Buchholz, Kurt
From: ROLAND HEMMETT [HEMMETT.ROLAND@epamail.epa.gov]
Sent: Monday, September 14, 1998 9:13 AM
To: buchholz@BATTELLE.ORG
Subject: Henry Lee's Comments

Question 1.

I believe the approach in the "Memo for the Record" lays out a logical approach to a difficult problem. I have not reviewed the Regional Matrix Levels, and so can not comment on them directly, though they are obviously a key component for the 4 compounds and dioxins. One limitation is the lack of more specific guidance for evaluating the dredge material as a "whole". I recognize this is very difficult, but you might try formalizing various approaches and see if they come up with reasonable conclusions. For example, you might use an approach similar to a toxic unit approach but using CBRs, WQCTLs values, or other residue effects levels. Each compound would be normalized to the fraction of the effects residue concentration (e.g., 0.2 of a CBR or WQCTL residue) for each compound. Then these fractions could be summed in these various ways: 1) compounds with broadly similar modes of action (e.g., metals vs. organics; all neutral narcotics) or 2) all compounds. Such an approach makes the assumption of simple additivity, which is open to criticism both for not accounting for synergistic interactions and for adding dissimilar toxicants and thereby overestimating actual toxicity. Nonetheless, this approach incorporates the full range of contaminants, their toxicity, how close they are to some estimate of an effects level, and makes a crude attempt at combined effects. Using trophic transport factors, the same approach could be applied to higher trophic levels. The total "bioaccumulative units" (or whatever) are probably better used in a comparative than absolute sense, so the test sediment would be compared to the reference and background.

There is more detail embedded in the text that is not captured in Figure 1. For example, the use of adjustment factors for some organics but not metals or comparison to background residues. I suggest that additional figures be drawn that give kinds details.

Question 3.

All eight factors are of some importance. The "phylogenetic diversity of the species" is the least important, as with the current procedure of using two species it is really a restatement of "the number of species".

See Question 1 for a possible approach for a framework to integrate various compounds. It is still important to have an integrated assessment because: 1) Matrix values exist for so few compounds, 2) the effects of at least some suites of contaminants are based on the summation of their residues (e.g., PAHs), and 3) the acceptability of a dredge material is based on different types of risks (e.g., direct impacts on benthos, biomagnification to human consumers) which requires some type of overall assessment.

Question 6A. & 6B.

The FDA action limits are useful as UPPER limits. They have some regulatory authority and so can be defended as a reason to deny (not pass) a dredge material. The only reason to omit them is if their inclusion causes confusion with applicants or the public that levels below the FDA limits are considered "safe".

Question 9.

A congener approach is much preferred to Aroclors though it is not practical to quantify all 209 congeners on a regular basis. Therefore, the best strategy is to derive ratios of total to a measured suite of congeners based on empirical studies and/or what is known about the original composition of the PCB mix and how the various congeners breakdown. I am not an analytical chemist, so can not comment directly on whether the doubling is the correct ratio. It is important, however, that the measured suite of PCB congeners include both the environmentally common ones (e.g., 154) and the less abundant but toxicologically important ones (e.g., co-planars).

Question 10.

The 28 day duration of the bioaccumulation test was a compromise between practicality and a duration that approached steady-state. Since establishing the 28-day test, new evidence indicates that a number of compounds, at least under some circumstances, do not approach within 80% of steady-state in 28 day. Since the purpose of the bioaccumulation tests is to estimate human and ecological risk, it is critical to have reasonable estimates of tissue residues in the field. Therefore, adjustments should be made on a compound-by-compound basis. Note that using an adjustment is not a "conservative" assumption (like use of a UF) but a method to correct for a lab artifact (duration). As the data become available, the need for an adjustment factor for all the organics should be evaluated. In particular, the use of an adjustment factor for the dioxins/furans needs to be considered. Rubinstein did some uptake work on 2,3,7,8-TCDD that could be used and there is probably more recent work.

Adjustment factors for metals are problematical. We did not see simple patterns when we reviewed metal BAFs (Boese and Lee, 1992), but there are cases when 28 days is insufficient to approach 80% of steady-state. I have not reviewed metal data since then, but at that time it was my belief that we did not have sufficient understand of matrix effects and/or physiological effects (e.g., incorporating metals into jaws) to derive precise adjustment factors. If a particular metal is of high concern, you should consider conducting longer term tests. As a general comment, the use of 10-day bioaccumulation test for metals is simply not scientifically defensible. We may not totally understand metal kinetics but there are abundant data (much of which was available when the Green Book was written) to show that 10 days often (usually?) does not approach steady-state and can even result in false negatives about identifying which metals are bioavailable. Question 11.

Because of their short life span of benthic species, B(a)P and other PAHs probably act as neutral narcotics rather than carcinogens, so the CBR approach is a more appropriate for ecological risk. (Note: Calculation of CBR really needs to be done on a molar basis rather than a concentration basis, as done on page 15. Fluorene and naphthlene can be added on a ppt basis only if they have same molecular weight.). Assessing human health risk is not my area of expertise. Based on what I know about toxic equivalents with dioxins, I suspect that the individual potency factors are not well known. Nonetheless, as with the dioxins, the sum is probably a better estimate of cancer risk than just evaluating the PAHs individually.

Question 12.

The problem with the PAHs may be more difficult than with the PCBs since there are multiple natural and anthropogenic sources. Therefore, the ratios among various PAHs is likely to vary more than among PCB congeners. The simplest approach would be to have a single ratio of total to the measured PAHs based on empirical studies. I believe a better approach would be to break down the PAHs into functional classes and have empirical ratios for each. The simplest functional classes would be high and low molecular weight but there may be other grouping that better capture the various local sources (e.g., petroleum vs. combustion vs pyrogenic PAHs). Note that this problem becomes more acute if the alkylated-PAHs are included.

Question 16.

I did not review Squibb et al (1991) in detail, however, it appears to have been a reasonably complete analysis. I suggest that rather than just listing TCDD (as in their Table 19), it is more appropriate to evaluate dioxins and furans. I did not see TBT mentioned. TBT is a possible contaminant to consider, especially if the area is used by the Navy. As a check on Squibb, you might consider going over NPDES reports of discharges into the area to determine if there are compounds with a high bioaccumulation potential that were not found.

Question 18.

Not really. For example, if a person's sodium level were 9-fold higher than average, they probably would be dead. Having said that, it is appropriate to use the extent of deviation from the reference site as a qualitative factor in a risk assessment especially if no residue-effects relationships are available.

I believe a more powerful ecological argument can be made by comparing the test sediment to the background at the disposal site. Assuming that the disposal site is not impacted (e.g., benthic analysis) and tissue residues in benthos and fish are acceptable (at least for the compound of concern), this is reasonable evidence that the test material would not cause undesirable effects if it did not exceed the background values. There are, however, a few caveats. First, is that the statistical tests have sufficient power to detect physiologically/ecologically relevant

differences so it is critical to consider both Type I and Type II errors. Given that the number of replicates is fixed, one approach is fix the statistical power and then adjust the alpha as needed. This puts the risk on the "discharger" rather than the "environment" and does not "reward" poor replication or high variance. Another approach is to rephrase the null hypothesis from "dredge material = background" to "dredge material > background", so that it is in the discharger's interest to have sufficient replication and low variance.

Second, with the benthos it is important to compare species with similar feeding habitats. Comparing a filter-feeding bivalve (e.g., Mercenaria) from the dredge site to Macoma might incorrectly indicate that the test sediment has a higher bioaccumulation potential than the background material. Third, in evaluating the high trophic levels, it is important to collect fish and megafauna (e.g., lobster) from the dredge site that are either territorial or are not highly mobile. In highly mobile species, unacceptable tissue levels could be diluted by time spend outside the dredge site.

Question 19.

As discussed under Question 18, the comparison to background tissue residues can be a defensible method of evaluating potential risk. Because of the various uncertainties, comparisons to background should not replace decisions based on individual residue criteria, if available.

If the data approach a normal distribution, the mean is best measure of central tendency. If the data approach a lognormal distribution, a geometric mean should be considered. An advantage of the mean is that the public understands it.

Question 20.

Yes, given the caveats discussed under Questions 18 and 19.