



GAO

Accountability * Integrity * Reliability

United States Government Accountability Office
Washington, DC 20548

Comptroller General
of the United States

DOCUMENT FOR PUBLIC RELEASE

The decision issued on the date below was subject to a GAO Protective Order. This redacted version has been approved for public release.

Decision

Matter of: Bonner Analytical Testing Co.

File: B-409586.2

Date: August 7, 2014

Laurence Schor, Esq., and David A. Edelstein, Esq., Asmar, Schor & McKenna, PLLC, for the protester.

Sara McGraw, Esq., Environmental Protection Agency, for the agency.

Jennifer D. Westfall-McGrail, Esq., and Edward Goldstein, Esq., Office of the General Counsel, GAO, participated in the preparation of the decision.

DIGEST

Agency's evaluation methodology was not improper where the methodology provided a rational basis for source selection and was consistent with the evaluation criteria set forth in the solicitation.

DECISION

Bonner Analytical Testing Company, of Hattiesburg, Mississippi, protests the disqualification of its proposal from consideration for award of a Superfund organics methods contract under request for proposals (RFP) No. SOL-HQ-12-00032, issued by the Environmental Protection Agency (EPA). The protester's proposal was disqualified for failing to achieve a passing score on a pre-award performance evaluation sample (PA-PES) analysis. Bonner argues that the agency's methodology for evaluating offerors' sample analyses was flawed and was improperly based on an unstated evaluation criterion.

We deny the protest.

BACKGROUND

The RFP, which was issued on November 21, 2013, anticipated the award of multiple indefinite-delivery/indefinite-quantity contracts for analytical testing services in support of the investigation and clean-up activities under the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 and the Superfund Amendments and Reauthorization Act of 1986. The data generated from the testing is to be used for a variety of purposes, including assessing the nature

and extent of contamination at a hazardous waste site, defining priorities for response based on risk to human health and the environment, identifying appropriate clean-up actions, and deciding when remedial actions are complete. Offerors were to propose prices for organic methods, inorganic methods, or both. RFP at 7. To be eligible for award, offerors had to perform a PA-PES analysis, which was designed to test their ability “to detect organic and/or inorganic target analytes of interest within established detection limits.” Id. at 93. The solicitation advised offerors that they should perform the analysis in accordance with pre-award performance evaluation instructions distributed with the samples, and that if an offeror received a score of less than 85 (out of 100) points, it would be considered ineligible for award. Id. at 93-94. The instructions accompanying the organic test sample set indicated that offerors were to prepare and analyze six different types of samples: trace volatiles in water matrix, semivolatiles in water matrix, semivolatiles--SIM [selected ion monitoring] in water matrix, low/medium volatiles in soil matrix, pesticides in soil matrix, and aroclors in soil matrix.

Bonner requested evaluation samples from EPA, tested them, and reported the results. By letter of April 16, 2014, EPA notified the protester that it had successfully passed the preliminary evaluation stage for the inorganic methods, but had failed the preliminary evaluation stage for organic methods, earning a score of only 78.6 on its sample analyses. Later the same day, the agency furnished Bonner with a copy of its organic methods scoring report.

The report explained that the agency used an algorithm for scoring the sample analyses that started with a score of 100 and then deducted points for various types of deficiencies, including: (1) failing to identify target analytes present in the sample; (2) identifying quantities of target analytes outside action limits; (3) identifying analytes not present in the sample; and (4) failing to identify non-target analytes in the sample. The report further explained that the action limits (or, more fully, action prediction interval limits) were calculated from the results submitted by participating laboratories (that is, the laboratories submitting proposals in response to the RFP here), and that the action limits were set at the 90 percent confidence level.¹ The report summarized the deficiencies identified in the protester’s analysis as follows:

- Number of target analytes not identified: 0
- Number of target analytes outside action limits: 19
- Number of TAL contaminants² over SOW limits: 1

¹ The protester explained that the 90 percent confidence level is the mean +/- 1.64 standard deviations. Protest at 5, n.5.

² According to the scoring report, TAL contaminants are analytes not added to the samples.

- Number of non-target analytes not identified: 1

Bonner Organic Methods Individual Laboratory Summary Report at 4. In other words, virtually all of the protester's point deductions resulted from its identification of target analytes at levels outside the action limits. Moreover, the report indicated that virtually all (i.e., 18 of 19) of the instances where Bonner identified target analytes outside the action limits occurred during its testing of the trace volatiles (TVOA) in water matrix.

In every instance in which the protester's results were outside the action limits, they exceeded the upper limit. For example, one of the trace volatile target analytes was carbon tetrachloride. The score report indicated that the low and high action limits for carbon tetrachloride, statistically calculated from the results submitted by participating laboratories, were 4.2 and 6.5 parts per billion (ppb) respectively, whereas the protester's reported result was 6.8 ppb.

Over the course of the two weeks following the protester's receipt of its score report, the protester sought additional information from the agency concerning the scoring of its organic methods samples. By email of May 2, the contracting officer notified the protester that its debriefing was concluded. Bonner protested to our Office on May 12.

DISCUSSION

Bonner argues that because the solicitation did not advise offerors that the agency would calculate the action prediction interval limits based on the PA-PES results of the offerors responding to the RFP here, the agency relied upon an unstated evaluation factor in employing this methodology. The protester contends that based on the solicitation language advising that the PA-PES test would be used to assess an offeror's ability to detect target analytes within established detection limits, and its own experience in the analytical testing industry, it reasonably understood that its results would be compared "either against a true value for target analyte concentration or against a statistically robust mean determined from a large number of laboratories." Protest at 7. Bonner maintains that there is no guarantee that the mean of the results submitted by participating laboratories will be an accurate estimate of the true value of the analyte concentrations--and thus, rather than testing the offeror's ability to detect target analytes (the agency's actual need), the agency's methodology merely establishes the extent of its conformance to other offerors' results.

Bonner further argues that had it known that EPA would compare its results to the mean of the results submitted by participating laboratories, it would have submitted a different set of results that would have passed the evaluation. The protester contends in this regard that because there is always variation in testing results, it is common practice in the industry for a testing lab to analyze a sample multiple times

and then select the data set that it considers most likely to represent the correct results. According to the protester, it analyzed the TVOA sample multiple times, and three of its six result sets would have passed the PA-PES evaluation as conducted by EPA. Bonner contends that it is well known in the analytical testing industry that less experienced labs tend to report analyte levels lower than the true value, and that if it had known that its data would be judged against the mean of the data of the other participating labs rather than the true value or a statistically robust mean, it would have submitted one of its lower sets of results. Bonner also contends--in an argument that is somewhat inconsistent with its above argument that the means were lower as a result of the participation in the competition of inexperienced testing firms--that approximately [deleted] offerors for organic methods under the solicitation here are incumbent contractors, and, as such, had an unfair advantage due to their (allegedly) superior knowledge of the agency's intended methodology. Bonner argues that because of this knowledge the incumbent contractors knew that their lower test results would be likely to pass.

Regarding the protester's argument that the agency used an undisclosed methodology to establish the action limits, we have previously distinguished between undisclosed evaluation methodologies and unstated evaluation criteria, finding that the two are not equivalent, and that it is not improper for an agency to rely upon an undisclosed evaluation methodology, so long as the methodology provides a rational basis for source selection and is consistent with the evaluation criteria set forth in the solicitation. Olympus Bldg. Servs., Inc., B-285351, B-285351.2, Aug. 17, 2000, 2000 CPD ¶ 178 at 5; ABB Power Generation, Inc., B-272681, B-272681.2, Oct. 25, 1996, 96-2 CPD ¶ 183 at 4. The protester's argument that the agency improperly relied upon an unstated evaluation factor is therefore denied.

In response to the protester's argument that the agency's methodology fails to provide a rational basis for source selection because--rather than establishing the extent to which an offeror has the ability to detect target analytes, it establishes only the extent to which the offeror's results conform to other offerors' results--the agency maintains that its methodology was consistent with industry standards. Moreover, it is clear from the record that the methodology suggested by the protester of comparing the protester's results to a statistically robust mean--which we understand to refer to a mean based on a large number of data points--was not available to the agency here because the PES solution being tested was new, and as a result, had not yet been subjected to extensive testing. That is, the only test results available to EPA other than the results received in response to RFP here were the results of five sample tests conducted by the lab that developed the solution to assess the testing solution's homogeneity.

As part of our development of the record here, we asked EPA why the results from the lab that developed the sample were not considered as part of the data set used to calculate the action limits. In response, the agency explained that because all of

these tests were performed by the same lab, using the same instrumentation, calibration reference standards, personnel, laboratory procedures, and operating conditions, they did not represent unique data sets. EPA Project Officer's Statement, July 9, 2014, at 3. According to the agency, combining the results from the labs responding to the RFP here with the results of the single lab testing for homogeneity skews the distribution of the data set being evaluated, and, as a result, the agency uses only the data from the former in the evaluation process. Id. at 8-9.

In reply, the protester agreed that the two sets of results should not have been combined in a single data set. Bonner argued, however, that the mean results from the tests conducted by the developing lab were consistently higher than the mean results from the labs responding to the RFP here. In Bonner's view, this result should have placed the agency on notice that the offeror lab results were too low and led it to investigate the discrepancy.

We are not persuaded that agency should have interpreted the difference in the means of the two sets of results as evidence that the results of the offeror labs were too low. In some instances, as conceded by the protester, the discrepancy between the two means is statistically insignificant--and even where it is not, the record fails to establish that the only reasonable explanation for the variation is that the results of the offeror labs were too low. Among other things, agency officials could have interpreted the discrepancy as an indication that in some instances, the developing lab's results trended on the high side.

Finally, we reject the protester's argument that the incumbent contractors may have known that the agency intended to evaluate their results against the mean results of labs responding to the RFP here. Even assuming that the incumbent contractors were on notice of the agency's intended methodology, we fail to see why such knowledge would have led them to conclude that other offerors were likely to submit their lower results and that they should thus do the same, absent collusion among the offerors, for which no evidence has been provided. Likewise, in connection with Bonner's argument that inexperienced labs are more likely to submit low results, there is nothing in the record to support the protester's speculation that the other labs responding to the RFP here are inexperienced--and, indeed, as previously noted, the protester itself acknowledges that several of the other labs are incumbent contractors. In our view, the fact that several of these labs hold incumbent contracts undercuts Bonner's argument that they are inexperienced.

The protest is denied.

Susan A. Poling
General Counsel