Comparison of methods for acute and chronic toxicity in marine sediments

Darrin Greenstein, Steven Bay, Brian Anderson[1], G. Thomas Chandler[2], J. Daniel Farrar[3], Charles Keppler[4], Bryn Phillips[1], Amy Ringwood[5] and Diana Young

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. A study was conducted to compare the sensitivity of several acute and sublethal toxicity methods and investigate their correlations with sediment chemistry and benthic community condition. Six sublethal methods (amphipod: *Leptocheirus plumulosus* 28-day survival, growth and reproduction; polychaete: *Neanthes arenaceodentata* 28-day survival and growth; benthic copepod: *Amphiascus tenuiremis*, 14-day life-cycle; seed clam: *Mercenaria mercenaria* 7-day growth; oyster: *Crassostrea virginica* lysosome destabilization; and sediment-water interface testing with embryos of the mussel *Mytilus galloprovincialis*) and two acute methods (10-day amphipod survival with *Eohaustorius estuarius* and *Leptocheirus plumulosus*) were used to test split samples of sediment from stations in southern California and San Francisco Bay. The life-cycle test with the copepod, *Amphiascus*, proved to be the most sensitive sublethal test and the most sensitive test overall. The *Leptocheirus* 10-day survival test was the most sensitive of the acute tests. In general, the sublethal tests were not more sensitive to sediments than the acute tests. Of the sublethal tests, only the *Amphiascus* endpoints and polychaete growth correlated with sediment chemistry. There was poor correspondence between the toxicity endpoints and indicators of benthic community condition. Differences in test characteristics such as mode of exposure, species-specific contaminant sensitivity, changes in contaminant bioavailability, and the 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.

Introduction

Acute sediment toxicity testing has been routinely conducted as part of monitoring and assessment programs, such as the USEPA's Environmental Monitoring and Assessment Program (Strobel *et al.* 1995). The toxicity tests are usually conducted on whole sediments using amphipod 10-day survival tests in accordance with standard protocols (USEPA 1994). 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 (Adams *et al.* 2005). Sublethal methods include embryo development tests and other tests with various life stages of animals having endpoints such as growth and reproduction in addition to survival. A wide variety of sublethal methods have been described (Lamberson *et al.* 1992); very few of these methods have been used commonly. Regularly used sublethal test methods include the amphipod *Leptocheirus plumulosus* 28-day growth and reproduction test (USEPA 2001), a 20-day polychaete growth test using *Neanthes arenaceodentata* (PSWQA 1995), pore water testing using echinoderm gametes or embryos (Carr and Nipper 2003), and a sediment-water interface (SWI)

[1]*University of California Davis, Monterey, CA*
[2]*University of South Carolina, Columbia, SC*
[3]*US Army Engineer Research and Development Center, Vicksburg, MS*
[4]*Marine Resources Research Institute, Charleston, SC*
[5]*University of North Carolina-Charlotte, Charlotte, NC*

test using sea urchin or mussel embryos (Anderson *et al.* 1996). Additional recently developed sublethal tests showing promise include the measurement of copepod reproduction (Chandler and Green 1996), juvenile clam growth (Ringwood and Keppler 1998), and oyster biomarker responses (Ringwood *et al.* 1998).

Because 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 (Anderson *et al.* 1998, Pinza *et al.* 2002). Few studies specifically designed to compare the relative attributes of various sublethal tests have been conducted. Studies conducted to date have only compared two or three methods to one another (DeWitt *et al.* 1997, Anderson *et al.* 1998, Green *et al.* 1999) or focused more on sublethal elutriate or pore water tests rather than whole sediment tests (Long *et al.* 1990). 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 concordance with benthic community impacts. Information on these factors is extremely limited for many sublethal tests.

This study was designed to investigate relative performance of several acute and sublethal test methods with 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, this study investigated the relationship between changes in benthic community condition and toxicity.

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., USEPA methods, ASTM methods). 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, *E. estuarius* and *L. plumulosus*.

The sediment samples tested were collected as part of two regional monitoring surveys, Southern California Bight 2003 Regional Monitoring Program (Figure 1) and the San Francisco Estuary Institute Regional Monitoring Program (RMP; Figure 2). The stations represented a wide range of expected contamination levels and habitat types for the purpose of targeting stations with expected low to moderate level of acute toxicity. Stations expected to have a high degree of acute toxicity were not included in this 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 geographical location, proximity to contamination sources and expected sediment grain size. The RMP sites had been monitored for approximately 10 years and were selected for their wide geographic distribution and a range of acute toxicity to amphipods.

Tests on split samples were conducted by laboratories with extensive experience using the various tests. The *L. plumulosus* and *N. arenaceodentata* testing was conducted at the Army Corps of

Table 1. Characteristics of the sublethal sediment toxicity methods.

Species	Taxon	Test endpoint(s)	Duration (days)
Mytilus galloprovincialis	mussel	sediment-water interface, embryo development	2
Mercenaria mercenaria	clam	growth	7
Crassostrea virginica	oyster	lysosomal destabilization	4
Leptocheirus plumulosus	amphipod	growth, reproduction, survival*	28
Neanthes arenaceodentata	polychaete	growth, survival*	28
Amphiascus tenuiremis	benthic copepod	reproduction, survival*	14

* Test endpoints that are secondary.

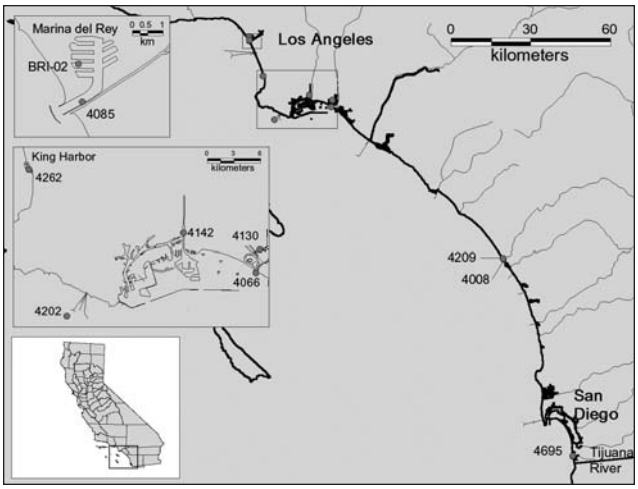

Figure 1. Location of southern California stations used for the sediment toxicity methods comparison study.

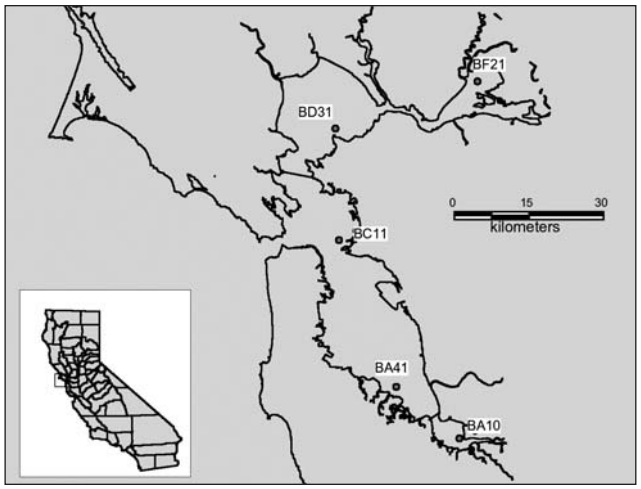

Figure 2. Location of San Francisco Bay stations used for the sediment toxicity methods comparison study.

Engineers, Research and Development Center, Environmental Laboratory in Vicksburg, MS. The *A. tenuiremis* assays were performed at the University of South Carolina in Columbia, SC. 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 in Charleston, SC. The sediment-water interface testing was conducted at the University of California, Davis, Marine Pollution Studies Laboratory in Carmel, CA. Ten-day *E. estuarius* acute survival tests were performed on sediment from each station. These acute tests were conducted by multiple laboratories as part of the regional monitoring efforts. The laboratories that performed the *Eohaustorius* tests on southern California stations participated in intercalibration procedures, which showed reasonable agreement between laboratories (Bay *et al.* 2005). The laboratory testing the San Francisco Bay 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.* 2006). Samples were also analyzed for organic and metals chemistry, total organic carbon (TOC), grain size and benthic infauna.

Sediments were collected in July through 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 Southern California Coastal Water Research Project (SCCWRP) within 24 hours of collection. The core samples were then transported with ice packs to the testing laboratory within 24 hours. Core samples from the San Francisco Bay stations were transported directly to the testing laboratory.

The subcores were shipped to the testing laboratory within 48 hours of collection and the sediment-water interface tests were initiated within 10 days of collection (Table 2). The whole sediment samples were shipped to the testing labs in two batches: one with six of the southern California stations, the other with the remaining four southern California 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 days (Table 2).

Toxicity testing

Eohaustorius estuarius 10-day survival

Ten day survival tests with *E. estuarius* were conducted using standard USEPA testing procedures (1994). 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

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

Station	*Eohaustorius*	Leptocheirus 10-Day	Leptocheirus 28-Day	SWI	*Mercenaria*	Lysosome	*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-02	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	-

were aerated and allowed to 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). 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-day survival

The experimental design followed guidelines set forth by the USEPA (1994). Sediment was added to each of 5 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* (500-750 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 recovered. The number of surviving organisms was counted and recorded.

Leptocheirus plumulosus 28-day survival, growth, and reproduction

The 28-day *L. plumulosus* experiments were conducted according to the guidelines provided by the USEPA (2001). 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 the other test methods (Table 2). Due to this confounding factor, the data for the 28-day *Leptocheirus* test are not presented.

Neanthes arenaceodentata 28-day survival and growth

The 28-day *N. arenaceodentata* experiments were conducted according to guidelines developed by the US Army ERDC (Bridges and Farrar 1997, Bridges *et al.* 1997). 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 hours light:dark. Organisms were obtained from Dr. Don Reish (California State University, Long Beach, CA). On day 0, one *N. arenaceodentata* (≤7-days old) was gently transferred to each replicate beaker. Water quality measurements (dissolved oxygen, pH, salinity and over-

lying 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® 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 pre-weighed pan and placed in a drying oven at 60°C for 24 hours. The pans were then removed, allowed to cool, and weighed to obtain the individual dry weight of worms for each replicate/animal.

Amphiascus tenuiremis 14-day lifecycle

Testing of the copepods (*Amphiascus tenuiremis*) followed the methods of Chandler and Green (1996). A sediment reference sample was collected from Oyster Landing, at North Inlet, SC. Stations BA41, BC11, BRI2, 4085, 4202, and 4262 were press-sieved through a 125-mm sieve in order to facilitate recovery of the copepods at the conclusion of the exposure. A larger sieve size was used for some of the larger grained stations in order to obtain a sufficient volume of sediment for testing. Sediment samples 4008 and 4695 were screened with a 250-mm sieve while 4209 and 4130 were sieved through 212- and 180-mm sieves, respectively. Sediment samples 4066 and 4142 were too sandy to pass a 250-mm sieve and could not be tested with the copepod method. A total of ten stations were tested with *Amphiascus*. Teflon 50-ml Erlenmeyer flasks with mesh-covered outflow holes were filled with 0.45 mm filtered, aerated seawater. Press-sieved sediment samples were then packed into Teflon syringes and slowly extruded onto the bottoms of their respective chambers (4 replicates per sediment sample). Adult non-gravid female and adult male copepods were then counted into each quadruplicate test chamber (25/sex). Chambers were placed in an incubator at 20°C under continuous dripping flow for 14 days with a 12:12 hour light:dark cycle. Chambers were fed every third day a mixture of frozen algal stock (10^7 cells of 1:1:1 *Isochrysis galbana*, *Phaeodactylum tricornutum*, and *Dunaliella tertiolecta*). 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. At the end of 14 days of exposure, copepods were collected using a 63-mm sieve. Samples were checked/counted for dead bodies. Copepods were stained with Rose Bengal and preserved in 5% borate-buffered formalin. Non-gravid adult females, gravid adult females, adult males, copepodites, nauplii, and clutch sizes were enumerated under a Nikon SMZ-U stereo dissection microscope. Two endpoints of the *Amphiascus* test were calculated: the number of copepodites produced and the realized offspring production (output of 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-day exposure to whole sediment (Ringwood and Keppler 1998, Keppler and Ringwood 2002). Sediment samples were pressed through a 500-mm sieve, homogenized, and 50-ml aliquots were placed into four replicate 250 ml beakers. The sediment was then overlain with clean 25 g/kg seawater. The replicates were gently aerated for the duration of the experiment, and the assays were conducted at room temperature (22°C to 25°C) for 7 days with a 16:8 hour light cycle. Juvenile clams (*Mercenaria mercenaria*) used for all experiments were obtained from Atlantic Littleneck Clam Farm, Charleston, SC. Clams were sieved through two mesh sizes (1.0 mm and 1.2 mm) to ensure that the clams were of a similar size range. Twenty-five clams were used for each replicate. Pre-assay wet weights of each clam group were taken for growth rate estimates and to ensure that all replicate groups had similar initial weights. Replicate subsets of clams were also counted, wet weighed, dried overnight, and reweighed to verify the wet:dry weight ratio used to estimate initial dry weights. Each replicate was fed on the first, third, and sixth days of the assay (50:50 mix of *Isochrysis galbana* and *Chaetoceros gracilis* for 20 x 10^6 cells/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 hours to depurate. Dead clams were counted and removed; percent mortalities were calculated. The surviving clams were counted and rinsed with distilled water to remove excess salts. Post-assay wet weights were determined, and clams were then dried for 48 hours (at 70°C). Each clam replicate was recounted and final dry weight per clam was determined. Initial dry weights were subtracted from final dry weights, and the results were expressed as growth rates

(mg/clam/day). Sediment pore water chemistry parameters (salinity, pH, and total ammonia – nitrogen (TAN) were measured for each sediment sample prior to use in any assay. Overlying water quality was also measured.

Crassostrea virginica 4-day lysosomal destabilization

The lysosomal destabilization assay was conducted following methods described in Ringwood *et al.* (1998). Sediment samples were homogenized, and 100-ml aliquots were placed into three replicate 1-L beakers. The sediment was topped with clean 25 g/kg seawater. The beakers were allowed to settle for 2 hours, and then 3 clean-scrubbed oysters were gently added to each replicate. Oysters (5.3 ± 0.7 cm) used for laboratory sediment exposures were collected from control sites and acclimated to laboratory conditions for at least 24 hours prior to the start of the experiment. The replicates were gently aerated for the duration of the experiment, and the assays were conducted at room temperature (22°C to 25°C) for 4 days with a 16:8 hour light cycle. Each replicate was fed on the first and third days of the assay (*Isochrysis* paste mixed into filtered sea water for 70 x 10^6 cells/replicate). The overlying water was not renewed during the exposure. Water quality parameters for both pore and overlying waters were measured in the same way as for the *Mercenaria* testing. Digestive gland tissue from the exposed oysters was diced and treated with trypsin to produce a cell suspension. A cell suspension aliquot was mixed with an equal aliquot of neutral red (NR) solution, placed on a microscope slide, and examined under a light microscope to evaluate NR retention by digestive gland cells containing lysosomes. At least 50 cells were scored as stable (NR retention in the lysosomes) or destabilized (NR leaking into the cytoplasm), and the data were expressed as the percentage of cells with destabilized lysosomes per oyster.

Mytilus galloprovincialis 2-day embryo development at the sediment-water interface

Exposure procedures followed those detailed by Anderson *et al.* (1996). One day prior to the start of the test, 300 mL of clean seawater (1 mm filtered, approximately 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 USEPA protocols (USEPA 1995) and added to the screen tubes. Mussels were exposed for 48 hours and sea urchins for 96 hours. 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 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 TOC. Analyses were conducted by a variety of laboratories participating in the regional monitoring programs and used standardized EPA recommended methods (Bight'03 Coastal Ecology Committee 2003, SFEI 2005). The laboratories had achieved acceptable comparability during pre-project intercalibration exercises and the data were subjected to rigorous post-survey 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 or a Horiba LA900 instrument. Sediment samples analyzed for all metal analytes except mercury were digested in strong acid according to the procedures described in EPA Method 3050B. Metals were quantified using either 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 organic chemistry analysis were solvent extracted using accelerated solvent extraction, sohxlet, or roller table. The extracts obtained were subjected to each laboratory's own clean-up procedures and analyzed by gas chromato-

graphic method (e.g., dual-column GC-ECD or GC-MS in the selected ion monitoring mode).

Benthic community analysis

A 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 species or 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; Ranasinghe *et al.* 2003). 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 (Thompson and Lowe 2004). The IBI uses a multimetric index to discriminate between impacted and reference areas.

Data analysis

Toxicity data were control normalized ((station value/control) x 100) to facilitate comparisons among the test methods. Statistical significance was tested using Student's t-test ($p \leq 0.05$) assuming unequal variance (Zar 1999). 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; Long *et al.* 1998) 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 (ERM) for DDT was not used in calculations because it has been found to be unreliable (Long *et al.* 1995). Relationships between sediment chemistry parameters or benthic community condition and toxicity response were analyzed using a non-parametric Spearman's 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 and had a low control normal-alive percentage. Because the difference between the control and sample response was very large, the data has 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 TAN). 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. For the SWI test, station BC11 had an overlying water ammonia concentration of 0.145 mg/L un-ionized ammonia, which is very near the EC50 (approximately 0.17 mg/L, unpublished data).

Comparisons among sublethal tests

A wide range in the percentage of stations identified as toxic by sublethal methods was observed (Figure 3). The highest percentage was for the copepod, *Amphiascus* that found 9 out of the 10 stations tested to be toxic, followed by *Neanthes* with 8 out of 15 stations. The proportion of stations identified as toxic was much lower for the remaining test methods, with the lowest percentage for the sediment-water interface testing which identified 1 out of 15 stations tested as toxic.

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 (Figure 4). Overall,

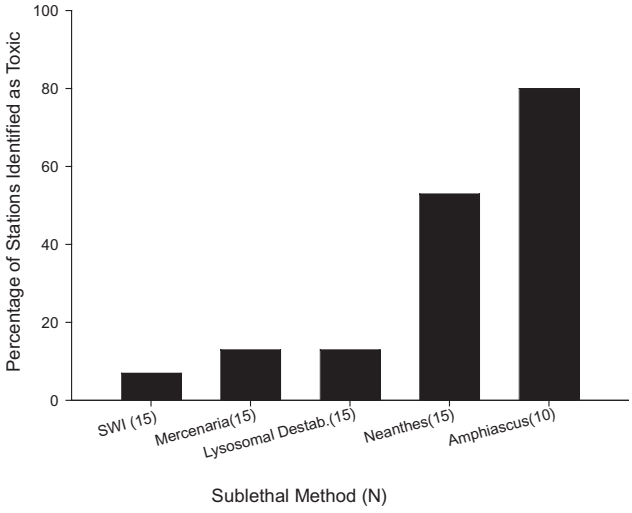

Figure 3. Percentage of stations that each sublethal method identified as toxic; number of stations tested is in parentheses.

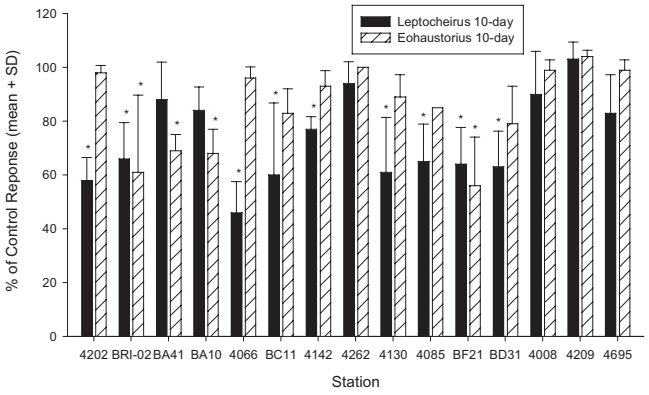

Figure 4. Results of *Eohaustorius* and *Leptocheirus* 10-day survival tests conducted stations in southern California and San Francisco Bay; stations marked with * are significantly different from control values ($p \leq 0.05$) and less than 80% of the control response.

the *Eohaustorius* method was near the mid-point of sensitivity relative to the sublethal tests. The *Leptocheirus* 10-day method identified 9 out of the 15 sites tested as toxic and was determined to be more sensitive than all but one of the sublethal methods.

Comparisons between sublethal and acute tests

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

The *Leptocheirus* 10-day test found a higher percentage 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

Table 3. Comparative ability of acute and sublethal toxicity test methods to detect toxicity in stations from southern California and San Francisco Bay; numeric values are expressed as percentage of stations tested.

	Acute Methods		Sublethal Methods				
Station	*Eohaustorius*	*Lepto* 10-Day	SWI	*Mercenaria*	Lysosome	*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 Toxic, *Eohaustorius* Not Toxic	--	--	7	13	13	27	70
% *Eohaustorius* Toxic, Sublethal Not Toxic	--	--	27	27	27	0	0
% Agree Toxic	--	--	0	0	0	27	20
%Agree Not Toxic	--	--	67	60	60	47	10
	--	--					
% Sublethal Toxic, *Leptocheirus* Not Toxic	--	--	0	7	0	27	40
% *Leptocheirus* Toxic, Sublethal Not Toxic	--	--	53	53	47	33	0
% Agree Toxic	--	--	7	7	13	27	50
%Agree Not Toxic	--	--	40	33	40	13	10

Y=Station identified as toxic
N=Station not identified as toxic
--=Station or comparison not tested

Leptocheirus acute test. Concordance was observed between the *Leptocheirus* 10-day 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 tests, a high percentage of stations (60% or more) not identified as toxic using these sublethal methods 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) were found with the combinations of *Leptocheirus* 10-day and *Neanthes* or *Amphiascus* methods (Table 3). The combination of the two acute tests (see Figure 4 where 11 out of fifteen stations were identified as toxic by one or both tests) 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). TOC values ranged from 0.02% to 2.9%.

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-02, 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 about half of the cases (Long *et al.* 1998).

With respect to *Eohaustorius* survival, both *Amphiascus* endpoints and *Neanthes* growth had significant 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 (8/15) or having an intermediate level of disturbance (5/15 stations at Level 2 or 3). Two stations (4066 and 4142) had 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 for Level-4 stations and highest for the Level-2 stations (Table 6).

Little correspondence was observed between changes in benthic community condition and toxicity for most of the test methods. *Leptocheirus* 10-day 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 through 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-day 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).

Ranking of stations to reflect sediment condition

Because most of the stations in this study had not been previously sampled, there was not a known gradient of expected sediment condition. To put the data into this context, the stations were ranked by a combination of chemical contamination and benthic

Table 4. Selected chemistry data from southern California and San Francisco Bay sediment samples on which toxicity tests were performed.

Station	4202	BRI-02	BA41	BA10	4066	BC11	4142	4262
Metal (mg/kg)								
Arsenic	8.51	13.0	4.51	4.10	1.04	3.98	0.975	4.03
Cadmium	6.56	0.335	0.166	0.134	0.05	0.335	0.225	0.57
Chromium	136	94.5	NA	NA	7.44	NA	5.40	45.5
Copper	5.14	362	29.9	23.7	7.46	39.3	5.74	3.18
Lead	30.0	113	17.4	11.3	4.73	29.7	4.32	38.9
Mercury	0.464	0.980	0.342	0.235	0.103	0.226	0.065	0.226
Nickel	29.0	41.6	58.2	46.9	4.05	65.9	4.06	21.2
Silver	1.9	1.98	0.138	0.254	0.65	0.143	0.28	0.67
Tin	NA	6.26	NA	NA	0.41	NA	0.46	NA
Zinc	180	382	90.5	69.5	22	108	49.0	92.2
TOC (%)	2.06	1.99	1.09	2.34	0.016	1.80	0.27	1.50
Sand (%)	39.3	7.9	20	44.0	100	22	62.5	56.5
Silt (%)	50.1	74.2	22	14.7	0	22	NA	35.7
Clay (%)	10.7	17.9	49	35.7	0	48	NA	7.8
∑PAHs (ug/kg)	678.1	76	1923	724	51.9	740.3	73.1	625.1
∑DDTs (ug/kg)	2301.3	2.2	0.22	0.64	1	0.63	ND	49.8
∑PCBs (ug/kg)	193.9	ND	2.53	2.33	ND	111.3	ND	66.0
Mean ERMq*	0.680	0.262	0.145	0.101	0.025	0.342	0.022	0.290
ERMq Ranking	1	4	7.5	10	12	2	13	3

*DDT concentrations not included in ERMq calculation.

Station	4130	4085	BF21	BD31	4008	4209	4695
Metal (mg/kg)							
Arsenic	7.04	11.6	8.54	7.50	2.5	1.5	1.12
Cadmium	0.845	1.74	0.204	0.216	0.14	ND	ND
Chromium	48.9	78.4	NA	NA	33.6	9.9	4.58
Copper	87.3	101	53.4	50.9	14.2	2.84	0.98
Lead	61.6	130	16.4	17.8	4.71	1.35	1.24
Mercury	0.405	0.41	0.272	0.238	0.08	0.02	0.02
Nickel	25.8	33.1	88.6	87.8	10.2	3.04	0.935
Silver	0.77	2.94	0.204	ND	0.67	0.24	0.22
Tin	3.8	6.26	NA	NA	1.56	0.48	0.3
Zinc	248	315	126	126	47.9	14.0	6.44
TOC (%)	2.04	2.93	1.37	1.33	0.67	0.038	ND
Sand (%)	44.2	30.4	1	9	53.8	97.5	100
Silt (%)	46.3	56.9	39	32	40.4	2.5	ND
Clay (%)	9.5	12.6	60	59	5.9	ND	ND
∑PAHs (ug/kg)	1205.7	578	582.4	450.2	11.7	ND	ND
∑DDTs (ug/kg)	9.9	14.6	0.78	1.42	1.3	ND	ND
∑PCBs (ug/kg)	15.7	22.6	0.80	0.75	ND	ND	ND
Mean ERMq*	0.168	0.242	0.145	0.143	0.037	0.012	0.008
	6	5	7.5	9	11	14	15

*DDT concentrations not included in ERMq calculation.

community health. To achieve this, the stations were ranked by their mean ERMq values (Table 4). The stations were similarly ranked by benthic community analysis results (Table 6). These two rankings were then summed and the stations re-ranked to get the combined effect. The data presented in Figure 5 show stations with the lowest rankings (highest chemistry and most degraded benthos) on the left and highest rankings on the right. Station 4202 had the highest concentrations of the most chemical con-

Table 5. Spearman rank correlations on selected sediment parameters and toxicity endpoints; shaded values are statistically significant (p ≤0.05).

	Eohaustorius Survival	Leptocheirus 10 Survival	Mytilus SWI	Mercenaria Growth	Crassostrea Lysosome	Neanthes Survival	Neanthes Growth	No. Copepodites	Realized Offspring
Arsenic	-0.60	-0.24	0.27	-0.15	-0.08	0.14	-0.54	-0.59	-0.81
Cadmium	-0.16	-0.40	0.26	-0.30	-0.10	0.13	-0.26	-0.21	-0.49
Copper	-0.79	-0.38	0.20	-0.35	-0.06	-0.05	-0.57	-0.83	-0.95
Lead	-0.37	-0.35	0.34	-0.31	-0.03	0.23	-0.39	-0.48	-0.84
Mercury	-0.60	-0.41	0.48	-0.14	-0.09	0.06	-0.51	-0.57	-0.74
Nickel	-0.84	-0.29	-0.39	-0.38	0.14	-0.02	-0.59	-0.87	-0.71
Silver	0.22	-0.09	0.53	0.01	-0.23	0.19	-0.08	-0.04	-0.46
Zinc	-0.55	-0.43	0.25	-0.30	-0.09	0.14	-0.48	-0.57	-0.84
TOC (%)	-0.44	-0.25	0.12	-0.27	0.07	-0.01	-0.42	-0.39	-0.66
Sand (%)	0.82	0.24	0.09	0.35	0.08	-0.07	0.65	0.93	0.79
Clay (%)	-0.82	-0.23	-0.32	-0.33	0.14	-0.03	-0.60	-0.88	-0.72
∑PAHs (ug/kg)	-0.49	-0.26	0.03	-0.35	0.22	-0.31	-0.49	-0.52	-0.49
∑DDTs (ug/kg)	-0.01	-0.33	0.12	-0.26	0.01	0.38	-0.32	-0.09	-0.37
∑PCBs (ug/kg)	-0.12	-0.30	-0.08	-0.19	0.34	0.06	-0.21	-0.07	-0.13
ERMq	-0.29	-0.27	0.02	-0.40	0.12	0.31	-0.45	-0.33	-0.37

Table 6. Incidence of toxicity within benthic index categories and Spearman's Rank Correlation values for toxicity test endpoints; shaded value is statistically significant (p ≤0.05).

	Benthic Index Category					
Test	Ref[1]	Level 2[2]	Level 3[3]	Level 4[4]	Levels 2-4	r[5]
Number of Stations	8	4	1	2	7	
Benthic Station Rank	11.5	5.5	3.0	1.5		
Mean ERMq	0.15	0.31	0.10	0.02	0.20	
	Incidence of Toxicity (%)					
Eohaustorius 10d Survival	12	50	100	0	42	-0.52
Leptocheirus 10d 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. Copepodites	83	100	na	na	100	-0.44

[1] Reference stations: BC11, 4262, 4085, BF21, BD31, 4008, 4209, 4695
[2] Level 2 (Loss of biodiversity): 4202, BRI-02, BA41, 4130
[3] Level 3 (Loss of community function): BA10
[4] Level 4 (Defaunation): 4066, 4142
[5] Correlation calculated using southern California data only

stituents and showed a toxic response to two of the sublethal test endpoints. It was ranked as having the worst sediment condition of all stations tested, even without consideration of the station's high DDT level. Station BRI-2, with high concentrations of three metals and Level 2 benthic designation, was ranked as having the second worst sediment condition. By comparison, Station 4085, with moderate levels of several chemicals, was ranked in the middle. Although stations 4066 and 4142 received Level-4 benthic designations, they fell in the middle of the ranks because their chemical concentrations were lower.

Discussion

The sensitivity of the toxicity 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 this 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-day test have shown that the sublethal endpoints from this test are not consistently more sensitive than acute amphipod tests to field and spiked sediments (DeWitt *et al.* 1997). Another study found that the acute *Ampelisca abdita* test was more sensitive than the *Leptocheirus* 28-day test, which was more sensitive than the *Neanthes* 28-day test (Kennedy *et al.* 2004). 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 were tested (Ringwood *et al.* 1996).

This 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.* (1999) found a greater incidence of toxicity with the *Amphiascus* life cycle method (73%) than with the *Ampelisca* 10-day 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 concentrations and reference benthic community condition, reflect chemical toxicity or the effects of potentially confounding factors such as ammonia or organic carbon.

Several factors may have accounted for the variation in sensitivity among methods observed in this 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 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 identst 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), indicating that variations in holding time or sediment handling among the tests and batches were not major confounding factors.

The most responsive of the acute and sublethal toxicity tests showed a general correspondence with the gradient of sediment condition described by a combination of the chemistry and benthic community data. The *Amphiascus* and *Neanthes* tests reflected the expected pattern of decreasing toxicity with improving sediment condition (Figure 5), as did both of the acute tests (Figure 4). These relationships were inconsistent for stations having intermediate rankings of sediment condition, indicating substantial uncertainty in the relationships among the different indicators of sediment quality. In addition to the sources of variability mentioned previously for the toxicity tests, measures of sediment chemistry and benthic community condition also have inherent uncertainty and sources of error that may have accounted for the inconsistent relationships.

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 (Long *et al.* 1995, Fairey *et al.* 2001). There were also significant correlations with grain size for each test. The chemistry values also correlated with grain size, and many of the chemical constituents also correlated with one another. These intercorrelations make determining whether toxicity is associated with chemistry or the confounding factor of

Table 7. Holding times, incidence of toxicity, and test method rankings (1 = highest frequency of toxic hits) for batches analyzed for sediment toxicity methods comparisons.

Test Method	Batch 1 (6 stations)			Batch 2 (9 stations)		
	Holding Time Range (days)	Stations Toxic (%)	Rank	Holding Time Range (days)	Stations Toxic (%)	Rank
Amphiascus	7-12	75	1	6-19	100	1
Leptocheirus 10-Day	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	-

grain size a difficult matter. Grain size is not known to be a confounding factor for *Eohaustorius* (USEPA 1994). 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. *Neanthes* have been tested in grain sizes ranging from 5 to 100% sand with no effects on either survival or growth (Dillon *et al.* 1993). These factors indicate that a likelihood of an association exists between sediment contamination and toxicity for these three methods in the current study, rather than a grain size effect.

The lack of correlations with sediment chemistry for some of the test methods may have several causes. There was little observed toxicity 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 between 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 did not have a significant influence on the results.

A strong relationship between the toxicity results and benthic community condition was not found in this study, suggesting that these indicators were responding to different aspects of sediment quality. Other studies have reported similar results. Analyses of Chesapeake Bay sediment toxicity using the *Leptocheirus* 10- and 28-day tests found a similar lack of correspondence with benthic community response (McGee *et al.* 2004). 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 (Ranasinghe *et al.*

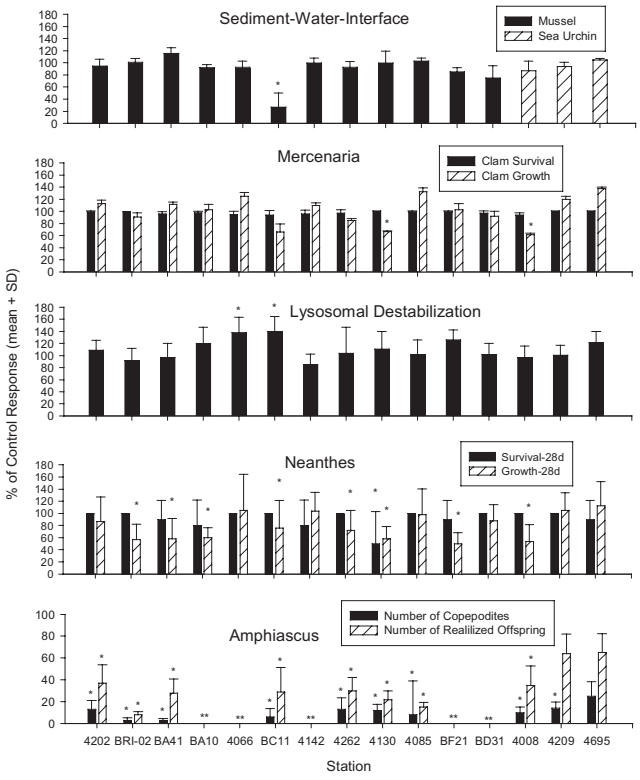

Figure 5. Results of sublethal sediment toxicity test methods conducted on samples from southern California and San Francisco Bay. Stations marked with * are significantly different from control values ($p \leq 0.05$). Stations marked with ** were not tested using that method.

2003). 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. These factors can affect 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 perfectly 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 community 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). 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 (Hyland *et al.* 2003). It is possible that variations in noncontaminant factors related to the diversity of habitats and sediment types included in this study may have influenced the benthic community results and confounded the ability to discern impacts due to toxicity.

This study and others have shown marked differences in sensitivity among toxicity tests that cannot be easily predicted on the basis of biological endpoint and mode of exposure. This diversity presents both a challenge and 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 (Chapman and Anderson 2005). 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 between tests can be used to help identify the cause of toxicity (USEPA 1993).

Literature Cited

Adams, W.J., A.S. Green, W. Ahlf, S.S. Brown, G.A. Burton Jr., D.B. Chadwick, M. Crane, R. Gouguet, K.T. Ho, C. Hogstrand, T.B. Reynoldson, A.H. Ringwood, J.D. Savitz and P.K. Sibley. 2005. Using sediment assessment tools and a weight-of-evidence approach. pp. 163-225 *in*: R.J. Wenning, G.E. Batley, C. Ingersoll and D.W. Moore (eds.), Use of Sediment Quality Guidelines and Related Tools for the Assessment of Contaminated Sediments. Society of Environmental Toxicology and Chemistry. Pensacola, FL.

Anderson, B.S., J.W. Hunt, M. Hester and B.M. Phillips. 1996. Assessment of sediment toxicity at the sediment-water interface. pp. 609-624 *in*: G.K. Ostrander (ed.), Techniques in aquatic toxicology. CRC Press Inc. Boca Raton, FL.

Anderson, B.S., J.W. Hunt, B.M. Phillips, S. Tudor, R. Fairey, J. Newman, H.M. Puckett, M. Stephenson, E.R. Long and R.S. Tjeerdema. 1998. Comparison of marine sediment toxicity test protocols for the amphipod *Rhepoxynius abronius* and the polychaete worm *Nereis* (*Neanthes*) *arenaceodentata*. *Environmental Toxicology and Chemistry* 17:859-866.

Bay, S., D. Greenstein and D. Young. 2006. Sediment Toxicity Indicator Report. Southern California Coastal Water Research Project. Westminster, CA.

Bay, S.M., T. Mikel, K. Schiff, S. Mathison, B. Hester, D. Young and D. Greenstein. 2005. Southern California Bight 2003 regional monitoring program: I. Sediment toxicity. Southern California Coastal Water Research Project. Westminster, CA.

Bight'03 Coastal Ecology Committee. 2003. Quality assurance manual. Southern California Coastal Water Research Project. Westminster, CA.

Bridges, T.S. and J.D. Farrar. 1997. The influence of worm age, duration of exposure and endpoint selection on bioassay sensitivity for *Neanthes arenaceodentata* (Annelida: Polychaeta). *Environmental Toxicology and Chemistry* 16:1650-1658.

Bridges, T.S., J.D. Farrar and B.M. Duke. 1997. The influence of food ration on sediment toxicity in *Neanthes arenaceodentata* (Annelida: Polychaeta). *Environmental Toxicology and Chemistry* 16:1659-1665.

Carr, R.S. and M. Nipper (eds.). 2003. Porewater toxicity testing: Biological, chemical, and ecological considerations. Society of Environmental Toxicology and Chemistry. Pensacola, FL.

Chandler, G.T. and A.S. Green. 1996. A 14-day harpacticoid copepod reproduction bioassay for laboratory and field contaminated muddy sediments. pp. 23-39 *in*: G.K. Ostrander (ed.), Techniques in aquatic toxicology. CRC Press. Boca Raton, FL.

Chapman, P.M. and J. Anderson. 2005. A decision-making framework for sediment contamination. *Integrated Environmental Assessment and Management* 1:163-173.

DeWitt, T.H., M.R. Pinza, L.A. Niewolny, V.I. Cullinan and B.D. Gruendell. 1997. Development and evaluation of standard marine/estuarine chronic sediment toxicity test method using *Leptocheirus plumulosus*. Battelle Marine Sciences Laboratory. Sequim, WA.

Dillon, T.M., D.W. Moore and A.B. Gibson. 1993. Development of a chronic sublethal bioassay for evaluating contaminated sediment with the marine polychaete worm *Nereis* (*Neanthes*) *arenaceodentata*. *Environmental Toxicology and Chemistry* 12:589-605.

Fairey, R., E.R. Long, C.A. Roberts, B.S. Anderson, B.M. Phillips, J.W. Hunt, H.R. Puckett and C.J. Wilson. 2001. An evaluation of methods for calculating mean sediment quality guideline quotients as indicators of contamination and acute toxicity to amphipods by chemical mixtures. *Environmental Toxicology and Chemistry* 20:2276-2286.

Farrar, J.D., A. Kennedy, G. Lotufo and C. McNemar. 2005. Comparative evaluation of the efficacy of acute and chronic sublethal sediment tests for assessing sediment quality. SETAC North America 26th Annual Meeting. Baltimore, MD.

Green, A.S., D. Moore and D. Farrar. 1999. Chronic toxicity of 2,4,6-trinitrotoluene to a marine polychaete and an estuarine amphipod. *Environmental Toxicology and Chemistry* 18:1783-1790.

Hyland, J.L., W.L. Balthis, V.D. Engle, E.R. Long, J.F. Paul, J.K. Summers and R.F. Van Dolah. 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. Environmental Monitoring and Assessment 81:149-161.

Kennedy, A., J.D. Farrar, J.A. Steevens and M. Reiss. 2004. Evaluation of the applicability of standard toxicity test methods to dredged material management. Society of Environmental Toxicology and Chemistry Annual Meeting. Portland, OR.

Keppler, C.J. and A.H. Ringwood. 2002. Effects of metal exposures on juvenile clams, *Mercenaria mercenaria*. *Bulletin of Environmental Contamination and Toxicology* 68:43-48.

Lamberson, J.O., T.H. DeWitt and R.C. Swartz. 1992. Assessment of sediment toxicity to marine benthos. pp. 183-211 *in*: G.A. Burton, Jr. (ed.), Sediment Toxicity Assessment. Lewis Publishers, Inc. Boca Raton, FL.

Long, E.R., M.F. Buchman, S.M. Bay, R.J. Breteler, R.S. Carr, P.M. Chapman, J.E. Hose, A.L. Lissner, J. Scott and D.A. Wolfe. 1990. Comparative evaluation of five toxicity tests with sediments from San Francisco Bay and Tomales Bay, California. *Environmental Toxicology and Chemistry* 9:1193-1214.

Long, E.R., L.J. Field and D.D. MacDonald. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. Environmental Toxicology and Chemistry 17:714-727.

Long, E.R., D.D. MacDonald, S.L. Smith and F.D. Calder. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environmental Management* 19:81-97.

Long, E.R., G.M. Sloane, G.I. Scott, B. Thompson, R.S. Carr, J. Biedenback, T.L. Wade, B.J. Presley, K.J. Scott, C. Mueller, G. Brecken-Fols, B. Albrecht, J.W. Anderson and G.T. Chandler. 1999. Magnitude and

extent of chemical contamination and toxicity in sediments of Biscayne Bay and vicinity. NOS NCCOS CCMA 141. National Oceanic and Atmospheric Administration. Silver Spring, MD.

McGee, B.L., D.J. Fisher, D.A. Wright, L.T. Yonkos, G.P. Ziegler, S.D. Turley, J.D. Farrar, D.W. Moore and T.S. Bridges. 2004. A field test and comparison of acute and chronic sediment toxicity tests with the estuarine amphipod *Leptocheirus plumulosus* in Chesapeake Bay, USA. *Environmental Toxicology and Chemistry* 23:1751-1761.

Pinza, M.R., J.A. Ward and N.P. Kohn. 2002. Results of interspecies toxicity comparison testing associated with contaminated sediment management. Sediment Management Annual Review Meeting. Seattle, WA.

Puget Sound Water Quality Authority PSWQA. 1995. Recommended guidelines for conducting Thompson, B. and S. Lowe. 2004. Assessment of macrobenthos response to sediment contamination in the San Francisco Bay estuary, California, USA. *Environmental Toxicology and Chemistry* 23:2178-2187.

US Environmental Protection Agency (USEPA). 1993. Methods for Aquatic Toxicity Identification Evaluations: Phase III Toxicity Confirmation Procedures for Samples Exhibiting Acute and Chronic Toxicity. EPA/600/R-92/081. US Environmental Protection Agency. Duluth, MN.

US Environmental Protection Agency (USEPA). 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, US Environmental Protection Agency. Narragansett, RI.

US Environmental Protection Agency (USEPA). 1995. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to west coast marine and estuarine organisms. Office of Research and Development. Cincinnati, OH.

US Environmental Protection Agency (USEPA). 2001. Methods for assessing the chronic toxicity of marine and estuarine sediment-associated contaminants with the amphipod *Leptocheirus plumulosus*. US Environmental Protection Agency. Washington, D.C.

Weston, D.P. 1996. Further development of a chronic *Ampelisca abdita* bioassay as an indicator of sediment toxicity. RMP Contribution #17. San Francisco Estuary Institute. Richmond, CA.

Zar, J.H. 1999. Biostatistical analysis (4th ed.). Simon & Schuster. Upper Saddle River, NJ.

Acknowledgements

The authors would like to thank the field crews from Weston Solutions, Marine Biological Consultants, and Aquatic Bioassay Laboratory for collecting sediment samples; the San Francisco Estuaries Institute for arranging for the collection of sediment; and Dr. Arthur Barnett for his valuable help in preparing this manuscript.