Executive Summary

This Test Method Evaluation Report, prepared by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), summarizes ICCVAM's evaluation of the validation status of five *in vitro* test methods proposed for assessing the potential pyrogenicity of pharmaceuticals and other products, as potential replacements for the *in vivo* rabbit pyrogen test (RPT). The five test methods are:

- The Human Whole Blood (WB)/Interleukin (IL)-1β *In Vitro* Pyrogen Test
- The Human WB/IL-1β *In Vitro* Pyrogen Test: Application of Cryopreserved (Cryo) Human WB
- The Human WB/IL-6 *In Vitro* Pyrogen Test
- The Human Peripheral Blood Mononuclear Cell (PBMC)/IL-6 *In Vitro* Pyrogen Test
- The Monocytoid Cell Line Mono Mac 6 (MM6)/IL-6 *In Vitro* Pyrogen Test

This report also provides ICCVAM's recommendations for current uses and limitations for each test method, as well as recommendations for standardized protocols, future studies, and performance standards. In support of this evaluation, ICCVAM prepared a draft Background Review Document (BRD) and ICCVAM draft test method recommendations, which were provided to an independent scientific peer review panel (Panel) and the public for consideration and comment. The ICCVAM draft BRD was prepared using data from validation studies that had been conducted by the European Centre for the Validation of Alternative Methods (ECVAM), a unit of the Institute for Health and Consumer Protection at the European Commission's Joint Research Centre. The ECVAM submission, prepared according to the ICCVAM submission guidelines (ICCVAM 2003), included five individual BRDs (i.e., one BRD for each test method), which summarized the validation studies for each of the five *in vitro* test methods.

The Panel met on February 6, 2007 to review the ICCVAM draft BRD for errors and omissions and to discuss the current validation status of the five *in vitro* test methods. The Panel also reviewed the extent that the information contained in the ICCVAM draft BRD supported the ICCVAM draft test method recommendations. In finalizing the test method recommendations presented here, ICCVAM considered the conclusions and recommendations of the Panel as well as comments from the public and its Scientific Advisory Committee on Alternative Toxicological Methods.

ICCVAM Recommendations: Test Method Uses and Limitations

Based on this evaluation, ICCVAM recommends that, although none of these test methods can be considered a complete replacement for the RPT for all testing situations for the detection of Gram-negative endotoxin, they can be considered for use to detect Gram-negative endotoxin in human parenteral drugs on a case-by-case basis, subject to validation for each specific product to demonstrate equivalence to the RPT, in accordance

with applicable U.S. Federal regulations (e.g., U.S. Food and Drug Administration [FDA] [*])[†]. When used in this manner, these methods should be able to reduce the number of animals needed for pyrogenicity testing. Pyrogenicity testing may involve more than slight or momentary pain or distress when a pyrogenic response occurs. Accordingly, alternative test methods must be considered prior to the use of animals for such testing, as required by U.S. Federal animal welfare regulations and policies. Therefore, these and other *in vitro* alternative test methods should be considered prior to the use of animals in pyrogenicity testing and should be used where determined appropriate for a specific testing situation. Use of these methods, once appropriately validated, will support improved animal welfare while ensuring the continued protection of human health.

ICCVAM developed a recommended standardized protocol for each test method based primarily on ECVAM standard operating procedures (SOPs). ICCVAM also provided recommendations for further research and development, optimization, and validation efforts. These recommendations should be helpful to various stakeholders (e.g., applicable U.S. Federal regulatory agencies, the international regulatory community, the pharmaceutical industry) for determining when these test methods might be useful.

The Panel concluded that the validation criteria were adequately addressed in the ICCVAM BRD to determine the usefulness and limitations of these test methods to serve as a substitute for the RPT to identify Gram-negative endotoxin on a case-by-case basis, subject to validation for that specific product. However, the Panel stated the performance of these test methods in terms of their reliability and relevance did not support this proposed use.

In March 2006, the ECVAM Scientific Advisory Committee (ESAC) endorsed a statement of validity for these five *in vitro* pyrogen test methods (see **Appendix E**). Like ICCVAM, ESAC concluded that these five methods can detect Gram-negative endotoxin in materials currently tested with the RPT, and, therefore, may be useful for regulatory decisions, subject to validation for that specific product. Both ICCVAM and ESAC also concluded that the currently available database does not support the use of these test methods to detect a wider range of pyrogens, as suggested in the original ECVAM submission. However, ESAC concluded that these tests "can currently be considered as full replacements for the evaluation of materials or products where the objective is to identify and evaluate pyrogenicity produced by Gram-negative endotoxins, but not for other pyrogens." ICCVAM has concluded that the current validation database for these test methods is inadequate to support such a definitive statement based on the ECVAM validation study design, which did not include biologics or medical devices and evaluated only a limited range and number of pharmaceutical products. Additionally, no RPT data were generated with the same test samples used in the *in vitro* test methods (i.e., parallel testing).

---

[*]Mechanisms exist for test method developers to qualify their method on a case-by-case basis. The use of any recommended method will be subject to product-specific validation to demonstrate equivalence as recommended by the FDA (e.g., U.S. Code of Federal Regulations (CFR) 21 CFR 610.9 and 21 CFR 314.50(d)(1)(ii)(a)).

[†]Substances other than endotoxin may induce the cellular release of IL-1β and/or IL-6. For this reason, users of these test methods should be aware that the presence of other materials might erroneously suggest the presence of endotoxin and lead to a false positive result.

Accuracy and Reliability

The accuracy of *in vitro* pyrogen test methods for detecting Gram-negative endotoxin was based on the results for 10 parenteral pharmaceuticals, each spiked with four concentrations of endotoxin (0, 0.25, 0.5, or 1.0 Endotoxin Units [EU]/mL, with 0.5 EU/mL tested in duplicate). As shown in **Table 1**, accuracy among the test methods ranged from 81% to 93%, sensitivity ranged from 73% to 99%, specificity ranged from 77% to 97%, false negative rates ranged from 1% to 27%, and false positive rates ranged from 3% to 23%.

Table 1 Accuracy of *In Vitro* Pyrogen Test Methods[1]

Test Method	Accuracy[2]	Sensitivity[3]	Specificity[4]	False Negative Rate[5]	False Positive Rate[6]
Cryo WB/IL-1β	92% (110/120)	97% (75/77)	81% (35/43)	3% (2/77)	19% (8/43)
MM6/IL-6	93% (138/148)	96% (85/89)	90% (53/59)	5% (4/89)	10% (6/59)
PBMC/IL-6	93% (140/150)	92% (83/90)	95% (57/60)	8% (7/90)	5% (3/60)
PBMC/IL-6 (Cryo)[7]	87% (130/150)	93% (84/90)	77% (46/60)	7% (6/90)	23% (14/60)
WB/IL-6	92% (136/148)	89% (79/89)	97% (57/59)	11% (10/89)	3% (2/59)
WB/IL-1β (Tube)	81% (119/147)	73% (64/88)	93% (55/59)	27% (24/88)	7% (4/59)
WB/IL-1β (96-well plate)[8]	93% (129/139)	99% (83/84)	84% (46/55)	1% (1/84)	16% (9/55)

Abbreviations: Cryo = Cryopreserved; EU/mL = Endotoxin units per milliliter; IL = Interleukin; MM6 = Mono Mac 6; PBMC = Peripheral blood mononuclear cells; WB = Whole blood

[1]Data shown as a percentage (number of correct runs/total number of runs), based on results of 10 parenteral drugs tested in each of three different laboratories. Samples of each drug were tested with or without being spiked with a Gram-negative endotoxin standard (0, 0.25, 0.5, or 1.0 EU/mL, with 0.5 EU/mL tested in duplicate).

[2]Accuracy = the proportion of correct outcomes (positive and negative) of a test method.

[3]Sensitivity = the proportion of all positive substances that are classified as positive.

[4]Specificity = the proportion of all negative substances that are classified as negative.

[5]False negative rate = the proportion of all positive substances that are falsely identified as negative.

[6]False positive rate = the proportion of all negative substances that are falsely identified as positive.

[7]A modification of the PBMC/IL-6 test method that uses Cryo PBMCs.

[8]A modification of the WB/IL-1β test method that uses 96-well plates instead of tubes for the test substance incubation.

Repeatability within individual laboratories was determined for each *in vitro* test method, using saline and various endotoxin spikes to evaluate the closeness of agreement among optical density (OD) readings for cytokine measurements at each concentration. The results indicated that the variability in OD measurements increased with increasing endotoxin concentration. However, the variability was low enough that the threshold for pyrogenicity could still be detected (i.e., the 0.5 EU/mL spike concentration could still be distinguished from the lower concentrations).

Reproducibility within individual laboratories was evaluated using three marketed pharmaceuticals spiked with various concentrations of endotoxin. Three identical, independent runs were conducted in each of the three testing laboratories, with the exception

of the Cryo WB/IL-1β test method[‡]. The correlations (expressed as percentage of agreement) between pairs of the independent runs (i.e., run 1 vs. run 2; run 1 vs. run 3; run 2 vs. run 3) were determined, and the mean of these three values was calculated. Agreement across three runs within a single laboratory ranged from 75% to 100%.

Reproducibility across all laboratories was evaluated in two different studies in which each run from one laboratory was compared to all other runs of another laboratory. The proportion of equally qualified samples provided a measure of reproducibility. In the first reproducibility study, three marketed pharmaceutical products were spiked with either saline control or various concentrations of endotoxin, and each sample was tested in triplicate in each of three different laboratories, except for the Cryo WB/IL-1β. In the second study, reproducibility was determined using the results from the 10 substances used in the accuracy analysis. Each drug was spiked with four concentrations of endotoxin and tested once in each of three laboratories. The extent and order of agreement among laboratories were similar in both studies: the WB/IL-1β test method showed the least agreement (57% to 58%), and the Cryo WB/IL-1β test method showed the most (88% to 92%).

ICCVAM Recommendations: Test Method Protocols

ICCVAM recommends standardized protocols for each test method that should be used for validation of specific products on a case-by-case basis for U.S. regulatory consideration. These recommended protocols, provided in **Appendix C**, are primarily based on ECVAM SOPs for each test method. ICCVAM has updated these protocols to address inadequacies identified by the Panel, including modifications to standardize essential test method components across the five *in vitro* test methods. These modifications are not expected to reduce or otherwise impact test method accuracy and reliability.

The Panel concluded that the information provided in the ICCVAM draft BRD supported the ICCVAM draft recommended protocols for these five *in vitro* test methods, as long as inadequacies identified by the Panel with respect to reliability and relevance are fully addressed.

ICCVAM Recommendations: Future Studies

ICCVAM recognizes that these test methods could be applicable for detection of a wider range of pyrogens (i.e., endotoxin and pyrogens other than endotoxin) and test materials, provided that they are adequately validated for such uses. Test materials that have been identified clinically as pyrogenic might be invaluable for use in future validation studies and might allow such studies to be conducted without the use of animals. Wherever possible, historical data generated with the same test samples in both *in vitro* and *in vivo* studies (i.e., parallel testing data) should be retrospectively evaluated, or *in vitro* testing should be performed in parallel with RPT and/or bacterial endotoxin tests (BET) conducted for

[‡]The ECVAM Cryo WB/IL-1β test method BRD stated that there was no direct assessment of intralaboratory reproducibility because such an evaluation was performed in the WB/IL-1 test method, and the authors assumed that variability would not be affected by the use of cryopreserved blood.

regulatory purposes[§]. Future validation studies should include the following considerations:

1. Both endotoxin-spiked and non-endotoxin spiked samples should be included. Non-endotoxin pyrogen standards should be characterized prior to their use in any study, if possible.
2. All aspects of the studies should comply with Good Laboratory Practices.
3. Future studies should include products that have intrinsic pro-inflammatory properties in order to determine if these tests can be used for such substances.
4. Optimally, a study that includes three-way parallel testing, with the *in vitro* assays being compared to the RPT and the BET, should be conducted to comprehensively evaluate the relevance and comparative performance of these test methods. These studies may be conducted with historical RPT data provided that the same substances (i.e., same lot) are tested in each method. Based on ethical and scientific rationale, any *in vivo* testing should be limited to those studies that will fill existing data gaps.
5. Test substances that better represent all categories of sample types (e.g., pharmaceuticals, biologicals, and medical devices) intended for testing by the methods should be included.
6. The hazards associated with human blood products should be carefully considered, and all technical staff should be adequately trained to observe all necessary safety precautions.
7. Formal sample size calculations should be made to determine the required number of replicates needed to reject the null hypothesis at a given level of significance and power. For reliability assessments, formal hypothesis testing is essential with the alternative hypothesis being no difference between groups.

The Panel agreed with ICCVAM that any future studies should be performed using the ICCVAM recommended test method protocols. The Panel also provided other suggestions and recommendations for future studies (see **Appendix A**). Like ICCVAM, the Panel also recognized that these test methods could be applicable to a wider range of pyrogens and test materials, provided that they are adequately validated for such uses.

ICCVAM Recommendations: Performance Standards

As indicated above, these test methods have not yet been adequately evaluated for their ability to detect Gram-negative endotoxin in parenteral pharmaceuticals, biological products, and medical devices compared to the RPT or the BET. For this reason, ICCVAM does not consider it appropriate at this time to develop performance standards that can be used to evaluate the performance of other test methods that are structurally and functionally similar.

[§]In order to demonstrate the utility of these test methods for the detection of non-endotoxin pyrogens, either an international reference standard is needed (as is available for endotoxin [i.e., WHO-LPS 94/580 *E. coli* O113:H10:K-]) or, when a positive non-endotoxin-mediated RPT result is encountered, this same sample should be subsequently tested *in vitro*.