication). To cover all the concerns about validation efforts of *in vivo* Comet assay, the Japanese Center for the Validation of Alternative Methods (JaCVAM) is organizing the international validation study in and after April, 2006, in cooperation with the U.S. National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), the European Centre for the Validation of Alternative Methods (ECVAM), and JEMS/MMS.

The purpose of this validation study is to evaluate the ability of the *in vivo* Comet assay to identify genotoxic chemicals as a potential predictor of rodent carcinogenicity, and as an alternative follow-up assay to more commonly used *in vivo* rodent Unscheduled DNA Synthesis (UDS) assay. The goal of this validation effort is to establish the OECD guideline on *in vivo* rodent alkaline Comet assay.

2. Background and Purpose

In the 3rd phase of the *in vivo* Comet assay validation study, three coded test chemicals and EMS, a positive control, were assayed in four leading laboratories in accordance with the Comet assay protocol-version 13, and generally comparable data were obtained. To obtain more consistent data between laboratories, the protocol was further optimized and revised to version 14 (Appendix 1). The version 14 protocol additionally involves 3 treatments (0, 24, and 45 h) of a test compound before collecting liver and glandular stomach tissue samples at 48 h after the first treatment. This allows for the detection of DNA damage using the Comet assay and of micronuclei in blood erythrocyte using the standard micronucleus assay in the same animals, which will significantly reduce the use of animals for *in vivo* genotoxicity testing.

In the 1st step of the 4th phase validation study, the purpose was to examine the extent of reproducibility and variability of assay results among laboratories using coded test chemicals and the positive control EMS, when experiments are conducted in accordance with the Comet assay protocol-version 14 (see the study plan: Appendix 2).

3. Experimental Period

May-December, 2009

4. Participant Laboratories

Thirteen laboratories* participated in the 1st step of 4th phase validation study, which include four leading laboratories[#] that have a lot of experiences of Comet assay and nine laboratories that passed our recruitment process for this phase of the validation study.

* Merck Research Laboratories[#] (code: Lab B), BioReliance[#] (Lab C), Huntingdon Life Sciences[#] (Lab D), Food and Drug Safety Center[#] (Lab E), The Institute of Environmental Toxicology (Lab F), Novartis Pharma (Lab G), AstraZeneca (Lab H), Sumitomo Chemical (Lab I), Mitsubishi Chemical Medience (Lab J), Janssen R&D (Lab K), Health Canada (Lab L), Covance (Lab M), and Bayer Schering Pharma (Lab N)

5. Success Criteria in the Study Plan (Appendix 2)

5-1. To obtain positive results in all positive control groups in all testing facilities.

5-2. To obtain consistent positive or negative results in testing facilities that examine

the same test chemical.

6. Materials and Methods

Outlines of the materials and the methods are described in this section, and the details are referred to the validation study protocol version 14 (Appendix 1) and the study reports prepared in each laboratory (Appendix 3).

A study protocol for Comet assay was made in each laboratory in accordance with the validation study protocol v.14. The experiments proceeded in each facility based on their own study protocol and SOP.

6-1. Animal species, strain, and sex

Rats were selected in this validation effort because of popular species in toxicology studies. Crl:CD(SD) male rats were used.

6-2. Test chemical, vehicle, dose level, and administration

Ethyl methanesulfonate (EMS) has been used as a positive control in a sequence of this validation effort, because it is a well-known genotoxic chemical for multiple organs. EMS was dissolved in physiological saline, and administered to rats at the dose level of 200 mg/kg twice (21 hr interval) with oral administration.

Four coded test chemicals were used in this study. The three of four were the same chemicals used in the 3rd phase validation study, i.e. EMS as well as the positive control, *N*-methyl-*N*-nitrosourea (MNU), and D-mannitol (MA), because VMT considered that examination with the same chemicals used in the 3rd phase validation study would provide useful information about the reproducibility on assay results between different phases of validation studies as well as between laboratories. Another chemical was 2-acetylaminofluoren (2-AAF) which was selected as a typical mutagen to produce bulky adducts, i.e. a different mode of action from the alkylating agents, EMS and MNU. The vehicle for a coded test chemical was selected in each testing facility. The dose levels were also decided in each facility based on the results of preliminary dose-finding study conducted in the testing facility, although VMT provided some toxicological information of published literatures such as LD_{50} values if available, to assist the dose selection in each laboratory. Each coded test chemical was administered to rats at the three dose levels in three times (24 and 21 hours interval) with oral gavage. This administration regimen was designed to allow the combination genotoxicity assay of micronucleus and comet, in consideration of the 3R's principle for animal use, although

investigation into micronucleus induction was optional and data on micronucleus are not included in this validation study report (if examined, the data may be included in study reports written in testing facilities; Appendix 3).

Each coded test chemical was examined in three or four laboratories to evaluate the extent of reproducibility and variability of assay results among laboratories using the coded test chemicals and the positive control EMS. The vehicle and the dose levels of test chemicals are summarized in Table 1.

Table 1 Test chemical, laboratory tested, vehicle, and dose levels

Test chemical	Lab. Code	Vehicle	Dose level (mg/kg/day)
EMS	Lab C	(Report not available)	75, 150, 300
	Lab J	Physiological saline	62.5, 125, 250
	Lab M	(Report not available)	100, 200, 400
	Lab N	Corn oil	75, 150, 300
MNU	Lab E	Physiological saline	25, 50, 100
	Lab F	Physiological saline	30, 60, 120
	Lab I	Physiological saline	50, 100, 200 [1)] Additional: 6.25, 12.5, 25
MA	Lab D	Purified water	500, 1000, 2000
	Lab G	Water	500, 1000, 2000
	Lab H	(Report not available)	500, 1000, 2000
	Lab L	Physiological saline	400, 800, 1600
2-AAF	Lab B	0.5% CMC aqua solution	250, 500, 1000 [2)]
	Lab H	Corn oil	75, 150, 300 [3)]
	Lab K	Corn oil	125, 250, 500

1) Animal death was found at 200 mg/kg/day. Cytotoxicity for the stomach was noted in all dose levels, and the additional study was done at 6.25, 12.5, and 25 mg/kg/day. There was no cytotoxicity at 6.25 mg/kg.

2) Lab B selected the highest dose in consideration of the amount of test chemical delivered. They suggested that the higher dose level should be evaluated because no animal toxicity was found at 1000 mg/kg/day.

3) Lab H knew the actual chemical name before the assay due to the information from customs (when JaCVAM sent test chemicals to overseas facilities, the customs checked it and then informed the facility of the chemical name), and selected the dose levels based on published toxicity information of 2-AAF. However, no toxicity was found in animals up to the highest dose level.

6-3. Organs analyzed

Liver and stomach (glandular stomach) are selected in this validation effort, because the former is the primary organ for the metabolism of absorbed chemicals, and the latter is a site of first contact of chemicals in orally administration. These organ analyses are recommended for screening purpose of genotoxic chemicals in the previous discussion in ICAW (2).

6-4. Data-acceptance criteria

6-4-1. Negative control

Means of %DNA in tail are 1-8% in the liver and 1-30% (preferably 1-20%) in the stomach.

6-4-2. Positive control in case of EMS, 200 mg/kg, once (or twice) p.o.

Effect (difference of means of % DNA in tail between groups of EMS and vehicle control) is statistically significant increases and is 5% or higher in the liver and the stomach (primary criteria); and Effect (ratio of means of % DNA in tail between groups of EMS and vehicle control) is 2-fold or higher in the liver and the stomach.

6-5. Data analysis

% DNA in tail was used as the primary endpoint of this validation study, because it is considered linearly related to the DNA break frequency over a wide range of damaged DNA levels (2).

Three conceptual key terms, "Endpoint", "Estimate" and "Effect" were defined and used in the data analysis. Briefly, Endpoint is defined as individual observed values for a parameter such as % DNA in tail. Estimate is defined as a mean calculated with values of Endpoint in each animal. Effect is defined as difference (hereafter designated as Effect (diff.)) or ratio (hereafter designated as Effect (ratio)) of a mean of Estimate between a negative control group and a treatment group. A general purpose of data analysis in validation studies is to investigate how large variation exists among data from several testing facilities, and Effect is considered as a good indicator to understand the variation of Comet assay parameters among testing facilities. VMT noticed through the 1st to the 3rd phase validation studies that Effect (diff.) was more effective for the comparison of variation than Effect (ratio), because Effect (ratio) depended on the magnitude of negative control values (i.e. lower negative control values produce higher Effect (ratio) easily) and would be often misread in the evaluation of responses induced with a test chemical administration. Therefore Effect (diff.) was mainly used to evaluate the assay results.

Dunnett's test (two-sided, $P<0.05$) and linear Trend test (two-sided, $P<0.05$) were

applied to Effect (diff.) in the groups of coded test chemicals. The two-sided analysis was used because both increases and decreases in the Comet parameter could be detected. A decreased % DNA in tail was expected to be a good index to detect cross-linking agents. For the positive control group, Student's t-test (one-sided, $P<0.025$) was applied to the Effect (diff.).

7. Results

In this section, outlines of the study results are described. The details are referred to the study reports written in each laboratory (Appendix 3) and the statistical analysis results by Dr. Takashi Omori (Appendix 4).

7-1. Control groups

Figs. 1 and 2 summarize the lab-orderly means of % DNA in tail (Estimate) in the vehicle and positive control EMS groups in the liver and the stomach, respectively.

7-1-1. Vehicle control groups

Figs. 3 and 4 are enlarged-figures of the lab-orderly Estimate in the vehicle control groups in the liver and the stomach, respectively. All the values in the stomach satisfied the preferable data acceptance criteria 1-20%, and those in the liver also met the data acceptance criteria 1-8% except for Lab C and one of two experiments in Lab H. The actual mean values in the liver were 0.9% in Lab C and 0.8% in Lab H.

7-1-2. Positive control groups

Figs. 5 and 6 show Effect (diff.) of mean %DNA in tail between the vehicle control group and the positive control group in the liver and the stomach, respectively. All of Effect (diff.) show statistically significant increases with Student's t-test (one-sided, $p<0.025$) both in the liver and the stomach, and also showed 5% or higher values. Therefore, it was judged that all the positive control values satisfied the primary data acceptance criteria.

Figs. 7 and 8 show Effect (ratio) of mean % DNA in tail between the vehicle control group and EMS group in the liver and the stomach, respectively. Since all Effects (ratio) were 2-fold or higher in both organs, it was judged that all the positive control values satisfied the data acceptance criteria.

7-2. Coded test chemical groups

7-2-1. EMS

EMS as a coded test chemical was evaluated in four laboratories. Figs. 9 and 10 show the mean % DNA in tail (Estimate) in the vehicle control group and the three dose-level groups in the liver and the stomach, respectively. The vehicle for dosing solution and the dose levels were decided in each laboratory independently based on the preliminary examination on suitable vehicle for the coded test chemical. A same vehicle physiological saline was selected in three of four laboratories (to be correctly described for Labs C and M when the reports are available), and the consistent dose levels were used based on the results of dose-finding study conducted in each facility.

All of Effect (diff.) showed statistically significant increases with dose-dependency in the liver and the stomach. The magnitude of responses in both organs was comparable in the same or similar dose levels among four laboratories, although higher values in the liver were observed in Lab M.

In histopathological examination, Lab N reported that hepatocellular hypertrophy was found in the liver of 300 mg/kg/day group, and inflammatory changes in the stomach of 200 and 300 mg/kg/day groups in the dose finding study. Lab J reported that hepatocyte necrosis (minimal and focal) in the liver was observed in one animal at 62.5 mg/kg/day, and mucosal erosion in the glandular stomach was observed in each one animal at 125 and 250 mg/kg/day. <To be described about histopathology in Labs C and M when the reports are available>

7-2-2. MNU

MNU was evaluated in three laboratories. Figs. 11 and 12 show the mean % DNA in tail (Estimate) in the vehicle control group and the three dose-level groups in the liver and the stomach, respectively. The treatment vehicle and the dose levels were decided in each laboratory independently as well as the studies with EMS. The same vehicle physiological saline was used in all laboratories, and the consistent dose levels were selected in all facilities based on the result of each dose-finding study, although animal death was found at the highest dose level in Lab I. Since Lab I selected 200 mg/kg/day as the highest dose level due to no clinical signs in animals at this dose for 3 days and death at 350 mg/kg/day in their dose-finding study, the dose selection would be justified.

All of Effect (diff.) showed statistically significant increases with dose-dependency in the liver and the stomach. The magnitude of responses in both organs was

comparable in the same or similar dose levels among three laboratories.

Regarding cytotoxicity, there was no treatment-related cytotoxic finding in the liver in histopathological examination. In the stomach, cytotoxic changes were observed in the histopathology as follows: congestion in lamina propria, degeneration/vacuolar in gastric pits, edema in submucosa and erosion at 50 and 100 mg/kg/day in Lab E; partial depletion of glandular epithelial cells and submucosal edema at all dose levels, and mucosal necrosis at 120 mg/kg/day in Lab F; and congestion/hemorrhage, degeneration/necrosis in mucosal epithelial cell, detachment in mucosal epithelial cell, glandular dilatation and infiltration in inflammatory cell findings at 50 mg/kg/day or above in Lab I. Lab F and Lab I also reported that almost all cells were hedgehog at the high dose levels (and also at the middle dose level in Lab F) in both organs (Appendix 3).

7-2-3. MA

MA was evaluated in four laboratories. Figs. 13 and 14 show the mean % DNA in tail (Estimate) in the vehicle control group and the three dose-level groups in the liver and the stomach, respectively. The treatment vehicle and the dose levels were decided in each laboratory independently. Water or physiological saline was used in all laboratories (to be correctly described for Lab H when the report is available), and the consistent dose levels were selected in all facilities based on the result of each dose-finding study, although Lab L set 1600 mg/kg/day as the highest dose level, because precipitations were noted in the dosing solution above 160 mg/mL (note: Lab L thought that the solubility limit to vehicle was one of the reasons to select the highest dose level, but it was not mentioned in our standard study protocol. VMT could not notice this misunderstanding of Lab L at this step of 4th phase validation study).

There was no statistically significant increase and dose-dependency in Effect (diff.) in any laboratories except Lab D which showed statistical significance with Dunnett's test in the stomach at the mid dose 1000 mg/kg/day.

No histopathology was examined in any laboratories due to negative judgment of Comet assay in all testing laboratories.

7-2-4. 2-AAF

2-AAF was evaluated in three laboratories. Figs. 17 and 18 show the mean % DNA in tail (Estimate) in the vehicle control group and the three dose-level groups in the liver and the stomach, respectively.

The treatment vehicle was decided in each laboratory independently. Lab B selected 0.5% CMC aqua solution as the vehicle because this chemical was not solved in saline but well suspended in 0.5% CMC aqua solution. Lab H selected corn oil based on the solubility information of 2-AAF, because this laboratory unexpectedly knew the actual chemical name before experiment as the customs informed the study director of the chemical name in their import process. Lab K also selected cone oil because this chemical was well suspended/soluble in this vehicle.

The dose levels were also determined in each laboratory. The dose-finding studies were conducted in Labs B and K. In Lab B, no toxic sign was observed up to 1000 mg/kg/day. However, as the amount of test chemical delivered was limited, Lab B examined the Comet assay up to 1000 mg/kg/day, not 2000 mg/kg/day which was recommended as an upper limit in the validation study protocol. In Lab K, 500 mg/kg/day was selected as the highest dose level of dose-finding study, based on the information from VMT, i.e., mouse oral LD50 is 810 or 1020 mg/kg. Apart from a (small) reduction in body weight, there were no clinical abnormalities up to 500 mg/kg/day. Hence, Lab K selected 500 mg/kg/day as the highest dose of Comet assay. Lab H did not examine the dose-finding study, because the study director knew the chemical name due to the reason mentioned above, and selected the dose levels based on the published toxic information of 2-AAF. Once a study director knows a chemical name and the toxic information is already available, the dose-finding study using animals would not be allowed due to ethical reasons. Therefore, the dose selection procedure in Lab H would be justified. As a result, there was no toxicity in animals up to 300 mg/kg/day.

In the stomach, there was no statistically significant increase and dose-dependency in Effect (diff.) in any laboratories. In the liver, Effect (diff.) showed statistically significant increases with dose-dependency in Lab K, but not in Labs B and H, except the lowest dose 75 mg/kg/day group of Lab H.

In Lab K, histopathology was examined for the liver due to a positive result in Comet assay. Periportal hepatocellular hypereosinophilia often associated with reduced hepatocellular glycogen was observed, which tended to be marginally increased in the highest dose group.

8. Discussion

EMS treatment as the positive control induced statistically significant increases in Effect (diff.) in both the liver and the stomach in thirteen laboratories. In addition,

Effect (diff.) and Effect (ratio) were 5% or higher and 2-fold or higher in both organs in all testing facilities, respectively, suggesting that the positive controls completely met the data acceptance criteria. Therefore, the first success criteria, i.e., to obtain positive results in all positive control groups in all testing facilities were satisfied in this validation study.

Some discussion would be needed about the slight deviation from the data acceptance criteria on negative control groups for the liver in Lab C (0.9%) and one experiment in Lab H (0.8%). The criteria of 1-8% in the liver are set to detect decreases in % DNA in tail assumed to be induced by DNA cross-linker type mutagens. Since EMS examined in Lab C and 2-AAF in Lab H were not cross-linkers, VMT considered that slightly lower values in the negative control groups would not affect the evaluation of both chemicals, and thus decided to accept the negative control values. Regarding the lower % tail DNA in the liver of negative control group, Lab I pointed out that it was difficult to keep 1% or higher means in negative control groups in the liver, because the values became lower experiment by experiment due to technical maturation of cell preparation. VMT advised Lab I that to keep 1% or higher values would be feasible by prolonged electrophoresis duration and/or higher temperature of alkaline solution below 10°C.

Of the four coded test chemicals, EMS, MNU, and MA were evaluated under the similar experimental designs in laboratories that examined the same test chemical, and gave very consistent Comet assay data and results among three or four laboratories, i.e., EMS and MNU were judged positive, and MA was judged negative in all laboratories that tested the same chemical. Although MA showed a slight but statistically significant increase in Dunnett's test for the stomach at the mid dose level 1000 mg/kg/day in Lab D, both Lab D and VMT judged the response biologically insignificant because of no dose dependency. Therefore, the second success criteria, i.e. to obtain consistent positive or negative results in testing facilities that examine a same test chemical were satisfied for those three chemicals. In addition, these results were quite consistent with those in the 3rd phase validation study, indicating that the robust reproducibility of assay results was confirmed between different phases of validation studies as well as between-laboratories.

Discussion would be needed about the results of 2-AAF. Negative results in both organs were obtained in Lab B, but Lab B suggested that the higher dose levels up to 2000 mg/kg should be examined because of no toxic sign in animals up to 1000 mg/kg/day. Lab H examined 2-AAF under the limited experimental conditions

mentioned above, and judged to be negative in both organs. In contrast, statistically significant increases in % DNA in tail were noted in the liver at all dose levels in Lab K, and a negative result was obtained in the stomach. Since the similar dose levels were resultantly used in three laboratories despite some problematic situations, VMT considered that it would be possible to assess the consistency of assay results in three laboratories with those data. The judgment of assay results was summarized as three negative calls in the stomach and one positive call and two negative calls in the liver, based on the statistical analysis. Thus, there is consistency between the assay results in the stomach, but seems inconsistency in the liver, although it would be still needed to examine the higher dose levels of 2-AAF in order to finalize the positive/negative judgment. Finally, VMT considered that 2-AAF was judged as inconclusive due to the limited testing conditions, but the data themselves would be comparable between three laboratories, because the magnitude of responses in the liver in Lab. K seemed very slight and no clear difference from the other two data sets. In addition, VMT decided that 2-AAF would be reexamined in the next step of the 4th phase validation study under the coded test chemical conditions.

Discussion is still remaining about how the results of histopathology (and % hedgehog) should be considered for interpretation of Comet assay results. In the previous discussion in IWGT, a consensus was obtained, i.e., histopathology is a golden standard to evaluate cytotoxicity in *in vivo* Comet assay (3). However, it has been still unclear how to use histopathology for interpretation of Comet assay results. In this study, the results with MNU may provide some important information for the interpretation. MNU treatment produced a lot of hedgehogs but no histopathological changes in the liver. In contrast, it produced not only increased hedgehogs but also necrotic changes in histopathology in the stomach. If hedgehogs are considered as cells with heavily damaged DNA and/or apoptotic cells, no histopathological finding but increased % of hedgehog may indicate that most cells receiving DNA damage continue to be alive due to protective functions such as DNA repair. Contrarily, necrotic findings without increased hedgehogs may indicate that the necrosis is related to cellular toxicity without primary DNA damage. In case of necrotic findings accompanied with increased % of hedgehog, it would be difficult to estimate whether the cytotoxic changes are induced by DNA damage or cellular toxic effects. More data are clearly needed to consider how to use data of histopathology and hedgehog for interpretation of Comet assay results.

Finally, based on the above discussion, VMT concluded that the overall reproducibility and variability of assay results were robustly confirmed among laboratories using the four coded test chemicals and the positive control EMS in this step, and decided to move to the 2nd step of the 4th phase validation study in order to investigate the predictive capability of *in vivo* Comet assay for chemical carcinogenicity.

9. References (need to be straightened)

1. Tice RR, Agurell E, Anderson D, Burlinson B, Hartmann A, Kobayashi H, Miyamae Y, Rojas E, Ryu JC, Sasaki YF. Single cell gel/Comet assay: guidelines for in vitro and in vivo genetic toxicology testing. Environ. Mol. Mutagen. 2000; 35:206-21.
2. Hartmann A, Agurell E, Beevers C, Brendler-Schwaab S, Burlinson B, Clay P, Collins A, Smith A, Speit G, Thybaud V, Tice RR. Recommendation for conducting the *in vivo* alkaline Comet assay. Mutagenesis 2003; 18:45-51.
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. Fourth International Workgroup on Genotoxicity Testing: result of the in vivo comet assay workgroup. Mutat. Res. 2007; 627:31-5.
4. Sasaki YF, Sekihashi K, Izumiyama F, Nishidate E, Saga A, Ishida K, Tsuda S. The comet assay with multiple mouse organs: comparison of comet assay results and carcinogenicity with 208 chemicals selected from the IARC monographs and U.S. NTP Carcinogenicity Database. Crit. Rev. Toxicol. 2000; 30:629-799.
5. Sekihashi K, Yamamoto A, Matsumura Y, Ueno S, Watanabe-Akanuma M, Kassie F, Knasmüller S, Tsuda S, Sasaki YF. Comparative investigations of multiple organs of mice and rats in the comet assay. Mutat. Res. 2002; 517:53-74.
6. Kirkland D, Speit G. Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens III. Appropriate follow-up testing *in vivo*. Mutat. Res. 2008; 654:114-132.
7. Nakajima M, Ueda M, Yamakage K, Nakagawa Y, Nakagawa M, Ohyama W, Omori T, Asano N, Hayashi M, Uno Y. Tissue sample preparation for *in vivo* rodent alkaline comet assay. Genes Environ. 2012; 34:50-4.