



UNITED STATES ENVIRONMENTAL PROTECTION AGENCY  
REGION IX  
75 Hawthorne Street  
San Francisco, CA 94105

NO0217.000223  
HUNTERS POINT  
SSIC NO. 5090.3

October 10, 2000

Mr. Richard Mach  
Southwest Division Naval Facilities  
Engineering Command  
1220 Pacific Highway  
San Diego, CA 92132-5180

**SUBJECT: DRAFT FINAL VALIDATION STUDY WORK PLAN, PARCEL F,  
HUNTERS POINT NAVAL SHIPYARD**

Dear Mr. Mach:

The Environmental Protection Agency (EPA) has completed review of the subject document dated September 13, 2000. Our comments are included in the enclosure. If you have any questions regarding these comments, please call me at (415) 744-2387.

Sincerely,

A handwritten signature in cursive script that reads "Sheryl Lauth".

Sheryl Lauth  
Remedial Project Manager

Enclosure

cc: Mr. Chein Kao, DTSC  
Mr. Brad Job, RWQCB  
Mr. Mike Wanta, TTEMI  
Ms. Karla Braesemle, Weston  
Ms. Claire Trombadore, EPA  
Mr. John Chester, City of SF  
Mr. Dave DeMars, Navy

**COMMENTS FROM EPA'S TECHNICAL SUPPORT BRANCH, DR. CLARENCE  
CALLAHAN REGARDING THE DRAFT FINAL VALIDATION STUDY WORKPLAN  
FOR PARCEL F, HUNTERS POINT NAVAL SHIPYARD**

**General Comments:**

The document is generally well done, but very complicated by the many parts and pieces. Although, the formal weight of evidence (WOE) approach for this project provides much more structure than we normally see in these documents, some kind of concise table or figure for the overall purpose might help any reader better understand the document. While the WOE approach provides a good decision process, we agree with the Navy that we cannot rely completely on the bright lines suggested by the WOE process. We would strongly suggest that the Navy provide in addition to the WOE bright lines a well thought out risk characterization discussion of the results of all of these studies.

**Specific Comments:**

1.0 Introduction.

p1, last par. Evaluation of the debris-lined shoreline areas is complicated by the fact that no samples can be taken from them during the confirmation effort. How will these areas be evaluated such that they may be "...found to be acting as an ongoing source of contamination to intertidal or subtidal sediments...?"

p2, Reference sites. The Navy has added some very useful plots of reference site data (Appendix B). This is a very useful addition to the document.

p21, Sediment Dynamics, Table 3-7. Data Quality Objectives for Evaluation of Sediment Dynamics.

Step 1. State the Problem. The reference here is to "Section 3.1" whereas, Section 2 states the objectives and Section 3 discusses data collection and analysis. The Section 2 reference is more appropriate.

Step 2. Is this the activities associated with the current and sedimentation study?

Step 4. We would suggest that the roman numerals somehow be identified with the areas of study and a figure in the text without which, there is little context to just the numbers.

p22, Table 3-8. Data Quality Objectives for Feasibility Study-Related Sediment Characterization. Again, shouldn't this be referenced to Section 2 and not Section 3?

Step4. Where are the "Questions 1 and 2" referred to in this step?

RWQCB, 2000 - There is no citation in the Section 4 References to this acronym, please provide a full citation.

Step 5. Not clear of the intended meaning of the statement, "Treatability test results for various dewatering and stabilization methods will be compared with each other to identify the most effective method for HPS sediments." Does this mean that the methods will be ranked using criteria for effectiveness of treatment? If so, what are these criteria?

Step 6. The two sentences in this step are not clearly written, please clarify.

Appendix A.

Meeting minutes. Page A-67 seems to be poorly centered, please provide a corrected copy of this material.

Appendix B.

The additional plots provided are very useful for examining the adequacy of the reference sites. What is the ambient threshold value?

pB-2, It appears that the chromium maximum at Bay Farm is greater than the other site and the ambient stations (Figure 1a). The mercury maximum is greater at Bay Farm than at the other site and the ambient station (Figure 1b). Silver shows an interesting distribution in that the mean at Bay Farm is much higher than Alcatraz and in the high range of the ambient data (Figure 1c). Total PAH Low shows a higher range at Alcatraz than at either Bay Farm or the ambient data (Figure 1c). DDD and DDE (4,4') show distributions that are higher than the ambient data range (Figure 1d). Generally, these data support the use of these two sites as reference sites, however, the Navy must be cautious about interpreting data for the contaminants listed above. Please show the results of the statistical analysis of these data to support their use for this project. The Navy should provide the plots and the results of the statistical analysis of the organotins in this document.

Macoma Tissue Data from 28-day Bioaccumulation Test. It appears that just about all of the data for both reference sites are greater in range or maximum value except for chromium, nickel and selenium. With only one sample representing the 1998 ambient data, it's difficult to maintain that these data are representative of ambient data.

pB-6, There are no PCB data or chlorinated pesticides (other than DDT) presented in this document for review. Please provide all of the PCB and other chlorinated pesticides data reviewed and evaluated.

Based on the small data sets available what other options do we have to establish a data set that might be representative of the Bay in the Hunters Point region? What's the down side of combining the three data sets for a single data set to be compared to site data? Yes, we know the Navy showed some statistical differences among the data sets, however, can we say that these data do not reflect the conditions where they were sampled and therefore, are representative of the range of conditions in the Bay area? Is this not the same argument about the pseudoreplication previously made for sediment bioassays?

B.4 Suggested Method for Evaluating Toxicity to Invertebrate Larvae in Support of SWDIV

## Programs

pB-15, What protocol will be used to perform the TIE, please provide a reference.

Invertebrate tissue. How will the invertebrate tissue data be used to address the uncertainty of bioaccumulation? We remain skeptical that using a grab (in a four hour period) will provide sufficient tissue for measuring tissue levels of Hunters Point contaminants. Please explain why a dredge or trawl cannot be used to collect tissue? We cannot accept that a small otter trawl (five to ten feet) would be destructive to the sediment surface at Hunters Point. An otter trawl even if only five feet in width would collect far more organisms than any kind of grab that the Navy is planning to use.

pB-22, What were the contaminants of concern in the Gray's Harbor 28 and 60 day comparative tests? How were they different from the "San Francisco Bay COPECs?"

pB-22, Recommendations. The second bullet states that there is no available laboratory evidence that demonstrates that contaminants of concern approach steady-state concentrations following longer exposure times (than 28 days). What do the Gray's Harbor test results show? Please provide a copy of this study.

pB-23, Comparison to reference. If the sediment and tissue chemistry data **do not** show differences in bioavailability between fine-grained and coarse-grained reference sediments, how will these data be treated?

pB-24, All pesticides should be considered priority, particularly those organochlorine compounds that bioaccumulate. We don't understand why the concentrations cannot be compared to reference sites as we defined in the WOE chart of June 30, 2000. The text in paragraph four on this page does not agree with the WOE chart under the column labeled "Macoma Bioaccumulation" and we suspect that the present information in that column refers to the upper trophic level comparisons.

Step 2: Evaluation of Risk to Higher Trophic Level Receptors.

While the calculation of HQ values are appropriate for the first level evaluation effort, an evaluation in the latter phases of an ERA is questionable. A comparison of the magnitude of HQ values adds a considerable amount of uncertainty to any decision derived from such an exercise. The level of uncertainty in exposure estimates is high because of the lack of site specific information e.g., site use factor, food ingestion rates and the actual food consumed by the receptors at the site. Although, we reviewed the bioaccumulation material in early August, we do not agree to any comparisons of  $HQ_{low} > 10$  or  $HQ_{high} > 1$ . There is no technical justification for assuming that a dose that is ten times the low TRV is unprotective and any dose approaching the high TRV should raise serious concern for the receptor.

The WOE chart should have another column to accommodate both Macoma bioaccumulation and the upper trophic evaluation. The WOE table should be revised.

pB-27, Food ingestion rates. Please include the calculations for ingestion rates using Nagy et al

(1999) formula in this document.

$C_{\text{prey}}$  Please explain why the deperated tissue level is proposed rather than the non-deperated tissue level.

Site Use Factor. The site use factor really involves the seasonal use of the site and the actual habitat area used by the receptor. Guessing at the area used by the receptor adds far more uncertainty to the process than the assumptions about seasonal use. Site use in area actually should be done by observation rather than estimation.

Describing the  $TRV_{\text{high}}$  as being “less conservative” is immaterial for the intended use of these data. The  $TRV_{\text{high}}$  is not intended to bound the estimated doses for a site, it is intended in all situations that are above this dose level that the site should go immediately to the FS or the site is a high positive, a problem or whatever label assigned to it. The  $TRV_{\text{high}}$  is a dose level that has a high probability of resulting in a significant risk. The  $TRV_{\text{high}}$  represents a value in the middle range of **probable effects**, not possible effects.

pB-28, Ancillary Evaluations for the Bioaccumulation Line of Evidence.

What is the “second evaluation” after a significant difference between the site and the reference area data? (See third step on pB-23).

How will the data be evaluated “...to assess the uncertainty associated with Steps 1 and 2 and assist in the interpretation of “gray” areas identified by the WOE evaluation?”

pB-33 and 34, Determination of Analytical Variation (noise):Correlation of Tissue Concentration with Bulk Sediment Measurements.

We would request that the Navy clarify the material in this section. The title of the section suggests that the Navy will identify and separate the noise from the meaningful data when examining sediment concentrations and tissue concentrations for Macoma. Although the scatter plots are instructive, we find it difficult to get a grasp of the variation of the tissue observations at a particular sediment concentration. The text states, “In most cases, although the sediment concentrations have considerable variation...” We cannot decipher the “variation” in the sediment observations. We can see a wide range of sediment concentrations, but there are no data to show any variance in the sediment concentrations. We can also see that tissue concentrations vary (about the mean) by a factor of two, however, a tabular presentation of these data would be more instructive.

Please clarify the statement, “This result suggests that observed tissue concentrations following a 28-day bioaccumulation exposure may not be a function of sediment concentration...” can be supported by the data presented. Are not the tissue concentrations due to active transport or is it defusion which the Navy seems to refute and wouldn't both be related to exposure to sediment concentration? The second phrase in this sentence, “...because the tissue data are tightly grouped and appear to represent the least possible analytical variation (noise) observed in analysis of tissue COPECs.” is not clear. We just don't understand what leads the author to state that the “data are tightly grouped” and how any statement can be made about analytical noise since no

data are presented on repeated measures or anything that shows an evaluation of “noise.” Finally, EPA does not see any data presented to arrive at a discussion of analytical variation surrounding replicate tissues. A “two-fold magnitude” may represent the variation about the mean for Macoma tissue when exposed to similar sediment tissue, but the Navy has not presented a convincing argument to support this position. Perhaps this material could be reworked to make the explanation clearer and to strengthen the connections between data presented and the establishment of positions.

pB-34, Laboratory Exposures in Oakland Harbor navigational Programs: Statistical Comparisons to Multiple References Sites.

The information in the three tables is not clearly presented and incomplete. Please add the station designation above the columns e.g., for Table 1, Magnitude (Oak.Harbor/R-AM) and for Table 2, Magnitude (Oak.Har/R-BF), etc. Also, the tables would be more complete if all the ratios were listed, those that were not significant in plain type and those that are significant in bold type. This would show what ratios were not significant rather than just indicating the lack of significance.

Based on this information, Although, the range of minimum ratios that are significant range widely, we would suggest that the cutoff point should be around 1.5 for all contaminants.

pB-40, Toxicity Identification Evaluation Procedures Associated with Sediment-Water Interface Larval Evaluations.

We don't understand the use of two protocols for the TIE studies rather than a single procedure for both sample collections. The SAIC work reported for Goss Cove seems to be the protocol that is closer to a traditional TIE protocol. There shouldn't be any difference between the protocol for a sample collected at depths more or less than 5cm, what is the decision to use two different protocols based on? Why is this test being performed on the suspended-particulate phase which dilutes the sediment sample rather than use a whole sample? We would suggest that measurements of COPECs in the original sediment sample be performed along with measurements of other conditions (pH, salinity, ammonia, sulfides), all of which are considered important to the interpretation of the TIE results. These data are considered baseline before any manipulation of the sediments for further testing without which the potential reduction of COPECs cannot be evaluated. Will COPEC analysis be done for any of the preparations other than the 100% SPP e.g., 10% or 50% of SPP or will these be considered nominal concentrations?

A minor point, but these kinds of tests produce exposure-response results not dose-response results, this should be changed in the text.

pB-50, Weight of Evidence, Step 2, Sediment Chemistry.

The addition of the number of contaminants above the ERM-Q should be used with caution because any group of contaminants above the ERM-Q can suggest more of a risk compared to another group, this feature has to be viewed very carefully.

While we support the strategy for using a ratio of estimated exposure and a qualified benchmark (i.e., HQ), a comparison of the magnitude of HQ values adds a considerable amount of

uncertainty to this decision process. With all of the uncertainty in exposure estimates to calculate the HQ, particularly when site specific information is not used, the HQ value is virtually useless to indicate potential risk. There is no technical justification for assuming that a dose that is ten times the low TRV is unprotective and any dose approaching the high TRV should not raise serious concern for the receptor. Any dose below the TRV<sub>low</sub> should be considered safe if sufficient data are available to estimate the exposure of a particular receptor. Any dose at or above the TRV<sub>high</sub> will definitely result in a problem if sufficient data are available to estimate the exposure of a particular receptor.

pB-53 and 54, Integrate all Endpoint Results for a Given Sample Location.

This is the area of the process that will result in much discussion and negotiation because this numerical process is replacing the Risk Characterization that is normally presented at this phase of the process. The Risk Characterization would present a discussion of the intensity and breadth of the estimated risk with a discussion of the potential for recovery. The Navy, perhaps recognizes this by making the statement, "It should be emphasized that the bright line criteria are a starting point for interpreting WOE results; the final decision about the actual remedial footprint will be made in Parcel F FS." which suggests that even with a bright line, the results from all lines of evidence will play a role in the final determination of the validation study.