APPENDIX G

WHITE PAPER

Discussion and Recommendations Related to *Arbacia punctulata* Whole Effluent Toxicity Testing Using Combined Effluent from the Bayamón, Puerto Nuevo and Bacardí Wastewater Treatment Plants

Prepared for

Bacardi Corporation

Prepared by

CH2MHILL

May 2007

Introduction

The wastewater treatment plant at the Bacardi Corporation (Bacardí) rum distillery in Cataño, Puerto Rico shares an ocean outfall with the Puerto Rico Aqueduct and Sewer Authority (PRASA) Bayamón and Puerto Nuevo regional wastewater treatment plants (RWWTPs). The combined effluent of the three facilities is discharged more than one-half mile offshore at a depth of 140 ft below mean sea level into dynamic ocean waters through a high-rate (>100:1 dilution) diffuser.

Bacardí and PRASA have submitted requests for National Pollutant Discharge Elimination System (NPDES) permit renewals for the three wastewater treatment facilities. The Puerto Rico Environmental Quality Board (EQB), in its statement of intent to issue a water quality certificate (WQC) for the existing NPDES permit for the Bacardi Corporation (Bacardí), required "a detailed description of the methodology to be utilized in the performance of the tests" for three sensitive marine test species used to evaluate possible short- and long-term effects of mixed effluent from the Bacardí, Bayamón, and Puerto Nuevo wastewater treatment plants (EQB, 2001).

Similarly, the new WQC that will be incorporated in the new NPDES permit (Permit No. PR0000591) may require acute and chronic whole effluent toxicity (WET) tests (bioassays) using the sheepshead minnow (*Cyprinodon variegatus*) and a mysid shrimp (*Mysidopsis bahia*), as well as chronic toxicity tests for the sea urchin *Arbacia punctulata (Arbacia)* using the existing EQB-approved WET test protocols. Per these protocols, the tests are performed on flow-proportional samples taken from the three effluents. A 24-hr composite sample is obtained from each facility; flow during the compositing period at each plant is recorded. The three effluent samples are sent to the bioassay laboratory with instructions about how to combine the samples in a proportional fashion based these flows. These flow-proportional composite samples are then used for WET testing and data evaluation.

The flow-proportional composite approach allows for an evaluation of whatever synergisms and/or antagonisms may be present in the three effluents in relation to the relative toxicity of the mixed effluent that is ultimately discharged to the marine environment. The WET test results are used by EQB to evaluate whether its receiving water toxicity requirements will be met at the edge of a small permitted mixing zone that is established around the outfall diffuser. A series of four tests are conducted during the first year of the permit, followed by annual testing during the remainder of the 5-year permit cycle to ensure that the relative toxicity of the effluent is not exceeding the receiving water requirements.

Bacardí has complied with similar requirements in its existing NPDES permit, reporting the WET results in terms of both statistical hypothesis testing and point estimates of relative toxicity for all three species: the minnow, the mysid shrimp, and the urchin. However, it has recently become apparent that the two data evaluation methods lead to very different conclusions in the case of the *Arbacia* tests.

In brief, the hypothesis-testing method relies on a No Observed Effects Concentration (NOEC) that is based on the statistical difference in variances between control and test populations of the organisms tested. The point estimate method uses a broader range of the WET test data to estimate (through interpolation) a sub-lethal biological response endpoint. Thus, the two methods may result in numerically different estimates of chronic endpoints.

An evaluation of published EPA guidance for WET test data interpretation indicates that the point estimation technique is preferred for purposes of regulatory compliance evaluations. This white paper is intended to clarify the most appropriate method to use for interpreting *Arbacia* test results with respect to both past and future WET test data obtained from the Bacardí, Bayamón, and Puerto Nuevo wastewater treatment plants. It discusses how NOECs derived from hypothesis testing frequently lead to "false positive" toxicity indications and summarizes key issues, presents case-specific data with respect to WET test findings and conclusions, questions whether statistical hypothesis testing should be used to evaluate the results of the chronic definitive bioassays conducted using *Arbacia*, and offers recommendations for what are considered to be appropriate WET test data evaluation methods when using *Arbacia* as a test organism.

Summary of *Arbacia* WET Test Results to Date

For the existing permit, the critical initial dilution (CID) and acceptable toxicity unit concentration (TUc) per the Puerto Rico Mixing Zone and Bioassay Guidelines are 142, equating to an acceptable NOEC of ≥ 0.70% effluent. For the renewal of the WQCs and NPDES permits for the Bacardí distillery and the Bayamón and Puerto Nuevo RWWTPs, which are expected to be issued by EQB and EPA, respectively, in 2007, the CID and compliance TUc are 104, equating to an acceptable NOEC of ≥ 0.96%.

Using the existing permit acceptable NOEC of ≥ 0.70% and statistical hypothesis testing to assess compliance, most of the tests indicate that the permitted chronic toxicity limit for *Arbacia* was not met. It is not clear whether the tests conducted on 8/29/2006 and 11/04/2006 complied at a NOEC of ≥ 0.70%. Using the anticipated NOEC of ≥ 0.96% for the new permit and statistical hypothesis testing to assess compliance, only the tests of 3/16/2006, 11/07/2006, 5/3/2007, 5/5/2007, and 5/17/2007 would have definitively complied with the Permit chronic toxicity limits.

These toxicity interpretations rely exclusively on statistical hypothesis testing to determine the NOEC (using Bonferroni's T-test), which is directly correlated to the degree of statistical variance in controls. Because this variance may be very small among control replicates, T-test results are purely statistically-based (i.e., based on statistical variance alone without respect to biological responses) and therefore are prone to "false positive" or Type I errors.

This is shown in Exhibit 1, where nine out of fourteen tests appear to be toxic (i.e., NOEC < 0.96% effluent) if evaluated by statistical hypothesis testing, but where using alternative EPA-approved (and preferred) data evaluation techniques (IC25[6] and biological significance testing) leads to the conclusion that there is no unacceptable toxicity indicated at the compliance TUc (or 0.96% combined effluent concentration).

In addition to the hypothesis testing-based NOECs, Exhibit 1 shows biologically-based NOEC values. These are based on an EPA test acceptability criterion that does not allow for a test to be considered valid if control fertilization rates are less than 70% (USEPA, 2002). Exhibit 1 also shows point estimates of chronic toxicity based on the IC25, which is commonly used and widely accepted by EPA and other regulatory agencies as a comparable

[6] The IC25 is the percent concentration of a test solution that results in a 25% inhibition of a measurable biological response – in this case fertilization success of *Arbacia* eggs.

value of the chronic toxicity threshold. In the case of the biological significance and the IC25 toxicity evaluations, all of the values are ≥ 0.96% effluent, suggesting that none of the tests indicate unacceptable levels of toxicity. It is noted that IC25 point estimates allow the use of all of the WET test response data to determine, through linear interpolation, the point at which the toxicity response is equal to the target value (i.e., a 25% inhibition of fertilization).

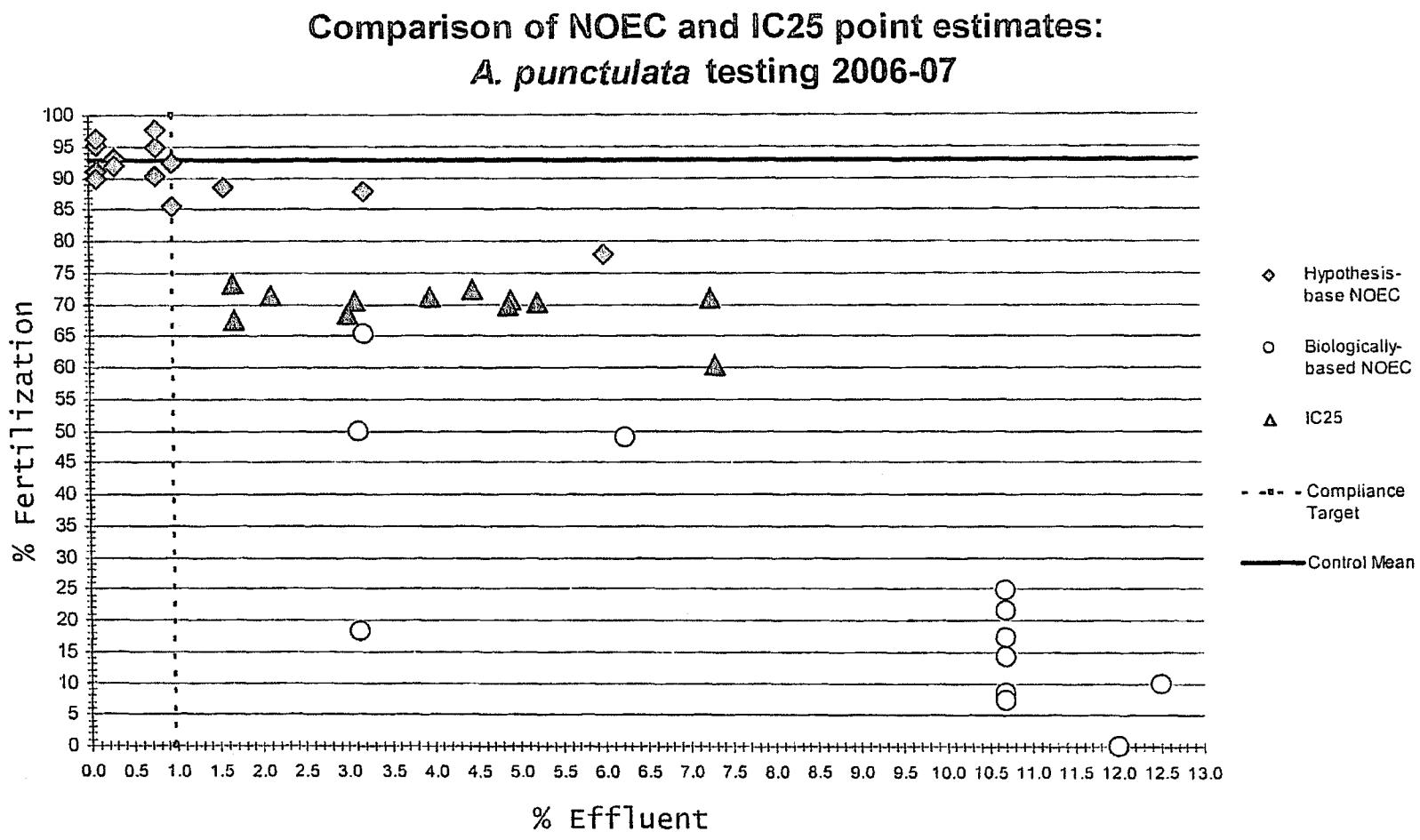
Exhibit 1. Summary of *Arbacia* Bioassay Tests Conducted to Date with Combined Bacardi/Bayamon/Puerto Nuevo Effluent, showing Comparison of NOECs with IC25 Point Estimates of Chronic Toxicity

Test Date	Hypothesis-based NOEC	% Fertilization	IC25	% Fertilization	Biologically-based NOEC	% Fertilization
2/16/06	0.78	95	7.25	71	12.5	10
3/16/06	6	78	7.31	60.4	3.13	50
8/29/06	<0.78	>68.8	1.68	67.7	3.13	18.25
11/4/06	<0.78	>91	1.67	73.3	6.25	48.9
11/7/06	1.56	88.6	3.97	71.3	10.7	8.6
4/17/07	0.29	93	3.09	70.5	3.2	65.4
4/19/07	<0.09	>90.8	2.12	71.6	10.7	7.4
4/21/07	<0.09	>91.5	4.47	72.3	12.0	0
5/1/07	0.09	91.1	4.92	70.7	10.7	14.2
5/3/07	0.96	92.4	14.8	69.6	35.5	1
5/5/07	3.2	87.8	14.4	67.9	35.5	4.3
5/15/07	0.09	89.8	4.88	70	10.7	24.8
5/17/07	0.96	85.5	3.01	68.5	10.7	17.3
5/19/07	0.29	92	5.23	70.4	10.7	21.5

Mean Control Fertilization = 92.9%

Exhibit 2 is a scatter plot showing percent fertilization and percent effluent for hypothesis-based NOECs, IC25 point estimates, and biologically-based NOEC values. It is clear that the only data points that appear to indicate non-compliant toxicity (i.e., are to the left of the 0.96% compliance target for the new WQC) are NOEC values derived from statistical hypothesis testing. Both the IC25 point estimates and the biologically-based NOEC data points do not provide evidence of unacceptable (non-compliant) effluent toxicity.

Exhibit 2. Percent Fertilization and Percent Effluent for Hypothesis-Based NOECs, IC25 Point Estimates and Biologically-Based NOEC Values



Alternative EPA-Approved *Arbacia* WET Test Data Evaluation Methods

EPA, in a recent evaluation of the WET test data developed by Bacardí in relation to its NPDES permit renewal and WQC applications, has assessed NOECs for *Arbacia* that are based solely on statistical hypothesis testing. The EPA interpretation of the data using that evaluative technique was that an unacceptable level of toxicity may exist in the effluent. Bacardí was ordered by EPA to perform a series of four additional tests at two-week intervals, according to the protocols in the existing NPDES permit, to obtain a more definitive evaluation of effluent toxicity.

As noted above, using statistical hypothesis testing to evaluate *Arbacia* WET test data is liable to introduce Type I errors because the percent fertilization variance within the control group replicates is normally very small. Thus, even a very small difference between the control group replicates and the effluent test group replicates would be calculated as statistically different from the variance for the control group, indicating an "effect" that is interpreted as "toxicity." This can either make it difficult to define a NOEC (as in the indeterminate <0.78 values in Exhibit 1) or may define a NOEC at an artificially low concentration that results in reported false positives for toxicity, and possibly erroneous findings of noncompliance with NPDES permit limits.

EPA has carefully addressed these and other issues related to toxicological data interpretation in several of its guidance documents. For example, in its 1991 *Technical Support Document for Water Quality-Based Toxics Control* (TSD; USEPA 1991), EPA compared results from hypothesis testing and point estimate endpoints such as the IC25 and concluded that:

> "Comparisons of both types of data indicate that a NOEC derived using the IC25 is the approximate analogue of a NOEC derived using hypothesis testing. For the above reasons, if possible, the IC25 is the preferred statistical method for determining the NOEC." (emphasis added)

Moreover, EPA (2000) specifically addresses effluent toxicity variability and states the following (on p. 6-4):

> "EPA recommends that point estimates be used to estimate effluent variability, to determine the need for limits, and to set permit limits. This is recommended whether the self-monitoring test results will be determined using hypothesis tests or point estimates. Point estimates have less analytical variability than NOECs using current experimental designs.... Point estimates make the best use of the whole effluent toxicity (WET) test data for purposes of estimating the coefficient of variation, long term average, and relative percent factors and calculating the permit limit." (emphasis added)

An EPA sponsored review committee was formed several years ago to assess this issue. The committee found that in the case of a species with low control variability, such as that exhibited by *Arbacia*, using only the NOEC derived from statistical hypothesis testing is problematic and may not be an effective approach for monitoring toxicity compliance and reporting. As a result of these issues EPA Region 1 modified the hypothesis testing approach to include the species test acceptability criteria (TAC) for determining permit compliance. This approach provides a more biologically relevant reporting endpoint for compliance evaluation. Documentation is provided at the following web page (http://www.epa.gov/region1/npdes/epa_attach.html)

under the link Marine Chronic Test Procedure and Protocol. The basis of the biological significance evaluation is that the TAC for control fertilization rate (>70% fertilization) is applied *in combination with* the statistical hypothesis testing results to determine the "biologically significant" effects concentrations (as opposed to only statistically-derived effects concentrations).

For its part, the Puerto Rico Water Quality Standards Regulation (PRWQSR) defines chronic toxicity testing and evaluation as follows:

> Chronic Bioassay
> Toxicity test designed to determine if the response to a stimulus such as, a total effluent, a specific substances, or combination of these has sufficient severity to induce a long-term effect that could linger for up to one-tenth of the life span of the organism. A chronic effect could be lethality, growth rate reduction, reproduction rate reduction, etc. A chronic bioassay shall be performed according to procedures described in "Mixing Zone and Bioassay Guidelines", approved by the Board.
>
> Chronic Effect
> Organism response to a stimulus, detected during a chronic bioassay, that comprises a stimulus that lingers or continues for a relatively long period of time, which could be of the order of one-tenth of the life span of the organism used in the test. A chronic effect could imply lethality, growth rate reduction, reduced reproduction rate, etc.
>
> Chronic Toxic Unit
> The reciprocal of the effluent dilution that causes no unacceptable effect on the test organisms by the end of the chronic exposure period, obtained during a chronic bioassay, as defined by the following equation:
>
> $$TUc = \frac{100}{NOEC}$$
>
> (The NOEC value should be expressed in terms of the percent (%) of the effluent in the dilution water).

It is noted that, although the PRWQSR chronic toxicity definition refers to a NOEC, it does not refer to a specific method by which a NOEC is to be obtained. It is further noted that the PRWQSR refers to the *Puerto Rico Mixing Zone and Bioassay Guidelines*, which are defined as follows:

> Technical guidelines developed by the Board which describe procedures, methods, models, techniques and organisms to be used to calculate the initial dilution; perform chronic and acute bioassays; to collect field data, or to establish the natural background concentration value, as required to verify compliance with inherent mixing zone conditions. These Guidelines are based on the following EPA publication: "Technical Support Document for Water Quality Based Toxics Control" and Users Guide to the Conduct and Interpretation of Complex Effluent Toxicity Tests at Estuarine/Marine Sites".7 The guidelines will be revised, as necessary, in accordance with updated versions of these documents or other documents released by EPA which directly impact the guidelines in effect at the time of publication of the final document.

There are several alternative EPA-approved methods that are available to evaluate compliance with toxicity criteria that do not rely solely on statistical hypothesis testing. These include

---

[7] It is noted that the most recent version of the Puerto Rico Mixing Zone and Bioassay Guidelines is a 1989 draft that predates the 2001 EPA Technical Support Document, and that advances in methods and technology in the last 17 years are therefore not reflected in the Guidelines. However, the Guidelines explicitly provide EQB with the ability to approve alternative methods.

biological significance evaluation (as described above used by EPA Region 1), IC25 point estimate evaluation, and test variability evaluation.[8] Of the three, the first two are in more common use for *Arbacia* fertilization tests. These are simply WET test data evaluation alternatives; they are not WET test protocol alternatives. The following subsections discuss these alternative methods. It is noted that these data evaluation alternatives should also be applicable to other Puerto Rico NPDES permits that use *Arbacia* as a test organism.

Biological Significance Evaluation

EPA Region 1 has recognized that evaluation of *Arbacia* fertilization tests using statistical hypothesis testing often results in putative statistically-based "toxicity effects" at effluent concentrations that are much lower than likely biological effects. When the fertilization success in the control group replicates varies by only small percentages, a statistically significant difference between the control and a test group could be interpreted as a "toxic" response, without respect to biological significance.

The EPA Region 1 website (see Marine Chronic Testing Methods, Section V: Test Methods, Item #16 in the Table of Recommended Test Conditions under "Acceptability of Test") stipulates that fertilization rates for the control group of replicates should be greater than 70%. For the purposes of evaluating permit compliance, if test group results yield fertilization rates greater than 70% (i.e., within the range of acceptable control group fertilization), but are shown to be statistically different from the control using hypothesis testing, those test group concentrations are not considered different from the control for the purposes of assessing toxicity (i.e., they are not biologically significant; see biologically-based NOEC data in Exhibit 1). In a test where that occurs, the NOEC concentration corresponds to the highest test group concentration that has a fertilization rate greater than or equal to 70%, without regard to whether it is statistically different from the control using hypothesis testing.

This combined hypothesis testing/biological significance method for *Arbacia* WET test data compliance evaluation is considered by EPA Region 1 to be a reliable approach and is preferred over the sole use of statistical hypothesis testing. Therefore, it is believed that evaluating the biological significance results for the Bayamón/Puerto Nuevo/Bacardí discharge system WET tests using this approach is a practical and acceptable means by which to evaluate compliance with toxicity criteria for *Arbacia*. This approach could replace statistical hypothesis testing alone as per EPA Region 1 data evaluation protocols.

IC25 Evaluation

Exhibit 1 also shows the IC25 point estimates for the *Arbacia* WET tests that have been conducted to date for the Bayamón/Puerto Nuevo/Bacardí discharge system. The IC25 is a commonly used, widely accepted point estimation technique that is calculated to estimate chronic toxicity thresholds. The IC25 method uses all of the WET test data as opposed to statistical hypothesis testing, which does not. As seen in Exhibit 1, if IC25 values were used to evaluate the data, all *Arbacia* chronic WET test results would have met permit compliance requirements of no chronic toxicity at the edge of the mixing zone at concentrations less than either the existing (0.70%) or anticipated future (0.96%) compliance targets.

---

[8] Test variability evaluation is discussed in Appendix 1 to this white paper.

In the preamble to its *Final Rule for Guidelines Establishing Test Procedure for the Analysis of Pollutants; Whole Effluent Toxicity Test Methods*, Fed. Reg. 69951-69972 (November 19, 2002) EPA states in two separate discussions:

> "EPA recommends the use of point estimation techniques over hypothesis testing approaches for calculating endpoints for effluent toxicity tests under NPDES Permitting Program."

(*Id.* at 69957 and 69958.) This statement is reiterated in EPA (2002). On Page 44, section 9, EPA states:

> "**NOTE: For the NPDES Permit Program, the point estimation techniques are the preferred statistical methods in calculating end points for effluent toxicity tests.**" (emphasis in original).

Therefore, it is believed that evaluating the IC25 point estimate for the Bayamón/Puerto Nuevo/Bacardí discharge system bioassays (or other NPDES permits requiring *Arbacia* testing) not only represents a reliable alternative with which to evaluate permit compliance relating to *Arbacia* test data, it is the preferred method of evaluation.

Summary

Arbacia is a species for which conventional statistically-based hypothesis testing alone typically fails to provide biologically meaningful results with respect to identifying toxicity for the purposes of permit compliance reporting. The problem stems largely from the very low variability in the control test fertilization responses. Because of this low variability, a very small difference between test dilutions and controls may be found to be statistically significant and interpreted as "toxic", when instead the results may lie within the range of the normal biological variability that is considered to be acceptable for the control replicates.

EPA (1991) and other subsequent EPA documents that address statistical variability, WET test analysis methodology, and NPDES compliance reporting provide insight and interpretive guidance that support a broader and more flexible evaluation of *Arbacia* WET test results than relying only on statistical hypothesis testing. In fact, EPA WET test evaluation guidance consistently recommends point estimation methods in preference to statistical hypothesis testing.

Conclusions and Recommendations

There are clearly problems inherent with using statistical hypothesis testing to evaluate toxicity data from *Arbacia* fertilization tests. EPA provides toxicity test evaluation guidance that explicitly recommends point estimate techniques as preferred alternatives to statistical hypothesis testing. Further, the PRQWSR and the associated Puerto Rico Mixing Zone and Bioassay Guidelines provide the flexibility to use alternative, EPA-approved approaches to compliance evaluations as they become available.

It is believed that a review of alternative methods for evaluating *Arbacia* test data and incorporating more appropriate agency-approved methods in new NPDES permits is warranted. Based on the above analysis, it is suggested that Bacardí (and PRASA) request that

EPA and EQB consider the following options as the basis for toxicity compliance evaluations for WET tests using *Arbacia*:

1. Use the IC25 point estimate methodology as the definitive toxicity evaluation.
2. Adopt the EPA Region 1 test acceptability criterion, using biological significance (i.e., the biologically significant NOEC as shown on Exhibit 1) in combination with statistical hypothesis testing.
3. Use both biological significance-based NOECs and IC25 point estimates to determine effluent toxicity using Arbacia data.

Options 1 or 2 are preferred, as they follow clear EPA guidance, and have already proven acceptable to EPA for use in NPDES permits for *Arbacia* WET test evaluation, and are therefore presumed to be acceptable (after careful review and evaluation) by EQB in light of the flexibility offered by the Puerto Rico Mixing Zone and Bioassay Guidelines. However, Option 3 is also acceptable and is consistent with EPA guidance concerning evaluation of acceptable whole effluent toxicity.

References

Environmental Quality Board. 2001. Intent to Issue Water Quality Certificates to Define and Authorize a Mixing Zone and Approve Compliance Plans. NPDES Permit No. PR0000591. Authorization to Bacardi Corporation to discharge under the NPDES System.

Federal Register. *Final Rule for Guidelines Establishing Test Procedure for the Analysis of Pollutants; Whole Effluent Toxicity Test Methods*, Fed. Reg. 69951-69972 (November 19, 2002)

National Association of Clean Water Agencies (NACWA). 2006. *Whole Effluent Toxicity (WET) NPDES Permit Testing and Limitations for Public Agencies.* White paper, January 2006.

USEPA. 2002. *Short-Term Methods For Estimating the Chronic Toxicity of Effluents and Receiving Water to Marine and Estuarine Organisms* (EPA-821-R-02-014, Third Edition).

USEPA. 2000. *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications under the National Pollutant Discharge Elimination System.* EPA 833-R-00-003, June 2000.

USEPA. 1991. *Technical Support Document for Water Quality-Based Toxics Control* (TSD), EPA/505/2-90/001. USEPA Office of Water.

Appendix A

Test Variability Evaluation when using Hypothesis Testing Methods

In the Preamble to its Final Rule, 67 Fed. Reg. at 69968, EPA states:

> "....to reduce the within-test variability and to increase statistical sensitivity when test endpoints are expressed using hypothesis testing rather than the preferred point estimation techniques, variability criteria must be applied as a test review step when NPDES permits require sublethal hypothesis testing endpoints (i.e., NOEC or LOEC) and the effluent has been determined to have no toxicity at the permitted receiving water concentration."

(67 Fed. Reg. at 69967 (emphasis added).) For tests for which in-test variability assessment is required, EPA defines this variability term as the percent minimum significant difference (PMSD). The Preamble to the EPA Final Rule states:

> "Within-test variability, measured as the percent minimum significant difference (PMSD), must be calculated and compared to upper bounds established for test PMSDs. Under this new requirement, tests conducted under NPDES permits that fail to meet the variability criteria (i.e., PMSD upper bound) and show "no toxicity" at the permitted receiving water concentration (i.e., no significant difference from the control at the receiving water concentration or above) are considered invalid and must be repeated on a newly collected sample."

(*Id.*) The EPA Final Rule did not include specific language requiring mandatory application of variability criteria for *Arbacia* fertilization tests, although a number of species with similar control test variability characteristics were defined. The Preamble to that Final Rule indicates that for the chronic methods that were not evaluated in the WET Interlaboratory Variability Study, EPA does not have sufficient data to support the implementation of mandatory variability criteria at this time.

Important to the issue of test variability, especially in the case of the *Arbacia* fertilization tests, are the following statements by EPA in the Preamble to the Final Rule:

> "Lower bounds on the PMSD are also applied, such that test concentrations shall not be considered toxic (i.e., significantly different from the control) if the relative difference from the control is less than the lower PMSD bound."

(*Id.* at 69957.) and

> "According to the proposed approach, any test treatment with a percentage difference from the control (i.e., [mean control response– mean treatment response]/ mean control response * 100) that is greater than the upper PMSD bound would be considered as significantly different; and any test treatment with a percentage difference from the control that is less than the lower PMSD bound would not be considered as significantly different."

(*Id.* at 69958.)

Because EPA, at the time of issuing its Final Rule, did not have sufficient data from an Interlaboratory Variability Study to develop variability criteria and PMSD bounds for the *Arbacia* fertilization test, there are no existing criteria with which to examine test variability. While test variability might prove to be an acceptable WET test data evaluation option for *Arbacia*, using it would require constructing a database that is not currently available. It is not believed that this approach is compatible with the current Bacardí and PRASA permit renewal schedules and it is further noted that there are other EPA-approved alternatives that are both appropriate and already in use for NPDES permit toxicity compliance evaluations for *Arbacia*.