FINDINGS REPORT

From A Critical Review of:

Full Life-Cycle Bioassay Approach to Assess Chronic Exposure of *Pseudodiaptomus forbesi* to Ammonia/Ammonium - Final Report

Dated August 31, 2011

Prepared by: Teh S, Flores I, Kawaguchi M, Lesmeister S, Teh C
Aquatic Toxicology Program, Department of Anatomy, Physiology, and Cell Biology, School of Veterinary Medicine, University of California Davis

This Critical Review Was Prepared By:

Pacific EcoRisk, Inc.
2250 Cordelia Rd.
Fairfield, CA 94534

This Critical Review Was Prepared For:

Larry Walker Associates
707 Fourth St.
Davis, CA 95616

Central Costa County Sanitary District
5019 Imhoff Place
Martinez, CA 94553

Submittal Date:
December 26, 2011

PACIFIC ECORISK
ENVIRONMENTAL CONSULTING & TESTING

1. Introduction

On behalf of the Central Contra Costa County Sanitary District (CCCCSD), Larry Walker Associates has contracted Pacific EcoRisk, Inc. (PER) to perform a critical review of the "Final Report: *Full Life-Cycle Bioassay Approach to Assess Chronic Exposure of Pseudodiaptomus forbesi to Ammonia/Ammonium*" authored by Teh S, Flores I, Kawaguchi M, Lesmeister S, and Teh C (dated August 31, 2011).). As requested by CCCCSD, the primary focus of this review were the experiments described as Subtasks 3-3 and 3-4-1 in the Teh *et al*. report; accordingly, our comments on these Subtasks are presented immediately below. However, in the course of reviewing the report, numerous other questions/problems/issues came to light for which comment also seemed necessary; these comments are therefore included as well, following the Subtask 3-3 and 3-4-1 comments.

2. Comments on Sub-Task 3-3 (Chronic [31-day] life cycle toxicity testing)

Comment #1. Teh *et al*.'s analysis of the number of nauplii and number of juveniles produced during the chronic (31-day) exposure is believed to be flawed at a very fundamental level. It is apparent in Teh et al.'s derivation of 'mean number of nauplii, juveniles, and adult *P. forbesi* produced per female' (in Teh *et al*.'s Table 11) and in the 'sum total number of nauplii, juvenile, and adult *P. forbesi* produced' (in Teh *et al*.'s Appendix III table) that they summed the counts of nauplii and juveniles that were counted on the progressive 2-3 day intervals (the raw data for these counts were provided in Teh *et al*.'s Appendix I) as if each new progressive count was of new individuals that had not been counted on the previous count day. So when 17 nauplii were counted in Control replicate A on Day 5 of the test, and 20 nauplii were counted on Day 7, and 17 were counted on Day 10, and so on, Teh *et al*. summed these up as if they were different nauplii that had been produced during the progressive 'count days'.

This would be correct had the nauplii and juveniles that were counted on each 'count day' been removed from the original replicate container and transferred to a new replicate container such that any nauplii or juveniles observed and counted in the original replicate containers on subsequent days would have been new organisms separate and distinct from the organisms that had been counted during the previous count day(s). Note that this approach would have created a logistical challenge, with a doubling of the number of experimental replicate beakers on Day 3 of the test (going from the original n=20 to n=40), a tripling of the beakers on Day 5 (n=60), a quadrupling of beakers on Day 7 (n=80), and so on and so on. This would then be compounded as nauplii that had transformed into juveniles would again need to be transferred to new replicates so as to allow observation of new juveniles produced by the remaining nauplii. The number of necessary beakers rapidly becomes logistically improbable.

However, it is not believed that this is what happened. Unfortunately, their report's inadequate description of test methodology is not explicit on this. However, it can be deduced from the nature of the study that the neonates were left in place in each replicate, as these were the source

of the subsequent juveniles, which were similarly left in place to serve as the source for the subsequent adults. This was confirmed by inquiry made with one of the other authors of the report (M Kawaguchi, pers. comm.). As a result, when 20 nauplii were counted in Control replicate A on Day 7, some (if not most) of these organism were the very same organisms that had been counted on the earlier Day 5 count, and the nauplii that were counted on Day 10 were some of the same as had been counted on Days 7 and Day 5.

This conclusion is also supported by the following observations made for closely-related congener *Pseudodiaptomus annandalei* (Golez et al. 2004):

1. hatching of the first brood of nauplii occurs within 24-hrs of spawning;
2. females produced new ovisacs at ~ 1 day intervals, again with hatching occurring within that 24-hrs;
3. "females that were isolated from males produced only two clutches of viable eggs". Additional ovisacs were produced (making it appear that the female is reproductive), but the "succeeding clutches of eggs were aborted or shed off within 48 hrs and never hatched out".

Of course, the reproductive biology of *P. forbesi* may differ from that of the congener *P. annandalei*; however, in the absence of contradictory empirical evidence, Occam's razor would dictate otherwise.

We are left to conclude that **Teh *et al*.'s reported results for 'total number' and 'mean number per female' for the nauplii and juveniles are incorrect, and that their analyses of that data are similarly incorrect**.

Interestingly, in Teh *et al*.'s analyses of the 'total number' and 'mean number per female' of adults produced during the study, the number of adults counted on each progressive 'count day' were **NOT** summed in similar fashion, with Teh *et al*. instead evaluating on the count data from a single 'count day' (Day 31).

Comment #2. While it is believed that Teh *et al*.'s count data are incorrect, let us assume for a moment that they are in fact correct. The organism counts using Teh *et al*.'s summation method are summarized in Table 1 below. When their juvenile count data are analyzed using CETIS (a statistical software specifically designed to analyze aquatic toxicity data), the NOEC and LOEC are shown to be 0.79 mg/L TAN and 1.62 mg/L TAN (Table 2 below), NOT the lower concentrations reported by Teh *et al*.

It should noted that CETIS is the statistical software most commonly used by toxicity testing labs to analyze toxicity test data, and is believed to be the statistical software used at the UC Davis Aquatic Toxicology Lab; indeed, Teh *et al*. used CETIS to analyze their Subtask 3-4-1 and Subtask 3-4-2 experimental data as evidenced in Appendices IV and V of their report.

It should also be noted that our assessment of problems with Teh *et al*.'s statistical analyses should not be interpreted as indicating that there was no effect resulting from the ammonia, but

simply that the experimental data do not support any differences that were observed as being statistically significant.

Table 1. Production on *Pseudodiaptomus forbesi* nauplii, juveniles, and adults (from Appendix I in Teh *et al*. report)

Test Treatment (mg/L TAN)	Test Replicate	Total # of *Pseudodiaptomus forbesi* Life Stage Counted			
		Nauplii[A]	Juveniles[A]	Adults[A] (counts made only on Day 31)	Adults[B] (counts made as for nauplii & juveniles)
Control	A	86	38	11	93
	B	100	73	26	178
	C	68	45	7	122
	D	75	52	3	52
0.36	A	60	27	0	1
	B	62	57	3	36
	C	83	79	18	167
	D	71	43	7	77
0.79	A	24	48	10	77
	B	64	31	4	45
	C	41	17	1	17
	D	52	22	8	77
1.62	A	47	1	0	0
	B	32	0	0	0
	C	46	14	5	28
	D	54	23	19	108
3.23	A	15	1	1	4
	B	39	1	1	6
	C	42	18	13	83
	D	30	13	5	34

A - For the nauplii and juveniles, Teh et al. summed the progressive counts on successive days as separate individuals; as explained in our review, this is believed to be erroneous, and is inconsistent with the counts of the "produced" adults which consist of the number of adults that were alive on Day 31 of the test.

B - Counts of "produced" adults using the summation of the progressive counts on successive days as separate individuals (as used by Teh et al. for the nauplii and juveniles); as explained in our review, this is believed to be erroneous.

Table 2. Comparative analyses of juvenile and adult production in the 31-day test (from CETIS analysis of juvenile data using Teh et al. summation method)

Statistical Endpoint	Juveniles		Adults	
	Teh et al. Analyses	CETIS Analyses	Teh et al. Analyses	CETIS Analyses
NOEC =	0.36 mg/L TAN	0.79 mg/L TAN	**<0.36 mg/L TAN**	**3.23 mg/L TAN**
LOEC =	0.79 mg/L TAN	1.62 mg/L TAN	**0.36 mg/L TAN**	**>3.23 mg/L TAN**
Chronic Value =	1.13 mg/L TAN	1.13 mg/L TAN	<0.36 mg/L TAN	>3.23 mg/L TAN

Chronic Value = geometric mean of NOEC and LOEC.

Comment #3. Teh *et al*.'s apparently erroneous statistical analysis of the adult data is even more significant (Table 2). Teh *et al*. reported that the NOEC and LOEC for adults were <0.36 mg/L TAN and 0.36 mg/L TAN, respectively. However, their inter-replicate variability for that endpoint is so high (CVs ranged from 70% to 150%) that even qualitative evaluation suggests otherwise. CETIS analysis indicates that the NOEC and LOEC are 3.23 mg/L TAN and >3.23 mg/L TAN.

Again, it should be noted that our assessment of problems with Teh *et al*.'s statistical analyses should not be interpreted as indicating that there was no effect resulting from the ammonia, but simply that the experimental data do not support any differences that were observed as being statistically significant. Certainly, the NOECs and LOECs resulting from this experiment should not be considered suitable for use in a regulatory framework.

Comment #4. While the apparent problems with Teh *et al*.'s counting method for the nauplii and juveniles would seem to preclude the experimental data from being able to be evaluated in any meaningful fashion, we believe that the data *can* be evaluated. If we are correct that the test organisms were left in the original beakers to subsequently develop into the progressive life stages, then the data are suggestive of the maximum number of organisms at a given life stage that occurred in any individual beaker, at least in regards to the counts that were made. It is acknowledged *a priori* that many assumptions must be made here (i.e., that additional organisms did not occur on the interim days that counts were not made). Accordingly, the resultant "maximum number of organisms observed in a given beaker" data must be acknowledged as not being definitive.

In examining the data in this fashion, it is readily seen that the maximum number of nauplii or juvenile in each replicate (or for each treatment) did not always fall on the same 'count day' (e.g., the maximum number of nauplii at the Control treatment [total number of nauplii = 70] were counted on Day 5, whereas the maximum number of nauplii at the 0.36 mg/L TAN treatment [total number of nauplii = 91] were counted on Day 7). Therefore, and in order to best characterize the best case 'total number' and 'mean number per female' for the nauplii and juveniles, the test result data summarized in the table below are the maximum counts of

organisms observed in each individual replicate, regardless of the 'count day'. These data are summarized in Table 3 below.

CETIS analysis of these "maximum count" data are summarized for the nauplii, juveniles, and adults in Tables 4a-c, below.

Table 3. Production on *Pseudodiaptomus forbesi* nauplii, juveniles, and adults using "maximum number of organisms observed in a given replicate on any given day of the test" (from Appendix I in Teh *et al*. report)

Test Treatment (mg/L TAN)	Test Replicate	*Pseudodiaptomus forbesi* Life Stage		
		Nauplii	Juveniles	Adults
Control	A	20	10	17
	B	20	20	30
	C	15	15	20
	D	15	15	12
0.36	A	21	15	1
	B	25	20	8
	C	25	30	30
	D	25	20	16
0.79	A	15	15	10
	B	20	10	12
	C	16	6	6
	D	20	10	15
1.62	A	15	1	0
	B	10	0	0
	C	10	6	5
	D	10	9	19
3.23	A	12	1	1
	B	7	1	1
	C	10	10	13
	D	10	6	8

Table 4a. Comparative analyses of nauplii production in the 31-day test (from CETIS analysis of data).

Statistical Endpoint	Teh *et al*. Counts	"Maximum # Organisms Observed"
NOEC =	0.36 mg/L TAN	0.79 mg/L TAN
LOEC =	0.79 mg/L TAN	1.62 mg/L TAN
Chronic Value =	0.53 mg/L TAN	1.13 mg/L TAN

Chronic Value = geometric mean of NOEC and LOEC.

Table 4b. Comparative analyses of juvenile production in the 31-day test (from CETIS analysis of data).		
Statistical Endpoint	Teh *et al*. Counts	"Maximum # Organisms Observed"
NOEC =	0.79 mg/L TAN	0.79 mg/L TAN
LOEC =	1.62 mg/L TAN	1.62 mg/L TAN
Chronic Value =	1.13 mg/L TAN	1.13 mg/L TAN

Chronic Value = geometric mean of NOEC and LOEC.

Table 4c. Comparative analyses of adult production in the 31-day test (from CETIS analysis of data).		
Statistical Endpoint	Teh *et al*. Counts	"Maximum # Organisms Observed"
NOEC =	3.23 mg/L TAN	1.62 mg/L TAN
LOEC =	>3.23 mg/L TAN	3.23 mg/L TAN
Chronic Value =	>3.23 mg/L TAN	2.29 mg/L TAN

Chronic Value = geometric mean of NOEC and LOEC.

3. Comments on Subtask 3-4-1 (Effects of ammonia on nauplii production over 3 days)

Comment #5. In this test, Teh *et al*. exposed individual gravid female copepods to TAN concentrations of 0 (control treatment), 0.38, and 0.79 mg/L for 3 days after which the number of nauplii produced were counted. The results of this test have been summarized in the table below.

From data reported in Teh et al.'s Table 12 and Appendix V:

Table 5. Effects of ammonia on *P. forbesi* production of nauplii over 3 days (Teh *et al*. Subtask 3-4-1).	
TAN Concentration (mg/L)	Mean # of Nauplii per Female
Control	7.6
0.38	5.5
0.79	5.4

The results from this test are somewhat troubling in that, while technically monotonically increasing as the ammonia concentration increases, no apparent concentration-response relationship is observed between the 0.38 mg/L treatment and the 0.79 mg/L treatment. One would expect that as the TAN concentration increases from 0.38 mg/L (a presumably toxic concentration) to 0.79 mg/L (a two-fold greater concentration), there should be an increase in the toxic response – this is a fundamental paradigm of toxicology.

We have already seen in the data evaluations presented above that there is variability in toxic responses made by these organisms. Indeed, in some cases, the variability has been so extreme as to preclude a meaningful statistical analysis (as in the case of the adult data from the 31-day test). The absence of the expected concentration-response in the current test (Table 5) suggests that variability in organism response is occurring (the CV was 48% in the 0.38 mg/L treatment) such that the treatment means may be deviating from the true population mean (in statistical terms, this is referred to as a "false positive" or a "false negative").

In the present case, it is impossible to determine which of the two test responses is deviating most from the true population mean response. However, it is worth noting that:

1. there were two replicates at the 0.38 mg/L treatment that had 10 nauplii (the highest number observed in ANY replicate) whereas there was only one replicate at the control treatment that had 10 nauplii, and
2. the CV at the 0.38 mg/L treatment was 48%, which was markedly higher than at the Control or 0.78 mg/L treatment.

This is suggestive that the variability at the 0.38 mg/L treatment was elevated and may have resulted in a false positive, such that the observed mean response of 5.5 nauplii per female was lower than the true population mean. If correct, then the conclusion(s) drawn from the test data may not reflect true conditions, and the true LOEC could be 0.79 mg/L, and not 0.38 mg/L. At a minimum, the absence of the expected concentration-response should cast enough uncertainty on the test results as to make them inappropriate for regulatory decision-making.

Comment #6. It is fortunate that multiple sets of test data from the study allow comparison of results between tests; for instance, the results of Subtask 3-4-1 can be compared to those generated in the earlier Subtask 3-3 (31-day) test in which gravid females were exposed to varying concentrations of TAN and counts of nauplii produced after 3 days were counted, but were also counted after 5 days and 7 days (recall that counts made on progressive count days are not believed to be all new organisms). The Subtask 3-3 data are summarized in Table 6 below, along with the data from Task 3-4-1.

If one were to "cherry-pick" the Day 3 data and exclude the additional data, then Teh *et al*.'s conclusion for the Subtask 3-4-1 might stand. However, by extending the observation period beyond 3 days, it becomes evident that not only is there no reduction in nauplii production at 0.36 mg/L TAN, but nauplii production actually appears to be *increased* relative to the control treatment (the maximum mean # of nauplii on Day 5 at the 0.36 mg/L TAN treatment is **31% greater** than the highest mean # of nauplii produced in the Control treatment on any of the count days). Furthermore, CETIS analysis indicates that there were no statistically significant reductions in nauplii production at the 0.36 mg/L (Table 7). Even if we use the count summation used by Teh et al., by extending the counts beyond 3 days, it becomes apparent that there is no statistically significant difference between the response at 0.36 mg/L TAN and the Control treatment. This certainly creates a very significant uncertainty over the results of the Subtask 3-4-1 test of the effects of ammonia on nauplii production over 3 days.

Table 6. Effects of ammonia on *P. forbesi* nauplii produced over 3 and 5 days.

Teh *et al.* Study Task	TAN Treatment (mg/L)	Mean Number of Nauplii per Female		
		Day 3	Day 5	Sum through Day 5 (Day 3 + Day 5)[A]
Subtask 3-4-1	Control	**7.6**	not counted	not counted
	0.38	**5.5**	not counted	not counted
	0.79	**5.4**	not counted	not counted
Subtask 3-3	Control-A	5.67	6.67	12.33
	Control-B	6.67	6.67	13.33
	Control-C	5	5	10
	Control-D	5	5	10
	treatment mean	**5.6**	**5.8**	**11.4**
	0.36-A	3	5	8
	0.36-B	2.33	8.33	10.67
	0.36-C	3.33	8.33	11.67
	0.36-D	3.33	3.33	6.67
	treatment mean	**3.0**	**6.3**	**9.3**
	0.79-A	0.33	1.67	2
	0.79-B	6.67	3.33	10
	0.79-C	2.67	2.67	5.33
	0.79-D	6.67	4	10.67
	treatment mean	**4.1**	**2.9**	**7.0**

A – These counts are made using method of Teh *et al.*, which assumes that the progressive counts on successive days are separate individuals; as explained in our review, this is believed to be erroneous.

Table 7. Comparison of nauplii production test results (all results expressed as mg/L TAN) (from CETIS analysis of data)

Statistical Endpoint	**Subtask 3-4-1**	**Subtask 3-3**				
	Day 3	Day 3	Day 5	Day 3 + Day 5[A]	Total (31 days)[A]	Total (31 days)[B]
NOEC =	<0.38	3.23	0.36	0.36	0.36	0.79
LOEC =	0.38	>3.23	0.79	0.79	0.79	1.62
Chronic Value =	<0.38	>3.23	0.53	0.53	0.53	1.13

Chronic Value = geometric mean of NOEC and LOEC.

A – These counts are made using method of Teh *et al.*, which assumes that the progressive counts on successive days are separate individuals; as explained in PER's review, this is believed to be erroneous.

B – These counts are made using what is believed to be the best remaining method: identifying the maximum number of nauplii observed on any given day for each replicate (this assumes that the individuals were left in the replicate beakers and were counted again and again on progressive days [i.e. repeated measures]).

It could be argued that this phenomenon is the result of ammonia having caused a delay in egg hatching, and the 31-day data are certainly suggestive of that. However, the only way to address that would have been to have some information from the scientific literature on the egg gestation period for this species, coupled with testing being performed under the current test conditions using females with egg sacs of the same age.

4. Comments on General Report Quality

Comment #7. The description of methodology is inadequate. Much of the experimental approach is not described at all, leaving it to the reader's imagination to assume what was actually done. Given the novel testing approach (i.e., these tests are acknowledged by the author as being "non-standard", and that the "test methods are considered developmental"), and the potential use of the information generated by this study by regulatory decision-makers, a more thorough description of the experimental methodology is essential.

Comment #8. QA measures that would typically be expected to be performed for testing performed as per the EPA manual cited by Teh *et al*. (e.g., evaluation of test data variability [e.g., PMSD evaluation, or in its absence, assessment of CV], reference toxicant testing) were not performed. As the potential exists for data variability and/or organism sensitivity to have profound effects on the test results and their interpretation, the absence of these evaluations and their complete omission from discussion in the report seriously compromises the use of the data or results in regulatory decision-making.

A standard QA element of NPDES testing, concurrent reference toxicant testing is even more important in a "special study" that might be used to generate information that could be used in regulatory decision-making. It is essential that the user(s) know if the organisms were responding to the toxicant in a typical fashion. If the particular batch of test organisms used in a given test were more sensitive to toxicant stress than is typical (which can only be determined by the reference toxicant test), then use of the data from that test should be qualified as it is not representative of how the population might be expected to respond. This might result from slightly impaired organism quality (e.g., organisms may have been unhealthy and/or injured), or from experimental technique (was the quality of the performance of the tests as competent as the "typical" performance).

If the batch of organisms used in the test were more sensitive, then that test might produce 'false positives'. Conversely, if the organisms used in the test were less sensitive, then the results of the test might be 'false negatives'. Without reference toxicant testing, one will never know whether or not one of these scenarios occurred.

Comment #9. Due to photosynthetic processes, autotrophic waters typically exhibit a diel pH cycle, with pH minima at dawn and pH maxima at mid-to-late afternoon. It is unclear whether the pH data reported by Teh et al. for ambient waters in the Sacramento River at Hood (Tables

1a and 1b) or for the sites in the Cache Slough complex (Table 2) are "snapshot" pH values measured at a particular time of day or whether they are the diel means. This distinction could be important, as Teh et al.'s subsequent decision to perform the ammonia toxicity tests at pH7.4 and pH7.8 may or may not best reflect the actual *in situ* pH exposure regime in these natural waters.

5. Comments on Persistent Issues/Problems: Test Solution Water Quality Characteristics

It should be noted at the outset that most of the comments on test solution water quality characteristics that are provided below should not be considered in and of themselves to necessarily compromise the validity or interpretation of this study's findings. However, they do raise questions regarding the performance by the lab in the basic preparation of test waters and their reporting of data.

Comment #10. Teh *et al*. describe the culture water as being "standard moderately hard fresh water", and that water quality was measured weekly and maintained with an alkalinity of 80 mg/L. In the EPA test guidelines, the final water quality alkalinity for "moderately hard water" should be in the range of 57-64 mg/L. While some variability is to be expected, having an alkalinity that is consistently around 80 mg/L is suggestive that the water was prepared incorrectly.

Comment #11. Teh *et al*. describe the culture water as having "<1 μg/L" ammonia. However, in Section 2A-2 of the report, it is stated that the ammonia analysis had a detection limit of 20 μg/L, and a reporting limit of 100 μg/L. It would therefore seem impossible to be able to report the culture water ammonia levels as "<1 μg/L".

Comment #12. In the Task 2A-1 description of "Methods for acute and chronic toxicity testing" it states that "moderately hard water was prepared according to methods published in EPA-821/R-02/013 and was used as culture and testing medium for all tests". They later describe some testing as being performed at 2.0 ppt salinity. Which is it?

If, as seems likely, the 2.0 ppt test water was prepared by adjusting the salinity of moderately-hard water upwards via the addition of salts, this needs to be described (i.e., what kind of salts?; does the salt(s) used result in a test solution salinity with the ion composition of saline Delta waters?).

Comment #13. The water quality data reported in Teh *et al*.'s Table 4 (attached below) seem highly questionable. For ***each*** of the six test treatments described, the alkalinity values are reported as exactly 50 mg/L and the hardness values are reported as exactly 206.7 mg/L. With 24+ years of experience in doing laboratory tests of this type (including literally thousands of test solution alkalinity and hardness analyses), the complete absence of variability seems very unusual (and unlikely). Even more unlikely is the reporting of the test solution conductivities as 975 ± 17 μMHOS for ***each*** of the six test solutions; conductivity measurements are even more

variable than are hardness and alkalinity measurements, and the absence of variability between treatments for this parameter seems questionable.

From the Teh *et al*. report:

Table 4. Summary of water chemistry and ammonia concentrations during acute toxicity testing

Nominal mgTAN/L	Temp (°C)	Alkalinity (mg/L)	conductivity (µHMOS)	DO (mg/L)	Hardness mg/L	pH	Salinity ppt	Measured mg TAN/L	Measured mg UIA/L
0	20	50	975 ± 17	>8	206.7	7.80±0.09	0.5	<0.1	0.000
1	20	50	975 ± 17	>8	206.7	7.80±0.04	0.5	0.87±0.01	0.022±0.000
2	20	50	975 ± 17	>8	206.7	7.79±0.03	0.5	1.60±0.00	0.039±0.000
4	20	50	975 ± 17	>8	206.7	7.79±0.02	0.5	3.20±0.00	0.078±0.000
6	20	50	975 ± 17	>8	206.7	7.79±0.02	0.5	4.75±0.05	0.116±0.001
8	20	50	975 ± 17	>8	206.7	7.79±0.01	0.5	6.25±0.05	0.156± 0.001

Comment #14. The description of "moderately hard water" from the EPA test manual cited by Teh *et al*. is shown below. Comparison of the alkalinity and hardness concentrations reported in Teh *et al*.'s Table 4 (see above) reveals that they do not fall within the expected water quality range for those parameters. Of particular note is Teh *et al*.'s reported hardness of 206.7 mg/L, which is over twice the EPA's 80-100 mg/L range. This suggests that the moderately hard water used in this test was prepared improperly.

From the EPA test guidance manual:

Water Type	Reagent Added (mg/L)[2]				Approximate Final Water Quality		
	$NaHCO_3$	$CaSO_4 \cdot 2H_2O$	$MgSO_4$	KCl	pH[3]	Hardness[4]	Alka-linity[4]
Very soft	12.0	7.5	7.5	0.5	6.4-6.8	10-13	10-13
Soft	48.0	30.0	30.0	2.0	7.2-7.6	40-48	30-35
Moderately Hard	96.0	60.0	60.0	4.0	7.4-7.8	80-100	57-64
Hard	192.0	120.0	120.0	8.0	7.6-8.0	160-180	110-120
Very hard	384.0	240.0	240.0	16.0	8.0-8.4	280-320	225-245

Comment #15. The alkalinity, hardness, and conductivity concentrations in Teh *et al*.'s Table 6 (attached below) are again reported as being identical for ***each*** of the six test treatments. And again, the absence of variability between treatments for these parameters seems unlikely.

From the Teh *et al*. report:

Table 6 Summary of water chemistry and ammonia concentrations during acute toxicity testing

Nominal mgTAN/L	Temp (°C)	Alkalinity mg/L	conductivity µHMOS	DO mg/L	Hardness mg/L	pH	Salinity ppt	Measured mgTAN/L	Measured mgUIA/L
0	20	20	580 ± 24	>8	140	7.40±0.000	0.5	<0.1	0.000
1	20	20	580 ± 24	>8	140	7.40±0.004	0.5	0.60±0.008	0.007±0.000
2	20	20	580 ± 24	>8	140	7.42±0.003	0.5	1.33±0.021	0.015±0.000
4	20	20	580 ± 24	>8	140	7.41±0.000	0.5	2.75±0.022	0.031±0.000
6	20	20	580 ± 24	>8	140	7.40±0.000	0.5	3.88±0.070	0.043±0.000
8	20	20	580 ± 24	>8	140	7.40±0.003	0.5	5.47±0.029	0.060±0.000

Comment #16. Comparison of the alkalinity and hardness concentrations reported in Teh *et al.*'s Table 6 (see above) reveals that they also do not fall within the expected water quality range for those parameters for EPA moderately hard water. Again, this suggests that the moderately hard water used in this study was prepared improperly.

Comment #17. Even more problematic is the fact that a comparison of the water quality values reported for the pH7.8 test (Teh *et al.*'s Table 4) and the pH7.4 test (Teh *et al.*'s Table 6) reveals that the concentrations for alkalinity, hardness, and conductivity diverge significantly between that prepared for the pH7.8 test and that prepared for the pH7.4 test. This extreme variability not only supports the assertion that preparation of the moderately hard water used in this study was incorrect, but that it was also inconsistent.

Comment #18. Although a detailed summary table of water quality characteristics was not provided for the test performed under Subtask 3-2 (determining the toxicity of ammonia to juvenile *P. forbesi* at different pH levels), the narrative description of "water chemistry for all treatments were maintained at 0.5 ppt (salinity), 75.6 ± 0.3 mg/L (alkalinity), 983.3 ± 2.7 µMHOS (conductivity), 163 ± 1.8 mg/L (hardness)" again implies that there was no variability in test solution water qualities between the test treatments.

Comment #19. As before, the alkalinity of 75.6 mg/L and hardness of 163 mg/L reported for the Subtask 3-2 test are inconsistent with the EPA ranges for moderately hard water.

Comment #20. In the acute toxicity test performed with '3-day old' nauplii (Subtask 3-4-2), the test was performed at a salinity of 2.0 ppt. This deviates from the approach that had been used in the previous acute tests performed as part of this study. Why was 2.0 ppt water used in this test?

6. Comments on Persistent Issues/Problems: Presentation of Test Data and Results

Comment #21. Tables and Figures are incorrect and/or misleading. Data reported in tables are apparently mislabeled and/or are misleading. For many of the test result descriptive figures, the data points (and error bars) represented in the figures do not appear to match the data in the tables nor the raw data provided in the appendices. In some cases, hypothetical modeled data (e.g., LC point estimates) are shown in the figures as if they were actual experimental data. Specific examples of these problems are described in detail in the comments below.

Comment #22. In Teh *et al*.'s Table 1A, the data in the fourth column are labeled as "$NH4^+$-N (TAN)". As TAN (total ammonia nitrogen) = $NH4^+$ + NH3, then the data in the column must be either $NH4^+$ or TAN, but cannot be both.

Comment #23. Although not critical to the understanding of the data intent, Tables 3a, 3b, and 3c should include the units of measurement for the abundance counts (e.g., organisms/L, etc.).

Comment #24. In Teh *et al*.'s Table 4, the UIA (un-ionized ammonia) is labeled as "measured". However, the description for Sub-Task 2A-2 indicates that the UIA concentrations were calculated based upon the measured TAN and pH levels.

This error also occurs in Table 6 and Table 8.

Comment #25. The test result "data points" purportedly represented in Figures 2A and 2B are not actual data from the study's experiments, but rather are hypothetical LC point estimates that have been modeled. Furthermore, it is not clear whether the linear equation and R^2 value shown with each graph were derived from actual experimental data or simply from the hypothetical modeled point estimates. This is unfortunate, as it misleads the non-toxicologist into perceiving that the data set is more comprehensive that it actually is, and that the fit of the data to the linearized model is tighter than it may actually be. With regards to the former, this is particularly problematic as the survival response never exceeds a >60% reduction relative to the Control treatment, even after extending the normal test duration from 96-hrs to 6 days. The result is that the nature (i.e., steepness) of the resulting modeled concentration response curve requires some degree of extrapolation.

This misleading presentation of "data" also occurs in in Figures 3A and 3B, and again in Figure 5. As before, this can mislead the non-toxicologist into perceiving that the data set is more comprehensive that it actually is, and that the fit of the data to the linearized model is tighter than it may actually be.

Comment #26. The 31-day test "mean total number of (nauplii, juveniles, or adults) per beaker" data shown in the graphs comprising Figures 6A-6C are not consistent with the data reported in Teh *et al*.'s Table 11 (each mean shown in Table 11 would be multiplied by 3 to calculate the

total number per beaker) or the data reported in Appendix III (the "sum total number" for each treatment in Appendix III would be divided by 4 to calculate the total number per beaker). Equally (if not more) important to note is that fact that the error bars shown around each treatment mean in Figures 6A-6C are not the true representation of variability around the means (and in fact, appear to be identical for each of the data points shown in the graphs). Again, this is VERY misleading, particularly for the adult data that, as stated above, are EXTREMELY variable.

7. Comments on Persistent Issues/Problems: General Test Methodology

Comment #27. It is stated that for the acute tests, the replicates consisted of 500 mL of test solution in a 600 mL beaker, each containing 20 organisms. It goes on to state that 80% of the test solution was renewed daily, but there is no description of how this is done.

There are a few ways in which such test solution renewal can occur: (1) the old test solution can be carefully poured out while trying to retain the test organisms within the beaker, or (2) the test solution can be siphoned out, typically using some type of screening apparatus to make sure that the test organisms are not sucked into the siphon. As stated by Teh et al. under his Task 2A description, observational staging of live copepods is "difficult because of their (small) size and speed in swimming", which is consistent with experimental work with copepods performed in the PER Lab. Accordingly, it would seem to be impractical to be able to perform test solution renewal by method #1 above due to the expected loss of organisms that would occur during the pouring out of the old medium. In method #2 above, the potential for the negative suction pressure of the siphon to trap organisms against the screen (potentially injuring the organisms) is a very real possibility. In this scenario, a non-lethal reduction in swimming performance could result in the organism being injured against the screen, potentially resulting in an artifactual mortality.

The large beaker and large test volume used by Teh *et al*. is best limited to fish and mysid organisms which are much larger and more amenable to the test solution renewal procedures described above. The test approaches for the similarly-sized *Ceriodaphnia dubia* from the acute and chronic EPA guidance manuals would have been more appropriate for use with the copepods.

Comment #28. During the tests, the copepods were fed a commercial algal mix daily prior to water renewal. It is not stated how long before the water renewal that the feeding occurred, but it should have been at least long enough for the organisms to feed (note – most acute tests are actually performed without any feeding in order to eliminate any artifacts that the presence of the food might otherwise have on the test results, but it is not known whether this is feasible with *P. forbesi* [i.e., can *P. forbesi* survive without feeding for 96 hrs?]). It seems reasonable to expect that the addition of the commercial food introduced ammonia into the test solutions [ammonia could be present as part of the nutrients that the commercial supplier used to grow the algae

and/or also could result from the senescence of any algal cells in the food medium), and that there may also have been some effect on the pH regime as well. Since the "old" test solution ammonia levels were taken prior to feeding, this is impossible to determine.

In the chronic tests, the organisms were fed the commercial algal food daily, but the test solution renewal was only performed as 2-3 day intervals, increasing the exposure time to any changes in ammonia and/or pH that might have accompanied the food additions.

Comment #29. One of the major problems in trying to conduct a toxicity test at a constant pH is the tendency of test solutions (including moderately hard water) to drift upwards (if the pH had been adjusted to <pH8) or downwards (if the pH had been adjusted to >pH8) during each 24-hr exposure period. It is exceedingly difficult to maintain constant pH at the initial adjusted level, and typically requires: (1) frequent manual pH re-adjustments during the day, (2) use of a buffer in the test solution, or (3) use of controlled CO_2 headspace. Our laboratory has performed literally dozens and dozens of ammonia toxicity studies over the years, and always must use some method of pH control to have a test solution hold a target pH. This problem is also acknowledged in the EPA TIE manual description of the "Graduated pH" TIE tests designed to identify ammonia (and other pH-labile toxicants) as a cause of toxicity, in which it is stated "Perhaps the greatest challenge faced in this graduated pH test is that of maintaining a constant pH in the test solution" and considerable detail is provided as to means that can be used to achieve pH control.

Despite this well-known problem, the test solution pH levels reported for each of the six test treatments in Teh *et al*.'s tests are remarkably stable with unusually low variability. Based upon my experience, the only way that this would be possible would be if the results being reported are ***only*** for the test solution pH levels measured in the fresh test solution (immediately prior to or at the time of first usage) or possibly that some type of pH control was used but not included in the description of test methodology (in which case, the variability would still be expected to be higher than what is reported).

It should be noted that in the only place in the Teh *et al*. report where test solution pH levels are clearly reported for both the beginning and end of each 24-hr exposure period (Teh et al.'s Table 8), the pH levels did in fact drift, from 7.0 to 7.29, from 7.4 to 7.59, and from 7.8 to 7.89. While generally supporting our assertion, I believe that the magnitudes of these drifts are still moderate at best.

In an attempt to provide my own validation, our lab prepared four replicates of correctly-prepared moderately-hard water, with 500 mL in a 600 mL beaker (identical to the replicate scale used in Teh et al.'s experiments). Two of these were adjusted to pH7.4 (via drop-wise addition of reagent-grade HCl), and the other two were similarly adjusted to pH7.8. These were maintained under a 16L:8D photoperiod at 20°C for 24 hrs after which the test solution pH levels were re-

measured. The results of this experiment follow, and clearly indicate that significant pH drift can occur:

pH Drift in Moderately-Hard Water over 24 Hrs.			
Nominal pH Level	Replicate	Initial Measured pH	pH After 24 Hrs
pH7.4	Rep A	7.40	7.98
	Rep B	7.40	7.87
pH7.8	Rep A	7.80	7.99
	Rep B	7.80	8.00

Depending upon the magnitude of the pH drift that likely occurred in Teh *et al*.'s pH7.4 and pH7.8 experiments, the final conclusions regarding toxicity of ammonia at a specific pH or calculations of UIA (un-ionized ammonia) based on the reported data could be suspect.

8. Comments on Task 3 ("Acute and chronic effects of ammonia on *P. forbesi*")

Comment #30. In the discussion of the physiological mechanisms of ammonia excretion, Teh *et al*. state that "high concentrations of IA (ionized ammonia) in the culture medium may compete with sodium ions flux, thereby diminishing body concentrations of this important sodium salt". However, ammonium ion (= IA) efflux and sodium ion influx are coupled via an ATPase active transport mechanism, for which the key determinants should be expected to be ammonium ion supply in the cell, sodium ion supply in the culture medium, and available ATPase. Furthermore, the external molar concentrations of ammonium ion relative to that of the concentrations of sodium should be expected to be small (likely by a factor of at least two orders of magnitude), such that the minor changes in the external ammonium ion concentration (relative to the external or internal sodium concentrations) would be more than offset by the much greater concentrations of sodium. Accordingly, we do not believe that this is significant.

9. Comments on Subtask 3-2 (Effects of pH on ammonia toxicity to juvenile *P. forbesi*)

Comment #31. In this test, the toxicity of ammonia to juvenile copepods at pH7.4 was determined for a TAN concentration of 3.6-3.7 mg/L. Interestingly, in the previous Subtask 3-1B, the toxicity of ammonia to juvenile copepods at pH7.4 was determined at the similar TAN concentration of 3.9 mg/L. The comparative survival results for these two tests are shown in the table below.

Comparative results of toxicity test of ammonia to juvenile copepods at pH7.4.		
Test	Measured TAN (mg/L)	Mean % Survival
Sub-Task 3-1B	3.9	20
Sub-Task 3-2	3.6-3.7	56.25

Although seemingly significant, the variability between the responses in these two tests should not be considered unusual, particularly for a novel test with a "non-standard" species. Indeed, it should be recognized *a priori* that different tests with the same species can exhibit variable responses between tests. This results from several factors, with differences in organism sensitivity being of obvious concern. This is precisely why the performance of concurrent reference toxicant testing is so essential in any type of testing for which the test results might be used for regulatory decision-making. As the comparative test results above illustrate, the variability being exhibited raises the question of "Which is right?" (or even, "Which is more likely to be right"). Without concurrent reference toxicant tests, this is difficult to ascertain.

Perhaps more troubling is the absence of any discussion by Teh et al. regarding the variability in their test response data, either between tests or within tests (i.e., inter-replicate variability). Without such acknowledgement, it is left for the non-scientist to assume that the data as presented are definitive. Moreover, it raises the question of whether the data from this study are adequate (or 'ready') for use in regulatory decision-making

FINAL COMMENT

It is important to note that this critical review is not intended to negate Teh *et al*.'s general observations that ammonia is toxic to naupliar, juvenile, and/or adult *P. forbesi* at elevated concentrations and that this toxicity is strongly influenced by pH. Indeed, the primary question of 'what are the effects of ammonia on *P. forbesi*' is relevant and Teh *et al*.'s study results certainly compel a more thorough examination of this. However, the problems associated with Teh et al.'s experimental methodology for Subtasks 3-3 and 3-4-1 and significant questions regarding the analysis of the resulting data do indicate that the quality of the work should preclude the resulting 'critical threshold' data (i.e., NOECs, LOECs, and point estimates [e.g., ECx, LCx, and ICx values]) from being used for regulatory purposes.

References Cited:

Golez MSN, Takahashi T, Ishimaru T, Ohnoa A (2004) Post-embryonic development and reproduction of *Diaptomus annandalei* (Copepoda: Calanoida). Plankton Biology & Ecology 51(1):15-25.