OMB Staff Working Comments on EPA's Dichloromethane (DCM) draft Toxicological Review (page numbers refer to the draft dated December 2009) **and Draft Charge to External Reviewers**

General Science Comments:

- We applaud EPA for providing such a highly technical document that presents some novel applications of multiple PBTK models and genotype information as well as some novel sensitivity analyses. EPA also presents many useful and informative alternative analyses. As such we encourage EPA to assemble an expert panel which includes multiple experts with a strong expertise in PBTK modeling and multiple experts with a strong statistical background. It may also be helpful to specifically encourage these reviewers to evaluate the appendices. Even though there are no specific charge questions on this, such an evaluation of the quantitative details may be useful. We note that while we did not review these in detail, a quick skim pointed us to some atypical findings (eg negative standard deviations). Thus it may be useful to ask a few of the more technical reviewers to take a detailed look at the modeling and statistics. It may be helpful to additionally include some specific questions taking comment on EPAs approach to the sensitivity analyses that are presented for both the cancer and noncancer modeling.

- As this document is cutting edge in its nature and in applications of some methodological approaches, to ensure a robust review and to decrease the potential for any conflicts of interest and to increase independence, we encourage EPA to follow the guidelines provided by the OMB Final Information Quality Bulletin for Peer Review (see http://www.whitehouse.gov/omb/memoranda/fy2005/m05-03.pdf) which we understand have been incorporated into EPA's Peer Review guidance. As per the recommendations for highly influential scientific assessments, the selection of reviewers should seek expertise and balance in addition to considering barring participation by scientists with a conflict of interest. Depending on the situation, consulting and contractual relationships with the Agency may raise issues of independence or conflict and thus EPA should consider all the factors identified in the Bulletin when putting together their expert panel.

- Page xix, EPA states that Section 6 is intended to convey the limitations of the assessment and to guide the risk assessor in ensuing steps. EPA also states that this section characterizes the overall confidence in the qualitative and quantitative aspects of the assessment. We note that this section does not provide a discussion of overall confidence in the quantitative values for either the reference values or the cancer values. Similarly, while uncertainties are discussed, the overall limitations are not. The quantitative derivations for noncancer values were based upon using the 1st percentile values in a population estimate as the point of departure. The cancer values were derived using a genotype which is specific for a sensitive subpopulation, but not necessarily representative of the total population. We thoroughly support EPA's evaluation of sensitive subpopulations. In addition, we note that for some analyses, for instance for regulatory purposes, EPA often is asked to additionally present best estimates. How does EPA suggest that these values, created to protect against the

most sensitive individuals be used in ensuing steps which also may ask for presentation of a best estimate? Discussion of this, and perhaps recommendations for multiple quantitative values, depending on use, would be consistent with the description of what Section 6 should be providing.

Specific Science Comments:

- Page 101, lines 3663-3677, in discussing the Serota study, EPA states that the authors concluded that DCM did not cause a carcinogenic response in mice and that increases were not dose related and were within the range of historical controls. EPA then notes their alternative conclusion which is that there are small but statistically significant increases. It would be useful for EPA to explain exactly what they did differently (perhaps a different statistical approach) to reach this conclusion. Also, how does EPA's determination that a maximum tolerated dose was not reach impact the cancer findings?

- Page 162, in discussing the evidence for mutagenicity, EPA may want to discuss the impact of the requirement for GST involvement for mutagenicity and how this may relate to exposure levels. As the CYP pathway is likely dominant at environmental exposure levels, it is possible that any effects seen are not due to GST involvement as DCM has less affinity for GST (as per discussion in Chapter 3) than the CYP pathway at levels below CYP saturation. EPA may also want to be clear about what dose ranges EPA expects there to be a mutagenic mode of action.

- Page 194, lines 5941-5944, it would be helpful to provide a citation for this most important sentence regarding tumors in mice. It should also be helpful to clarify who was responsible for the statistical analysis (as in some cases the EPA analysis is different from that of the authors). It may also be helpful to have separate discussions for both the oral and inhalation routes as the data cancer sets are not equally strong for both. It seems as though evidence of oral cancer risk comes from liver tumors in male mice. EPA may want to clarify how this evidence is consistent with the EPA 2005 Cancer Guidelines suggestions for classifications. EPA may also want to clarify which exposure routes are supported by the finding of 'some evidence' in the occupational studies.

- Page 194, beginning on line 5959, as per comments above, it may be helpful for EPA to clarify which dose ranges the finding of a mutagenic mode of action apply to as activation of the GST route seems to be a requirement. Similar discussion would be helpful throughout section 4.7.3.1 and 4.7.3.2.

- Page 220, line 6671, EPA states that the best fitting model was based on the lowest Akaike's Information Criterion (AIC). It would be useful for EPA to explain how they determined if the AIC's were significantly different from each other. If they are essentially the same (eg 38.2 vs 33), it is unclear that simply picking the lowest value is a robust statistical approach. It may be more appropriate to average those BMD values that have statistically similar AIC values. A charge question asking reviewers to comment and provide guidance on how to best combine values from multiple

models, when the AIC values are similar, may be useful. EPA may also want to note that EPA 2000b is a draft document and does not represent official agency guidance.

- Page 221, figure 5-2, at the end of the flow chart says "recommend lower percentile". It would be helpful for EPA to clarify what is meant by this (eg 1%, 5%, 10% etc) and in the text EPA should provide a justification for this choice.

- Page 222, line 6731, EPA states that the first percentile value was used. We support EPA's presentation of an analysis which protects the most sensitive individuals. EPA says this was chosen because it provided a stable estimate. Does this mean that all other percentiles were unstable? Please address why only mean, 5th and 1st percentiles and not other percentiles are presented. EPA should discuss and if possible provide citations to any general guidance these choices are based upon. If this is a pure policy call to ensure that risks at the 99th percentile are not underestimated EPA should clearly state this here and throughout the document when describing the level of protectiveness provided by the RfD and RfC values. As per comments above (in general comments), while we support the analysis of sensitive populations, it may be useful for EPA to carry forward and present values based on the mean as this may be more useful for some regulatory purposes. It may also be useful for EPA to have a charge question on this important choice. [Note similar comments apply to page 242 and the discussion of the RfC]

- Page 223 discusses the choice of using a CYP metric over a GST or AUC metric for noncancer quantification. It would be useful to have a charge question asking the expert reviewers to comment on the scientific validity and justification for this important choice.

- Page 226/227, in discussion EPA says the model with the lowest AIC was chosen. We note that this value was 183.61. EPA did not use models with values of 183.74 or 183.81 or up to 185. How did EPA determine that 183.61 was statistically different from 183.74 and other AIC values? As table 5-2 shows that the BMDL10 values vary greatly (37-77) with a less than 2 point change in AIC value, it is unclear why EPA did not simply average the BMDL10 values.

- Page 227 discusses that EPA used a scaling factor because clearance of DCM metabolites is slower in the humans compared to the rat. As DCM is generally thought to be more toxic in rodents than humans, does this make sense from a 30,000 foot perspective? More discussion of the impacts of the scaling factor and why it makes sense would be helpful. EPA may also want to include specific charge questions about the choice of scaling factor for the noncancer and cancer values (each treated separately).

- Page 228, EPA's derivation of the RfD is provided in table 5-3. However the details of this derivation (including application of specific Uncertainty Factors) is only presented in footnote E of the table. It may be helpful to clearly present the formula

showing clearly in text which specific UF's are applied to the mean to derive the RfD. [Note similar comments apply to table 5-7 and the RfC derivation]

- Page 229, based on what we know about DCM, considering the PBTK modeling EPA has conducted, would we expect a 3 fold different in toxicokinetic disposition among humans? More clarity and description of the scientific rationale for applying this factor would be helpful.

- Page 230, EPA states that a database uncertainty factor of 3 is justified based on possible neurodevelopmental toxicity and the lack of a 2-generation reproductive study. Since EPA understands so much regarding the kinetics of DCM (as per all the PBTK modeling) it may be helpful to clarify why the inhalation 2-generation study can't be used to inform the need for this uncertainty factor.

- Page 230, we note that in 2000, ATSDR applied UF's of 100 (more than EPA's chosen value of 30) to derive a chronic oral MRL of 0.06 using the same Serota dataset. We note that in section 5.1.6, we see that EPA's previous derivation was the same as ATSDR. In reading the EPA document, we should be able to determine what is causing this 100 fold difference, yet it is a bit unclear. It would be useful in this section for EPA to explain what the driver is for the current difference in derivations as EPA does in detail for the IUR. It may also be informative to explain why EPA is no longer supporting the previous approach and is thus choosing to not even present it as an alternative derivation.

- Page 240, EPA chose the log-probit model for the RfC based on the lowest AIC value. As per comments above, it is unclear if these values are statistically different in a meaningful way. In this case, as per table 5-6 a 0.29 change in AIC value leads to a change of 371 in BMDL10 values.

- Page 245, it is unclear why a database UF of 10x is applied to the RfD, as compared to the RfC which is 3x, when there is a 2-generational study available.

- Page 250, beginning on line 280, identifies immunotoxicity as a critical data uncertainty. This seems a bit surprising as in the discussion of the database uncertainty factors for the RfD and RfC, this is never mentioned as a major justification for the uncertainty factor, although it is mentioned in passing. It may also be useful to explain how this data gap leads to the justification of a higher uncertainty factor for the RfD. [Please note similar comments in Section 6 as well]

- Page 254, in discussing the CYP2E1 kinetics, EPA states that the studies indicate that any error is no greater than three and is thus a reasonable level of uncertainty. For uncertainties like this, which EPA can provide ballpark estimates for, it may be useful for EPA to provide a summary table that lays out all the uncertainties and the directional and quantifiable impact they may have on the derivation of the reference values. Such a visual table may be very useful in allowing users of the values to

understand quantitatively and qualitatively the overall level of uncertainties that exist and their potential cumulative impact.

- Page 255, line 240, EPA states that the impact of the model uncertainty appears "relatively small" for the noncancer assessment but "quite large" for the cancer assessment. It may be useful for EPA to use specific numeric values as it is unclear to the readers what is meant by these terms. Similarly, as per comments above, it may be useful to present these values in a table. [Note similar concerns regarding the discussion of uncertainties in cancer modeling, page 302-305, a similar table would be helpful]

- Page 258, line 492, EPA states that uncertainties "should not markedly affect the values..". It may be useful for EPA to clarify, with numerical values, what is meant by this comment. Showing these values in a table would also be informative. A similar clarification is also needed on line 524 when EPA states that the value "is quite similar."

- Page 263, lines 620-623, EPA discusses their determination that it may not be reasonable to apply a correction for multiple comparisons given the lack of independence of the groups and the focus on the liver. More clarity on why EPA suggests that a correction may not be reasonable would be helpful. Why wouldn't it be reasonable? It may also be useful to have a charge question on this as it is a critical departure from what the study authors determined.

- Page 268, line 7718, EPA states that the multi-stage model was used for dose response modeling. It may be useful to explain why EPA chose this model, as opposed to other available models. [Similar question on page 282 for the IUR derivation]

- Page 269, table 5-12, in looking at this table, it is clear that to derive the BMD values, as per footnote b, EPA dropped dose groups. EPA's justification for dropping dose groups should be clearly explained. A robust statistical approach involves determining what data are relevant and sound before beginning a modeling exercise. Dose groups should not be dropped simply to improve model fit. If no model fits well to all the available, relevant, and statistically sound data, EPA provide a justification for why a dose group was dropped. A charge question on this approach would be useful. We note that in section 5.4.1.7, using an administered dose approach, the modeling dose not require EPA to drop dose groups to achieve good model fit.

Editorial Comments (with Scientific Impacts):

- Page 85-88, in the summary table (4-9 and 4-10), the results present only point estimates and not the 95% confidence intervals. In some cases, if readers see the confidence intervals, they will realize that the SMR's or OR's may not be statistically significant and thus this would change the interpretation of the data. Thus we suggest including the full ranges. EPA may also want to take a look at the characterization in

the results column to ensure that it represents whether or not findings of elevated risks were statistically significant.

- Similar to the comment above, in table 4-28, EPA may want to clarify if the increases and decreases are statistically significant.

- Page 124, section 4.3's title refers to oral and inhalation studies, however one of the first studies discussed (Raje, 1988) is a subcutaneous injection study. Suggest moving this to a different section. A similar change is suggested in section 4.3.1.1 which should cover only gavage studies.

- In many tables, for example 4-26 through 4-28, the authors/citations for the studies should be provided. It may be helpful to check this throughout the document.

- Page 182, line 5686-5689, considering that we can never prove a negative, this sentence is a bit awkward. Suggest reframing in the positive with the appropriate caveats about the uncertainties and strength of the supporting data.

- Page 200, line 6086, typo in citation.

- Page 217, line 6606, EPA refers to the hepatic effects as "the critical dose-dependent noncancer effects". It may be helpful to explicitly clarify if by critical EPA means these are adverse effects (or perhaps EPA means they are precursor effects). Using the IRIS definition of a critical effect, it could be interpreted either way.

- Page 219, in figure 5-1 legend (and in other similar figures), EPA may want to clarify that the closed circles refer to other exposure concentrations used in the studies. Otherwise readers may think that the NOAEL values are extrapolated.

- Page 263, line 638, please clarify what is meant by "enhanced genotoxicity."

- Page 278, line 7910, please clarify what is meant by "a weaker trend." Was this trend statistically significant?

- Page 287, line 8094, please clarify what is meant by "slightly higher". It may be useful to bring the quantitative values forward from the appendices.

- Page 294, section 5.4.3, we note that section 5.4.3 appears to have some overlap with the first section5.4.2.8 (which appears on page 291 and again on page 292, but with a different name). EPA may want to consider putting the details of section 5.4.3 in an appendix.

- Page 299, table 5-26, justification for the data set. The last portion of this section discusses support provided by epidemiological studies. It is unclear how this provides justification for use of the animal data. In the justification for the dose-response

modeling, it may be useful for EPA to provide scientific rationale for the choice of the multistage model. Stating that it is "consistently used" seems a bit weak.

- Page 302, line 8428, it may be useful to clarify what is meant by "little modeling uncertainty". Can this be quantitatively defined? Did EPA try to fit any other models?

- Page 302, line 8436, EPA states that mutagenicity is the most plausible mode of action. Since this implies that the mode of action is not known, this is a bit contradictory to EPAs application of the ADAFs.

- Page 306, line 8542, instead of saying "about an order of magnitude", it may be helpful to present the specific values.

Comments on the Draft Charge:

(Note: some suggestions for charge questions are provided in comments in the above sections. Many of those comments have not been reiterated here, but should be considered as equally important in ensuring a rigorous peer review of this highly technical document.)

- Since the development of Agency Information Quality (IQ) guidelines required by statute, many agencies have been using charge language that tracks with the standards of their own IQ guidelines. For example, such language often focuses on whether or not the information in question is accurate, clear, complete, transparently and objectively described, and scientifically justified. We believe it may be useful for EPA to follow a similar approach and incorporate some of the language from your IQ guidelines into the formulation of the charge questions.

- For both the RfD and RfC, EPA transparently provides derivations for other candidate values. It may be useful to ask the peer reviewers to comment on the scientific justification for these derivations and provide input as to whether they agree with EPA's determination to not use the values.

- A1a, suggest asking a more rigorous question than whether or not the model "adequately" represents the Toxicokinetics. For instance, EPA could ask whether the inputs and assumptions represent the best available scientific data. Similarly, in this section and elsewhere throughout the charge, instead of asking if something is appropriately considered, EPA could also ask reviewers if the scientific information is objectively presented and used considering all the available information that EPA could choose from.

- A1b, discusses scaling at $BW^{3/4}$, however we note that for quantification, EPA appears to use $BW^{0.25}$. It may be helpful for EPA to clarify this and to also ask specific questions about the value that was used for quantification. EPA may also want to specifically ask peer reviewers to comment on whether or not they believe there is scientific support for the rationale (in addition to simply asking if the description is transparent and objective in presentation).

- B1, B5, C3 and C5, it may be helpful to specifically ask reviewers to comment on any other studies that may have be an appropriate choice for derivation of the RfD, RfC, OSF and IUR respectively.

- B2 and B6, in this question it may be helpful to specifically state whether or not EPA finds the critical effect to be an adverse effect or a precursor effect (see comments above in general comments). It may also be helpful for EPA to provide reviewers with the IRIS definition of a critical effect, as each reviewer may interpret it differently if they are not provided explicit guidance. EPA may also want to consider asking a charge question about the description.

- B4 and B8, as with other previous charges, it may be helpful to specifically describe the uncertainty factor choices in the charge question and ask reviewers to comment on each choice. In addition, on page 316, and elsewhere, EPA discusses how the toxicodynamic uncertainty factor is protective of sensitive populations. As EPA is already using the first percentile in the population distribution for the point of departure, it may be informative to ask the expert reviewers to comment on whether the cumulative effect of all EPA's decision points, in conjunction with the application of chosen uncertainty factors, provides a plausible quantitative estimate for the RfD and RfC.

- On page 252, EPA discusses the choice of the equation for the CYP pathway, and CYP:GST ratios. As this choice leads to some uncertainty (described on page 254 as 'a reasonable level of uncertainty'), it may be helpful to have a specific charge question asking about this important choice.

- C1, page 194 of the toxicological review (and perhaps elsewhere in chapter 5 and 6), provides a finding of "likely to be carcinogenic" by the inhalation and oral routes. Yet the charge asks about carcinogenicity by "all routes of exposure". If EPA is going to make a finding relevant to all routes of exposure, this should be clarified in the toxicological review. In particular, EPA may want to include discussion of the dermal exposure route and evidence which exists to support carcinogenicity via this route. In addition, as the weight of evidence is not similar for the oral and inhalation routes (and likely dermal), it may be very useful to specifically ask the reviewers about the strength of evidence to support the classification for each route of exposure separately. Finally, as there is so much epidemiological data discussed in the toxicological review, and this human database is quite robust, compared to other chemicals, EPA may want to specifically ask the peer review experts to comment on EPA's determination to not rely predominantly on the human data for both characterization and quantification.

- C2, in table 4-32, many of the in vivo assays are negative and the ones that are not negative are positive at doses well above the CYP saturation levels. Similarly, EPA states, on page 162, that most of the in vitro assays show positive results when there was GST activity and this is a requirement for the observation of genotoxic effects.

At environmental exposure levels, it is not clear that the GST pathway would play a dominant role as the CYP pathway is talked about as having a 10 fold higher affinity. EPA should consider asking about the mutagenic mode of action taking dose levels into account. This is very relevant as EPA is concerned about lower dose environmental exposures and it would be helpful to explicitly ask the expert reviewers about the mode of action in this dose range. It is quite possible that at different dose ranges, different modes of action prevail.

- It may be useful for EPA to ask reviewers to comment generally on figure 5-14, EPA's overall approach for deriving the OSF's and IUR's. Such a question would allow them to comment on each aspect of the approach and provide any specific comments.

- For the OSF, EPA derives the tumor risk factor using allometric scaling and a scaling factor of one. EPA should take comment on their choice of using the allometric scaling based on the rate of GST metabolism. EPA states on page 269, "The question of the role of specific metabolites, and particularly how these metabolites are transformed or removed is a key question affecting the choice of a scaling factor to be used in conjunction with the internal dose metric based on rate of GST metabolism." A similar type of question, regarding the scaling approach, would also be useful for the IUR derivation.

- For the OSF, EPA chooses liver-specific allometric scaling over whole-body scaling. It may be helpful to ask peer reviewers to comment on this decision. A similar type of question would also be useful for the IUR derivation.

- For the OSF, EPA presents, in section 5.4.1.6, 5.4.1.7, and 5.4.1.9 alternative derivations. It may be helpful for EPA to ask expert reviewers to specifically comment on these derivations and EPA's decision to not use them as a primary estimate. A clear endorsement from reviewers, of EPA's primary choice, would help to strengthen confidence in the overall assessment. A similar type of question would be useful for the IUR derivation and presentation of alternatives as well.

- Page 287, EPA discusses summing risks across tumors. It may be useful to ask the reviewers to specifically comment on EPA's determination and subsequent approach, to treating the two types of cancers (liver and lung) as occurring independently.

- C4 and C6, EPA is basing the OSF and IUR on the GST +/+ genotype, which represents the most sensitive of the groups. We applaud EPA for presenting this evaluation of sensitive subpopulations. EPA may want to ask peer reviewers if this value should also be used for assessments where EPA is looking for a best estimate of risk. While EPA typically does not like to underestimate risk and errs on the side of health protection, using an OSF and IUR which are derived to be protective of the most sensitive populations seems like a change for cancer quantification. Thus a charge question on this would be quite helpful. This is an important science policy decision and asking expert reviewers to comment on the scientific aspects (relating to

plausibility and relevance to different populations) will be very helpful. Similarly, it may be useful to describe for the expert reviewers and the public any policy guidance EPA has to describe this choice. In addition, EPA may want to consider presenting median values to be representative of the total population (using the mixed and representative GST genotype values) for use in situations where risk managers would like to understand best estimates of risk.

- EPA may want to consider taking comment on section 5.4.4. EPA could frame it to take comment on whether or not reviewers find that a mutagenic mode of action is certain enough at environmental exposure levels such that the ADAF application is warranted.