SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 27, No. 4, pp. 933–944, 2008

Printed in the USA
0730-7268/08 $12.00 + .00

# COMPARISON OF METHODS FOR EVALUATING ACUTE AND CHRONIC TOXICITY IN MARINE SEDIMENTS

DARRIN GREENSTEIN,*† STEVEN BAY,† BRIAN ANDERSON,‡ G. THOMAS CHANDLER,§ J. DANIEL FARRAR,‖ CHARLES KEPPLER,# BRYN PHILLIPS,‡ AMY RINGWOOD,†† and DIANA YOUNG†

†Southern California Coastal Water Research Project, 3535 Harbor Boulevard Suite 110, Costa Mesa, California, 92626, USA
‡Department of Environmental Toxicology, University of California, Davis, California, 95616, USA
§Department of Environmental Health Sciences, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, 29208, USA
‖U.S. Army Engineer Research and Development Center, Waterways Experiment Station, 3909 Halls Ferry Road, Vicksburg, Mississippi, 39180
#South Carolina Department of Natural Resources, Marine Resources Research Institute, 217 Fort Johnson Road, Charleston, South Carolina, 29412, USA
††Biology Department, University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, North Carolina, 28223, USA

(*Received* 3 *May* 2007; *Accepted* 26 *October* 2007)

**Abstract**—Sublethal test methods are being used with increasing frequency to measure sediment toxicity, but little is known about the relative sensitivity of these tests compared to the more commonly used acute tests. The present study was conducted to compare the sensitivity of several acute and sublethal methods and to investigate their correlations with sediment chemistry and benthic community condition. Six sublethal methods (amphipod: *Leptocheirus plumulosus* survival, growth, and reproduction; polychaete: *Neanthes arenaceodentata* survival and growth; benthic copepod: *Amphiascus tenuiremis* life cycle; seed clam: *Mercenaria mercenaria* growth; oyster: *Crassostrea virginica* lysosome destabilization; and sediment–water interface testing with mussel embryos, *Mytilus galloprovincialis*) and two acute methods (amphipod survival with *Eohaustorius estuarius* and *L. plumulosus*) were used to test split sediment samples from stations in California. The test with *Amphiascus* proved to be the most sensitive sublethal test and the most sensitive overall, identifying 90% of the stations as toxic. The *Leptocheirus* 10-d test was the most sensitive of the acute tests, identifying 60% of the stations as toxic. In general, the sublethal tests were not more sensitive to sediments than the acute tests, with the sublethal tests finding an average of 35% of the stations to be toxic while the acute found 44%. Of the sublethal tests, only the *Amphiascus* endpoints and *Neanthes* growth significantly ( $p \leq 0.05$) correlated with sediment chemical concentrations. Poor correspondence occurred between the toxicity endpoints and the indicators of benthic community condition. Differences in test characteristics such as mode of exposure, species-specific contaminant sensitivity, changes in contaminant bioavailability, and influence of noncontaminant stressors on the benthos may have been responsible for variation in response among the tests and low correspondence with benthic community condition. The influence of these factors cannot be easily predicted, underscoring the need to use multiple toxicity methods, in combination with other lines of evidence, to provide an accurate and confident assessment of sediment toxicity.

**Keywords**—Sediment toxicity  Sublethal toxicity  Multiple species  Acute toxicity

## INTRODUCTION

Acute sediment toxicity testing has been routinely conducted as part of monitoring and assessment programs, such as the U.S. Environmental Protection Agency's (EPA's) Environmental Monitoring and Assessment Program [1]. The toxicity tests are usually conducted on whole sediments using amphipod 10-d survival tests in accordance with standard protocols [2]. Sublethal testing has been conducted on a much more limited basis, but there is increased interest in using sublethal methods due to the assumption that they are more sensitive to contaminated sediments than the acute methods [3]. Sublethal methods include embryo development and other tests with various life stages of animals and endpoints such as growth and reproduction, in addition to survival. A wide variety of marine sublethal methods have been described [4]; however, very few of these methods are commonly used. Marine sublethal test methods that are commonly used include the amphipod *Leptocheirus plumulosus* 28-d growth and reproduction test [5], a 20-d polychaete growth test using *Neanthes arenaceodentata* [6], pore-water testing using echinoderm gametes or embryos [7], and a sediment–water interface (SWI) test using sea urchin or mussel embryos [8]. Recently developed sublethal tests showing promise include copepod reproduction [9], juvenile clam growth [10], and oyster biomarker responses [11].

Because marine sublethal toxicity methods have been used less commonly, there are questions regarding the practicality, reproducibility, and sensitivity of these test methods in comparison to acute methods already in use [12,13]. Few studies specifically designed to compare the relative attributes of marine sublethal tests have been conducted. A large study comparing 10 species with multiple sublethal endpoints has been conducted for freshwater sediment toxicity [14]. Marine studies conducted to date have only compared two or three methods [12,15,16] or focused more on sublethal elutriate or pore-water tests rather than whole sediment tests [17]. Important factors to consider in the selection and interpretation of toxicity tests include the degree of exposure to whole sediment, the relative sensitivity to sediment contaminants, and the level of concor-

* To whom correspondence may be addressed (darring@sccwrp.org).
Published on the Web 12/11/2007.

Table 1. Characteristics of the sublethal sediment toxicity methods included in the comparison study

| Species | Taxon | Test endpoint(s) | Duration (d) | Reference |
|---|---|---|---|---|
| *Mytilus galloprovincialis* | Mussel | Embryo development at sediment-water interface | 2 | [8,27] |
| *Mercenaria mercenaria* | Clam | Growth | 7 | [10,25] |
| *Crassostrea virginica* | Oyster | Lysosomal destabilization | 4 | [11] |
| *Leptocheirus plumulosus* | Amphipod | Growth, reproduction, survival[a] | 28 | [5] |
| *Neanthes arenaceodentata* | Polychaete | Growth, survival[a] | 28 | [23,24] |
| *Amphiascus tenuiremis* | Benthic copepod | Reproduction, survival[a] | 14 | [9] |

[a] Test endpoints that are secondary.

dance with benthic community impacts. Information on these factors is extremely limited for many sublethal tests.

The present study was designed to investigate the relative performance of several acute and sublethal test methods with marine whole sediments. Three specific points were examined. First, the relative sensitivity of the toxicity test methods was compared. Sensitivity was defined as the relative ability of a test method to detect toxicity in a sample. Sensitivity comparisons were made between acute and sublethal methods and among the sublethal methods. Second, the relationship between sediment chemical concentrations and toxicity of each method was examined. Third, the present study investigated the relationship between changes in benthic community condition and toxicity.

## MATERIALS AND METHODS

Six candidate whole sediment sublethal methods were selected (Table 1). These methods appeared to be technically feasible and had data available that indicated some level of sensitivity to contaminated sediments. Methods were first selected from established methods that had been published by government or scientific agencies (e.g., the U.S. EPA and the American Society for Testing and Materials). Additional methods were selected from the scientific literature and recommendations by toxicologists with experience in sediment quality assessment. Acute amphipod testing was also conducted for comparison with sublethal methods using two species, *Eohaustorius estuarius* and *L. plumulosus*.

The sediment samples tested were collected as part of two regional monitoring surveys, Southern California Bight 2003 Regional Monitoring Program [18] (Fig. 1) and the San Francisco Estuary Institute Regional Monitoring Program [19] (Fig. 2). In general, the stations represented a wide range of geographical locations and expected contamination levels and habitat types, for the purpose of targeting stations with expected low to moderate levels of acute toxicity. Stations expected to have a high degree of acute toxicity were not included in the present study because they would be less effective in eliciting different sublethal responses among the tests. The stations from Southern California were selected to include a range of proximity to contamination sources and expected sediment grain size. The San Francisco Estuary Institute Regional Monitoring Program sites had been monitored for approximately

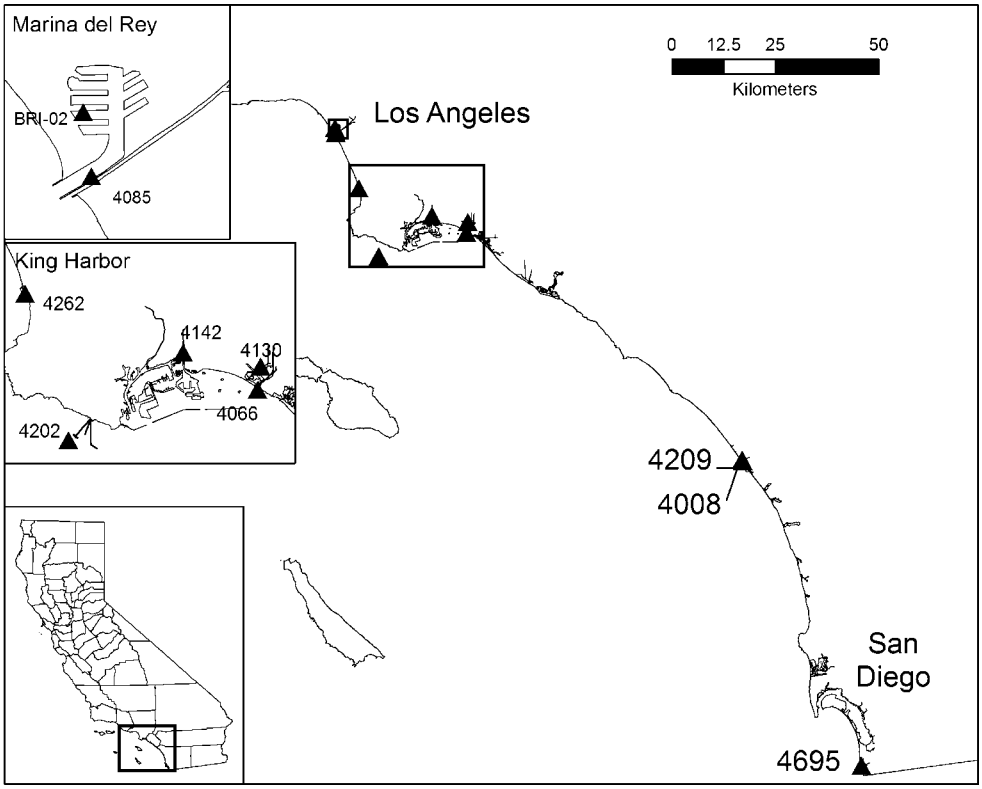


Fig. 1. Location of Southern California, USA, stations used for the sediment toxicity methods comparison study.

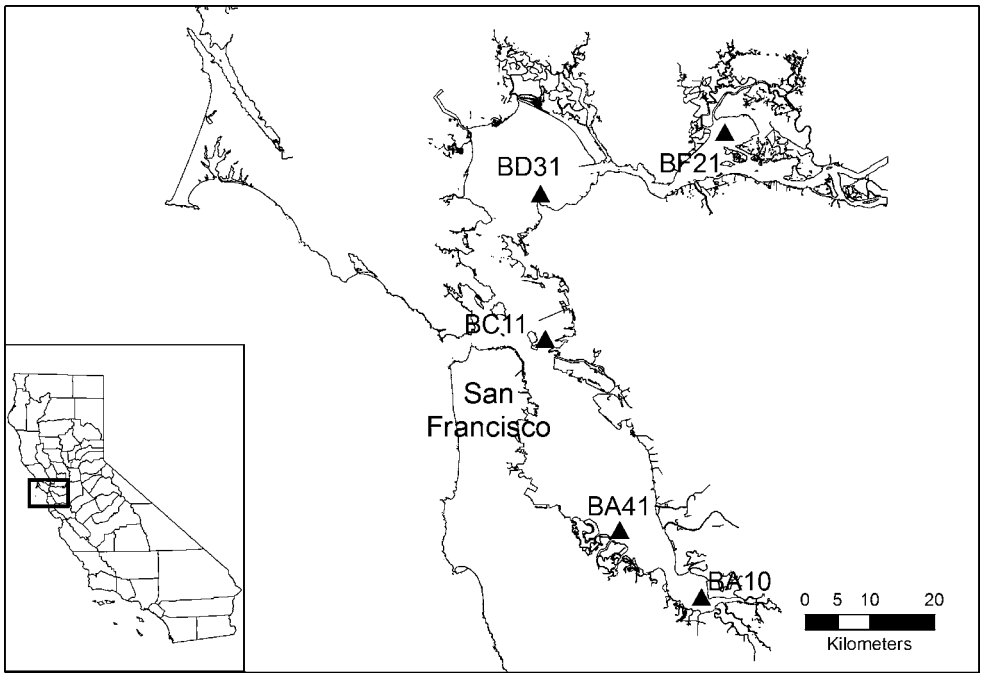


Fig. 2. Location of San Francisco Bay, California, USA, stations used for the sediment toxicity methods comparison study.

10 years and were selected for their range of acute toxicity to amphipods.

Tests on split samples were conducted by laboratories with extensive experience using the various tests. In some cases, these laboratories were the only ones with any experience conducting a given test. Each laboratory provided its own control sediment against which statistical comparisons were made. Logistical and time constraints prevented the use of a common control known to be acceptable for all of the methods. The *L. plumulosus* (both 10 and 28 d) and *N. arenaceodentata* testing was conducted at the U.S. Army Corps of Engineers, Research and Development Center, Environmental Laboratory in Vicksburg, Mississippi. The *A. tenuiremis* assays were performed at the University of South Carolina (Columbia, SC, USA). The *M. mercenaria* growth test and *C. virginica* lysosomal destabilization procedures were performed at the South Carolina Department of Natural Resources, Marine Resources Research Institute (Charleston, SC, USA). The SWI testing was conducted at the University of California, Davis, Marine Pollution Studies Laboratory (Carmel, CA, USA). Ten-day *E. estuarius* acute survival tests were performed on sediment from each station. These acute tests were conducted by multiple laboratories not included in the previously given list and preceding the other methods as part of the regional monitoring efforts. The laboratories that performed the *Eohaustorius* tests on Southern California stations participated in intercalibration procedures, which showed high agreement among laboratories with a Kendall coefficient of 1.0, indicating a perfect agreement on rankings of toxicity of blind sediment samples [20]. The laboratory testing the San Francisco Bay, California, USA, stations did not participate in this intercalibration. A summary of the characteristics of all of these test methods can be found in Bay et al. [21]. Samples were also analyzed for organic and metals chemistry, total organic carbon, grain size, and benthic infauna.

Sediments were collected in July and August 2003. A Van Veen grab was used to collect whole sediment from the surface (top 2 cm) and subcores. Surface sediment was obtained from multiple grabs at each site, composited, transferred to plastic containers, and stored at 5°C. Sediment–water interface subcores were also collected from the Van Veen grab by inserting a polycarbonate core tube into the sediment to a depth of 5 cm and capping the bottom and top of the tube. All sediment samples were transported to the Southern California Coastal Water Research Project (Costa Mesa, CA, USA) within 24 h of collection. The core samples were then transported with ice packs to the testing laboratory within 24 h. Core samples from the San Francisco Bay stations were transported directly to the testing laboratory. The subcores were received by the testing laboratory within 48 h of collection, and the SWI tests were initiated within 10 d of collection (Table 2).

The whole sediment samples were shipped to the testing laboratories in two batches: one with six of the Southern California samples, the other with the remaining four Southern California samples and all five samples from San Francisco stations. Before shipment of each batch, all of the sediment from each station was placed in a large polycarbonate bowl and homogenized with a polycarbonate spoon. Samples for each laboratory were then aliquoted into polyethylene containers and shipped overnight with sufficient quantities of ice packs to maintain temperature at 5°C. Holding time between collection and testing of the composites varied from 6 to 116 d due to laboratory scheduling conflicts (Table 2). Recommended storage time for sediment toxicity tests is less than eight weeks [22]. Data for the present study was not used if holding time exceeded 60 d.

### *Toxicity testing*

Eohaustorius estuarius *10-d survival.* Ten-day survival tests with *E. estuarius* were conducted using standard U.S. EPA testing procedures [2]. Sediment samples were press-sieved through a 2-mm mesh screen and homogenized in the laboratory before testing. Sediment was placed in 1-L glass jars to a depth of 2 cm. The samples were aerated and allowed to

Table 2. Holding times (number of days) for sediment samples tested with acute and sublethal toxicity methods[a]

| Station | *Eohaustorius* | *Leptocheirus* 10-d | *Leptocheirus* 28-d | SWI | *Mercenaria* | *Crassostrea* | *Neanthes* | *Amphiascus* |
|---|---|---|---|---|---|---|---|---|
| Batch 1 | | | | | | | | |
| 4066 | 27 | 26 | 116 | 6 | 13 | 26 | 32 | — |
| 4130 | 26 | 26 | 116 | 6 | 13 | 26 | 32 | 12 |
| 4142 | 27 | 26 | 116 | 6 | 13 | 29 | 32 | — |
| 4008 | 11 | 22 | 112 | 10 | 9 | 25 | 28 | 8 |
| 4209 | 11 | 22 | 112 | 10 | 9 | 22 | 28 | 8 |
| 4695 | 10 | 21 | 111 | 9 | 8 | 24 | 27 | 7 |
| Batch 2 | | | | | | | | |
| 4202 | 13 | 41 | 90 | 6 | 21 | 37 | 58 | 19 |
| 4262 | 12 | 40 | 89 | 5 | 20 | 36 | 57 | 18 |
| BRI-2 | 14 | 28 | 77 | 1 | 8 | 24 | 45 | 6 |
| 4085 | 7 | 28 | 77 | 1 | 8 | 24 | 45 | 6 |
| BA10 | 8 | 36 | 85 | 1 | 16 | 32 | 53 | — |
| BA41 | 11 | 39 | 88 | 4 | 19 | 35 | 56 | 17 |
| BC11 | 13 | 41 | 90 | 6 | 21 | 34 | 58 | 19 |
| BD31 | 13 | 41 | 90 | 6 | 21 | 34 | 58 | — |
| BF21 | 15 | 43 | 92 | 8 | 23 | 36 | 60 | — |

[a] SWI = sediment–water interface; — = station not tested.

equilibrate overnight before the addition of 20 adult amphipods to each of five replicates. All of the laboratories obtained amphipods from Northwestern Aquatic Sciences (Yaquina Bay, OR, USA). The exposures took place at 15°C, at a salinity of 20 g/kg with constant lighting. The amphipods were not fed and the water was not renewed during the exposures. At the end of the exposure, the sediment from each jar was sieved and the surviving animals were counted and recorded. Water quality measurements (dissolved oxygen, pH, salinity, and overlying water ammonia) were determined at day 0 and prior to test termination.

Leptocheirus plumulosus *10-d survival.* The experimental design followed guidelines set forth by the U.S. EPA [2]. Sediment was added to each of five replicate 1-L beakers to obtain a 2 cm depth. Sediment was then overlain with 20 g/kg synthetic seawater. Temperature was maintained at 25°C with constant illumination, and the beakers were aerated during the exposure. At day 0, 20 *L. plumulosus* (500- to 750-μm sieve size class) obtained from in-house cultures were gently transferred to each replicate beaker. The amphipods were not fed and the water was not renewed during the exposures. Water quality measurements (dissolved oxygen, pH, salinity, and overlying water ammonia) were determined at day 0 and prior to test termination. On day 10, the sediment in each beaker was sieved and the surviving amphipods were recovered. The number of surviving organisms was counted and recorded.

L. plumulosus *28-d survival, growth, and reproduction.* The 28-d *L. plumulosus* experiments were conducted according to the guidelines provided by the U.S. EPA [5]. Due to conflicts in the laboratory schedule and a test failure, the samples for this test method were held for a much longer period than those of the other test methods (Table 2). Due to this confounding factor, the data for the 28-d *L. plumulosus* test are not presented.

Neanthes arenaceodentata *28-d survival and growth.* The 28-d *N. arenaceodentata* experiments were conducted according to guidelines developed by the U.S. Army Corps of Engineers, Research and Development Center [23,24]. Sediment was added to 10 replicate 300-ml tall-form beakers to obtain the required depth of 2 cm. Sediment was then overlain with 30 g/kg synthetic seawater and gently aerated. Temperature was maintained at 20°C, and light cycle was set at 12:12 h light:dark. Organisms were obtained from California State University (Long Beach, CA, USA). On day 0, one *N. arenaceodentata* (≤7 d old) was gently transferred to each replicate beaker. Water quality measurements (dissolved oxygen, pH, salinity, and overlying water ammonia) were determined at day 0, prior to test termination, and in three replicates per sample weekly. Water was changed in each beaker once per week after water quality parameters were measured. Each beaker was provided 2 mg of Tetramarin® (Tetra Sales, Blacksburg, VA, USA) once per week and 2 mg of Tetramarin plus 2 mg of alfalfa once per week. On day 28, the sediment contained in each beaker was sieved and surviving worms were recovered, counted, and recorded. Surviving worms from each replicate were put on a preweighed pan and placed in a drying oven at 60°C for 24 h. The pans were then removed, allowed to cool, and weighed to obtain the individual dry weight of worms for each replicate and animal.

Amphiascus tenuiremis *14-d life cycle.* Copepod testing (*A. tenuiremis*) followed the methods of Chandler and Green [9]. A sediment reference sample was collected from Oyster Landing at North Inlet, South Carolina, USA. Stations BA41, BC11, BRI-2, 4085, 4202, and 4262 were press-sieved through a 125-μm sieve prior to testing in order to facilitate discrimination of assayed copepods from muddy particulates at the conclusion of exposure. A larger presieve size was used for some of the larger-grained stations in order to obtain a sufficient volume of sediment for testing. Copepod discrimination in these sandy, less than 250-μm coarse-grained samples was not difficult because sand grains are translucent, whereas muddy sediments are not. Sediment samples 4008 and 4695 were screened with a 250-μm sieve, while 4209 and 4130 were sieved through 212- and 180-μm sieves, respectively. Sediment samples 4066 and 4142 were too sandy to pass through a 250-μm sieve and could not be tested with the copepod method. A total of 10 stations were tested with *Amphiascus*. Teflon® 50-ml Erlenmeyer flasks with mesh-covered outflow holes were filled with 0.45 μm of filtered, aerated seawater and 12 ml of sediment per replicate ($n$ = 4 replicates per station). Twenty-five adult nongravid female and adult male copepods were then counted into each quadruplicate test chamber and incubated with 3-d

feedings at 20°C under continuous dripping flow for 14 d. Water quality parameters (dissolved oxygen, pH, and salinity) were measured every third day. Overlying water ammonia was measured once at the end of each exposure period. After 14 d, sediments were sieved and examined to determine copepod survival. Live copepods were stained with Rose Bengal and preserved in 5% borate-buffered formalin. Copepod life stages and clutch sizes were enumerated. Two chronic reproductive endpoints were calculated: the number of copepodites produced in 14 d and the realized offspring production (i.e., output of all new animals normalized to the number of females surviving at the end of the test).

Mercenaria mercenaria *7-day growth.* The clam tests measured growth during a 7-d exposure to whole sediment [10,25]. Sediment samples were homogenized and pressed through a 500-mm sieve, and 50-ml aliquots were placed into four replicate 250-ml beakers and overlain with clean 25 g/kg seawater. Sediment pore-water chemistry parameters (salinity, pH, and total ammonia–nitrogen) were measured for each sample at the start of the exposures. The replicates were gently aerated for the duration of the experiment, and the assays were conducted at room temperature (22–25°C) with a 16:8-h light cycle. The same water chemistry parameters were measured in the overlying water at the beginning and end of each assay. Clams (25 per replicate) were presieved and preweighed to ensure that they were of a similar size range. The preassay wet and dry weights were also used for final growth rate estimates. Each replicate was fed on the first, third, and sixth days of the assay (50:50 mix of *Isochrysis galbana* and *Chaetoceros gracilis;* $20 \times 10^6$ cells per replicate). The overlying water was not renewed during the exposure. At the end of the exposures, clams were sieved from the sediments and placed in fresh 25 g/kg seawater for approximately 2 h to depurate. Dead clams were removed and percentage mortalities were calculated. The surviving clams were rinsed with distilled water to remove excess salts, then dried for 48 h (70°C). Each clam replicate was counted, and the final dry weight per clam was determined. Initial dry weights were subtracted from final dry weights, and the results were expressed as growth rates (micrograms per clam per day).

Crassostrea virginica *4-d lysosomal destabilization.* Sediment samples were homogenized, 100-ml aliquots were placed into three replicate 1-L beakers, and each was overlain with clean 25 g/kg seawater. The beakers were allowed to settle, and then three clean-scrubbed oysters were gently added to each replicate. Oysters ($5.3 \pm 0.7$ cm) were collected from control sites and acclimated to laboratory conditions for at least 24 h. The replicates were gently aerated for the duration of the experiment, and the assays were conducted at room temperature (22–25°C) with a 16:8-h light cycle. Each replicate was fed on the first and third days of the assay (*Isochrysis* paste mixed into filtered seawater; $70 \times 10^6$ cells per replicate). The overlying water was not renewed during the exposure. Water quality parameters for both pore and overlying waters were measured in the same manner used in the *Mercenaria* testing. The lysosomal destabilization assay was conducted following methods described in Ringwood et al. [11,26]. Briefly, digestive gland tissue from individual oysters was disaggregated in a $Ca^{2+}$, $Mg^{2+}$ free saline and trypsin solution, then added to an equal volume of neutral red solution. Using a light microscope, a minimum of 50 cells were scored as stable (neutral red retention in the lysosomes) or destabilized (neutral red leaking into the cytoplasm). Data were expressed as the percentage of cells with destabilized lysosomes per oyster.

Mytilus galloprovincialis *2-d embryo development (SWI).* Exposure procedures followed those detailed by Anderson et al. [8]. One day prior to the start of the test, 300 ml of clean seawater (1 μm filtered, ~34 g/kg) was added over the sediment to each of five replicate core tubes. Samples were then aerated overnight to equilibrate. On test day 0, water quality samples were collected from the core tubes, and tubes containing a 25-μm screen were placed on the sediment surface. The screen was approximately 1 cm above the sediment. Mussel embryos were unavailable to test stations 4008, 4209, and 4695, so sea urchin embryos were used instead. Embryos were prepared following U.S. EPA protocols [27] and added to the screen tubes. Mussels were exposed for 48 h and sea urchins for 96 h. Exposures were carried out at 15°C with gentle aeration. Water quality parameters of dissolved oxygen, total ammonia, pH, and salinity were measured at the beginning and end of exposure periods. Temperature was measured continuously. The exposures were terminated by removing the screen tube, rinsing the embryos into a vial, and adding formalin to fix and preserve embryos. The samples were then examined microscopically for normal embryo development. Data were expressed as percentage normal-alive. This endpoint was calculated by dividing the number of normal embryos by the initial number of embryos inoculated into the chambers.

### *Chemical analysis*

Sediment samples were analyzed for a suite of parameters that included metals, organics, grain size, and total organic carbon. Analyses were conducted by a variety of laboratories participating in the regional monitoring programs and used standardized U.S. EPA–recommended methods [19,28]. The laboratories had achieved acceptable comparability during preproject intercalibration exercises, and the data were subjected to rigorous postsurvey review. Quality assurance samples were included in each sample batch and included method blanks, duplicates, matrix spikes, and a certified reference material. Sediment particle size was measured by light-scattering technology using either a Coulter LS230 (Beckman Coulter, Fullerton, CA, USA) or a Horiba LA900 (Horiba Instruments, Irvine, CA, USA) instrument. Sediment samples analyzed for all metal analytes except mercury were digested in strong acid according to the procedures described in U.S. EPA method 3050B (www.epa.gov/SW-846/pdfs/3050b.pdf). Metals were quantified using inductively coupled plasma mass spectrometry, inductively coupled plasma emission spectroscopy, flame atomic absorption, or graphite furnace atomic absorption. Mercury was analyzed using cold vapor atomic absorption spectroscopy. Samples for polycyclic aromatic hydrocarbons (PAHs) and chlorinated pesticide analysis were solvent extracted using accelerated solvent extraction, a Soxhlet extractor, or a roller table. The extracts obtained were subjected to each laboratory's own cleanup procedures and analyzed by a gas chromatographic method (e.g., dual-column gas chromatograph–electron capture device or gas chromatograph–mass spectrometer in the selected ion monitoring mode).

### *Benthic community analysis*

A single, separate grab sample was taken for benthic community analysis at all stations. The contents of the grab were washed through a 1.0-mm screen, and all of the retained animals identified to the lowest possible taxon. Different benthic

indices were used to assess community status for the San Francisco Bay and Southern California stations because of habitat differences between the two regions that affect species composition. The benthic community condition of the Southern California stations was assessed using the Benthic Response Index (BRI) [29]. The BRI is the abundance-weighted average pollution tolerance score of organisms occurring in a sample. The Index of Biotic Integrity (IBI) was used to determine benthic community condition for the San Francisco Bay stations [30]. The IBI uses a multimetric index to discriminate between impacted and reference areas. Both of these indices have been calibrated to the individual habitats that they represent and therefore are expected to be highly indicative of conditions at those locations. These indices have been compared to individual metrics and other multimetric indices and have been found to be most representative of their individual habitats (J. Ranasinghe et al., Southern California Coastal Water Research Project, unpublished data). Both the BRI and the IBI use a final station categorization based on a scale of one through four. A rating of one indicates a community similar to that expected under reference conditions, while a rating of four indicates a highly degraded benthic community structure.

*Data analysis*

Statistical significance was tested using Student's *t* test ($p \leq 0.05$) assuming unequal variance [31]. For sublethal methods having more than one endpoint, if either or both endpoints were significantly different from control, the station was designated as toxic.

The mean effects range median quotient (ERMq) [32] was calculated for each station to integrate a subset of the analyzed chemicals into a value that is predictive of toxic effects. The effects range median for DDT was not used in calculations because it has been found to be unreliable [33]. Relationships among sediment chemistry parameters or benthic community condition and toxicity response were analyzed using a nonparametric Spearman rank correlation.

## RESULTS

The experimental batches for all toxicity data presented passed test control acceptability criteria, with the exception of one SWI batch with *Mytilus*. That batch contained the only sample with a significant toxic response for the SWI test but also had a low control normal-alive percentage. Because the difference between the control and sample response was very large, the data have been included.

Water quality measurements made during testing indicated that the values were within acceptable range for the majority of sample–test combinations. For the *Mercenaria* test, station 4130 exhibited elevated pore-water ammonia (37.5 mg/L of total ammonia–nitrogen). Although the tolerance of *Mercenaria* to ammonia is not known, there is correlative evidence that the ammonia level in the sample may have been the cause of toxicity [10]. For the SWI test, station BC11 had an overlying water ammonia concentration of 0.145 mg/L of un-ionized ammonia, which is very near the 50% effect concentration (~0.17 mg/L, D. Greenstein, unpublished data).

*Comparisons among sublethal tests*

A wide range in the percentage of stations identified as toxic by sublethal methods was observed (Fig. 3). The highest percentage was for the copepod, *Amphiascus*, which found 9 of the 10 stations (90%) tested to be toxic, followed by *Neanthes* with 8 out of 15 stations (53%). Of the nine stations found to be toxic by *Amphiascus*, three were not found to be toxic by *Neanthes* (Table 3). The percentage of stations identified as toxic was much lower for the remaining test methods, ranging from 13% for *Mercenaria* and *Crassostrea* to 7% for the SWI method.

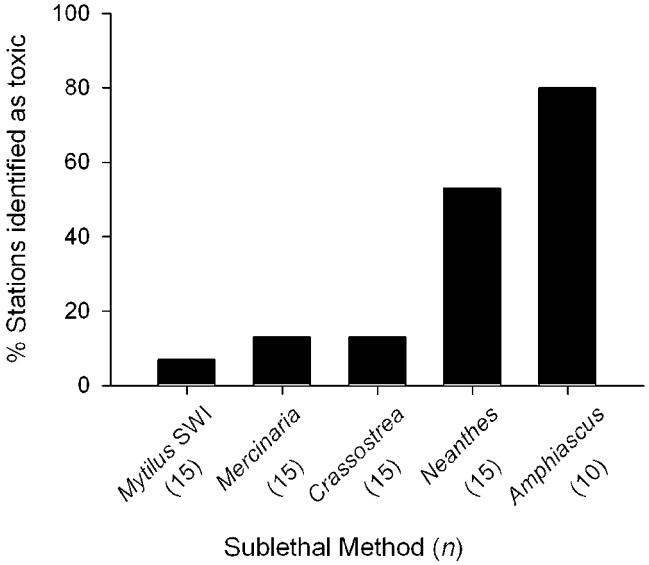


Fig. 3. Percentage of stations that each sublethal method identified as being toxic. Number of samples tested is in parentheses. SWI = sediment–water interface.

*Comparisons between acute tests*

The *Eohaustorius* method was the less sensitive of the two amphipod acute protocols, identifying 4 out of 15 stations tested as toxic (Table 3). Overall, the *Eohaustorius* method was near the midpoint of sensitivity relative to the sublethal tests. The *Leptocheirus* 10-d method identified 9 of the 15 sites tested as toxic and was more sensitive than all but one of the sublethal methods.

*Comparisons between sublethal and acute tests*

The *Neanthes* and *Amphiascus* sublethal tests detected toxicity at 27 and 70% of the stations where the *Eohaustorius* acute test did not; while in no cases did *Eohaustorius* demonstrate toxicity where either of the other two tests did not (Table 3). Alternatively, the *Eohaustorius* test identified a higher percentage of stations (27%) as toxic than did the SWI (7%), *Mercenaria* (13%), and *Crassostrea* (13%) tests. The *Eohaustorius* test identified toxicity in 27% of the samples that these other three tests classified as nontoxic.

The *Leptocheirus* 10-d test found a higher percentage (60%) of toxic stations than all of the sublethal methods except for the *Amphiascus* test (Table 3). The *Amphiascus* test found four stations (40%) to be toxic that were not identified as toxic by the *Leptocheirus* acute test. Concordance was observed between the *Leptocheirus* 10-d test and the *Amphiascus* test for the remaining stations, with both tests identifying five stations as toxic and one not. The *Neanthes* test found four stations to be toxic that were not identified by the *Leptocheirus* acute test. However, five stations were identified as toxic in the *Leptocheirus* acute test but not identified as toxic in the *Neanthes* test. In comparison with the *Mercenaria*, *Crassostrea*, and SWI sublethal tests, a high percentage of stations

Table 3. Comparative ability of acute and sublethal sediment toxicity test methods to detect toxicity in stations from Southern California and San Francisco Bay, USA[ab]

| | Acute methods | | Sublethal methods | | | | |
|---|---|---|---|---|---|---|---|
| Station | *Eohaustorius* | *Leptocheirus* | SWI | *Mercenaria* | *Crassostrea* | *Neanthes* | *Amphiascus* |
| 4202 | N | Y | N | N | N | N | Y |
| BRI-2 | Y | Y | N | N | N | Y | Y |
| BA41 | Y | N | N | N | N | Y | Y |
| BA10 | Y | N | N | N | N | Y | — |
| 4066 | N | Y | N | N | Y | N | — |
| BC11 | N | Y | Y | N | Y | Y | Y |
| 4142 | N | Y | N | N | N | N | — |
| 4262 | N | N | N | N | N | Y | Y |
| 4130 | N | Y | N | Y | N | Y | Y |
| 4085 | N | Y | N | N | N | N | Y |
| BF21 | Y | Y | N | N | N | Y | — |
| BD31 | N | Y | N | N | N | N | — |
| 4008 | N | N | N | Y | N | Y | Y |
| 4209 | N | N | N | N | N | N | Y |
| 4695 | N | N | N | N | N | N | N |
| % Sublethal indicating toxic versus *Eohaustorius* indicating not toxic | | | | | | | |
| | — | — | 7 | 13 | 13 | 27 | 70 |
| % *Eohaustorius* indicating toxic versus sublethal indicating not toxic | | | | | | | |
| | — | — | 27 | 27 | 27 | 0 | 0 |
| % Agree toxic | — | — | 0 | 0 | 0 | 27 | 20 |
| % Agree not toxic | — | — | 67 | 60 | 60 | 47 | 10 |
| % Sublethal indicating toxic versus *Leptocheirus* indicating not toxic | | | | | | | |
| | — | — | 0 | 7 | 0 | 27 | 40 |
| % *Leptocheirus* indicating toxic versus sublethal indicating not toxic | | | | | | | |
| | — | — | 53 | 53 | 47 | 33 | 0 |
| % Agree toxic | — | — | 7 | 7 | 13 | 27 | 50 |
| % Agree not toxic | — | — | 40 | 33 | 40 | 13 | 10 |

[a] Numeric values are expressed as the percentage of stations tested.
[b] SWI = sediment–water interface; N = station not identified as toxic; Y = station identified as toxic; — = station or comparison not tested.

(60% or more) identified as nontoxic were identified as toxic using the *Leptocheirus* acute test.

Combining the data from an acute test and a sublethal test or the two acute tests provided more information regarding toxicity than conducting just one test of either type. The greatest sensitivities (most toxic stations detected by one or both tests) were found with the combinations of *Leptocheirus* 10-d and *Amphiascus* (9 of 10) or *Neanthes* (13 of 15) methods (Table 3). The combination of the two acute tests (11 of 15) was found to be nearly as sensitive as these acute–sublethal combinations.

### *Chemistry*

Sediment physical parameters were very wide ranging, with grain sizes that were nearly 100% fines (silt + clay) to 100% sand (Table 4). Total organic carbon values ranged from 0.02 to 2.93%.

Sediment contaminant concentrations also were variable among stations (Table 4). Three stations had elevated chemistry compared to the other stations. Station 4202, on the Palos Verdes shelf, had a very high concentration of total DDTs. Station BRI-2, in Marina Del Rey, had low concentrations of organic contaminants but substantial concentrations of copper, lead, and zinc. Station 4085 contained intermediate concentrations of several metals and organics. Based on the mean ERMq calculations, all of the stations tested fell into what would be considered the low to moderate range of contaminant concentrations, with all mean quotients less than 0.7 (Table 4). Five samples had mean ERMq values below 0.1, a level not expected to be toxic. The mean quotients for the remaining stations fell between 0.11 and 1.0, a range that has been found to be toxic in approximately 50% of samples [32].

*Eohaustorius* survival, both *Amphiascus* endpoints, and *Neanthes* growth all had statistically significant ($p \leq 0.05$) Spearman correlations with sediment chemistry (Table 5). Correlations with various metals, but none with organics, were observed. All of the significant correlations were negative, indicating that as the concentration increased the endpoint decreased (e.g., decreased survival or growth). All toxicity test methods that correlated with chemistry also had significant correlations with sediment grain size. The chemical constituents that correlated with toxicity also correlated with the grain size parameters.

### *Benthic community*

A range of benthic community condition was observed among the stations. Most stations were classified as being in reference condition (8 of 15) or having an intermediate level of disturbance (5 of 15 stations at level 2 or 3). Two stations (4066 and 4142) had the most degraded (level 4) designations (Table 6), which indicated severe effects to the benthic community. The variations in benthic community condition did not correspond with the sediment contamination gradient. The average mean ERMq of all stations in each benthic condition category was lowest (0.02) for level 4 stations and highest (0.31) for the level 2 stations (Table 6).

Little correspondence was observed between changes in benthic community condition and toxicity for most of the test

Table 4. Chemical characteristics of Southern California and San Francisco Bay, USA, sediment samples on which toxicity tests were performed[a]

| Station | Arsenic (mg/kg) | Cadmium (mg/kg) | Chromium (mg/kg) | Copper (mg/kg) | Lead (mg/kg) | Mercury (mg/kg) | Nickel (mg/kg) | Silver (mg/kg) | Tin (mg/kg) | Zinc (mg/kg) | TOC % | Sand % | Silt % | Clay % | ΣPAHs µg/kg | ΣDDTs µg/kg | ΣPCBs µg/kg | Mean ERMq[b] | ERMq ranking |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4202 | 8.5 | 6.6 | 136 | 5 | 30.0 | 0.46 | 29.0 | 1.9 | NA | 180 | 2.06 | 39 | 50 | 11 | 678 | 2,301.3 | 193.9 | 0.68 | 1 |
| BRI-02 | 13.0 | 0.3 | 94 | 362 | 113.0 | 0.98 | 41.6 | 2.0 | 6.3 | 382 | 1.99 | 8 | 74 | 18 | 76 | 2.2 | ND | 0.26 | 4 |
| BA41 | 4.5 | 0.2 | NA | 30 | 17.4 | 0.34 | 58.2 | 0.1 | NA | 90 | 1.09 | 20 | 22 | 49 | 1,923 | 0.2 | 2.5 | 0.14 | 7.5 |
| BA10 | 4.1 | 0.1 | NA | 24 | 11.3 | 0.24 | 46.9 | 0.2 | NA | 70 | 2.34 | 44 | 15 | 36 | 724 | 0.6 | 2.3 | 0.10 | 10 |
| 4066 | 1.0 | 0.1 | 7 | 7 | 4.7 | 0.10 | 4.0 | 0.6 | 0.4 | 22 | 0.02 | 100 | 0 | 0 | 52 | 1.0 | ND | 0.02 | 12 |
| BC11 | 4.0 | 0.3 | NA | 39 | 29.7 | 0.23 | 65.9 | 0.1 | NA | 108 | 1.80 | 22 | 22 | 48 | 740 | 0.6 | 111.3 | 0.34 | 2 |
| 4142 | 1.0 | 0.2 | 5 | 6 | 4.3 | 0.06 | 4.1 | 0.3 | 0.5 | 49 | 0.27 | 62 | NA | NA | 73 | ND | ND | 0.02 | 13 |
| 4262 | 4.0 | 0.6 | 46 | 3 | 38.9 | 0.23 | 21.2 | 0.7 | NA | 92 | 1.50 | 56 | 36 | 8 | 625 | 49.8 | 66.0 | 0.29 | 3 |
| 4130 | 7.0 | 0.8 | 49 | 87 | 61.6 | 0.40 | 25.8 | 0.8 | 3.8 | 248 | 2.04 | 44 | 46 | 10 | 1,206 | 9.9 | 15.7 | 0.17 | 6 |
| 4085 | 11.6 | 1.7 | 78 | 101 | 130.0 | 0.41 | 33.1 | 2.9 | 6.3 | 315 | 2.93 | 30 | 57 | 13 | 578 | 14.6 | 22.6 | 0.24 | 5 |
| BF21 | 8.5 | 0.2 | NA | 53 | 16.4 | 0.27 | 88.6 | 0.2 | NA | 126 | 1.37 | 1 | 39 | 60 | 582 | 0.8 | 0.8 | 0.14 | 7.5 |
| BD31 | 7.5 | 0.2 | NA | 51 | 17.8 | 0.24 | 87.8 | ND | NA | 126 | 1.33 | 9 | 32 | 59 | 450 | 1.4 | 0.8 | 0.14 | 9 |
| 4008 | 2.5 | 0.1 | 34 | 14 | 4.7 | 0.08 | 10.2 | 0.7 | 1.6 | 48 | 0.67 | 54 | 40 | 6 | 12 | 1.3 | ND | 0.04 | 11 |
| 4209 | 1.5 | ND | 10 | 3 | 1.4 | 0.02 | 3.0 | 0.2 | 0.5 | 14 | 0.04 | 98 | 2 | ND | ND | ND | ND | 0.01 | 14 |
| 4695 | 1.1 | ND | 5 | 1 | 1.2 | 0.02 | 0.9 | 0.2 | 0.3 | 6 | ND | 100 | ND | ND | ND | ND | ND | 0.01 | 15 |

[a] TOC = total organic carbon; ΣPAHs = total polyaromatic aromatic hydrocarbons; ΣPCBs = total polychlorinated biphenyls; ERMq = effects range median quotient; NA = not analyzed; ND = not detected.

[b] DDT concentrations are not included in the ERMq calculation.

methods. *Leptocheirus* 10-d survival was the only test to consistently detect toxicity at the level 4 stations (Table 6). Most of the stations that did show toxicity were in the reference or level 2 categories for benthic community condition. Three of the test methods (*Eohaustorius*, *Leptocheirus*, and *Amphiascus*) showed an increased incidence of toxicity among all impacted stations (levels 2–4 combined) compared to stations classified as having a reference benthic condition. Correlations of BRI values for the Southern California stations showed that only the *Leptocheirus* 10-d test method had a significant correlation with benthic community condition (Table 6). The correlation coefficients were negative for all but the *Crassostrea* lysosome method, indicating that as the BRI value increased the toxicity endpoint value decreased (i.e., survival or growth decreased).

## DISCUSSION

Sediment toxicity testing is an important tool in the assessment of sediment condition. The selection of species used for testing can affect the outcome of the assessment [34]. Factors that should be considered in species selection include relative sensitivity, contact with sediment, availability of organisms, ecological importance, geographical distribution, sediment physicochemical tolerance, and peer review of methods [21,35]. For some of the newer sublethal marine sediment toxicity tests, much of the information regarding these criteria was previously unavailable. The present study was able to provide information regarding relative sensitivity, response compared to the benthos, and response compared to sediment physicochemical properties.

The sensitivity of the toxicity test methods were variable within the two broad categories of tests evaluated, indicating that general classifications of the tests as either acute or sublethal do not reliably represent their relative sensitivity. For example, the most sensitive test in the present study was the sublethal *Amphiascus* life cycle method, but the acute *Leptocheirus* survival test was more sensitive than any of the other sublethal tests compared. This variation in sensitivity between sublethal and acute tests is consistent with other studies, suggesting that the relative sensitivity of acute and sublethal tests to whole sediment samples varies according to the combination of tests and sample types evaluated. Comparative studies using the *Leptocheirus* 28-d test have shown that the sublethal endpoints from this test are not consistently more sensitive than acute amphipod tests to field and spiked sediments [15]. Another study found that the acute *Ampelisca* test was more sensitive than the *Leptocheirus* 28-d test, which was more sensitive than the *Neanthes* 28-d test [36]. In contrast to the results of the present study, the *Mercenaria* test was found to be more sensitive than the acute *Ampelisca* survival test when sediment samples from the Carolinian Province (USA) were tested [37].

The present study's finding that *Amphiascus* was the most sensitive method overall is consistent with other studies indicating the high sensitivity of this life cycle test. Tests using sediments from Biscayne Bay, Florida, by Long et al. [38] found a greater incidence of toxicity with the *Amphiascus* life cycle method (73%) than with the *Ampelisca* 10-d survival test (7%). The high sensitivity, chronic exposure, and multiple endpoints that are characteristic of this test are desirable qualities; however, more investigation is needed to determine whether the high level of response associated with this test of Southern California samples, having low contaminant con-

Table 5. Spearman rank correlations on selected sediment parameters and toxicity endpoints[ab]

| | | | | | | | | *Amphiascus* | |
|---|---|---|---|---|---|---|---|---|---|
| | *Eohaustorius* survival | *Leptocheirus* survival | *Mytilus* SWI | *Mercenaria* growth | *Crassostrea* Lysosome | *Neanthes* survival | *Neanthes* growth | No. of copepodites | Realized offspring |
| Arsenic | −0.604 | −0.239 | 0.274 | −0.145 | −0.080 | 0.136 | −0.542 | −0.585 | −0.806 |
| Cadmium | −0.155 | −0.401 | 0.264 | −0.295 | −0.099 | 0.132 | −0.264 | −0.206 | −0.488 |
| Copper | −0.786 | −0.375 | 0.196 | −0.354 | −0.059 | −0.051 | −0.565 | −0.829 | −0.952 |
| Lead | −0.366 | −0.350 | 0.337 | −0.306 | −0.025 | 0.233 | −0.390 | −0.482 | −0.842 |
| Mercury | −0.596 | −0.406 | 0.476 | −0.143 | −0.093 | 0.059 | −0.514 | −0.572 | −0.742 |
| Nickel | −0.836 | −0.289 | −0.386 | −0.382 | 0.136 | −0.022 | −0.594 | −0.866 | −0.709 |
| Silver | 0.220 | −0.089 | 0.533 | 0.012 | −0.225 | 0.188 | −0.080 | −0.043 | −0.455 |
| Zinc | −0.549 | −0.434 | 0.250 | −0.301 | −0.085 | 0.138 | −0.476 | −0.567 | −0.842 |
| TOC (%) | −0.440 | −0.250 | 0.119 | −0.268 | 0.070 | −0.012 | −0.424 | −0.390 | −0.661 |
| Sand (%) | 0.820 | 0.237 | 0.091 | 0.349 | 0.081 | −0.069 | 0.653 | 0.933 | 0.794 |
| Clay (%) | −0.823 | −0.229 | −0.320 | −0.326 | 0.139 | −0.032 | −0.596 | −0.881 | −0.717 |
| ΣPAHs(μg/kg) | −0.491 | −0.259 | 0.032 | −0.354 | 0.222 | −0.314 | −0.490 | −0.520 | −0.486 |
| ΣDDTs(μg/kg) | −0.013 | −0.333 | 0.123 | −0.264 | 0.014 | 0.382 | −0.320 | −0.086 | −0.365 |
| ΣPCBs(μg/kg) | −0.124 | −0.295 | −0.078 | −0.192 | 0.339 | 0.062 | −0.211 | −0.066 | −0.125 |
| ERMq[c] | −0.288 | −0.268 | 0.018 | −0.402 | 0.124 | 0.306 | −0.449 | −0.329 | −0.370 |

[a] Underlined values are significant ($p \leq 0.05$).
[b] SWI = sediment–water interface; TOC = total organic carbon; ΣPAHs = total polyaromatic aromatic hydrocarbons; ΣPCBs = total polychlorinated biphenyls; ERMq = effects range median quotient.
[c] Does not include DDT data.

centrations and reference benthic community condition, reflect chemical toxicity or the effects of potentially confounding factors such as ammonia or organic carbon. The possible influence of grain size is discussed later. It should also be noted that the *Amphiascus* test was only used on 10 of the 15 stations in the present study.

Several factors may have accounted for the variation in sensitivity among methods observed in the present study, including mode of exposure, species-specific sensitivity to contaminants, and the influence of confounding factors. The mode of exposure varied greatly among tests; those tests with the longest exposure duration and the most direct contact with the sediment (i.e., *Amphiascus* and *Neanthes*) tended to be most sensitive. For the SWI method, which was least sensitive, the organisms are in the water column directly above the sediment and exposed for a relatively short period of time to only those contaminants diffusing into the overlying water. These differences in exposure method and sample response can be used advantageously to investigate the mode of contaminant exposure or identify the cause of toxicity.

Differences in contaminant sensitivity among test methods have been documented for some of the test species and may have influenced the results of the present study. Several studies have compared the *Leptocheirus* 10- and 28-d tests and the *Neanthes* 28-d test to various chemicals and found varying patterns of response. The *Neanthes* test was more sensitive than *Leptocheirus* to sediments contaminated with metals or the explosive trinitrotoluene; both of these sublethal tests were more sensitive than the acute *Leptocheirus* test to polychlorinated biphenyls, yet the *Leptocheirus* acute method was more sensitive to PAH-contaminated sediments than was *Neanthes* [39]. Comparisons among acute tests using *Ampelisca*, *Eohaustorius*, and *Rhepoxynius* showed that *Eohaustorius* was the most sensitive to DDT while *Ampelisca* and *Rhepoxynius* were more sensitive to cadmium [40]. Sediment contaminant mixtures varied among the stations in the present study; with

Table 6. Incidence of toxicity within benthic index categories and Spearman rank correlation values for toxicity test endpoints[ab]

| | Benthic index category | | | | | |
|---|---|---|---|---|---|---|
| Test | Ref[c] | Level 2[d] | Level 3[e] | Level 4[f] | Levels 2–4[g] | $r$[h] |
| No. of stations | 8 | 4 | 1 | 2 | 7 | |
| Mean ERMq | 0.15 | 0.31 | 0.10 | 0.02 | 0.20 | |
| | | | Incidence of toxicity (%) | | | |
| *Eohaustorius* 10-d survival | 12 | 50 | 100 | 0 | 42 | −0.52 |
| *Leptocheirus* 10-d survival | 50 | 75 | 0 | 100 | 71 | −0.64 |
| *Mytilus* SWI | 12 | 0 | 0 | 0 | 0 | −0.27 |
| *Mercenaria* growth | 12 | 25 | 0 | 0 | 14 | −0.20 |
| *Crassostrea* lysosome | 12 | 0 | 0 | 50 | 14 | 0.04 |
| Neanthes growth | 50 | 75 | 100 | 0 | 57 | −0.12 |
| *Amphiascus* No. of copepodites | 83 | 100 | NA | NA | 100 | −0.44 |

[a] Underlined values are statistically significant ($p \leq 0.05$).
[b] ERMq = effects range median quotient; SWI = sediment–water interface; NA = not analyzed.
[c] Reference condition stations: BC11, 4262, 4085, BF21, BD31, 4008, 4209, 4695.
[d] Level 2 (low disturbance): 4202, BRI-02, BA41, 4130.
[e] Level 3 (moderate disturbance): BA10.
[f] Level 4 (high disturbance): 4066, 4142.
[g] Combination of levels 2–4. Incidence of toxicity is the percentage of stations within these levels found to be toxic.
[h] Correlation calculated using Southern California data only.

Table 7. Holding times (days), incidence of toxicity, and test method rankings (high frequency of toxic hits to low) for the two batches of samples analyzed for sediment toxicity methods comparisons[a]

| | Batch 1 (6 stations) | | | Batch 2 (9 stations) | | |
|---|---|---|---|---|---|---|
| Test method | Holding time range | % Stations toxic | Rank | Holding time range | % Stations toxic | Rank |
| *Amphiascus* | 7–12 | 75 | 1 | 6–19 | 100 | 1 |
| *Leptocheirus* 10-d | 21–26 | 50 | 2 | 28–43 | 67 | 2.5 |
| *Neanthes* | 27–32 | 33 | 3.5 | 45–60 | 67 | 2.5 |
| *Mercenaria* | 8–13 | 33 | 3.5 | 8–23 | 0 | 5 |
| *Crassostrea* | 22–26 | 17 | 5 | 24–37 | 11 | 4 |
| *Mytilus* SWI | 6–9 | 0 | — | 1–8 | 11 | — |
| *Eohaustorius* | 10–27 | 0 | — | 7–15 | 44 | — |

[a] The sediment–water interface (SWI) and *Eohaustorius* methods were not tested in these batches and therefore are not ranked.

differences of up to two orders of magnitude in metals, polychlorinated biphenyls, and PAH concentrations and up to three orders of magnitude in DDT. These differences may have contributed to the variation in response among the test methods.

Variations in holding time or sediment handling that occurred among the laboratories are potential confounding factors that may have altered the toxicity of the samples through changes in bioavailability or chemical composition. Guidance for sediment holding times varies between two and eight weeks [22]. All experiments for which the data was used in the present study had holding times of 60 d or less. Dillon et al. [41] found that the toxicity of sediment gradually increased over time but then appeared to stabilize. The nature and magnitude of such effects was not determined in the present study, but an analysis of the data indicates that the patterns of relative sensitivity observed among the test methods were independent of holding time. For example, holding times were shortest and similar for the SWI and *Amphiascus* methods, yet these two tests had very different patterns of response to the samples (Table 7). The patterns of relative response among the tests were also similar for the two batches of whole sediment tested (e.g., *Amphiascus* most sensitive and *Mercenaria* and *Crassostrea* usually least sensitive). These facts suggest that variations in holding time or sediment handling among the tests and batches were not likely major confounding factors.

Significant correlations with chemistry concentrations were found in the present study for the *Eohaustorius* survival, *Amphiascus* reproduction, and *Neanthes* growth tests. Relationships have been similarly documented in many other studies for a variety of test organisms and form the basis for empirical sediment quality guidelines [33,42]. Significant correlations also were found with grain size for each test. The chemistry values also correlated with grain size, and many of the chemical constituents correlated with one another. These intercorrelations make determining whether toxicity is associated with chemistry or the confounding factor of grain size a difficult matter. Grain size is not known to be a confounding factor for *Eohaustorius* [2]. Grain size should not have been an issue for *Amphiascus* because all samples were sieved to remove large particles and optimize the sediments for the animals. Differences in sediment organic carbon linked to grain size can influence the nutritional quality of test sediments to *Amphiascus* since it is a general deposit-feeding copepod that ingests fine detritus and associated bacteria (G.T. Chandler, unpublished data). Copepods were fed algae to slight excess during these bioassays to minimize nutrition-related differences in response. *Neanthes* have been tested in grain sizes ranging from 5 to 100% sand with no effects on either survival or growth [43]. These data indicate that an association between sediment contamination and toxicity, rather than a grain size effect, is likely for these three methods in the present study.

The lack of correlations with sediment chemistry for some of the test methods may have several causes. Little toxicity was observed for many of the tests, making the detection of correlations difficult. In addition, no measure of bioavailability of chemical constituents was made for the sediments, adding uncertainty regarding the actual chemical dose received by the test animals. Sediment chemistry analyses do not quantify all possible toxicants, so it is possible that unmeasured chemical constituents or interactions among compounds may have caused the observed toxicity. Another potential source of uncertainty is toxicity from confounding factors such as ammonia or sulfides. While the sensitivity of some of the test methods to these factors is uncertain, water quality data from the tests show that dissolved ammonia concentrations were low and below concentrations of concern for most of the samples, indicating that these factors probably did not have a significant influence on the results.

A strong relationship between toxicity results and benthic community condition was not found in the present study, suggesting that these indicators were responding to different aspects of sediment quality. The lack of replication of benthic community samples may account for some of the poor correspondence. Other studies have reported similar results. Analyses of Chesapeake Bay (USA) sediment toxicity using the *Leptocheirus* 10- and 28-d tests found a similar lack of correspondence with benthic community response [44]. A statistically significant correlation between *Eohaustorius* mortality and benthic community impact was found for Southern California embayment sediments, but the relationship accounted for only 10% of the variation in community condition [29]. Toxicity tests differ from the in situ benthic environment in many aspects, such as the exposure duration, species type, and laboratory handling of the sediment. In depositional areas where contaminant input has abated, benthic communities may thrive while the sediment just below may be toxic in a laboratory test. These factors can affect the contaminant bioavailability or the sensitivity of the response and may account for the relatively high frequency of toxicity detected in samples containing an unimpacted benthic community. It is not possible for toxicity tests to replicate environmental exposure conditions or provide a substitute for the assessment of biological effects on resident organisms; these tests are intended to provide a measure of potential contaminant effects that is complementary to chemical and biological measures.

The effects of noncontaminant factors on the benthic com-

munity analyses may have also influenced the correlation analyses with toxicity. Changes in benthic community condition did not correspond with increasing contamination levels, as represented by the mean ERMq (Table 6). The stations having the most disturbed benthos had the lowest mean ERMq. This finding contrasts with studies in other regions of the United States that have shown an increase in the incidence of degraded benthos within the mean ERMq range present among Southern California samples [45]. It is possible that the benthic community results were influenced by variations in noncontaminant factors related to the diversity of habitats (e.g., salinity and depth) and sediment types included in the present study and that the results, combined with the small sample size, confounded the ability to discern impacts due to toxicity.

The present study and others have shown marked differences in sensitivity among toxicity tests that cannot be easily predicted on the basis of biological endpoint (lethal vs sublethal) and mode of exposure (duration, whole sediment, or SWI). This diversity presents both a challenge and an opportunity for sediment toxicity evaluation. The challenge lies in selecting the most appropriate tests for use in a particular study. Variations in relative sensitivity related to contaminant type and uncertainties in the interpretation of chemistry and benthic community data suggest that the use of a single test method, selected on the basis of high sensitivity to a subset of samples, is unlikely to provide a complete or confident assessment of toxicity. Data from multiple toxicity tests that represent a diversity of species, endpoints, and exposure modes, in addition to sediment chemistry and benthic community analyses, are needed to assess sediment quality to the level of confidence needed to support management decisions [46]. The use of multiple toxicity tests and several lines of evidence increases the complexity for data interpretation, as there will be conflicts among the results leading to uncertainty in the assessment's conclusion. Numerous approaches have been developed to integrate and interpret disparate results, including the quantitative approaches (e.g., ranks, means, and frequency of toxicity), tabular decision matrices, and best professional judgment [46,47]. The use of a diverse suite of toxicity tests also provides an opportunity to improve understanding of the causes of sediment toxicity, as differences in the patterns or symptoms of response among tests can be used to help identify the cause of toxicity [48].

*Acknowledgement*—The authors would like to thank the field crews from Weston Solutions, Marine Biological Consultants, and Aquatic Bioassay Laboratory for collecting sediment samples in Southern California; the San Francisco Estuary Institute for collection and analysis in San Francisco Bay; and Arthur Barnett for his valuable assistance in preparing this manuscript. The present study was funded in part by agreement 01-274-250-0 with the California State Water Resources Control Board.

## REFERENCES

1. Strobel CJ, Klemm DJ, Lobring LB, Eichelberger JW, Alford-Stevens A, Potter BB, Thomas RF, Lazorchak JM, Collins GB, Graves RL. 1995. Environmental monitoring and assessment program (EMAP) laboratory methods manual estuaries, Vol 1—Biological and physical analyses. EPA 620/R-95/008. U.S. Environmental Protection Agency, Office of Research and Development, Narragansett, RI.
2. U.S. Environmental Protection Agency. 1994. Methods for assessing the toxicity of sediment-associated contaminants with estuarine and marine amphipods. EPA 600/R-94/025. Office of Research and Development, Narragansett, RI.
3. Adams WJ, Green AS, Ahlf W, Brown SS, Burton GA Jr, Chadwick DB, Crane M, Gouguet R, Ho KT, Hogstrand C, Reynoldson TB, Ringwood AH, Savitz JD, Sibley PK. 2005. Using sediment assessment tools and a weight-of-evidence approach. In Wenning RJ, Batley GE, Ingersoll C, Moore DW, eds, *Use of Sediment Quality Guidelines and Related Tools for the Assessment of Contaminated Sediments.* Society of Environmental Toxicology and Chemistry, Pensacola, FL, USA, pp 163–225.
4. Lamberson JO, DeWitt TH, Swartz RC. 1992. Assessment of sediment toxicity to marine benthos. In Burton GA Jr, ed, *Sediment Toxicity Assessment.* Lewis, Boca Raton, FL, USA, pp 183–211.
5. U.S. Environmental Protection Agency. 2001. Methods for assessing the chronic toxicity of marine and estuarine sediment-associated contaminants with the amphipod *Leptocheirus plumulosus*. EPA 600/R-01/020. Office of Research and Development, Washington, DC.
6. Puget Sound Water Quality Authority. 1995. Recommended guidelines for conducting laboratory bioassays on Puget Sound sediments. Olympia, WA, USA.
7. Carr RS, Nipper M. 2003. Historical overview of porewater toxicity testing. In Carr RS, Nipper M, eds, *Porewater Toxicity Testing: Biological, Chemical, and Ecological Considerations.* Society of Environmental Toxicology and Chemistry, Pensacola, FL, USA, pp 1–10.
8. Anderson BS, Hunt JW, Hester M, Phillips BM. 1996. Assessment of sediment toxicity at the sediment–water interface. In Ostrander GK, ed, *Techniques in Aquatic Toxicology.* CRC, Boca Raton, FL, USA, pp 609–624.
9. Chandler GT, Green AS. 1996. A 14-day harpacticoid copepod reproduction bioassay for laboratory and field contaminated muddy sediments. In Ostrander GK, ed, *Techniques in Aquatic Toxicology.* CRC, Boca Raton, FL, USA, pp 23–39.
10. Ringwood AH, Keppler CJ. 1998. Seed clam growth: An alternative sediment bioassay developed during EMAP in the Carolinian Province. *Environ Monit Assess* 51:247–257.
11. Ringwood AH, Conners DE, Hoguet J. 1998. Effects of natural and anthropogenic stressors on lysosomal destabilization in oysters *Crassostrea virginica*. *Mar Ecol Prog Ser* 166:163–171.
12. Anderson BS, Hunt JW, Phillips BM, Tudor S, Fairey R, Newman J, Puckett HM, Stephenson M, Long ER, Tjeerdema RS. 1998. Comparison of marine sediment toxicity test protocols for the amphipod *Rhepoxynius abronius* and the polychaete worm *Nereis* (*Neanthes*) *arenaceodentata*. *Environ Toxicol Chem* 17:859–866.
13. Pinza MR, Ward JA, Kohn NP. 2002. Results of interspecies toxicity comparison testing associated with contaminated sediment management. *Abstracts*, Annual International Conference on Soils, Sediments and Water, Amhust, MA, USA, October 21–24, p 1024.
14. Burton GA Jr, Ingersoll CG, Burnett LC, Henry M, Hinman ML, Klaine SJ, Landrum PF, Ross P, Tuchman M. 1996. A comparison of sediment toxicity test methods at three Great Lake areas of concern. *J Gt Lake Res* 22:495–511.
15. DeWitt TH, Pinza MR, Niewolny LA, Cullinan VI, Gruendell BD. 1997. Development and evaluation of standard marine/estuarine chronic sediment toxicity test method using *Leptocheirus plumulosus*. Battelle Marine Sciences Laboratory, Sequim, WA, USA.
16. Green AS, Moore D, Farrar D. 1999. Chronic toxicity of 2,4,6-trinitrotoluene to a marine polychaete and an estuarine amphipod. *Environ Toxicol Chem* 18:1783–1790.
17. Long ER, Buchman MF, Bay SM, Breteler RJ, Carr RS, Chapman PM, Hose JE, Lissner AL, Scott J, Wolfe DA. 1990. Comparative evaluation of five toxicity tests with sediments from San Francisco Bay and Tomales Bay, California. *Environ Toxicol Chem* 9:1193–1214.
18. Bight'03 Coastal Ecology. 2003. Southern California Bight 2003 regional monitoring survey (Bight'03): Coastal ecology workplan. Southern California Coastal Water Research Project, Westminster, CA, USA.
19. San Francisco Estuary Institute. 2005. The San Francisco estuary regional monitoring program for trace substances (RMP). 2003 Annual monitoring results. Oakland, CA, USA.
20. Bay SM, Mikel T, Schiff K, Mathison S, Hester B, Young D, Greenstein D. 2005. Southern California Bight 2003 regional monitoring program. I. Sediment toxicity. Technical Report 451. Southern California Coastal Water Research Project, Westminster, CA, USA.

21. Bay S, Greenstein D, Young D. 2007. Evaluation of methods for measuring sediment toxicity in California bays and estuaries. Technical Report 503. Southern California Coastal Water Research Project, Costa Mesa, CA, USA.
22. U.S. Environmental Protection Agency. 2001. Methods for collection, storage and manipulation of sediments for chemical and toxicological analyses. EPA 823/B-01/002. Technical Manual. Office of Water, Washington, DC.
23. Bridges TS, Farrar JD. 1997. The influence of worm age, duration of exposure and endpoint selection on bioassay sensitivity for *Neanthes arenaceodentata* (Annelida: Polychaeta). *Environ Toxicol Chem* 16:1650–1658.
24. Bridges TS, Farrar JD, Duke BM. 1997. The influence of food ration on sediment toxicity in *Neanthes arenaceodentata* (Annelida: Polychaeta). *Environ Toxicol Chem* 16:1659–1665.
25. Keppler CJ, Ringwood AH. 2002. Effects of metal exposures on juvenile clams, *Mercenaria mercenaria*. *Bull Environ Contam Toxicol* 68:43–48.
26. Ringwood AH, Conners DE, Hoguet J, Ringwood LA. 2005. Lysosomal destabilization assays in estuarine organisms. In Ostrander GK, ed, *Techniques in Aquatic Toxicology,* Vol 2. CRC, Boca Raton, FL, USA, pp 287–300.
27. U.S. Environmental Protection Agency. 1995. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to west coast marine and estuarine organisms. EPA 600/R-95/136. Office of Research and Development, Cincinnati, OH.
28. Bight'03 Coastal Ecology Committee. 2003. Southern California Bight 2003 regional monitoring survey (Bight'03). Quality Assurance Manual. Southern California Coastal Water Research Project, Westminster, CA, USA.
29. Ranasinghe JA, Montagne DE, Smith RW, Mikel TK, Weisberg SB, Cadien DB, Velarde RG, Dalkey A. 2003. Southern California Bight 1998 regional monitoring program. VII. Benthic macrofauna. Technical Report 382. Southern California Coastal Water Research Project, Westminster, CA, USA.
30. Thompson B, Lowe S. 2004. Assessment of macrobenthos response to sediment contamination in the San Francisco Bay estuary, California, USA. *Environ Toxicol Chem* 23:2178–2187.
31. Zar JH. 1999. *Biostatistical Analysis.* Simon & Schuster, Upper Saddle River, NJ, USA.
32. Long ER, Field LJ, MacDonald DD. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. *Environ Toxicol Chem* 17:714–727.
33. Long ER, MacDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environ Manag* 19: 81–97.
34. American Society for Testing and Materials. 2002. Standard guide of designing biological tests with sediments. E 1525. In *Annual Book of ASTM Standards,* Vol 11.05. West Conshohocken, PA, pp 877–901.
35. Ingersoll CG, Ankley GT, Benoit DA, Brunson EL, Burton GA, Dwyer FJ, Hoke RA, Landrum PF, Norberg-King TJ, Winger PV. 1995. Toxicity and bioaccumulation of sediment-associated contaminants using freshwater invertebrates: A review of methods and applications. *Environ Toxicol Chem* 14:1885–1894.
36. Kennedy A, Farrar JD, Steevens JA, Reiss M. 2004. Evaluation of the applicability of standard toxicity test methods to dredged material management. *Abstracts*, Society of Environmental Toxicology and Chemistry 25th Annual Meeting, Portland, OR, USA, November 14–18, pp 30.
37. Ringwood AH, Holland AF, Kneib RT, Ross PE. 1996. EMAP/NS&T pilot studies in the Carolinian Province: Indicator testing and evaluation in the Southeastern estuaries. NOAA technical memorandum NOS ORCA 102. National Atmospheric and Oceanic Administration, Silver Spring, MD, USA.
38. Long ER, Sloane GM, Scott GI, Thompson B, Carr RS, Biedenback J, Wade TL, Presley BJ, Scott KJ, Mueller C, Brecken-Fols G, Albrecht B, Anderson JW, Chandler GT. 1999. Magnitude and extent of chemical contamination and toxicity in sediments of Biscayne Bay and vicinity. NOAA technical memorandum NOS NCCOS CCMA 141. National Oceanic and Atmospheric Administration, Silver Spring, MD, USA.
39. Farrar JD, Kennedy A, Lotufo G, McNemar C. 2005. Comparative evaluation of the efficacy of acute and chronic sublethal sediment tests for assessing sediment quality. *Abstracts*, Society of Environmental Toxicology and Chemistry North America 26th Annual Meeting, Baltimore, MD, USA, November 13–17, pp 362.
40. Weston DP. 1996. Further development of a chronic *Ampelisca abdita* bioassay as an indicator of sediment toxicity. RMP Contribution 17. San Francisco Estuary Institute, Richmond, CA, USA.
41. Dillon TM, Moore DW, Jarvis AS. 1994. The effects of storage temperature and time on sediment toxicity. *Arch Environ Contam Toxicol* 27:51–53.
42. Fairey R, Long ER, Roberts CA, Anderson BS, Phillips BM, Hunt JW, Puckett HR, Wilson CJ. 2001. An evaluation of methods for calculating mean sediment quality guideline quotients as indicators of contamination and acute toxicity to amphipods by chemical mixtures. *Environ Toxicol Chem* 20:2276–2286.
43. Dillon TM, Moore DW, Gibson AB. 1993. Development of a chronic sublethal bioassay for evaluating contaminated sediment with the marine polychaete worm *Nereis* (*Neanthes*) *arenaceodentata*. *Environ Toxicol Chem* 12:589–605.
44. McGee BL, Fisher DJ, Wright DA, Yonkos LT, Ziegler GP, Turley SD, Farrar JD, Moore DW, Bridges TS. 2004. A field test and comparison of acute and chronic sediment toxicity tests with the estuarine amphipod *Leptocheirus plumulosus* in Chesapeake Bay, USA. *Environ Toxicol Chem* 23:1751–1761.
45. Hyland JL, Balthis WL, Engle VD, Long ER, Paul JF, Summers JK, Van Dolah RF. 2003. Incidence of stress in benthic communities along the U.S. Atlantic and Gulf of Mexico coasts within different ranges of sediment contaminations from chemical mixtures. *Environ Monit Assess* 81:149–161.
46. Chapman PM, Anderson J. 2005. A decision-making framework for sediment contamination. *Integr Environ Assess Manag* 1:163–173.
47. Burton GA Jr, Chapman PM, Smith EP. 2002. Weight of evidence approaches for assessing ecosystem impairment. *Hum Ecol Risk Assess* 8:1657–1673.
48. U.S. Environmental Protection Agency. 1993. Methods for aquatic toxicity identification evaluations: Phase III toxicity confirmation procedures for samples exhibiting acute and chronic toxicity. EPA 600/R-92/081. Duluth, MN.