EXECUTIVE SUMMARY

This Background Review Document (BRD) reviews available data and information regarding the validation status of the Bovine Corneal Opacity and Permeability (BCOP)[1] test method for identifying ocular corrosives and severe irritants. The test method was reviewed for its ability to predict ocular corrosives and severe/irreversible effects as defined by the U.S. Environmental Protection Agency (EPA) (EPA 1996), the European Union (EU) (EU 2001), and the United Nations (UN) Globally Harmonized System (GHS) of Classification and Labelling of Chemicals (UN 2003). The objective of this BRD is to describe the current validation status of the BCOP test method, including what is known about its accuracy and reliability, the scope of the substances tested, and the availability of a standardized test method protocol.

The information summarized in this BRD is based on publications obtained from the peer-reviewed literature, as well as unpublished information submitted to the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) in response to two *Federal Register* (*FR*) Notices requesting high quality *in vivo* rabbit eye test data and *in vitro* ocular irritation data for BCOP, the Isolated Chicken Eye (ICE), the Isolated Rabbit Eye (IRE), and the Hen's Egg Test – Chorioallantoic Membrane (HET-CAM) test methods. An online literature search identified 18 publications that contained BCOP test method results and protocol information; of these publications, detailed *in vivo* data were obtained for five studies. Submitted BCOP and detailed *in vivo* data for three additional studies allowed for an evaluation of test method accuracy[2] and reliability[3] for a total of eight studies.

Other published and unpublished BCOP test method studies are reviewed in **Section 9.0** (Other Scientific Reports and Reviews). This section discusses BCOP studies that could not be included in the performance analyses because of the lack of appropriate study details or test method results and/or the lack of appropriate *in vivo* rabbit eye reference data.

The BCOP assay is an *in vitro* eye irritation test method using isolated bovine eyes from cattle that have been slaughtered for meat or other purposes. In the BCOP assay, opacity is determined by the amount of light transmission through the cornea, and permeability is determined by the amount of sodium fluorescein dye that passes through all corneal cell layers. Both measurements are used to calculate an *In Vitro* Irritancy Score, which is used to assign an *in vitro* irritancy classification for prediction of the *in vivo* ocular irritation potential of a test substance. More recent additions/endpoints to the BCOP assay are assessment of corneal swelling or hydration, and histological assessment of morphological alterations in the

[1] Exposure of the isolated bovine cornea to irritants can produce opacity and/or an increase in permeability to sodium fluorescein dye. Both of these endpoints can be quantified and used to evaluate the potential eye irritation of substances.

[2] (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. It is a measure of test method performance and one aspect of "relevance". The term is often used interchangeably with "concordance."

[3] A measure of the degree to which a test method can be performed reproducibly within and among laboratories over time. It is assessed by calculating intra- and inter-laboratory reproducibility and intralaboratory repeatability.

cornea (Bruner et al. 1998; Ubels et al. 1998; Cooper et al. 2001; Jones et al. 2001). When histological assessment is added to the BCOP assay, the type and depth of corneal injury can be evaluated, as well as whether the tissue damage is permanent (e.g., damage to the endothelium) (Gran et al. 2003). Therefore, a histopathological assessment can be useful to discriminate borderline cases (i.e., substances that produce results that preclude assignment to a single category) or to identify ocular damage that does not produce opacity or permeability in isolated cornea. Histology also is used for new chemistries or formulas that have not been well characterized in the BCOP assay, for known chemistries with delayed effects or where the mode of action cannot be easily predicted, and for known chemistries when a complete characterization of damage is needed.

U.S. Federal regulatory agencies were surveyed to determine whether BCOP test method data have been considered for regulatory use where submission of testing data is required. The EPA and the Food and Drug Administration (FDA) responded that BCOP data have been submitted to their respective agencies. The EPA Office of Pesticide Programs (OPP) received and reviewed BCOP data submitted in support of two new products (formulations). A labeling decision was made by the EPA for the two new products; however, hazard classification and labeling of these products was not based solely on the results of the submitted BCOP data. The FDA Center for Drug Evaluation and Research (CDER) has accepted BCOP data, on a case-by-case basis, for topically-applied products and more than 25 oral and inhalation products, but not for any ocular products. These substances or products were not formally classified for ocular irritation potential by the FDA.

The BCOP assay is currently used by several U.S. and European companies (e.g., pharmaceutical, personal care, and household cleaning product companies) as an in-house screen to assess the ocular irritation potential of a wide range of substances for which there could be accidental exposures in the workplace or home. The test method is used in the following ways:

1. For workplace safety applications to assess the irritancy of synthetic intermediates, various ingredients of a product, or the final product during the manufacturing process (Sina 1994).
2. For product safety applications to assess cosmetics, pharmaceuticals, soaps, household and industrial cleaners, personal care products, and other types of product formulations (Swanson et al. 1995; Casterton et al. 1996; Chamberlain et al. 1997; Harbell and Curren 1998; Cater et al. 2002; Cuellar et al. 2003; Bailey et al. 2004).

For example, it has been reported that some companies perform the assay as an in-house screen of industrial raw materials and intermediates; materials that give a BCOP score of 25 or higher are labeled as irritants with no further testing. Materials considered nonirritating based on the BCOP assay are tested *in vivo* to confirm the *in vitro* results (Chamberlain et al. 1997). In another company, the BCOP assay is used to evaluate both non-registered household products and registered household disinfectants, pesticides and repellents (Cuellar N and Swanson J, personal communication). For non-registered household products, BCOP data from new product formulations are usually matched with relevant benchmark formulations for which the ocular irritation potential is well characterized; *in vivo*

confirmatory testing is generally not performed. For registered products, use of the BCOP assay is limited to product development issues and worker safety at this company.

The BCOP test method protocols used in the various studies considered in this BRD are similar, but not identical. The essential principles of the test method protocol include isolating and culturing the bovine cornea, treating the isolated cornea with a test substance, collecting opacity and permeability data, and evaluating the data in relation to a prediction model. However, given the various uses and applications of the BCOP test method by different investigators and laboratories, and the evolution of the test method over time, a number of laboratory-specific differences have been noted regarding the conduct of the test method. Variations in the publicly available BCOP protocols include different instrumentation to evaluate opacity, different prediction models or *in vitro* classification systems, and differences in the use of positive controls, among other methodological variations.

A total of 161 substances and formulations were evaluated in the eight studies, of which 69 were commercial products or formulations. A variety of chemical and product classes have been tested in the BCOP assay. The chemical classes with the greatest amount of *in vitro* BCOP data are alcohols, carboxylic acids, esters, formulations, heterocyclic compounds, hydrocarbons, ketones, and onium compounds. The formulations tested include hair shampoos, personal care cleansers, detergents, bleaches, insect repellents, petroleum products, and fabric softener. Other chemical classes tested include amines, ethers/polyethers, inorganic and organic salts, and organic sulfur compounds. The most common product classes tested in the BCOP assay are chemical/synthetic intermediates, cleaners, drugs/pharmaceuticals/therapeutic agents, petroleum products, solvents, shampoos, and surfactants. Other product classes tested include detergents, pesticides, plasticizers, reagents, bactericides, and insect repellents.

Some of the published *in vivo* rabbit eye test data on the substances used to evaluate the accuracy of BCOP for detecting ocular corrosives and severe irritants was limited to average score data or a reported irritancy classification based on a laboratory specific classification scheme. However, detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal at 24, 48, and 72 hours and/or assessment of the presence or absence of lesions at 7, 14, and 21 days were necessary to calculate the appropriate EPA (1996), EU (2001), and GHS (UN 2003) ocular irritancy hazard classifications. Thus, a portion of the test substances for which there was only limited *in vivo* data could not be used for evaluating test method accuracy as described in this BRD.

Only a few of the reports provided original *in vitro* test result data. However, summary *in vitro* data were available for all of the test substances evaluated, such that they could be assigned *in vitro* irritancy classifications for comparison to the available *in vivo* reference data.

The accuracy evaluation of the BCOP test method was limited to the substances evaluated in eight *in vitro-in vivo* comparative studies. The ability of the BCOP test method to correctly identify ocular corrosives and severe irritants, as defined by the EPA (1996), the EU (2001),

and the GHS (UN 2003) was evaluated using two approaches. In the first approach, the accuracy of BCOP was assessed separately for each *in vitro-in vivo* comparative study. In the second approach, the accuracy of BCOP was assessed after pooling data across *in vitro-in vivo* comparative studies that used similar protocols and the same method of data collection. While there were some differences in results among the three hazard classification systems evaluated (i.e, EPA [EPA 1996], EU [EU 2001], and GHS [UN 2003]), the accuracy analysis revealed that BCOP test method performance was comparable among the three hazard classification systems. The overall accuracy of the BCOP test method ranged from 79% to 81%, depending on the classification system used. Sensitivity and specificity ranged from 75% to 84% and 79% to 81%, respectively. The false positive rate ranged from 19% to 21%, while the false negative rate ranged from 16% to 25%.

The accuracy analysis also indicated that alcohols are often overpredicted (50% to 56% [7/14 to 9/16] false positive rate, depending on the classification system used) in the BCOP test method. Ketones (40% [4/10]), carboxylic acids (38% to 44% [3/8 to 4/9]), and heterocyclic compounds (33% [2/6]) also had high false positive rates. Although there were a small number of underpredicted substances (4 to 5), alcohols (2) were most often underpredicted by the BCOP test method.

With regard to physical form of the substances overpredicted by the BCOP test method, 18 to 20 were liquids and two were solids. Considering the proportion of the total available database, liquids (90/120 to 92/124) appear more likely than solids (30/120 to 32/124) to be overpredicted by the BCOP test method.

With regard to physical form of the substances underpredicted by the BCOP test method, five were solids and one was a liquid. Despite the proportion of the total available database indicated above, solids (42% to 50% false negative rate) appear more likely than liquids (4% to 5% false negative rate) to be underpredicted by the BCOP test method.

Exclusion of three discordant classes (i.e., alcohols, ketones and solids) from the data set resulted in an increased accuracy (from 81% to 92%), a decreased false positive rate (from 20% to 12%), and a decreased false negative rate (from 16% to 0%).

The 35 substances labeled as surfactants were rarely underpredicted by the BCOP test method for substances classified as severe by the EU (EU 2001) and GHS (UN 2003) classification systems (i.e., R41 or Category 1), as evidenced by the false negative rates ranging from 7% to 8%. Substances classified as severe (i.e., Category I) by the EPA classification system (EPA 1996) were more often underpredicted (false negative rate of 23%). However, although the available database was smaller (n = 7 to 9), substances labeled as pesticides were more often underpredicted by the BCOP test method (false negative rates ranging from 40% to 50%).

Considering the comparable proportion of acidic and basic underpredicted substances (18% to 30% [2/11 to 3/10] vs. 23% to 33% [3/13 to 3/9]), there was little difference among the underpredicted substances for which pH information was available. However, it is noted that

pH information was available for only a portion of the 40 to 43 severe irritant substances (i.e., Category 1, Category I, or R41) in the database for each classification system.

Finally, with respect to the GHS classification system only, the seven underpredicted substances were more likely to be substances classified *in vivo* based on persistent lesions (false negative rate of 23% [3/13]), rather than on severe lesions (false negative rate of 17% [4/24]).

A quantitative assessment of intralaboratory data (*In Vitro* Irritancy Scores) from three studies (Southee 1998; Dr. Sina's submission; Dr. Van Goethem's submission) provides an indication of the extent of intralaboratory repeatability of the BCOP test method for substances predicted as severe eye irritants. For the 16 substances evaluated in the Southee (1998) study, the median %CV for *In Vitro* Irritancy Scores for replicate corneas ranged from 11.8 to 14.2 for the three laboratories. For the 29 substances evaluated by Dr. Sina, the within experiment mean and median %CV values for *In Vitro* Irritancy Scores were 71 and 35, respectively. The dataset provided by Dr. Sina included 10 substances with low *In Vitro* Irritancy Scores around the background range of the assay (< 3.5), contributing to the increased variability of this dataset. However, the range of %CV values for the five substances predicted as severe irritants (*In Vitro* Scores > 55.1) in this study is 1.1 to 13. For the 52 substances evaluated by Dr. Van Goethem in the Gautheron et al. (1994) study, the median %CV for *In Vitro* Irritancy Scores for replicate corneas was 18.1%, comparable to the results obtained with the data from Southee (1998).

A quantitative assessment of intralaboratory data (*In Vitro* Irritancy Scores) from two studies (Gettings et al. 1996; Southee 1998) indicates the extent of intralaboratory reproducibility of the BCOP test method for substances predicted as severe eye irritants. For the Gettings et al. (1996) study, the between experiment (n = 3) mean and median %CV values for permeability values were 33.4 and 29.0, respectively, for 25 surfactant-based personal care cleaning formulations. For the Southee (1998) study, the mean %CV values for *In Vitro* Irritancy Scores for the 16 substances tested two or more times in Laboratory 1, Laboratory 2, and Laboratory 3 ranged from 12.6 to 14.8 for the three laboratories, while the median %CV values ranged from 6.7 to 12.4.

A qualitative assessment of the data provided for multiple laboratories in three studies (Gautheron et al. 1994; Balls et al. 1995; Southee 1998) provides an indication of the extent of interlaboratory reproducibility. In an assessment of interlaboratory reproducibility of hazard classification (EPA, EU, or GHS), the five participating laboratories for the Balls et al. (1995) study were in 100% agreement in regard to the ocular irritancy classification for 40 to 41 (67% to 68%) of the 60 substances tested *in vitro* in the study, depending on the classification system used. The extent of agreement between testing laboratories was greatest for substances identified from *in vivo* rabbit eye data as corrosives or severe irritants when compared to any other combination of *in vivo* and *in vitro* results (76% to 86% of the accurately identified severe substances were shown to have 100% classification agreement among testing laboratories). For the study by Gautheron et al. (1994), regardless of the classification system used, there was 100% agreement in regard to the ocular irritancy classification for 35 (69%) of the 51 substances, which were tested in either 11 or 12

laboratories. For the study by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification for 15 (94%) of the 16 substances, regardless of the classification system used. Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds, and such product classes as surfactants, organic solvents, chemical intermediates, detergents, and pesticides.

A quantitative evaluation of interlaboratory reproducibility was conducted for three studies (Gautheron et al. 1994; Balls et al. 1995; Southee 1998) by performing a %CV analysis of *In Vitro* Irritancy Scores obtained for substances tested in multiple laboratories. For the Gautheron et al. (1994) study, the 17 substances predicted as severe in the BCOP assay had mean and median %CV values of 36% and 17%, respectively, for results obtained in either 11 or 12 laboratories. For the Balls et al. (1995) study, the 32 substances predicted as severe in the BCOP assay had mean and median %CV values of 25% and 22%, respectively, for results obtained in five laboratories. For the Southee (1998) study, the mean and median %CV values for the *In Vitro* Irritancy Scores of the 16 substances were 32.4% and 22.8%, respectively, for three laboratories.

As stated above, this BRD provides a comprehensive summary of the current validation status of the BCOP test method, including what is known about its reliability and accuracy, and the scope of the substances tested. Raw data for the BCOP test method will be maintained for future use, so that these performance statistics may be updated as additional information becomes available.