5.0 IRE TEST METHOD DATA AND RESULTS

5.1 Description of the IRE Test Method Protocols Used to Generate Data

The ocular irritancy of a wide variety of test substances was determined in four reports using the IRE test method and compared to results obtained in the *in vivo* Draize rabbit eye irritation test (CEC 1991; Balls et al. 1995; Gettings et al. 1996; Guerriero et al. 2004). Individual eye test data was obtained for all of these reports. In these reports, the protocols used to generate the *in vitro* IRE test method data varied in the number of endpoints measured and in the time the measurements were taken. For example, Gettings et al. (1996) only evaluated corneal swelling, whereas Balls et al. (1995) evaluated corneal opacity and corneal swelling. In the CEC (1991) study, corneal opacity, corneal swelling, and fluorescein retention were reported. In Guerriero et al. (2004), corneal opacity, corneal swelling, and fluorescein penetration with an assessment of epithelial integrity were evaluated. Variations in the protocols used to generate IRE test data in these studies for specific endpoints are shown in **Appendix A**.

5.1.1 CEC Collaborative Study (1991)

The "Collaborative Study on the Evaluation of Alternative Methods to the Eye Irritation Test" was sponsored by the Commission of the European Communities (CEC) and published in 1991. One study in this report evaluated 21 chemically diverse test substances using the IRE test method conducted by three independent laboratories. The IRE data was obtained using a standard IRE protocol (Burton et al. 1981) for the measurement of corneal swelling (percent increase in corneal thickness relative to pretreatment value) and assessment of corneal opacity (score of 0 to 4) over a period of 30 minutes to four hours, but fluorescein retention (score of 0 to 4) at 30 and/or 240 minutes was added. However, each laboratory used an independent irritancy prediction model to evaluate overall *in vitro* severity based on a scale of I (non-irritating or no EU label) to IV (comparable to EU R41 label) (EU 2001). The final *in vitro* irritancy rating (A; mild to D; most severe) was then compared to published *in vivo* ocular data ranked according to severity of injury using the EU (EU 2001) classification system.

5.1.2 Balls et al. (1995)

In the EC/HO International validation study, 59 test substances were evaluated using the IRE test method. In this study, data from the IRE test method was generated by four separate laboratories and consisted of measurement of corneal opacity and corneal swelling at 1 and 4 hours. Corneal opacity was evaluated using the Draize scoring system (scale of 0-4). Corneal swelling was determined by measurement of pretreatment corneal thickness and calculation of the relative increase in corneal thickness at each time point, but only the one and four hour values were used in the analysis. The effects of the test substance on epithelial integrity and the degree of fluorescein penetration were not reported in this validation study. However, while the EC/HO study only acquired data from two endpoints (corneal opacity and swelling), two authors recognized that the use of additional endpoints (i.e., fluorescein penetration and assessment of endothelial integrity) might significantly increase the accuracy of the irritancy potential assigned to a test substance. A negative, untreated control was used in each experiment, but a positive control was not included.

5.1.3 Gettings et al. (1996)

The CTFA conducted an evaluation of *in vitro* alternatives comparing 41 *in vitro* test endpoints including the IRE test method to the Draize rabbit eye irritation test. In this study, 25 surfactant-based formulations were tested representing cosmetic and personal care products. The endpoints evaluated were corneal opacity and corneal swelling (0.5, 1, 2, 3, and 4 hours following application of the test substance). Fluorescein penetration was evaluated one hour after application of the test substance. Histological assessment of the corneal tissue was determined on tissue obtained four hours after application. Corneal opacity was measured macroscopically and by slit-lamp examination. Corneal thickness was measured before application of the test substance and at each time point afterwards. The percentage increase in corneal thickness was calculated relative to the pretreatment value and expressed as corneal swelling. The mean swelling of each treatment group was then calculated and compared to *in vivo* data. The data were transformed into four groups based on "no significant swelling" to "maximal response" and compared also to data obtained in the *in vivo* rabbit eye test.

5.1.4 Guerriero et al. (2004)

The IRE test method protocol used by Guerriero et al. (2002, 2004) to compare *in vitro* and *in vivo* test results is outlined in **Appendix A** of this BRD. Corneal opacity and area of opacity were assessed macroscopically and by slit-lamp examination at 0.5, 1, 2, 3 and 4 hours after application of the test substance. Corneal swelling was calculated as the percentage increase in corneal thickness at each time point relative to a pre-application value. Fluorescein uptake and assessment of epithelial integrity were determined at four hours. The decision criteria described in **Appendix A** were used to classify a test substance as either a severe irritant or a nonsevere irritant.

5.2 Data Obtained to Evaluate the Accuracy and Reliability of the IRE Test Method

NICEATM staff requested original data from IRE test method studies. A *Federal Register* notice (Vol. 69, No. 57, pp. 13859-13861; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), requesting original IRE test method data, was published on March 24, 2004. In addition, authors of published IRE studies were contacted to request original IRE test method data for the respective publications. As a result of these efforts, data for the 59 substances evaluated in the Balls et al. (1995) study using the IRE test method were obtained in electronic format. In addition, Frederick Guerriero at GlaxoSmithKline kindly provided the IRE test method data and the individual *in vivo* rabbit eye test data from the Guerriero et al. (2004) study. In addition, GlaxoSmithKline generously provided the actual chemical names and MSDS sheets of the 30 GSK test substances used in the study that were originally reported as numbered (i.e., 1 to 30) generic substitutions (e.g., substituted aniline, substituted pyridine, aromatic acetanilide). No other IRE test method data was received. Therefore, this evaluation is based, in part, on data obtained from the published literature. Furthermore, given the lack of availability of the original records for these studies, the testing laboratory's summary judgment regarding the outcome of each study cannot be evaluated. The availability of notebooks or other material containing original data is unknown.

5.3 Description of the Statistical Approaches Used to Evaluate the Resulting Data

Statistical approaches used to analyze the *in vitro* data obtained in the IRE test method protocol were discussed in **Section 2.2.12**. Evaluation of the accuracy and reliability of the *in vitro* data obtained using the IRE test method in comparison to *in vivo* data has been approached differently by various authors. In all of these reports, none of the data were compared to the GHS (UN 2003), EPA (1996), or EU (EU 2001) regulatory classification systems. Therefore, as discussed in **Sections 6.0** and **7.0**, the results of these studies were reanalyzed to assess the accuracy and reliability of the test method, when compared to these classification schemes.

5.3.1 CEC Collaborative Study (1991)

In this study of 21 test substances, *in vitro* irritation was correlated only to the EU-based irritation classification (EU 2001). It is not clear what, if any, statistical methods were used to compare these scores. No correlation coefficients were provided in the report. However, raw data for the IRE studies were provided for each test substance with an *in vivo* EU classification based on literature evaluation. This data was sufficient to permit an analysis of performance based on EU classification only.

5.3.2 Balls et al. (1995)

In this study, 59 test substances were ranked by degree of irritancy according to the mean MMAS values and the standard deviations from the mean according to ECETOC (1992). The data was arranged in a matrix containing the identity of the test substance, the *in vivo* eye irritation data and each alternative method endpoint. Data analysis was carried out on the full data set and on subsets based on solubility criteria. Correlation coefficients were determined to assess interlaboratory reproducibility of alternative data (27 test index scores) and the relationship between *in vivo* eye irritation potential versus the endpoint for each alternative method. To conduct the proposed accuracy analysis, the *in vivo* data were reclassified using the GHS (UN 2003), EPA (EPA 1996) and EU (EU 2001) irritancy classification systems.

5.3.3 Gettings et al. (1996)

In this study of 25 surfactant-based formulations, statistical analysis was divided into three components. The distributional characteristics of the Draize test and *in vitro* test results were determined, concordance analysis was used to assess the extent of association of the Draize test and *in vitro* scores, and regression analysis was used to predict Draize scores (MAS values) from *in vitro* test results. The IRE results were compared to an *in vivo* irritation score based on the FHSA regulatory classification system in which test substances are classified as irritants or nonirritants without further separation. To conduct the proposed analysis, the CTFA *in vivo* data were reclassified according to the GHS (UN 2003), EPA (1996) and EU (EU 2001) regulatory classification systems.

5.3.4 Guerriero et al. (2004)

In this study of 44 test substances, *in vitro* irritation was correlated to EU-based irritation classification (EU 2001). It is not clear what, if any, statistical methods were used to compare these scores. No correlation coefficients were given. However, adequate information to assign at least one regulatory irritancy classification (EU [EU 2001]) to each

test substance was given in the publication; these classifications were used in an analysis of performance. Additional *in vivo* data from these studies were provided on request that permitted assignment of GHS (UN 2003) and EPA (EPA 1996) regulatory classifications and additional retrospective accuracy analyses.

5.4 Summary of Results

A summary of results used to evaluate test method accuracy is shown in **Appendix C**. This appendix, sorted by substance tested, provides the name of the substance tested, the Chemical Abstract Service Registry Number (CASRN), the concentration tested, the calculated score, the irritation classification of the test substance, and the literature source. No attempt was made to identify the source and purity of a test substance if the authors did not provide such information. If available, a CASRN was entered for each test substance. This identifier was obtained from various sources, including the publication and the National Library of Medicine's ChemID database. All substances with the same CASRN were listed under the same name, regardless of the synonym that was used in the original publication.

5.5 Use of Coded Chemicals and Compliance with GLP Guidelines

Ideally, all data supporting the validity of a test method should be obtained and reported in accordance with GLP guidelines and with the use of coded chemicals (OECD 1998; EPA 2003a; 2003b; FDA 2003). The data quality was evaluated by a review of the methods section in literature references only and in the submitted reports. The data quality presented in the reviewed literature references can only be evaluated by what was provided in the published reports. Based on the available information, the Balls et al. (1995) study, the Gettings et al. (1996) study, and the Guerriero et al. (2004) study were reportedly conducted according to GLP guidelines (see **Section 8.0**). As described in **Section 3.4**, based on the available information, coded chemicals were employed in the Balls et al. (1995) and Gettings et al. (1996) studies. The Guerriero et al. (2004) study was performed in a GLP-compliant laboratory in which the test substances are individually coded. In the CEC pilot collaborative study, there was no indication that the studies were conducted under GLP compliance, although at least two of the three testing laboratories are known to run GLP-compliant studies. Although test substances were numbered in these studies as indicated in **Section 3.4**, it is not known if the testing laboratories were blinded with respect to the numbering.

5.6 Description of "Lot-to-Lot" Consistency, Time Frame of Studies and Testing Laboratories

Ideally, the lot-to-lot consistency of test substances is evaluated to ensure that the same substance, with the same physicochemical properties, is being evaluated over the duration of the study. A description of the procedures used to evaluate and control the lot-to-lot consistency was described in three of the published reports (CEC 1991; Balls et al. 1995; Gettings et al. 1996), but was applicable to interlaboratory studies using the same test substance. For the studies described in this BRD, substances were only tested once in each study, and therefore, lot-to-lot consistency within a study was not applicable. No attempt

was made to review the original records and assess the procedures used to evaluate different batches of tested substances.

5.6.1 CEC Collaborative Study (1991)

In the CEC collaborative IRE study, each of the 21 test substances came from a single batch with purity greater than 97% and was supplied by the Aldrich Chemical Company Limited (UK) via the FRAME (Nottingham, UK).

The testing laboratories were TNO-CIVO Institutes (Zeist, Netherlands), Shell Research Ltd. (Sittingbourne, UK), and the Instituut voor Hygeine en Epidemiologie (IHE) (Brussels, Belgium).

5.6.2 Balls et al. (1995)

The *in vivo* reference test substances used in the EC/HO Study were readily available chemicals of high consistency and purity (90 to >99.5%) with known stability in storage over time. The substances were obtained from the same source when feasible. Otherwise, substances with specifications as close as possible to the original were used. Coding, labeling, dispensing, and storage procedures were tightly controlled to avoid confusion of test substances during preparation, storage and delivery.

Four laboratories were selected for each of the nine test methods. The laboratories chosen for the IRE test method were:

- ESL, Unilever Research, Sharnbrook, Bedford, United Kingdom (Dr. Lesley Earl) – *Lead Laboratory*
- Shell Research Ltd, Sittingbourne, Kent, United Kingdom (Dr. John Gardner)
- Zeneca CTL, Macclesfield, Cheshire, United Kingdom (Dr. Richard Lewis)
- IHE, Ministry of Public Health of Environment, Brussels, Belgium (Dr. GA Jacobs)

5.6.3 Gettings et al. (1996)

The test substances used in the CTFA validation study were individual generic formulations designed to represent a variety of product types. Test substances used in the CTFA validation study were carefully controlled with respect to documentation, coding, labeling, dispensing and transfer. Documentation included manufacturer, lot and/or batch number, amount of test substance, and condition upon receipt. The substances were prepared in bulk and dispensed using stringent requirements to maintain consistency for delivery to the individual testing laboratories.

5.6.4 Guerriero et al. (2004)

In the GSK/Safe-Pharm Laboratories study, information on the individual test substances with respect to lot-to-lot variation was not given. However, for 30 of the 44 pharmaceutical process materials used as test substances, the *in vivo* data and the *in vitro* data were obtained on the same chemicals within the same time frame. Substances were analyzed for quality control purposes and reasonable purity (> 90%) was attained, but the analytical data were not immediately available for all substances at the time the study was performed (Guerriero F, personal communication). With the exception of the ECETOC (1998) test substances,

identical lots and batch numbers of test substances were used for both the *in vitro* and *in vivo* studies, performed in the same laboratory.

The testing laboratories were GSK (United States and United Kingdom) and SafePharm Laboratories (Derbyshire, United Kingdom).

5.7 Availability of Data not Submitted for External Audit, If Requested

An attempt to obtain the original study records for the IRE data was made by NICEATM. The original study records were not readily available for any of the studies; thus, it appears unlikely that such data are available for an external audit.