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.).

6-9. Histopathology

Regarding the evaluation of cytotoxicity, the "Gold Standard" for assessing levels of necrosis and apoptosis when an *in vivo* Comet assay gave positive results is concluded to be histopathology, which is a consensus of IWGT meeting in 2005 (3), and it is also pointed out that there is a need to standardize ways to present histopathological findings in the meeting (3). In this validation study, when increased % DNA in tail was observed in the liver and/or stomach, the organ(s) was histopathologically examined by pathologists of testing laboratory, except for retests of A4204, A4206 and A4219 which were examined by pathologists in Biosafety Research Center (BSRC).

Based on the IWGT consensus and suggestion (3), VMT decided that histopathology results should be primarily considered to interpret a relationship between positive findings in Comet assay and cytotoxicity of test chemicals. Increased % hedgehog was another candidate for cytotoxicity evaluation, but it was not adopted in this validation study although the data were collected, because increased % hedgehog would indicate not only cellular toxicity but also DNA damage.

In histopathological examination, VMT considered that necrosis and/or degeneration (or findings indicating degenerative changes) would be main indicators of cellular toxicity. In contrast, apoptosis would indicate both cellular toxicity and DNA damage. Thus necrosis and/or degeneration were mainly focused for interpretation of positive findings in Comet assay.

To decide standardized ways to present histopathological findings, VMT firstly reviewed histopathological data of test chemicals showing negative results in Comet assay. In most of negative compounds, there seemed no histopathological finding indicating cytotoxicity related to chemical treatment, except for the following cases.

1) A4215: single cell necrosis in glandular stomach mucosa (grade: minimal; incidence: two of five (2/5) rats at 25 mg/kg, and 5/5 rats at 50 and 100 mg/kg)
2) A4217: single cell necrosis in hepatocytes (minimal to slight; 5/5 rats at 500 mg/kg)
3) A4235: eosinophilic degeneration/necrosis in hepatocytes (slight; 5/5 rats at 2000 mg/kg)

4) Histopathology of stomach in many chemicals (e.g., A4223, A4232, A4233, ···) showed ulcer/erosion in the glandular stomach or forestomach, and the grade of lesions was minimal to slight (or mild). Since these findings are often observed in rats incidentally by stress, they may not always indicate cytotoxicity.

The above findings indicate that minimal to slight (or mild) histopathological changes related to cytotoxicity would not affect increased % DNA in tail in the liver and the stomach. Therefore, VMT decided that the grade and incidence of the histopathological findings should be considered for the interpretation. For example, when one of five animals showed minimal to slight (or mild) necrosis and/or degenerative findings in a target organ, such weak changes would not affect Comet assay results. In addition, study directors and/or pathologists in most of participant laboratories did not consider that single cell necrosis or single cell death with grade of minimal to slight (or mild) would affect Comet assay results. VMT has agreed to the interpretation considered in those laboratories.

The above criteria were presented to participants in Kyoto meeting held on September 12-13, 2011, and accepted. Based on those criteria, VMT and the meeting-participants interpreted the significance of positive findings in Comet assay under the blind-condition of chemical names (Appendix 6).

6-10. % hedgehog

The frequency of hedgehogs, % hedgehog was determined per sample, based on the visual scoring of 100 cells per sample. No statistical analysis was applied to % hedgehog.

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 4) and the statistical analysis results by Dr. Takashi Omori (Appendix 5).

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 there seems no tendency within/between-laboratory variability. Those in the liver also met the data-acceptance criteria 1-8% except for one experiment with a coded test chemical A4236 in Lab I (the actual mean value in the liver was 0.7%). In the liver, there seems very slight between-laboratory variability, but no within-laboratory variability, indicating that experiments were well-controlled in each laboratory. In the stomach, there seems no clear within/between-laboratory variability.

Figs. 5 and 6 show the vehicle-orderly Estimate in the vehicle control groups in the liver and the stomach, respectively. In the liver, the Estimate seems no different between four vehicles, water, saline, 0.5%CMC, and corn oil. In the stomach, in contrast, the values seem slightly higher in corn oil compared with the other watery vehicles. <This comparison will be re-discussed with results of statistics later.>

As a deviation from the validation study protocol v.14.2, DNA was not stained with SYBR Gold but with ethidium bromide (EtBr) in Lab M due to careless mistake, i.e., Lab M routinely uses EtBr and then used it in this validation study, too. Thus the control data were separately shown in Fig. 7. Since EtBr is generally used for staining DNA in Comet assay, VMT decided to accept the data in Lab M, although this deviation should be taken into consideration for the data review. The negative control values in both organs satisfied the data acceptance criteria.

7-1-2. Positive control groups

Figs. 7 and 8 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. Regarding the magnitude of increased % tail DNA in both organs, there seems very slight between-laboratory variability, but no within-laboratory variability, indicating that experiments were well controlled in each laboratory.

As mentioned in the section 7-1-1., DNA was stained with EtBr, not SYBR Gold in Lab M, and thus the control data were separately shown in Fig. 9. The positive control values in both organs satisfied the data acceptance criteria.

7-2. Justification of test chemical evaluation

From the viewpoint of data-acceptance criteria, VMT considered that all the data on coded test chemicals would be acceptable for the evaluation of predictive capability of *in vivo* Comet assay. Some deviations were noted as mentioned in the section 7-1., but they were considered noncritical and not to affect the evaluation of coded test chemicals with Comet assay.

However, at Huntingdon meeting held on February 7-9, 2011, three deviations from the validation study protocol were pointed out as shown in the following items 1) to 3). In addition, at Kyoto meeting held on September 12-13, 2011, an issue on histopathological analysis was noted as the following item 4). After those meeting, all the issues were solved as mentioned in each item.

1) Test chemical code A4205 should be retested, because no toxicity was observed in the highest dose level of 100 mg/kg. This issue happened due to a misunderstanding about the validation study protocol, i.e., A4205 was soluble up to 10 mg/mL in saline which is the first choice of vehicles in the validation study protocol, and Lab L misunderstood that the solubility limit was one of the reasons to select the highest dose level (since dosing volume was designated as 10 mL/kg, the highest dose level was set at 100 mg/kg). Lab L accepted the request of retest, and reexamined A4205 up to 2000 mg/kg (upper limit dose of the validation study protocol) with corn oil as a vehicle.
2) Test chemical code A4217 should be retested, because no toxicity was observed in the highest dose level of 500 mg/kg. In accordance with the request, Lab N retested A4217 with the dose level of 750 mg/kg. However, animal death was observed at the dose level, indicating that A4217 showed extremely severe toxicity with slight increase in the dose levels. Therefore, 500 mg/kg of A4217 was considered to be the maximum tolerated dose, and the Comet assay data up to 500 mg/kg were justified for the evaluation.
3) Test chemical code A4219 should be retested due to equivocal judgment in the statistical analysis conducted in VMT, although Lab M judged A4219 as positive based on their overall analysis. Since Lab M could not accept the request of retest, Lab O reexamined A4219.
4) Test chemical codes A4204 and A4206 showed increased % DNA in tail in the liver in all dose levels, but histopathology for the liver was only examined at the highest dose levels, and cytotoxic changes were observed. In those cases, histopathology data were needed for the treatment groups of mid and low dose levels, too. VMT asked Lab L the possibility of retests, but Lab L could not accept it. Lab O on behalf of Lab L retested A4204 and A4206.

Finally, from all the points of view to justify this validation study, VMT concluded that all the data on coded test chemicals including results of retests would be acceptable for the evaluation of predictive capability of *in vivo* Comet assay.

7-3. Coded test chemical groups

Table 2 show the summary of results on all test chemicals. The table includes three types of judgment on Comet assay results, i.e., lab judgment, VMT judgment, and final judgment. Lab judgment is that each testing facility calls the Comet assay results as "Positive" or "Negative" based on their own statistical analysis results and histopathological consideration, and some laboratories also consider their own historical control range. VMT judgment is simply conducted based on the statistical analysis results without consideration of histopathology, and it is shown as "Increase", "Decrease", "Equivocal", or "No change" in % DNA in tail. "Increase" means a statistically significant increase in both Dunnett's and linear Trend tests, "Decrease" means a statistically significant decrease in both statistical tests, "Equivocal" means a statistically increase or decrease in either statistical test, and "No change" means no statistical significance in both statistics. Final judgment was decided through the discussion with all participants in Kyoto meeting under a blind-condition with coded test chemicals, based on the consideration of both statistics and histopathology results, and it is shown as "Positive" or "Negative" (or "Equivocal"). When a statistically significant increase in % DNA in tail is noted in either the liver or the stomach, it is simply judged to be "Positive". Appendix 6 is a modified material originally used in Kyoto meeting, which includes the summary data on test chemicals showing increased % tail DNA.

7-3-1. Genotoxic carcinogens

12 of 19 genotoxic carcinogens tested induced a statistically significant increase % tail DNA in the liver and/or the stomach, and VMT judgment for the 12 chemicals was "Increase". In 10 of the 12 chemicals, the final judgments were "Positive", which were coincident with the lab judgment. However, two chemicals, acrylonitrile and thioacetamide were inconsistent between the lab judgment and the final one.

Regarding acrylonitrile, in the first experiment of laboratory, it showed an equivocal response in the liver (a statistically significant increase in linear Trend test, but not in Dunnett's test) and a decrease in the stomach (a statistically significant decrease in both statistics). In the retest for the liver (not done for the stomach), it showed a

statistically significant increase in both statistics in the liver. Based on the statistics, VMT judged the response "Increase" in the liver, and "Decrease" in the stomach. Acrylonitrile, however, was judged "Negative" in the testing facility, because the increased % tail DNA in the liver was within the historical control range of testing facility (note: the decreased % tail DNA in the stomach was not discussed). No histopathological change related to the chemical treatment was found in the liver. In the glandular stomach, edema in lamina propria/submucosal layer was found at 31.3 and 62.5 mg/kg/day (grade: very slight or slight). In forestomach examined at 62.5 mg/kg/day, degeneration/necrosis in squamous cells (grade: moderate), edema in lamina propria/submucosal layer (grade: moderate), degeneration (vacuolar) in muscle fiber of muscular layer (grade: slight), and ulcer (non-graded) were found As the final judgment, acrylonitrile is judged to be "Positive" in the liver in consideration of the reproducible results in two independent experiments and no cytotoxic finding in the liver. Lab judgment in consideration of their own historical control range was considered inappropriate, because data from experiments conducted with the lab-original study protocol were included to establish the historical control rage.

Thiacetamide induced a statistically significant increase % tail DNA in the liver and stomach, and the VMT judgment was "Increase" in both organs. However, it was judged "Negative" in the testing facility, because cytotoxicity was noted in both organs in the histopathological examination and/or increased % hedgehogs (liver). In the histopathology for the liver, hepatocellular degeneration (grade: minimal to moderate), central vein phlebitis (minimal to moderately severe), increased mitosis (minimal or slight), and bile duct hyperplasia (minimal) were found in all treatment groups. In the stomach, no abnormality was noted at 19, 38 mg/kg/day, and erosion/ulcer (one of 6 rats, no-graded), and forestomach gastritis (one of 5 rats, no-graded) were found at 75 mg/kg/day. VMT considered that clear histopathological changes indicating cytotoxicity were observed in the liver, but not in the stomach. Therefore, this chemical is considered to be "Positive" in the stomach as the final judgment.

Five of 19 genotoxic carcinogens tested failed to induce a statistically significant increase or decrease in % tail DNA in the liver and the stomach. One of 19, 4,4'-oxydianiline showed a statistically significant decrease in % DNA in tail in the stomach, but no change in the liver.

Another one, sodium arsenaite induced increases in % tail DNA in the liver, but the statistical significance was only noted with the linear Trend test in Lab M, and with

Dunnett's test in Lab O. This chemical was considered to produce marginal increase in % tail DNA in the liver in two independent experiments, and VMT finally judged this chemical "Equivocal".

In % hedgehog analysis, only two "Positive" chemicals in the liver, 1,2-dimethylhydrazine dihydrochloride and *N*-nitrosodimethylamine, and one "Positive" chemical in the stomach, methyl methanesulfonate clearly increased hedgehogs in the positive organs with dose-dependency, but the other test chemicals including "Positive" chemicals were considered not to clearly affect % hedgehog.

7-3-2. Genotoxic non-carcinogens

One of six genotoxic non-carcinogens tested, 2,6-diaminotoluene induced a statistically significant increase in % tail DNA in the liver, and the VMT judgment was "Increase" in the liver. In histopathology for the liver, no abnormality was noted at 150 mg/kg/day, and mitoses in hepatocyes and hemopoiesis were less frequently or absent at 300 and 600 mg/kg/day. Because of no clear cytotoxicity in the liver, the final judgment was "Positive", which was coincident with the lab judgment. No changes of % hedgehog were noted in all test chemicals in this category.

7-3-3. Non-genotoxic carcinogens

One of seven non-genotoxic carcinogens tested, chloroform induced a significant % tail DNA response in the liver (significant in Dunnett's test in the 250 and 500 mg/kg/day groups) with dose-dependency. In histopathology for the liver, the following chamges were observed: centrilobular hepatocellular single cell necrosis (grade: +), centrilobular hepatocellular necrosis (one of 5 rats, grade: +), centrilobular hepatocellular vacuolation (grade: + or ++), centrilobular hepatocellular enlargement/granular change (grade: + or ++), centrilobular inflammatory cell infiltration (grade: + or ++), and centrilobular hemorrhage (one of 5 rats, grade: ++) at 250 mg/kg/day; and centrilobular hepatocellular single cell necrosis (grade: +), centrilobular hepatocellular necrosis (one of 5 rats, grade: ++), centrilobular hepatocellular vacuolation (grade: + or ++), centrilobular hepatocellular enlargement/granular change (grade: + or ++), centrilobular inflammatory cell infiltration (grade: +), and centrilobular hemorrhage (one of 5 rats, grade: ++) at 500 mg/kg/day. Since serious cytotoxic findings such as moderate grade (++) of hepatocellular necrosis and vacuolation followed by inflammatory cell infiltration and hemorrhage were observed with dose-dependency, the increased % tail DNA in the liver was considered to be related to cytotoxicity, and thus the final

judgment was "Negative", which was coincident with the lab judgment. As a result, all test chemicals in this category were evaluated as "Negative" as the final judgment. No clear changes of % hedgehog were noted in all test chemicals in this category.

7-3-4. Non-genotoxic non-carcinogens

One of eight non-genotoxic non-carcinogens tested, *t*-butylhydroquinone (*t*-BHQ) induced a significant % tail DNA response in the liver, and the VMT judgment was "Increase" in the liver. However, the lab judgment was "Negative", because the increased % DNA in tail was within their historical control range which was established in testing conditions based on the validation study protocol (Brian Burlinson, personal communication). Finally, *t*-BHQ was judged to be "Positive" in the liver based on the statistical analysis results, but VMT decided to note in this validation study report that the chemical was judged to be "Negative" in the testing facility. No changes of % hedgehog were noted in all test chemicals in this category.

8. Discussion

The purpose of the 2nd step of 4th phase international validation study was to investigate the predictive capability of *in vivo* rat alkaline Comet assay against carcinogenicity of test chemicals. In general, the predictive capability of a screening genotoxicity assay for carcinogens is simply described as "positive sensitivity" for carcinogens, and "negative specificity" for non-carcinogens. Such simple approach is, of course, possible, but may not fit the evaluation of predictive capability in this validation study, because we already know that test chemicals have two different aspects of biological/toxicological properties, i.e., genotoxic properties and carcinogenic properties, although it is still possible to calculate the positive sensitivity and the negative specificity only for the carcinogenic properties. The most appropriate approach to evaluate the predictive capability using this validation study results would be to focus the consistency with expected assay results determined before the validation study (see the section 6-4.). Therefore, we discuss the predictive capability of *in vivo* rat alkaline Comet assay from the aspect of four categories of test chemicals, i.e. genotoxic carcinogens, genotoxic non-carcinogens, non-genotoxic carcinogens, and non-genotoxic non-carcinogens.

For genotoxic carcinogens, it was expected to call positive judgment. 12 of 19 genotoxic carcinogens tested were judged "Positive", and six of 19 were "Negative",

529 and one was "Equivocal". Since 13 of 19 genotoxic carcinogens are considered to
530 show at least alert of positive call, the positive sensitivity for genotoxic carcinogens
531 would be calculated to be 68%. In contrast, for non-genotoxic non-carcinogens, it
532 was expected to give negative judgment, and the negative specificity for
533 non-genotoxic non-carcinogens is calculated to be 88%, since 7 of 8 chemicals were
534 judged to be "Negative". For genotoxic non-carcinogens and non-genotoxic
535 carcinogens, it was preferably expected to call negative. The negative specificity for
536 genotoxic non-carcinogens is calculated to be 83%, since 5 of 6 chemicals were
537 judged to be "Negative". That for non-genotoxic carcinogens is calculated to be
538 100%, since all (seven) chemicals were judged to be "Negative". Therefore, the
539 concordance for expected assay results is calculated to be 80%, since 32 of 40 test
540 chemicals gave the expected assay results, indicating that *in vivo* rat alkaline Comet
541 assay would have a good capability to identify genotoxic chemicals as a potential
542 predictor of rodent carcinogenicity.
543
544 Since this step of validation study is the final stage of a series of validation efforts
545 on *in vivo* rat alkaline Comet assay, it should be referred to another purpose through
546 all phases of the validation studies, i.e., to evaluate the ability of the *in vivo* Comet
547 assay as an alternative follow-up assay to more commonly used *in vivo* rodent
548 Unscheduled DNA Synthesis (UDS) assay. The genotoxic carcinogens used in this
549 study included five UDS-positive chemicals, 2-acetylaminofluorene (2-AAF),
550 2,4-diaminotoluene, 1,2-dimethylhydrazine dihydrochloride, methyl
551 methanesulfonate and *N*-nitrosodimethylamine, and four UDS-negative chemicals,
552 acrylonitrile, *o*-anisidine, 1,3-dichloropropene and 4,4'-oxydianiline. Four of the
553 five UDS positives were judged to be "Positive" in this study, but 2-AAF was
554 "Negative". Two of the four UDS negatives, acrylonitrile and 1,3-dichloropropene
555 were judged to be "Positive" in this study. Although the number of chemicals tested
556 is limited, VMT concluded that Comet assay could be at least equal or maybe more
557 sensitive to detect genotoxic carcinoges which could be detected by UDS assay.
558
559 The above discussions are simply focused on the consistency between Comet assay
560 results in this validation study and genotoxicity/carcinogenicity of test chemicals. To
561 evaluate the predictive capability of Comet assay more accurately, it is clearly
562 needed to consider more detailed biological/toxicological properties of test
563 chemicals, i.e., genotoxic and/or carcinogenic mode of actions (MoA) of test
564 chemicals, especially for genotoxic carcinogens judged to be "Negative" in this

study. For such discussion, it should be taken into account that the liver and the stomach (glandular stomach) were only analyzed in this validation study. The reasons why those organs were selected in this validation study was that our validation study protocol was designed for screening purpose against genotoxicity and/or carcinogenicity without consideration of the target organ specificity. The liver is the primary organ for the metabolism of absorbed chemicals, and the stomach is a site of first contact of chemicals in orally administration, which are recommended for the screening purpose of *in vivo* Comet assay in the previous discussion in ICAW (2). Actually, in this validation study, the following chemicals were judged to be "Positive" in the liver and/or the stomach, although both organs were not targets in the rodent carcinogenicity studies: acrylonitrile, azidothymidine, cadmium chloride, *p*-chloroaniline, and thioacetamide (Table 2). In addition, we had decided a rule before starting this step of validation study that target organ specificity of carcinogenicity might be considered for the interpretation of negative results about genotoxic carcinogens, but it should be minimized (see the section 6-4.). Therefore, the following discussion for negative chemicals of genotoxic carcinogens is mainly focused on whether or not the negative results are rational when considering known genotoxic and/or carcinogenic MoA.

1) 2-Acetylaminofluorene (CASRN: 53-96-3; 2-AAF)

2-AAF is a well-known genotoxic carcinogen. The positive results are reported in almost all *in vivo* genotoxicity assays, i.e., rat liver UDS test, rat bone marrow micronucleus (MN) test, mouse Comet assay in the colon, liver, kidney and lung, and gene mutation assay in BigBlue mouse liver (note: reference papers on genotoxicity assay results are described in Appendix 3, as well as the following chemicals discussed). It is reported that the oral administration to animals induced tumor for many organs such as the liver, urinary bladder, and kidney in many animal species (8). To induce genotoxicity and carcinogenicity, metabolic activation of 2-AAF is required, and the critical activation pathway is know to convert 2-AAF to the *N*-hydroxy derivative followed by mainly *N*-SO_4 derivative (9).

In this validation study, this chemical was judged to be negative in both liver and stomach. 2-AAF was also examined in the 1st step of 4th phase validation study, and a positive result (i.e., statistically significant increases in both Dunnett's and Trend tests) in the liver was reported in only one of three laboratories testing this chemical, and the overall judgment of 2-AAF would be

considered negative in this validation study. As well as this validation study result, it is reported that 2-AAF failed to increase DNA migration in the rat liver when administered orally (10). Interestingly, in the same report, it is mentioned that clearly increased DNA migration was noted in the liver when 2-AAF was administered by i.p. injection, indicating that the route of administration affects the genotoxic potential in the liver (10). However, considering the overall genotoxic MoA of 2-AAF, it is unclear why 2-AAF failed to increase DNA migration in the liver under the conditions of this validation study and also in the previous report. When considering that the dose level usually used in UDS assay is 50 mg/kg or below based on the reference article in OECD-TG486 (11), the lower dose levels could be considered to understand the negative results in this validation study.

2) *o*-Anisidine (CASRN: 90-04-0)

This chemical is known to be oxidatively activated by peroxidase and cytochrome P-450 (CYP), and DNA adducts are identified as deoxyguanosine adducts formed from a metabolite, *N*-(2-methoxyphenyl)hydroxylamine. Rats were treated i.p. with *o*-anisidine (0.15 mg/kg daily for 5 days) and DNA from several organs was analyzed by ^{32}P-postlabeling. Two *o*-anisidine-DNA adducts were detected in urinary bladder (4.1 adducts per 10^7 nucleotides), the target organ of carcinogenicity, and, to a lesser extent, in the liver, kidney and spleen (12). In *in vivo* genotoxicity assays, this chemical is reported to be positive in the kidney, bladder, lung, stomach and colon in rat Comet assay, and in the bladder in gene mutation assay with BigBlue mice (see Appendix 3). In contrast, negative results are reported in the liver in the BigBlue mouse gene mutation test and the rat liver UDS test, and also in the mouse bone marrow MN test (see Appendix 3). The carcinogenicity was evaluated with *o*-anisidine hydrochloride by administering in feed to Fischer 344 rats at 5000 or 10000 ppm and B6C3F1 mice at 2500 or 5000 ppm, for 103 weeks then observed for 1 or 2 additional weeks. It is concluded that *o*-anisidine hydrochloride was carcinogenic for rats and mice, inducing transitional-cell carcinomas or papillomas of the bladder in both rats and mice and in both sexes of each species, transitional-cell carcinomas of the pelvis of the kidney in male rats, and follicular-cell tumors of the thyroid in male rats (13).

Based on the above reports, the primary target organ in carcinogenicity studies using *o*-anisidine hydrochloride is considered to be the urinary bladder. The

target-organ specific effects of genotoxicity would be supported by a series of data that the DNA adducts are formed mainly in the bladder and lesser in the liver, kidney, and spleen, and the positive results are obtained in the urinary bladder in the gene mutation assay using TG mice and in the rat Comet assay (also positive in the kidney), and, in contrast, the negative results in the rat liver UDS assay, the mouse bone marrow MN assay, and the liver gene mutation assay using TG mice. Positive results in the lung, colon and stomach, non-target organ in carcinogenicity studies, are also reported in the rat Comet assay, but the increases in tail migration were noted in a limited testing condition, i.e., increased only at 8 hr after the single oral administration of 1000 mg/kg but not at 3 and 24 hrs, and not examined the dose-dependency, indicating that the positive findings in non-target organs might be considered questionable. Therefore, when considering the higher target organ specificity of this chemical, the negative judgment in this validation study may be considered rational. Reversely, the negative result might indicate a limitation of our standard protocol using the liver and the stomach that, when carcinogenicity is noted in organs other than the liver or the gastrointestine, at least in the urinary bladder, the target organ(s) should be examined with Comet assay if necessary to investigate genotoxic MoA in the carcinogenic target organ(s).

3) Benzene (CASRN: 71-43-2)

Two-year carcinogenesis studies of benzene were conducted in F344/N rats and B6C3F1 mice. Doses of 0, 50, 100, or 200 mg/kg body weight benzene in corn oil (5 ml/kg) were administered by gavage to male rats, 5 days per week, for 103 weeks. Doses of 0, 25, 50, or 100 mg/kg benzene in corn oil were administered by gavage to female rats and to male and female mice for 103 weeks. There was clear evidence of carcinogenicity of benzene for both sexes in rats and mice. For male and female rats, benzene caused increased incidences of Zymbal gland carcinomas and squamous cell papillomas and squamous cell carcinomas of the oral cavity. In male rats, benzene also caused increased incidences of squamous cell papillomas and squamous cell carcinomas of the skin. For male and female mice, benzene caused increased incidences of malignant lymphomas, Zymbal gland squamous cell carcinomas, alveolar/bronchiolar carcinomas and alveolar/bronchiolar adenomas or carcinomas (combined). In male mice, harderian gland adenomas, and squamous cell carcinomas of the preputial gland were also increased. For female mice, benzene also caused increased incidences

673 of ovarian granulosa cell tumors, ovarian benign mixed tumors, and carcinomas
674 and carcinosarcomas of the mammary gland. Dose-related lymphocytopenia was
675 observed for male and female F344/N rats and male and female B6C3F1 mice
676 (14).
677 It is generally agreed that the toxicity of inhaled benzene results from its
678 biotransformation to reactive species. Benzene is metabolized in the liver by
679 cytochrome P-4502E1 (CYP2E1) to its major metabolites: phenol, hydroquinone,
680 and catechol. The intermediate benzene oxide can also undergo ring opening to
681 trans-trans muconic acid. Although there is a scientific consensus that
682 metabolism of benzene is required for resultant toxicity and carcinogenic
683 response, the role of a metabolite or metabolites of benzene in producing these
684 adverse effects is controversial and more research data are needed to better
685 define sequelae of pathogenesis following exposure to benzene and its
686 metabolites. Current evidence indicates that benzene-induced myelotoxicity and
687 genotoxicity result from a synergistic combination of phenol with hydroquinone,
688 muconaldehyde, or catechol (15).
689 The bone marrow, Zymbal gland, and Harderian gland all contain peroxidases,
690 which can activate phenols to toxic quinones and free radicals. Sulfatases, which
691 remove conjugated sulfate and thus reform free phenols, are also present at high
692 levels in these target organs. The selective distribution of these two types of
693 enzymes in the body may explain the accumulation of free phenol,
694 hydroquinone, and catechol in the bone marrow and the resulting differences in
695 target organ toxicity of benzene metabolites in humans and animals (15).
696 Molecular targets are considered mainly tubulin, topoisomerase II and histones,
697 and less DNA (oxidation and adduct formation). In genotoxicity studies,
698 benzene exposure has been shown to induce aneuploidy in dividing cells,
699 presumably through inhibition of tubulin assembly during mitosis. However,
700 benzene exposure has failed consistently to induce point mutation in
701 genotoxicity studies (15).
702 In *in vivo* genotoxicity assays, oral administration of benzene showed positive
703 results in the rat bone marrow MN test, and the rat and mouse Comet assay in
704 many organs including the liver and stomach. In contrast, inhalation of benzene
705 failed to increase mutation frequency in the liver using BigBlue mouse gene
706 mutation assay (see Appendix 3).
707 In this validation study, benzene is judged to be negative in the liver and
708 stomach, and it is inconsistent with the previous report of Comet assay (5).

However, considering the known genotoxic and carcinogenic MoA, the negative result might be more likely and rational.

4) Busulfan (CASRN: 55-98-1)

It is reported that this chemical is a direct-acting bi-functional alkylating agent that binds to cellular macromolecules including DNA, RNA, and proteins. Mono-adduct formation, intrastrand cross-links, and DNA-protein cross-links are reported. In rats, i.v. administration of busulfan for one year was reported to induce a variety of tumors in male rats, although the experiments could not be evaluated in IARC due to incomplete reporting. In mice, the i.p. administration induced lymphomas in two studies, but did not increase the incidence of tumors in two other studies. The i.v. administration to mice increased the incidences of thymic and ovarian tumors (16).

The above report indicates that the major genotoxic MoA of busulfan is cross-link formation (16). When considering that simple mono-functional alkylating agents such as EMS or MMS clearly increased % tail DNA in this validation study, the "Negative" result of busulfan indicates that this chemical would give DNA damage through cross-linking formation rather than mono-adduct formation *in vivo*. It is reported that cross-linking agents mitomycin C and cisplatin decreased % tail DNA compared with the negative control (17, 18), and thus we expected that cross-linking agents might decrease % tail DNA in this validation study. However, cross-linking agents, busulfan and cisplatin used in this validation study failed to decrease % tail DNA, indicating that cross-linking agents could not be detected in *in vivo* Comet assay, at least under the conditions of this validation study. Longer electrophoresis duration might be needed to detect the decreased % tail DNA, because the decrease was reported in the longer duration with kidney cells (18).

Anyway, this chemical was optionally examined with the micronucleus assay using peripheral blood in Lab C, and they reported that increased micronuclei were observed in all dose groups (Appendix 4). This finding would be rational when considering the genotoxic MoA of busulfan.

5) Hydroquinone (123-31-9)

The 2-year carcinogenicity study was conducted by administering 0, 25 or 50 mg/kg in deionized water by gavage to F344/N rats, 5 days/wk. Nearly all male rats and most female rats in all vehicle control and exposed groups had

745 nephropathy, which was judged to be more severe in high-dose male rats.
746 Hyperplasia of the renal pelvic transitional epithelium and renal cortical cysts
747 were increased in male rats. Tubular cell hyperplasia of the kidney was seen in
748 two high-dose male rats, and renal tubular adenomas were seen in low- and
749 high-dose male rats. Mononuclear cell leukaemia in female rats occurred with
750 increased incidences in the dosed groups. B6C3F1 mice were given 0, 50 or 100
751 mg/kg on the same schedule as rats. Compound-related lesions observed in the
752 liver of high-dose male mice included anisokaryosis, syncytial alteration and
753 basophilic foci. The incidences of hepatocellular neoplasms, primarily adenomas,
754 were increased in dosed female mice. Follicular cell hyperplasia of the thyroid
755 gland was increased in dosed mice (19).
756 All renal tubule adenomas and all cases of renal tubule atypical hyperplasia
757 occurred in areas of severe or end-stage chronic progressive nephropathy and the
758 neoplasms were not otherwise confined to any particular part of the kidney. It is
759 likely that the mode of carcinogenic action of hydroquinone is exacerbation of
760 this natural disease process.
761 Hydroquinone is a metabolite of benzene. It is mutagenic *in vitro* and *in vivo*,
762 having caused genotoxicity or chromosomal aberrations in rodent bone-marrow
763 cells, although the Ames test showed a negative result. At least a portion, if not
764 all, of the chromosomal effects are caused by interference by hydroquinone or
765 its metabolites with chromosomal segregation, probably due to interaction with
766 mitotic spindle proteins (20). Although the dose routes used to demonstrate these
767 effects in almost all of the studies *in vivo* were not p.o., but i.p. or s.c. injection,
768 the major genotoxic MoA is considered to be interaction with mitotic spindle
769 proteins (aneugenic effects). There were five genotoxicity studies by the oral
770 route. These included a mouse bone-marrow cell MN test in which a weak,
771 marginally positive response was obtained following a single oral dose of 80
772 mg/kg body weight. The remaining oral route studies showed no significant
773 effect. They included a mouse bone-marrow cell MN test in which there was no
774 genotoxic activity after exposure to a diet containing 0.8% hydroquinone for 6
775 days; two ^{32}P-post-labeling assays, one with targets of Zymbal gland, liver, and
776 spleen in SD rats, and the other with the kidney as target in F344 rats; and the
777 last oral assay was for 8-hydroxydeoxyguanosine adducts in F344 rat kidney
778 DNA. Thus, the evidence (and the database) for any genotoxic effect *in vivo* is
779 sparse and none has been observed in kidney (20).
780 While glutathione conjugates could be responsible for the tumor induction,

careful histology seems to show that the most actively toxic of several glutathione compounds tested, 2,3,5-triglutathion-S-yl hydroquinone, targets a very specific region of the kidney, the outer stripe of the outer medulla, whereas hydroquinone-associated adenomas are more randomly distributed and occur in the cortex as well as the medulla. A non-genotoxic MoA that involves exacerbation of a spontaneously occurring rodent renal disease, chronic progressive nephropathy (CPN), is proposed and evaluated. This disease is particularly prominent in male rats and the evidence is consistent with an absence of any human counterpart. Therefore, the increased incidence of renal tubule adenomas in hydroquinone-dosed male rats is without human consequence (20).

When considering the above genotoxic and carcinogenic MoA, the negative result in this validation study would be rational.

6) 4,4'-Oxydianiline (CASRN: 101-80-4)

A bioassay of this chemical for possible carcinogenicity was conducted by feeding diets containing 200, 400, or 500 ppm of the test chemical to F344 rats and 150, 300, or 800 ppm to B6C3F1 mice for 104 weeks. 4,4'-Oxydianiline was carcinogenic for male and female F344 rats, inducing hepatocellular carcinomas or neoplastic nodules and follicular-cell adenomas or carcinomas of the thyroid. 4,4'-Oxydianiline was also carcinogenic for male and female B6C3F1 mice, including adenomas in the harderian glands, hepatocellular adenomas or carcinomas in both sexes, and follicular-cell adenomas in the thyroid of females (21).

This chemical is considered a mutagen due to the positive results in Ames and *in vitro* chromosome aberration tests. Positive results in mice are also reported in the micronucleus test (i.p. injection) and the Comet assay (p.o.). However, a negative result is reported in the rat liver UDS test (p.o.). On the other hand, goitrogenic (anti-thyroid agent) effects of this chemical were also reported for rats and mice in the 90-day toxicity studies (21). The authors indicate that rats or mice receiving anti-thyroid compounds may develop tumors of thyroid gland. They also point out that the goitrogenic and carcinogenic effects of this chemical may be related to the structural similarity between the compound and thyroxin, and nuclear binding sites for thyroxin have been demonstrated in the rat liver and pituitary. In addition, 4,4'-oxydianiline was tested in a neonatal rat liver focus model, and showed clear evidence of hepatocarcinogenicity (22). While

the chemical did not show initiating activity in neonatal models, promoting activity, as indicated by increased number, size, or volume fraction of histochemically detected hepatic foci of cellular alteration, was evident for the chemical with previously demonstrated hepatocarcinogenicity.

In this validation study, the negative results in both liver and stomach were obtained. Since this chemical clearly has genotoxic potential, the negative results in UDS and Comet assays by using rats, and initiation assay of neonatal rat model indicate that rats would be less sensitive for detecting genotoxicity of this chemical compared with mice. On the other hand, the previous reports on carcinogenicity indicate that the carcinogenic MoA of this chemical might be related to the goitrogenic effect for the thyroid and the tumor promoting effect for the liver. Therefore, it is also speculated that contribution of mutagenicity of this chemical might be minimal to the carcinogenicity in rats. Further investigation of genotoxicity using rats would be needed to clarify the rat carcinogenic MoA, e.g., transgenic rat gene mutation assay or rat liver micronucleus assay.

As a summary of above discussion, the five chemicals except for 2-AAF may be justified in showing negative results because of the following reasons: *o*-anisidine due to the higher target-organ specificity for urinary bladder; benzene and hydroquinone due to the aneugenic effects; busulfan due to the cross-linking effects; and 4,4'-oxydianiline due to the possibility of goitrogenic effects on rat carcinogenicity. In other words, those negative results may indicate some limitations of *in vivo* rat Comet assay using the liver and the stomach that the assay cannot detect aneugens and cross-linkers, and target organs of carcinogenicity should be analyzed to investigate whether or not genotoxic MoA is included in carcinogenic MoA. In addition, a combination Comet and MN assay would be practically recommended for screening purpose of the *in vivo* genotoxic potential of test chemicals, because the five chemicals except for *o*-anisidine are reported to be detected with the rodent MN assays.

The within- and/or between-laboratory variability of assay results was already examined in the 1st to 3rd phase validation studies conducted in four or five lead-laboratories, and variability found in the early phase disappeared by the refinement of study protocol. By using the version 14 study protocol, no clear between-laboratory variability was noted, which was robustly confirmed in the 1st

step of 4th phase validation study conducted in 13 laboratories. The within/between-laboratory variability was also evaluated in this study with the data of negative and positive control groups. No clear within-laboratory variability was observed, indicating that the experiments would be well controlled in each testing facility. Although slight between-laboratory variability was noted, the variation was within the range of data-acceptance criteria except for one negative control value of the liver in one laboratory. Those results indicate that consistent Comet assay results could be obtained within and/or between laboratories as far as the experiments are conducted with the version 14 (actually 14.2) study protocol.

Regarding cytotoxicity evaluation, histopathology was used in this validation study based on the previous consensus of IWGT meeting (3). Histopathology criteria for interpretation of positive findings in Comet assay were originally proposed by VMT based on histopathological findings in the negative test chemicals of Comet assay and independent interpretation of each test facility. Then the interpretation method was discussed and concluded by all participants in Kyoto meeting under the blind-condition of test chemicals (Appendix 6). As a result, increased % tail DNA in the liver observed in thioacetamide and chloroform were interpreted to be related to severe cytotoxicity. Both chemicals are well-known hepatotoxicants, and the liver carcinogenic MoA of chloroform is reported to be cytotoxicity followed by regenerative hepatocyte proliferation (23). Thus the interpretation of positive findings in Comet assay using histopathology would be justified at least for chloroform. In contrast, it is recommended that positive findings in Comet assay should be interpreted to be relevant to *in vivo* genotoxicity even if weakly cytotoxic changes such as single cell death/necrosis are observed in histopathology, because many genotoxic carcinogens showing significant increases in % tail DNA in this validation study induced such slight cytotoxicity in histopathology.

In contrast to histopathology, clear increases in % hedgehog were found in only three genotoxic carcinogens, 1,2-dimethylhydrazine dihydrochloride, methyl methanesulfonate, and *N*-nitrosodimethylamine. There were no relationships between positive judgment in Comet assay or histopathological findings and increased % hedgehog, indicating that % hedgehog would not be a good indicator for cytotoxicity evaluation. Presumably, increased % hedgehog might be an indicator to detect severe genotoxicity, because increased % hedgehog was only observed in the treatment of three genotoxic carcinogens showing higher increases

889 in % tail DNA in this validation study, and also in the treatment of
890 *N*-methy-*N*-nitrosourea which also showed very high increases in % tail DNA in the
891 4th phase-1st step validation study as well as this validation study. <Neutral
892 diffusion assay may be discussed if some data are available in the lab's study
893 reports.>
894
895 Another discussion point would be the significance of decreased % tail DNA in the
896 stomach, which were observed in the treatment with acrylonitrile (genotoxic
897 carcinogen), 4,4'-oxydianiline (genotoxic carcinogen), 8-hydroxyquinoline
898 (genotoxic non-carcinogen), ethionamide (non-genotoxic non-carcinogen), and
899 sodium chloride (non-genotoxic non-carcinogen). Since those chemicals are not
900 cross-linking agents as far as we know, and typical cross-linking agents, busulfan
901 and cisplatine failed to decrease % tail DNA in this validation study, decreased %
902 tail DNA found in this validation study would not indicate cross-linking genotoxic
903 MoA. A possible explanation to interpret the significance may be that cytotoxic
904 effects of above chemicals decreased % tail DNA in the stomach. For example,
905 acrylonitrile induced cytotoxicity for the stomach in histopathology, and sodium
906 chloride was selected for this validation study because of a well-known cytotoxic
907 agent for the stomach. Regarding a possible mechanism of decreased % tail DNA, it
908 was pointed out that, when cytotoxicity was observed in the stomach, the epithelial
909 cells would be eliminated and then cells mainly recovered from the stomach for
910 Comet analysis would be the basal cells which show lower values of % DNA in tail
911 (Sachiko Kitamoto, personal communication at Kyoto meeting). Since
912 histopathology was not always examined for the stomach in this validation study
913 because it was done in the case of increased % tail DNA, further investigation of
914 histopathology for the stomach would be needed to clarify the relationship between
915 decreased % tail DNA in the stomach and the histopathological changes.
916
917 Based on the discussion mentioned in the above three paragraphs, VMT has noticed
918 that there may be two different aspects about the relationship between changes of %
919 tail DNA and cytotoxicity, i.e., serious cytotoxicity would increase % tail DNA in
920 the liver but might decrease it in the stomach. Another consideration about those
921 phenomena may be that cytotoxicity might not affect at lease increase in % tail
922 DNA, because cells damaged by serious cytotoxicty would be theoretically
923 non-scorable cells in the image analysis of alkaline-conditioned Comet assay.
924 Therefore, it may be suggested that more careful interpretation about increased or

925 decreased % tail DNA would be needed when severe cytotoxic changes are observed
926 in histopathology. To help the careful interpretation, it may be useful to examine test
927 chemicals having such toxicological properties by using other methods of
928 genotoxicity analysis, e.g., DNA-adduct formation, gene mutation assay with
929 transgenic animal models.
930
931 It should be also discussed how to use historical control data, especially for
932 considering the biological significance when increased % DNA in tail is noted in the
933 treatment with a test chemical. This topic has already discussed in IWGT meeting
934 (24), and it is suggested that the distribution of historical control data should be set
935 with negative control data from at least 10 (preferably 20) independent experiments
936 conducted with a same study protocol and testing conditions (after technical
937 maturation). Although the distribution of historical control data was not considered
938 to evaluate results of this validation study, it would be practically possible to use the
939 distribution in each testing facility that is established based on the IWGT consensus,
940 for the interpretation of biological significance of increased % DNA in tail.
941
942 We selected only the liver and the stomach in this validation study due to the use of
943 screening purpose, but it would be needed to discuss the possibility about
944 application of the methodology to other organs. A key-methodology, cell preparation
945 method used for the liver and the stomach would be basically applicable to other
946 organs as they are, because organs are largely classified as parenchymatous organs
947 represented by the liver and hollow organs represented by the stomach. Since the
948 methods other than cell preparation in the study protocol are, of course, applied to
949 other organs as they are, our Comet assay protocol would be applicable to all organs.
950 In case of the application to other organs, it is necessary to collect sufficient
951 historical control data in each testing facility with vehicle controls and a positive
952 control such as EMS (see the above paragraph about how to establish historical
953 control data).
954
955 Finally, VMT would like to suggest how the *in vivo* rodent alkaline Comet assay
956 should be conducted to obtain a toxicologically significant result, because many
957 researchers, at least in Japan, feel that Comet assay often gives false-positive results
958 for genotoxicity evaluation. VMT has noticed through the progress of this validation
959 effort that lower % tail DNA in the negative control group would sometimes
960 produce questionable (probably toxicologically insignificant) increases in % tail