HSIA

Halogenated Solvents Industry Alliance, Inc.

1530 Wilson Boulevard, Arlington, VA 22209; Tel: 703-875-0683

Toxicological Review of Dichloromethane (Methylene Chloride)

In Support of Summary Information on the Integrated Risk Information System (IRIS)

Draft Released March 2010

Comments of the Halogenated Solvents Industry Alliance

Submitted by Paul H. Dugard, PhD

Docket Number: EPA-HQ-ORD-2010-0224

General:

EPA NCEA is to be congratulated for a comprehensive collection of the toxicological and epidemiological information for dichloromethane (DCM) and clear presentation of the data in summary tables. The previous risk assessment for DCM released by EPA in 1987 was a landmark assessment since it was the first by the Agency to employ PBPK modeling to develop cancer potency terms. This most recent risk assessment is also noteworthy since the PBPK models have greater sophistication. In particular, the human model now incorporates known human variability in CYP 2E1enzyme activity and the three phenotypes now known to exist in the human population for the enzyme considered related to carcinogenicity, glutathione S-transferase T1 (GST-T1). These factors are incorporated in a Markov Chain-Monte Carlo process to develop model parameters that are considered to represent variability in the human population better than non-probabilistic approaches. Since additional information available after 1987 continues to support GST-T1 as the enzyme that generates an active moiety in the induction of mouse liver and lung tumors, HSIA fully supports the approach used by EPA and believes that this risk assessment will also be recognized as a landmark in the application of PBPK modeling in the interpretation of toxicological data.

There are specific concerns that HSIA has identified and the most significant of these are outlined below within the framework of the draft charge questions to the external peer review panel (released in March 2010). The concerns relate to EPA's interpretations that differ from those of the original study authors, findings of adverse effects that are simply not supported by the evidence and certain procedures that may be inappropriate and lead to unreasonably conservative outcomes.

Note: References are as in the draft IRIS assessment except for Dell et al. (1999) which is included in this submission for the convenience of the peer review panel.

Chemical Specific Charge Questions:

(A) PBPK Modeling

Overview comment: The developments in the models for rat mouse and human all appear rational and founded in reliable information. We have not attempted to review the details in depth and we feel comfortable that the scientists, both within the Agency and without, are well recognized for expertise in the field of PBPK modeling. For that reason, we accept that the PBPK modeling is truly state of the art and reliable.

1. Rat PBPK model

a) To the extent that we have reviewed the detail, there are no concerns regarding the PBPK model itself and the modifications introduced by EPA.

b) We support the dose metric for RfC and RfD based on total metabolism of DCM via the CYP2E1 pathway provided that the appropriate toxicological end-point is liver toxicity (this is discussed below). If the effect to be used for RfC and RfD calculation is altered, for example to neurotoxicity or COHb formation, the dose metric will need to be reconsidered.

It is apparent that EPA NCEA has adopted the science policy that, where the dose metric is in terms of formation of metabolites, interspecies conversion is based on a bodyweight to the power 0.75 (b. wt.$^{0.75}$) factor to account for clearance potentially being slower in man than in the laboratory animal. This factor is substantial – 7-fold for mice and 4-fold for rats. We are not convinced that this default assumption is warranted and it seems to be a rather crude treatment to apply when sophisticated PBPK models are employed. Can the models really not provide a calculated interspecies conversion? We look forward to the opinion of the peer review panel, especially since the policy has general implications beyond DCM.

2. Mouse PBPK model

a) To the extent that we have reviewed the detail, there are no concerns regarding the PBPK model itself.

b) We fully support dose metric, tissue-specific GST-T1 metabolism, for calculations based on mouse liver and lung tumors. As expressed in 1. b) above, we consider the default b.wt.$^{0.75}$ conversion to allow for differences in clearance between rodent and man to be undesirable and believe that the PBPK models should allow a more realistic interspecies conversion.

3. Human PBPK model

To the extent that we have considered the detail of the human PBPK model and the modifications of the model published by David et al. (2006), the EPA approach appears to be appropriate and to reflect the sophistication that can be achieved in today's PBPK modeling.

A science policy issue that is of concern is the use of the 1st percentile of the distribution of human equivalent doses and concentrations for the calculation of RfD and RfC. This appears to be an emerging general policy when probabilistic PBPK models are employed. The concern is that by taking the 1st percentile rather than the "average" human, a double counting of variability/uncertainty has been introduced because of the conservative nature of the other elements involved in the RfD and RfC calculations. This was a concern expressed in comments by Lorenz Rhomberg PhD, FATS (Gradient Corp.) on the recent draft IRIS assessment for trichloroethylene. We reproduce his comments here since they are directly relevant to DCM:

"Comment #1 - The allowance for inter-human PK variability double counts and misconstrues the nature of the dose-response curve.

There are two questions about the allowance for human variability in metabolic activation. The first, addressed elsewhere in these comments, is whether the extent of variability has been reliably estimated. The second, addressed here, is how allowance for variability has been entered in to the RfD/C calculations. It would appear that allowance for human variability has been double-counted because inter-individual variability is built in to the tolerance distribution-based dose-response curve.

The method employed in the document is to set a point of departure (PoD) on the animal-based dose-response curve, using central estimates of "standard rat" internal doses as the dose measure. That is, inter-individual PK variation among rats, even though it exists, was not estimated and not considered in the dose-response curve estimation. For non-cancer endpoints, the dose-response curve is interpreted as a tolerance distribution – as the cumulative distribution of individual sensitivity variation. The reason that some animals respond at a given (externally applied) dose and others do not is that some have their individual tolerances exceeded while others do not, and higher doses exceed the individual tolerances of a greater fraction of the variable population, thereby yielding higher disease incidences.

Some of this variation is in PK, and so to some extent, the rats that respond do so because they are more vulnerable owing to their individual PK variation that makes them have a higher proportionality of internal to external dose. The contribution of this effect is captured in the fitted dose-response curve, which also reflects variation in sensitivity for other, non-PK reasons, but the contributions of PK variation are already incorporated, and are not readily split out without some attempt to characterize PK variation among individual rats.

The rat dose-response curve is then used to determine a PoD by finding a dose that yields a low predicted response, say 1%. Because the dose scale is measured in *average* internal dose among the rats, the dose associated with a 1% response level is the average internal dose for rats such that 1% of them are expected to have their individual tolerances exceeded. For the sake of argument, if we hypothetically say that there is absolutely no inter-rat variation in PK, then all the rats in a hypothetical experiment at the 1% response dose will have the same internal dose, and which rats respond and which do not will be ruled entirely by variation in pharmacodynamic (PD) sensitivity to this fixed internal dose. But, if one instead hypothesizes that variation in sensitivity is entirely ruled by *PK* variation (with no contribution of PD variability) then the 1% of rats responding are that same 1% that are most sensitive owing to their PK variation – that is, they are the 99^{th} percentile of the internal dose distribution.

The reality is somewhere in between, with both PK and PD variability contributing to variation in ability to tolerate the dose. But without characterization of PK variation among individual rats, we have no way to split the components out (though there is the conventional split between PK and PD that we apply to Uncertainty Factors).

Staying with the hypothetical case that sensitivity variation is all in PK, then the only reason to make further allowance for *human* PK variation is if variation in PK among humans is greater than variation among rats, and even then the correction should only be for the degree to which it is greater – that is, the ratio of the 99^{th} percentile in humans *versus* the 99^{th} percentile in rats rather than the ratio of the 99^{th} to the 50^{th} percentile in humans.

The hypothetical case of pure PK dependence of sensitivity variation is made to clarify the argument, but in the real case of contributions from both PK and PD, the principle illustrated still applies. There is some mix of influence of PK- and PD-based sensitivity among the responding rats, and the effect of this is captured in the fitted dose-response curve, for which the dose variable is the average internal dose. That internal dose is likely higher on average among the 1% of rats responding, because of the contribution of PK to their sensitivity; but, since this is unmeasured, all the analysis can say is that when a group of rats is dosed at a given external level, the average internal dose among them has some level estimated by the rat PBPK model. In view of the (unknown) contribution of PK to sensitivity and the (unknown) degree to which PK varies among rats, there is some (unknown) degree to which some rats have higher-than-average internal doses and thereby have an increased response probability (which is dictated by PD sensitivity to internal dose levels).

When the rat PoD is extrapolated to a human PoD based on average PK in the two species, it implicitly assumes that the mix of PK and PD, and the extent of inter-individual variation in PK, are the same in humans as in the rats. If one then makes a correction for the difference between the 50^{th} percentile of PK in humans and the 99^{th} percentile (as the draft reassessment does) it essentially implicitly assumes that all of the variation in sensitivity reflected in the dose-response curve is attributable to PK alone.

If one assumes that the mix of PK and PD influence is similar across species, then the correct correction is the ratio of 99^{th} percentiles across species, but since the 99^{th} percentile in rats is not estimated, this cannot be calculated. If one cannot assume that the mix of PK and PD is the same, then it is doubly impossible to calculate a correction.

The method that has been employed in the draft reassessment seems to implicitly assume that all of the dose-response in rats is attributable to PD (and this drives the PoD down as far as possible in internal-dose terms) and that all of the dose-response in humans is attributable to PK (and this drives the sensitive human allowance down as far as possible). The net result is to yield an RfC that is overcorrected for human inter-individual variation to a degree that is not possible to know with the analyses available."

This, again, is a policy issue which has implications beyond DCM and trichloroethylene. It seems that application of the 1st percentile in conjunction with other conservative elements contributes to improbably (and unnecessarily) low RfC and RfD values.

(B) Noncancer Toxicity of Dichloromethane

The charge questions will be addressed initially but additional comments regarding end-points not within the direct scope of the charge questions will also be covered.

Oral reference dose (RfD)

HSIA considers that "animals with liver foci/areas of alteration" from the 2 year drinking water study in F344 rats (Serota et al 1986a) to be an inappropriate basis for the calculation of an RfD and that the $7x10^{-3}$mg/kg-day RfD that results is too low relative to human experience.

End-point selection: EPA has reviewed the evidence for liver effects in workers exposed to high levels of DCM by inhalation and no convincing adverse effects have been found. Also, the "animals with liver foci/areas of alteration" designation in the Serota et al. (1986a) study may well represent effects of low toxicological concern (note 70% incidence in control males and 51% incidence in control females). Since it seems unlikely that DCM at low dose levels will induce effects in the human liver, there may be more appropriate end-points for setting an RfD. One such could be the increase of carboxyhemoglobin (COHb) that results from DCM metabolism via the CYP2E1 pathway. The standard for carbon monoxide itself could be used and the exposure to DCM that adds COHb equivalent to the CO standard could be back-calculated to yield an RfC (or suitable precursor term) for a dose route conversion using PBPK modeling to result in an RfD. This calculation would have the advantages that information from human studies only would be required and the generation of COHb is considered an effect of significance in man. Another end-point for which there is human data available is neurotoxicity. Two studies, one by Cherry et al. (1983) and the other by Lash et al. (1991) have been used as examples of "alternative" bases for calculating RfC values. Although inhalation exposures were involved in these studies, dose route conversion could be effected using PBPK modeling. The advantages are that only human data is used although there are limitations in the studies.

Calculation of RfD (based on "% animals with liver foci/areas of alteration" in the Serota et al [1986a] 2 year drinking water study in F344 rats): The pattern of the dose response in male rats (see Table 5-2 in the IRIS draft) seems unlikely to have yielded a better fit in the benchmark dose models than in the female rats, but we assume the curve fitting results are correct. Thus the logistic model applied to male rat data is the starting point for RfD calculation. A sequence of calculations then drive the RfD to an irrationally low value: the point of departure in the rat data is the $BMDL_{10}$ (not the BMD), the internal dose equivalent to the BMDL is then calculated and converted to a human internal dose employing the $b.wt.^{0.75}$ presumption of equivalence discussed above (4-fold reduction), the 1st percentile term for the human internal dose is used (as discussed above, a relatively small factor of approx. 2-fold in this case), an overall uncertainty

factor of 30 was applied (3- fold for each of: possible toxicodynamic differences between species, variability in toxicodynamics between humans, database uncertainties – discussed later). Thus a long term study with a clear NOAEL of 5 mg/kg-day leads to an RfD of 7µg/kg-day for a low-concern effect in rats that seems highly unlikely to occur in humans without repeated very high levels of exposure (based on human evidence). This is not a rational outcome.

Inhalation reference dose (RfC)

HSIA considers "hepatocyte vacuolation" in the long term inhalation study in Sprague Dawley rats to be an inappropriate end-point for the calculation of an RfC and the calculation based on that end-point is unduly conservative.

End-point selection: EPA has reviewed the evidence for liver effects in workers exposed to high levels of DCM by inhalation and no convincing adverse effects have been found. Also, the "animals with hepatocyte vacuolation" designation in the Nitschke et al (1988a) study may well represent effects of low toxicological concern (note 59% incidence in control females). Since it seems unlikely that DCM at low dose levels will induce effects in the human liver, there may be more appropriate end-points for setting an RfD. One such could be the increase of carboxyhemoglobin (COHb) that results from DCM metabolism via the CYP2E1 pathway. The standard for carbon monoxide itself could be used and the exposure to DCM that adds COHb equivalent to the CO standard could be back-calculated to yield an RfC. This calculation would have the advantages that information from human studies only would be required and the , generation of COHb is considered an effect of significance in man. Another end-point for which there is human data available is neurotoxicity. Two studies, one by Cherry et al. (1983) and the other by Lash et al. (1991) have been used as examples of "alternative" bases for calculating RfC values. The advantages for the neurotoxicity studies are that only human data is used although there are limitations in the studies. EPA has listed a number of other effects with calculated RfC values and some of these are addressed later in this review of RfC calculations.

Calculation of RfC (based on "hepatocyte vacuolation in a long term inhalation study in Sprague Dawley rats [Nitschke et al. 1988a]). The results for female rats were employed in the calculation and benchmark dose modeling showed the best fit to be a log-probit treatment. A sequence of conservative treatments was then applied to derive the RfC. These were: The internal dose at the $BMDL_{10}$ (not BMD) in the rat was converted to a human equivalent internal dose assuming a $b.wt.^{0.75}$ equivalence (a factor of 4-fold), the 1st percentile for human dose is applied (as described above – a 3-fold factor), an overall uncertainty factor of 30 was then applied (includes 3-fold for each of: toxicodynamic differences between rat and man, toxicodynamic differences among humans, and 10-fold for database uncertainties). The result of the calculation is an RfC of $0.2mg/m^3$. This is derived from a long term study in which a clear NOAEL of 200 ppm ($700mg/m^3$) was established for an effect of low concern that is highly unlikely to occur in humans without very high repeated exposures. This is not a rational outcome.

Uncertainty factors

The concern regarding use of the 1st percentile for human metabolism of DCM in conjunction with uncertainty factors is explained in detail above. A specific uncertainty factor for "database uncertainty" of 3 (more correctly $10^{0.5}$) is applied in the calculation of RfD and a factor of 10 has been applied in calculations of RfC. This is applied despite the extensive information available and appears to depend upon the absence of specific information on developmental neurotoxicity and immunotoxicity following chronic exposure. EPA has misinterpreted the available evidence relating to immunotoxicity. The results for DCM regarding mouse lung resistance to infection and bactericidal activity (Aranyi et al. 1986) cannot be regarded as indicating an effect. The single 3 hour inhalation exposure to 50 ppm and the 3 hour exposures to 50 ppm on each of 5 days are clearly without effect. The supposed positive result for a single 3 hour exposure to 100 ppm is based on concurrent control values that show low mortality and high bactericidal activity. The values for the DCM-exposed animals are consistent with most control results. The Aranyi et al. (1986) publication is included with this submission so that the peer review panel may draw its own conclusions. The other study of DCM immunotoxicity was reported by Warbrick et al. (2003) and was designed to fill a "priority data need" identified by the Agency for Toxic Substances and Disease Registry (ATSDR). As such, the protocol and study report were subject to rigorous peer review by expert immunotoxicologists from EPA, NIEHS, NIOSH and three non-governmental scientists. This study to EPA guidelines employed a single high dose (5,000 ppm) in a 28-day inhalation study in male and female rats and it was considered to be a screen answering the question "is there a need to investigate the immunotoxicity of DCM in further studies?" The answer accepted by the peer reviewers was that no such need existed. No "chronic assay" for immunotoxicity has been designed. Thus the available evidence does not suggest that a significant uncertainty exists for this end-point.

ATSDR has considered whether a priority data need exists for a developmental neurotoxicity (DNT) study on DCM. The available neurotoxicity information is not considered a strong enough driver for the DNT study to be required, and the use of animals could not be justified.

Thus the end-points identified by EPA in support of the "database" uncertainty factors of 3 or 10 do not justify their application.

Potential points of departure for candidate RfCs (Table 5-8)

The EPA-preferred RfC based on hepatocyte vacuolation in female rats has been discussed above. We believe that renal tubular degeneration was reduced in DCM-treated rats in the Mennear et al. (1988)/ NTP (1986) study and therefore not a basis for RfC calculation. The Raje et al. (1988) study has not been available to us but the result (in a genotoxicity study) stands in contrast to the findings in a multigeneration reproduction study (Nitschke et al. 1988b). The interpretation of the Aranyi et al. (1986) study has been discussed above: not clearly positive in an acute exposure, but clearly no effect in repeat exposures at 50 ppm and thus the absence of an effect means this is not a candidate for RfC calculation. Use of the Warbrick et al. (2003) study is unjustified – there were no adverse DCM effects in this study. The human neurotoxicity studies of Lash et al. (1991) and Cherry et al. (1983) do not suggest DCM is a significant

neurotoxicant but may address a more relevant end-point than others listed. However, as discussed above, a "database uncertainty factor" greater than 1 cannot be justified.

(C) Carcinogenicity of Dichloromethane

1. Cancer classification

Interpretation of epidemiological information. In general, EPA has tended to criticize the contribution of occupational cohort studies and has failed to acknowledge, in full, the weaknesses of the case control studies. Rather than attempt a detailed discussion within these comments, we recommend that the peer review panel considers the published review by Dell et al. (2003) that we include as part of this submission. A study which is used as "suggestive" evidence for an association between DCM exposure and astrocytic brain cancer is that of Heinemann et al. (1994); this study has also stimulated a more intensive analysis of brain cancer in other studies. The principal concern regarding this study is the job exposure matrix upon which exposures are assessed. The bases for assigning DCM exposures and the grading of the exposures are not explicit even in the paper dedicated to describing the framework of the job exposure matrix (Gomez et al. 1994). However, a "high probability of exposure" was linked to the occupations of painting, paint or varnish manufacture, ship or boat building and repair, and electronics manufacture. HSIA does not recognize any of these occupations as carrying a high probability of exposure to DCM. These supposed high probability occupations are also considered to involve high intensity exposures as are those in roofing and pharmaceutical manufacture. As expressed in the publication, it appears that exposures to DCM may have been grossly misclassified which would render the marginal results uninterpretable. Since the job exposure matrix has been used in many other studies and is still in use today, a full investigation of the matrix may be in order. It is likely that the results of the Heinemann et al. study have led to an over-interpretation by EPA of results for brain cancer in the studies reported in Tomenson et al. (1997) and Cohort 1 of Hearne and Pifer (1999) where a small number of brain cancers show SMRs greater than one but, in both cases, the 95% confidence intervals include the null.

Although a possible association between DCM and some of the less frequent human malignancies cannot be ruled out with certainty, the lack of consistency, absence of associations in well defined cohorts having experienced high exposures suggest that the carcinogenic hazard of DCM to man is low or non-existent.

2. Mutagenic mode of action.

Although a number of studies *in vitro* and some *in vivo* have shown positive genotoxicity, these results have often been achieved with exaggerated dose levels or specialized activation systems. The presumption that a product of the GST-T1 pathway is responsible for positive results is borne out by the pattern of results. The split between positive results in the mouse and negative results in the rat is compatible with species differences in GST-T1 activity. Thus a genotoxic mechanism may be implied in the mouse that may not be operative in the rat *in vivo* and is even less likely to be significant in man, based on species differences in GST-T1 activity.

Quantitative cancer assessment – oral exposure

3. Study selection

The Serota et al. (1986b) 2 year drinking water study in mice is not a suitable study for calculating an oral slope factor for cancer. For many years, this study has been regarded as not indicating a carcinogenic response, and EPA's reinterpretation of the results for male mice is unconvincing. The liver tumor incidence data are shown in Table 5-11 (information drawn directly from the study report). The range and mean of historical control liver tumor incidences are: range 5 - 40%, mean 17.8% for the laboratory. The range of liver tumors observed in NTP studies in control B6C3F1 mice at the time of the DCM study reached an even higher level (range 7 - 58%, mean 32%). EPA has adjusted the p-value for statistical significance from that used in the original study report and pairwise comparisons show that the incidences of liver tumors in the top three dose levels could be considered statistically significantly different from that in controls. However, this cannot overcome the broader factors in the interpretation. EPA also points to the "marginally increased trend test (p=0.058)" – this is not statistically significant despite the normal propensity for the Cochran-Armitage trend test to show statistical significance. Thus the pattern of liver tumor incidence in male mice does not show a credible dose-related pattern and the incidences for DCM-dosed mice lies comfortably within the historical control range for the laboratory. For these reasons, we accept the interpretation of the original investigators and find that DCM was not shown to induce mouse liver tumors in this study.

4. Calculation of oral slope factor (OSF)

The reasons for considering the interpretation of the Serota et al. (1986b) mouse drinking water study unsuitable for calculating OSF are further supported by the calculations themselves. The pattern of mouse liver tumor incidences is clearly not ideal for benchmark dose modeling. In their manipulation of the data, the incidence in the top dose mice had to be dropped from the calculation of benchmark dose. Perhaps most telling is the contrast between the OSF calculated from the drinking water study versus that obtained in a dose route conversion derived from the more clearly positive NTP inhalation study – with comparable treatments, the value for OSF from the drinking water study preferred by EPA is 1.7×10^{-3} per mg/kg-day whereas that from the inhalation study is 1.2×10^{-4} per mg/kg-day. The best explanation for this discrepancy is the unsuitability of the data from the Serota et al. mouse drinking water study for the calculation of OSF. We recommend that a dose route conversion from mouse inhalation study data be employed based on the preferred calculation for inhalation unit risk discussed below.

Quantitative cancer assessment –inhalation exposure

5. Study selection

HSIA agrees with EPA's selection of the NTP (1986) 2 year inhalation study as the basis for calculation of the inhalation unit risk (IUR) for DCM. Although there may be reasons to believe that mouse lung tumors are not relevant to man, we accept the combination "lung or liver tumor" incidence in female mice as the basis for calculations.

6. Calculation of inhalation unit risk (IUR)

We agree that the average daily mass of DCM metabolized via the GST-T1 pathway per unit volume of liver and lung is an appropriate dose metric. EPA has, again, employed a b.wt.$^{0.75}$ factor for determining equivalent internal dose in man versus mouse (a factor of 7-fold). This is based on potentially slower clearance of metabolites in man. Since this is an important policy development, HSIA would appreciate the peer reviewers' opinion on whether the 7-fold factor or a factor of 1 is the more appropriate. We accept that the IUR should be based on the human GST-T1 $^{+/+}$ phenotype. HSIA prefers that the factor of equivalence for internal dose between mouse and man be set at 1 but, as mentioned earlier, we are surprised that the clearance cannot be explicitly defined from existing information. Use of the factor 1 yields an IUR of 1.9×10^{-9} per $\mu g/m^3$.