

2.0 HET-CAM TEST METHOD PROTOCOL COMPONENTS

2.1 Overview of How the HET-CAM Test Method is Conducted

The HET-CAM protocol was first described by Luepke (1985). According to the original test method, fertilized hen's eggs are incubated, under optimized conditions, for nine days. On the 10th day, the eggs are opened and the CAM exposed. Then, 0.3 mL of the test substance is applied to the surface of the CAM. After a 20-second exposure period, the CAM is rinsed with 5 mL of water. The CAM is evaluated for development of irritant endpoints (hyperemia, hemorrhage, and coagulation) at 0.5, 2, and 5 minutes after rinsing off the test substance. Irritant effects in the CAM blood vessels and albumen are subjectively assessed and a score is assigned based on the time required for development of each endpoint. The scores are totaled to yield a total irritation score for the tested substance (maximum score of 21).

Since the initial description of the HET-CAM test method, several studies have been conducted to evaluate the feasibility of using HET-CAM as a complete replacement for the *in vivo* rabbit ocular test. Most of these reports describe a HET-CAM test method protocol that is similar, but not identical, to the original protocol. These differences include the breed of hen from which eggs are obtained, the endpoints evaluated, data collection procedures, and methods used to analyze the data. To date, no single HET-CAM test method protocol has gained wide acceptance as a standard protocol.

2.2 Description and Rationale for the Test Method Components

Currently, there is no widely accepted, single HET-CAM test method protocol for detecting ocular corrosives and severe irritants. The essential principles of the test method protocol include exposing the CAM, treating the CAM with a test substance, observing the development of endpoints, and evaluating the data in relation to a prediction model. However, given the various uses and applications of the HET-CAM test method by different investigators and laboratories, and the evolution of the assay over time, a number of laboratory-specific differences have been noted regarding the conduct of the test method. Variations in the publicly available protocols include different prediction models or *in vitro* classification systems, and differences in the use of positive controls, among other methodological variations. These test method protocol differences are described in detail in **Section 2.2.1**, where variations in specific test method components for the HET-CAM test method are discussed.

The following sections describe in detail the major components of the HET-CAM test method protocol. Similarities and differences in the test method components of available HET-CAM protocols are discussed. For many of these components, no rationale for inclusion in the HET-CAM was provided in the published literature; in such cases, historical use is considered the rationale. **Appendix A** provides a comparative summary of test method components for all protocols reviewed.

2.2.1 Materials, Equipment, and Supplies Needed

2.2.1.1 *Hen Breed*

A limited number of hen breeds have been used for obtaining eggs for HET-CAM studies. Most studies used the White Leghorn breed of hens for egg production. Specific strains of White Leghorn hens that were used included Shaver Starcross 288A, Lohmanns Select LSL, and Leghorn SA31. Additional breeds of chickens used to supply eggs were White Essex and Lohmann Brown.

No rationale was provided in the reports for the selection of a specific hen breed or strain. Furthermore, no information was provided in the reports that indicated that there were differences in the results obtained from studies using eggs from different hen breeds. Therefore, while a formal study to determine the optimal hen breed to use for these studies was not found in the published scientific literature, use of the White Leghorn breed appears to produce consistent results. It also is noted that White Leghorns and their descendants are the most numerous today and are prodigious egg layers (Anonymous 1996).

2.2.1.2 *Criteria for Egg Use*

Published HET-CAM protocols are inconsistent in describing the criteria for selecting eggs for use. Most reports fail to note specific criteria (such as age of eggs, weight of eggs) that should be met prior to use in the test method. Those reports that do note criteria for egg selection and use typically state that the eggs weighed between 50 and 80 grams (Luepke 1985; CEC 1991; Bagley et al. 1992; de Silva et al. 1992; Spielmann 1995). Additional criteria noted for egg selection include (1) viability, as determined by candling the eggs, and (2) age (eggs were not older than seven days) (Gilleron et al. 1996, 1997; Budai et al. 1997).

While no rationale was provided in the scientific literature for the use of these criteria, studies indicate that excessively large or small eggs may have less than optimal embryonic growth and development while eggs with cracked or thin shells may affect retention of the necessary moisture for proper embryonic development (Lowther D, personal communication).

2.2.1.3 *Egg Incubation Conditions (Temperature, Relative Humidity, Egg Rotation During Incubation, and Length of Incubation)*

Temperature: There is little variation in the incubation temperature used for the eggs. A majority of incubation temperatures ranged from 37 to 38°C (e.g., Luepke 1985; Kalweit et al. 1987; de Silva et al. 1992; Hagino et al. 1993; Gilleron et al. 1997; Spielmann et al. 1997; Schlage et al. 1999). The most common temperature used for incubation was $37.5 \pm 0.5^\circ\text{C}$ (Luepke 1985; Blein et al. 1991; CEC 1991; Bagley et al. 1992; Rougier et al. 1992; INVITTOX 1992; Spielmann 1995).

Relative Humidity: The relative humidity used during the incubation period either was not reported or differed among the studies. A majority of reports indicate that the average relative humidity used during incubation ranged from 40% to 65%. The most commonly reported average relative humidity was 62.5%, with standard deviations of 1.5% or 7.5% (Luepke 1985; Blein et al. 1991; CEC 1991; Bagley et al. 1992; Rougier et al. 1992; INVITTOX 1992; Spielmann 1995; Gilleron et al. 1996). Other reported relative humidity

values include about 70%, 60% to 70%, 80%, and 80% to 90% (Hagino et al. 1991, 1993, 1999; Kojima et al. 1995; Budai et al. 1997; Budai and Várnagy 2000; Brantner et al. 2002; Demirci et al. 2003, 2004).

Egg Rotation During Incubation: Some of the published reports discussed rotation of the eggs during incubation. Information about rotation frequency, however, was infrequently provided. When reported, eggs were rotated either once an hour or twice an hour (van Erp et al. 1990; Hagino et al. 1993, 1999; Kojima et al. 1995; Brantner et al. 2002). Dr. med Horst Spielmann indicated that eggs incubated for the HET-CAM test method were rotated five times per day (Spielmann H, personal communication). Generally, where reported, the eggs were rotated up to the day prior to application of the test substance (Spielmann 1995; Gilleron et al. 1996; Budai et al. 1997; Budai and Várnagy 2000).

Length of Incubation: In most studies, the eggs were incubated for either 9 or 10 days¹ (e.g., Luepke 1985; Kalweit et al. 1987; Sterzel et al. 1990; CEC 1991; Spielmann 1995). Several of these studies reported that the eggs were evaluated for viability by a candling technique on the day prior to use in the test method. Those eggs that were considered defective or nonviable were excluded from the study. Viable eggs were returned to the incubator, with the larger end placed upwards, and not rotated for the remainder of the incubation period. In contrast to the typical 9- or 10-day incubation period, three references reported an incubation period of approximately three days (Brantner et al. 2002; Demirci et al. 2003, 2004). Interestingly, these same investigators also used high relative humidity conditions (80% to 90%) compared to the majority of test method protocols reviewed.

Published sources indicate that proper development of the chicken embryo occurs when the incubation temperatures range from 35.6 to 38.9°C for still-air incubators and 37.5 to 38.1°C for forced-air incubators. The optimal relative humidity ranges, from 50% to 60%, were not dependent on the temperature of the incubator or the incubator type (Clauer 2000; Smith 2000; Anonymous 2002, 2004a, 2004b). The core range of relative humidity used was from 55% to 60%.

Hand rotation of the eggs on the eight days prior to use in the test method prevents attachment of the CAM to the shell (Spielmann 1995). Candling of the eggs prior to the last day ensures that the eggs used in the test method are viable and have developed normally during the incubation period. After establishing the viability of the eggs, placing the eggs back into the incubator, with the large end facing upwards, ensures accessibility to the CAM (Spielmann 1995).

In the embryo, the allantois begins to emerge from the hindgut of the chicken on incubation/gestation days four or five, but does not fuse with the chorion until incubation/gestation day six to form the chorioallantoic membrane (Sinn-Harlon 1998b; Clauer 2002). Therefore, based on the development of the embryo, the earliest incubation/gestation day that could be used in the HET-CAM test method is incubation day seven. The use of a nine-day incubation period is based on chicken embryo development and

¹ The difference in the days of incubation usually depended upon whether investigators defined the first day of incubation as Day 0 or Day 1.

international regulations that define when a chicken-based test is no longer considered an *in vitro* test method. International regulations have provisions for the protection of animals used for experimental or other scientific purposes. Some provisions indicate the time in which a test method using an animal embryo or fetus is considered an animal and therefore protected by the regulations. According to some of these regulations, a bird is considered a protected animal (and therefore the test is considered an *in vivo*, and not *in vitro* test) when greater than half of the gestation or incubation period has elapsed (day 10.5 for the 21 day incubation period of the chicken embryo) (Animals [Scientific Procedures] Act 1986; EU 1986). In the United States, the Public Health Service Policy, with which all National Institutes of Health (NIH)-funded research projects must comply, covers all live vertebrate species. The NIH Office of Laboratory Animal Welfare has provided written guidance in this area, interpreting "live vertebrate animal" to apply to avians only after hatching (Kulpa-Eddy J, personal communication; NIH 2000).

It has been proposed that at incubation day nine, the embryonic differentiation of the chicken central nervous system is sufficiently incomplete that suffering from pain perception is unlikely to occur (MSPCA 2005; Liebsch M, personal communication). Evaluations suggest that there are few sensory fibers present at day nine in the avian embryo and that there is significant development of the sensory nerve ending between incubation days 11 and 14 (Romanoff 1960). Studies also have suggested that the extraembryonal vascular systems (e.g., yolk sac, CAM) are not sensitive to pain (Rosenbruch 1997; Spielmann H, personal communication). Combined, these studies suggest that at incubation day nine there is little to no pain perceived by the developing embryo during the conduct of the HET-CAM test method.

2.2.1.4 Rotary Saw

Many of the test method protocols reviewed discussed a procedure for opening the eggshell prior to application of test substances. These protocols described using a dentist rotary saw, small rotary saw, or scissors to scratch the eggshell around the air space or cut away at the eggshell. The shell was then removed to expose the inner membrane. Only one literature reference provided information on the type of saw used to scratch the eggshell. Wilson and Steck (2000) described using a Dremel™ tool (Multi-Pro model 225T2) with a coarse 25 mm cutting disk to cut away the eggshell around the marked air space. A rationale for using this specific tool was not provided in the report.

The use of a rotary saw or tool with a cutting disk appears to be the most efficient way to remove a portion of the eggshell and allow access to the inner membrane and CAM.

2.2.2 Dose-Selection Procedures, Including the Need for Any Dose Range-Finding Studies or Acute Toxicity Data Prior to Conducting a Study

Dose-selection procedures are not relevant to this *in vitro* test method. As described in **Section 2.2.4.2**, test substances are applied as neat chemicals (solids or liquids) if possible. Substances should be diluted with a preferred solvent, if technical limitations require the evaluation of a diluted test substance.

2.2.3 Endpoints Measured

A variety of endpoints were evaluated in the HET-CAM test method protocols reviewed. The original protocol described by Luepke (1985), as well as many additional protocols, evaluated the CAM for development of hemorrhage (blood from a ruptured vessel), hyperemia (increased blood flow), and coagulation (presence of blood clots) after application of the test substance (Blein et al. 1991; Hagino et al. 1991, 1993, 1999; Bagley et al. 1992; Rougier et al. 1992; de Silva et al. 1992; Kojima et al. 1995; Doucet et al. 1999). Other endpoints that were evaluated to assess the irritancy potential of test substances include injection (mild hemorrhage), vasoconstriction (narrowing of the vessels), dilation (expansion of the vessels), and lysis (disintegration of the vessels) (e.g., Luepke and Kemper 1986; Sterzel et al. 1990; CEC 1991; Spielmann 1995; Macián et al. 1996; Gettings et al. 1996; Budai and Várnagy 2000). The rationale for evaluation of these specific endpoints was not provided.

A combination of three endpoints was typically evaluated by each test method protocol. However, the combination of evaluated endpoints was not consistent. Most test method protocols evaluated the development of hemorrhage and coagulation in combination with one of the other endpoints noted. No rationale was provided for the combination of endpoints used in the various HET-CAM protocols.

One test method protocol called for the evaluation of the development of an “anti-angiogenic” effect (Demirci et al. 2003, 2004). This protocol was used to evaluate the anti-inflammatory and anti-angiogenic properties of various test substances. Therefore, the endpoint selected for this test method protocol was likely unique for these effects.

Prevalidation studies conducted in Germany evaluated the reproducibility of various HET-CAM endpoints. These studies indicated that some of the endpoints (e.g., injection, hyperemia) were not reproducible within or between laboratories (Spielmann H, personal communication); the most reproducible endpoints were hemorrhage, lysis, and coagulation. Therefore, the endpoints that appear to be the most reproducible (i.e., lysis, coagulation, and hemorrhage) are the endpoints that appear to be evaluated the most often during the performance of the HET-CAM test method.

2.2.4 Exposure Parameters

2.2.4.1 *Test Substance Exposure Amount or Volume*

The quantities of the test substance used in the HET-CAM test method were provided for liquids and solids. However, there was no consensus on the amount or volume of the test substance used in the protocols and no rationale was provided for the amounts used. For liquids, many of the studies used 0.2 mL or 0.3 mL of test substance (e.g., Luepke 1985; Kalweit et al. 1987; CEC 1991; Hagino et al. 1991; Gettings et al. 1994). Additional volumes used in the reviewed protocols were 0.2 to 0.3 mL, 0.1 to 0.2 mL, and 0.1 mL (Luepke and Kemper 1986; Gettings et al. 1996; Lönnroth et al. 1999). For solid forms of test substances, amounts between 0.1 g and 0.3 g were used (e.g., Luepke 1985; CEC 1991; Hagino et al. 1993; Gilleron et al. 1996, 1997).

No rationale was provided in the reports for the selection of specific test volumes. Furthermore, no information was provided in the reports that indicated that there were differences in the results obtained from studies using different volumes and/or amounts. Therefore, while a formal study to determine the optimal test volume and/or amount was not found in the published scientific literature, use of amounts and volumes that sufficiently cover the CAM appear to produce consistent results.

2.2.4.2 *Concentration Tested*

A variety of test substance concentrations have been evaluated in the HET-CAM test method. Concentrations tested for liquid and solid test substances have ranged from 1% to neat. Several test method protocols indicated that colored or turbid test substances, in particular, were diluted to allow for evaluation of the CAM after application of the test substance.

Historical use generally supports testing substances at a variety of concentrations. However, it is recognized that concentrations may require adjustment for certain chemical or product classes depending on the physical form and/or color of the test substance.

2.2.4.3 *Application of Test Substance to CAM*

Application of test substances to the CAM depended on the physical form of the test substance. Several HET-CAM test method protocols did not provide information on the quantity of solids applied to the CAM, suggesting that all test substances were tested at a standard amount or were solubilized prior to application (e.g., Kalweit et al. 1987; Vives et al. 1997). For liquids or solutions, most protocols indicated that the substances were placed directly onto the CAM (e.g., Kalweit et al. 1987; Vives et al. 1997). For those studies where solid forms of the test substances were used, different application techniques were noted. Most of the protocols indicated that the solid test substances were placed directly onto the CAM. However, a few studies identified the use of a rubber ring or similar device to contain the test substances to a specific area of the CAM (Gilleron et al. 1996, 1997; Hagino et al. 1999). The test substance applicator (TSA) described in Gilleron et al. (1996, 1997) consisted of a perlon mesh locked between two Teflon rings. The substance was placed in the ring and after treatment, it was removed to allow for evaluation of the CAM.

Two protocols required producing agarose pellets containing the test substance, which was then applied to the CAM surface (Demirci et al. 2003, 2004).

The only rationale located in the reports on the methodology used during the application of the test substance was for the use of a device to confine a test substance to a specific location on the CAM. The rationale for such a device was to reduce variability in the area covered by different test substances. Additionally, there was some discussion in the reports that at least 50% of the CAM surface should be covered with solid, paste, or particulate test substances to ensure high quality results. While a formal study to determine the optimal test volume and/or amount was not found in the published scientific literature, direct application of substances to the CAM (with or without a device that would contain the substances to a specific location) appears to produce consistent results.

2.2.4.4 *Test Substance Exposure Duration*

The exposure duration of the test substance differed among published reports. Several protocols noted that after a 20-second exposure of the CAM, all test substances were rinsed (typically with water) from the CAM (e.g., Luepke 1985; Luepke and Kemper 1986; Lawrence et al. 1990; Hagino et al. 1999). One protocol indicated that test substances were rinsed from the CAM after a three-minute exposure (Balls et al. 1995). Some test method protocols indicated that only insoluble or colored, opaque, or turbid substances were rinsed after a user defined time period (Spielmann et al. 1993, 1996; Spielmann 1995). This distinction suggests that the exposure duration for clear and/or soluble test substances was greater.

Rationales for the various exposure durations and discussion of optimal exposure duration were not located. Historical use of these exposure periods and criteria for rinsing suggest that these procedures are appropriate to assess the irritancy potential of test substances (Spielmann H, personal communication).

2.2.5 Known Limits of Use

A review of the available literature indicates that there are some test substance characteristics that could have an impact on the accuracy and/or reliability of the HET-CAM test method. The accuracy of irritancy prediction of test substances that are: (a) colored or turbid, or (b) that adhere to the CAM may be compromised. Such substances may impair clear visualization of the CAM during the study and thus could lead misclassification of the substance. Therefore, the usefulness of the test method for testing such substances undiluted may be limited.

2.2.6 Nature of the Response Assessed

The endpoints evaluated in the HET-CAM test method were evaluated by visual inspection of the CAM surface after test substance application. Depending on the type of data compiled from the test method, the response could be considered qualitative or semi-quantitative.

Several test methods evaluated if the endpoint of interest had developed at a specified time point and/or the severity of the response elicited by the test substance and recorded a qualitative score.

Other test methods observed the CAM after application of the test substance and recorded the time (in seconds) when an effect occurred. These data are considered semi-quantitative since the development of an endpoint is subjective and dependent upon interpretation of the investigator, but the data collected (the time after test substance application when an endpoint developed) is a quantitative value.

No rationale was provided in the reports for the selection of a specific response assessed. Therefore, while evaluating either type of response (qualitative or semi-quantitative) may be appropriate for evaluating ocular toxicity it appears that historical evaluation of semi-quantitative data is more accurate.

2.2.7 Appropriate Controls and the Basis for their Selection

2.2.7.1 *Negative Controls*

Of those test method protocols that identified an appropriate negative control, the most commonly identified substance was 0.9% NaCl (Vinardell and Macián 1994; Budai et al. 1997; Spielmann et al. 1997; Lönnroth et al. 1999; Budai and Várnagy 2000). An agarose pellet also was identified by one set of studies as a negative control (Demrici et al. 2003, 2004). No rationale was provided for the choice of negative control substance used. It appears that the commonly used negative controls for the HET-CAM test method offer no distinct advantages or disadvantages, except for their use for substances with differing solubilities.

However, it is clear that a negative control is useful in the HET-CAM test, so that nonspecific changes in the test system can be detected. This type of control also provides a baseline for the assay endpoints, and ensures that the assay conditions do not inappropriately result in an irritant response. Any of the commonly used negative controls is acceptable as long as the same negative control is used consistently within a laboratory.

2.2.7.2 *Positive Controls*

Of those test method protocols that identified an appropriate concurrent positive control, the most commonly identified substances were sodium dodecyl sulfate (SDS), laureth-8-sulfate, and sodium hydroxide (NaOH) (e.g., Reinhardt et al. 1987; Sterzel et al. 1990; INVITTOX 1992; Spielmann 1995; Macián et al. 1996; Budai et al. 1997; Spielmann et al. 1997; Vives et al. 1997). When used as positive controls, these substances were used at relatively low concentrations. SDS was used at a concentration ranging from 0.1% to 1%, with 1% being the most commonly used concentration. NaOH was typically used at a concentration of 0.1M. Additional positive controls that were identified included benzalkonium chloride (0.01% to 100%), *N,N*-dimethylformamide (DMF; 0.1% to 100%), and imidazole (0.1% to 100%) (Gilleron et al. 1996). No rationale was provided in the literature reference for the choice of positive control substances used.

Review of *in vivo* data provides some conflicting information about some of the test substances used as positive controls for the *in vitro* test method. Evaluation of data from the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) database (ECETOC 1998) indicates that 3% sodium lauryl sulfate (SLS; also referred to as SDS) is a nonirritant according to the EU (2001) and GHS (UN 2003) classification systems and a Category III irritant according to the EPA (1996) classification system (see **Sections 1.0** and **4.0** for descriptions of various regulatory classification systems). Based on this data, it is likely that a 1% SDS solution (which is the concentration most typically used as a positive control) also would be considered a nonirritant according to the EU (2001) and GHS (UN 2003) classification systems and at most a Category III irritant according to the EPA (1996) classification system. Evaluation of the *in vivo* NaOH data indicates that a 1% solution is classified as an R36 irritant according to the EU classification system (EU 2001), a Category 2B irritant according to the GHS classification system (UN 2003), and a Category III irritant according to the EPA classification system (EPA 1996). Based on this data, it is likely that a 0.1 N NaOH solution (which is approximately a 0.4% solution) would be classified as a nonirritant or mild irritant by these different classification systems.

The rationale for the use of these positive controls, which are likely nonirritants or mild irritants *in vivo*, was not provided in the various study reports. Furthermore, there is no discussion in the reviewed literature of the use of different positive controls in the HET-CAM test method depending on the purpose of the test method.

Based on historical use in the HET-CAM test method, 1% SDS or 0.1 M NaOH are the most commonly used positive controls. Inclusion of a known severe ocular irritant substance in each experiment as a positive control demonstrates the functional adequacy of the test method and the consistency of laboratory operations in accurately identifying ocular corrosives and severe irritants. A positive control not only ensures the integrity of the test system and its proper execution, but also provides a measure of test method performance over time.

2.2.7.3 *Solvent Controls*

Although testing substances directly or undiluted is preferred, some HET-CAM test method protocols used a solvent to dilute the test substance. Of those studies that used solvents, there was little variability in which ones were used. The most commonly identified solvents were distilled or sterilized water, 0.9% sodium chloride (NaCl), and olive oil (INVITTOX 1992; Vinardell and Macián 1994; Spielmann 1995; Gilleron et al. 1996; Macián et al. 1996; Doucet et al. 1999). Carboxymethyl cellulose was used in one study (Macián et al. 1996) for water insoluble test substances. Two additional solvents described were cell culture media (MEM, 2 mM L-glutamine, 100 IU/mL penicillin, 100 mg/ml streptomycin, 5% fetal bovine serum) and 2.5% agarose (Lönnroth et al. 1999; Demrici et al. 2003, 2004). Special solubilization methods were described only for the study that listed cell culture media as the preferred solvent (Lönnroth et al. 1999). No rationale was provided for the choice of solvent used in the various studies.

It is clear that a solvent control is useful in the HET-CAM test, so that changes in the test system due to a solvent can be detected. This type of control also provides a baseline for the assay endpoints, and ensures that the assay conditions do not inappropriately result in an irritant response. Any of the commonly used solvent controls is acceptable.

2.2.7.4 *Benchmark Substances*

None of the study reports reviewed indicated the use of the benchmark controls. In the European Commission (EC)/British Home Office (HO) study (Balls et al. 1995), a reference standard (a 5% solution of Texapon ASV, an anionic surfactant) was used to evaluate the relative irritancy potential of the substances tested.

Benchmark substances are often used during the testing of substances of unknown toxicity potential. The toxicity of the benchmark substance is generally well characterized (i.e., adequate human or animal toxicity data are available). A benchmark is selected to match the chemical or product type of the unknown substance, and is used to set an upper or a lower limit of response against which the unknown is compared (Harbell and Curren 2002). Benchmark substances are often selected from a list of reference chemicals for the assay and have the following properties:

- consistent and reliable source(s)
- structural and functional similarity to the class of the substance being tested
- known physical/chemical characteristics
- supporting data on known effects in the *in vivo* rabbit eye test
- known potency in the range of the desired response

They are useful for evaluating the ocular irritancy potential of unknown chemicals of a specific chemical or product class, or for evaluating the relative irritancy potential of an ocular irritant within a specific range of irritant responses.

2.2.8 Acceptable Range of Control Responses and the Basis for the Acceptable Ranges

2.2.8.1 *Negative/Solvent Controls*

HET-CAM studies using 0.9% NaCl as a negative control, were conducted with and without the use of a TSA (Vanparys and Van Goethem 2005). The use of a TSA was described in Gilleron et al. (1996, 1997) (see **Section 2.2.4.3**). Over 90 tests using the TSA and three tests without using TSA were provided. As shown in **Table 2-1**, time to development of endpoints and the overall irritation scores calculated were consistent and classified as nonirritants for all tests (for additional information see **Section 7.0**).

Table 2-1 Means and Standard Deviations of Negative Control Test Substances Evaluated With and Without Use of the Test Substance Applicator

0.9% NaCl	Hemorrhage ¹ (mean ± SD)	Lysis ¹ (mean ± SD)	Coagulation ¹ (mean ± SD)	<i>In Vitro</i> Score (mean ± SD) ²
With TSA (n=92)	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Without TSA (n=3)	0 ± 0	0 ± 0	0 ± 0	0 ± 0

¹Mean values of time until development of identified endpoint.

²*In Vitro* irritation score calculated as described in Kalweit et al. (1987, 1990). See **Section 2.2.12.1** for further details on this analysis method.

It would seem appropriate to establish an upper limit of both opacity and permeability for the negative or solvent control. Negative and solvent controls must produce the anticipated response to ensure the test system is functioning properly and that the specific test is valid.

2.2.8.2 *Positive Controls*

HET-CAM studies using DMF as a positive control were conducted with and without the use of a TSA (for further information see **Section 2.2.4.3** and **Section 7.0**; Vanparys and Van Goethem 2005).

With the studies that were conducted with the TSA, the hemorrhage endpoint was evaluated inside the TSA and outside the TSA. Of note, the time of development of the hemorrhage endpoint inside the TSA was significantly lower than the time to development of the hemorrhage endpoint outside the TSA (see **Table 2-2**). The reason for the difference is not clear. Two proposed reasons for the difference in time to development are (1) the vessels outside the TSA may open more easily than those under the TSA, or (2) once the liquid is applied it the liquid accumulates around the edge of the TSA rather than between the TSA and CAM.

Table 2-2 Comparison of Means and Standard Deviations of Positive Control Test Substances Evaluated With and Without Use of the Test Substance Applicator

Positive Control	Hemorrhage ¹ (mean ± SD)	Lysis ¹ (mean ± SD)	Coagulation ¹ (mean ± SD)	<i>In Vitro</i> Score (mean ± SD) ²
With TSA ³ (n=69)	0.02 ± 0.17	6.93 ± 0.03	8.82 ± 15.77	15.77 ± 0.19
With TSA ³ (n = 10)	3.36 ± 0.32	6.54 ± 0.19	8.81 ± 0.04	18.71 ± 0.38
Without TSA (n = 2)	4.00 ± 0.13	6.84 ± 0.05	8.76 ± 0.08	19.60 ± 0.15
Imidazole (n=15)	4.50 ± 0.39	6.84 ± 0.08	8.66 ± 0.17	20.00 ± 0.45

¹Mean values of time until development of identified endpoint.

²*In Vitro* irritation score calculated as described in Kalweit et al. (1987, 1990). See **Section 2.2.12.1** for further details on this analysis method.

³Hemorrhage endpoint in studies described in the first row were evaluated inside the TSA while hemorrhage endpoint in studies described in the second row was evaluated outside the TSA.

In addition to DMF, 100% imidazole also was evaluated as a positive control. In all studies with imidazole a TSA was used. As shown in **Table 2-2**, time to development of each endpoint evaluated and the overall irritation score calculated were consistent and classified as irritants for all tests with the substance.

HET-CAM studies using 1% SDS and 0.1 N NaOH also were provided upon NICEATM request (Spielmann and Liebsch 2005). Using the mean values determined for these studies, the overall irritation score calculated (according to the method of Kalweit et al. 1987, 1990) for these substances classified them as irritants (see **Table 2-3** and **Section 7.0**).

Table 2-3 Means and Standard Deviations of Positive Control Test Substances

Positive Control	Hemorrhage ¹ (mean ± SD)	Lysis ¹ (mean ± SD)	Coagulation ¹ (mean ± SD)
1% SDS (n=377)	14.69 ± 5.36	35.18 ± 17.15	--- ²
0.1 N NaOH (n=336)	8.96 ± 4.96	35.60 ± 24.71	48.04 ± 34.56

¹Mean values of time until development of identified endpoint.

²It was indicated that 1% SDS does not produce coagulation in the CAM after application. However, in the studies conducted coagulation was identified in a single study. In these evaluations, the non-existing data was calculated with an arbitrary value of "0". Therefore, the calculation of a mean value for the coagulation endpoint was not meaningful.

Positive controls are typically used as one of the criteria for determination of a valid test. If the positive control value falls within the accepted range, the test is considered valid. If the positive control value falls outside of the accepted range, the test may need to be repeated.

2.2.8.3 *Benchmark Controls*

Benchmark substances may be useful in demonstrating that the test method is functioning properly for detecting the ocular irritancy potential of chemicals of a specific chemical class or a specific range of responses, or for evaluating the relative irritancy potential of an ocular irritant. Therefore, benchmark substances should produce an irritation response that is within acceptable limits of historical data.

2.2.9 Nature of the Data to be Collected and the Methods Used for Data Collection

Review of the literature indicates that the nature of the response assessed from the HET-CAM test method varied. The nature of the data collected depends on whether a qualitative response or semi-quantitative response was evaluated (see **Section 2.2.6**). In general, the HET-CAM test method protocols evaluated the CAM for (1) the time to development of observed endpoints after exposure to the test substance, (2) the highest or lowest test substance concentration needed to produce a minimal response on the CAM after exposure to the test substance, or (3) the maximum severity of response observed after exposure to the test substance. Some test method protocols described evaluating the CAM for two of the three of these responses. The following sections describe the three responses and the data collected by each, as described in the literature.

Visual inspection of the CAM is the only procedure described for collecting data.

2.2.9.1 *Time to Development of Observed Endpoints After Exposure to the Test Substance*

In studies that monitored the development of endpoints, two different data collection procedures were followed:

(A) One procedure evaluated the development of the response (i.e., endpoint) at specific time points after the test substance was applied to the CAM. In these reports, the blood vessels on the CAM were examined and scored for the presence of, for example, vessel lysis, coagulation, and hemorrhage at 0.5, 2, and 5 minutes after exposure to the test substance. The assigned score was time dependent (see **Table 2-4**); the earlier the endpoint developed in the 5-minute observation period, the higher the resulting score assigned to that endpoint (e.g., hyperemia developing before 0.5 minutes was assigned a score of 5 while hyperemia developing between two and five minutes was assigned a score of 1). The assigned scores were weighted with the highest scores being assigned to the coagulation endpoints. There was no rationale provided in the study reports on the scoring system or for the weighting in the scores used. Individual values for the observed endpoints were then used to assign the irritation potential of the test substance.

(B) Another procedure used to evaluate the time to development of endpoints after exposure to the test substance was to continually observe the CAM during the 5-minute observation period and record (typically in seconds) the time at which each of the endpoints developed (e.g., Kalweit et al. 1987, 1990; Sterzel et al. 1990; CEC 1991; Spielmann et al. 1991; Macián et al. 1996; Spielmann et al. 1996; Gilleron et al. 1997; Schlage et al. 1999). Therefore, three separate time values were obtained and recorded for each egg (one time value for each endpoint). Individual values for the observed endpoints were then used to determine the irritation potential of the test substance.

Table 2-4 HET-CAM Scoring System

Endpoint	Score at Different Observation Times		
	0.5 min	2 min	5 min
Hyperemia	5	3	1
Hemorrhage	7	5	3
Coagulation	9	7	5

Adapted from Luepke (1985).

None of the evaluated test method protocols provided guidance and/or information on the identification of any endpoints or what constituted development of each of the endpoints.

2.2.9.2 *Lowest Test Substance Concentration Needed to Produce a Minimal Response on the CAM After Exposure to the Test Substance*

Several test method protocols identified the use of a threshold concentration in evaluating the irritation potential of the test substance (e.g., Spielmann et al. 1993; Spielmann 1995). The lowest test substance concentration that produced a slight to weak irritant response on the CAM was recorded. The test method protocols that evaluated the threshold concentration did not provide guidance on what constituted a weak or slight response on the CAM. The identified test substance concentration was then used in combination with other evaluated responses to evaluate the irritation potential of the test substance.

2.2.9.3 *Severity of Response Observed After Exposure to the Test Substance*

One set of test method protocols evaluated the severity of each of the endpoints at a single time point. The severity of each of the endpoints was evaluated and scored based on a user-defined scoring scale; the scores ranged from 0 (no effect) to 3 (severe irritant effect) (Balls et al. 1995; Steiling et al. 1999). There was no guidance in either set of test method protocols as to the decision criteria for the different scores.

Historical use of all types of methods of collection appear useful in identification of ocular corrosives and severe ocular irritants. It is noted that collection of the time to development of each endpoint (see **Section 2.2.9.1 (B)**) allows investigators to use a variety of analysis methods to assess irritancy classification (see **Section 2.2.12**).

2.2.10 Type of Media in Which Data are Stored

One of the reviewed test method protocols noted that data were entered into and stored in electronic format on a computer (Spielmann et al. 1991). However, no additional information was provided. No other information on the type of media in which data are stored was noted in the published literature, but it can be assumed the data were written on data sheets and stored in a study book.

It would seem appropriate that data from the HET-CAM be stored and archived in a manner consistent with international Good Laboratory Practices (GLP) guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003). GLP guidelines are nationally and internationally recognized

rules designed to produce high-quality laboratory records. These guidelines provide a standardized approach to report and archive laboratory data and records, and information about the test protocol, to ensure the integrity, reliability, and accountability of a study (EPA 2003a, 2003b; FDA 2003).

2.2.11 Measures of Variability

There is little to no discussion in the published literature on how variability of response among replicate eggs within an experiment or among experiments conducted in the same laboratory is evaluated. Several of the studies indicated that the mean of the scores are taken as the final score. However, whether this is the mean of replicate eggs or replicate experiments is not clear.

Calculation of the mean score and SD provides the user with information on the performance of the test method. These values allow for an assessment of the performance of the test conducted and whether the observed variability between replicates is greater than would be considered acceptable.

2.2.12 Statistical or Nonstatistical Methods Used to Analyze the Resulting Data

As shown in **Section 2.2.9**, several types of data can be obtained from the HET-CAM test method. Therefore, depending on the data collected, different analysis methods were used to evaluate the irritancy potential of tested substances in the test method protocols. For data that evaluated the time to development of observed endpoints after application of the test substance to the CAM (**Section 2.2.9.1**), an Irritation Score (IS) or a “Q-Score” was determined. Alternatively, a mean detection time for the appearance of coagulation (mtc) was determined. For data that evaluated the lowest test substance concentration needed to produce a minimal response on the CAM after application of the test substance (**Section 2.2.9.2**), the Irritation Threshold Concentration (ITC) was determined. For data that evaluated the severity of response observed after application of the test substance (**Section 2.2.9.3**), an “S-Score” or Severity Irritation Score (SIS) was calculated.

2.2.12.1 *Irritation Scores*

A majority of the test method protocols calculated a score (referred to as irritation score, irritation index, or irritation potential) that represented the irritation potential of the test substance based on endpoint development. This score (referred to in this document as the IS value) could be determined by a variety of mathematical models. Several of the test method protocols did not provide the model used to determine the IS. However, the two main mathematical models are discussed below.

For those test method protocols that assigned a score to each of the endpoints evaluated, depending on the time the endpoint develops (see **Section 2.2.9.1 [A]**), the values assigned to each endpoint were totaled to give an overall IS value for the test substance (referred to in this document as the “IS[A] analysis method”). The final IS value ranged from 0 (for test substances that do not induce development of any of the observed endpoints) to 21 (for test substances that induce development of all three endpoints within 30 seconds of application of the test substance).

For those test method protocols that noted the time that a specific endpoint was observed (see **Section 2.2.9.1 [B]**), an IS score could be calculated using the general formula (referred to in this document as the “IS[B] analysis method”):

$$\left(\left(\frac{(301 - \text{Hemorrhage time})}{300} \right) \times 5 \right) + \left(\left(\frac{(301 - \text{Lysis time})}{300} \right) \times 7 \right) + \left(\left(\frac{(301 - \text{Coagulation time})}{300} \right) \times 9 \right)$$

where:

Hemorrhage time = time (in seconds) of the first appearance of blood hemorrhages

Lysis time = time (in seconds) of the first appearance of vessel lysis

Coagulation time = time (in seconds) of first appearance of protein coagulation

The IS, when calculated using this formula, has a maximal value of 21.

There is no explanation of the source for the multiplication factors of 5, 7, and 9 in the above formula (Kalweit et al. 1987). However, it is likely that these values were selected to correspond to the highest score value possible for each of the endpoints in the IS(A) analysis method (see **Section 2.2.9**).

When development of hyperemia, injection, or another endpoint was evaluated in place of vessel lysis, the time point for the alternative endpoint replaced the lysis time point. There was no specific description in any of the test method protocols of statistical or nonstatistical methods used to analyze dose-response relationships described for either of the analysis methods.

Despite the fact that both analytical methods yield IS values that range from 0 to 21, care should be taken if values from these two methods are compared. The time to endpoint development ranges required to obtain IS values are different for each analytical method. For example, a maximum IS value of 21 can be obtained for the IS(A) analysis method when all three endpoints develop within 30 seconds after exposure of the CAM to the test substance. Comparatively, the same IS value can only be obtained by the IS(B) analysis method when all three endpoints develop within one second after exposure of the CAM to the test substance.

2.2.12.2 *Q-Score*

The Q-Score represents a comparison of the irritation potential of the test substance with that of a reference substance. To determine the Q-Score, the irritation potential of both the test substance and reference substance are evaluated. The irritation potential could be determined using any method, but typically are expressed as IS values. Of the test method protocols that evaluated the Q-Score, one did not provide the mathematical model used to determine the IS and the other used a weighted differential model. The IS of the test substance was then compared to the IS of the reference standard to calculate a ratio for the values, which was then used to assess the irritation potential of the test substance.

2.2.12.3 Mean Detection Time

For those test method protocols that noted the time that a specific endpoint was observed (see **Section 2.2.9.1 [B]**), the mean detection time for the appearance of an endpoint was determined. Mean detection times for the development of hemorrhage, lysis, and coagulation were evaluated, based on the times for three tested eggs (Spielmann et al. 1996).

2.2.12.4 ITC

Several test method protocols described using a combination of IS value and ITC to evaluate the irritation potential of a tested substance. The ITC was defined as the lowest concentration required to produce a slight or weak response on the CAM after application of the test substance. No definition was provided in the test method protocols for the terms “slight response” or “weak response.” No data analysis or manipulation was required for this value, since the ITC was strictly the test substance defined as a percentage of the volume tested.

2.2.12.5 S-Score

Another set of test method protocols calculated a term defined as the S-Score. This score represents the highest total score for any endpoint evaluated for a test substance. For this score, the severity scores assigned for each endpoint (which ranged from 0 to 3 and were assigned at a single user-defined time point) were totaled for all of the replicate eggs evaluated per test substance to produce an endpoint total score. Therefore, there were three endpoint total scores per test substance. The endpoint that yielded the highest score was used as the final test substance S-Score. Many of the test method protocols that evaluated the irritation potential of test substances using this analysis method advocated the use of six eggs per test substance. In such cases, the maximal S-Score is 18.

According to this analysis method, the endpoints used to develop the S-Score could be different for different test substances. Thus, the S-Score could be defined by the development of coagulation for one test substance and hyperemia for another test substance.

2.2.12.6 SIS

A single test method protocol used this method to assess the anti-irritation potential of test substances (Demirci et al. 2003, 2004). The SIS is based on the potency of the anti-angiogenic effect produced by the test substance. Using the scoring range of 0 to 2, the ability of the test substance to produce an anti-irritant effect was noted. The number of eggs scored with a 0, 1, or 2 was inserted into the following formula to calculate the SIS:

$$\text{AverageScore} = \frac{\text{Number of Eggs(Score2)} * 2 + \text{Number of Eggs(Score1)} * 1}{\text{TotalNumber of Eggs(Score0,1,2)}}$$

2.2.13 Decision Criteria and the Basis for the Prediction Model Used to Classify a Test Chemical as a Severe Eye Irritant

A review of the test method protocols indicates that there is not a single set of decision criteria that has been used to classify test substances as producing a positive (i.e., irritant) or negative (i.e., nonirritant) response. Depending on the type of data collected and the method used to analyze the data, various irritation classification schemes and decision criteria have

been developed. Most of these schemes and criteria were developed by individuals and do not typically correlate *in vitro* scores with *in vivo* irritation classification schemes, such as those used by the EPA (1996), EU (2001), and GHS (UN 2003). However, as detailed below, some correlations were noted.

2.2.13.1 Irritation Classification Based on IS

Many of the test method protocols that calculated IS values (either by the IS[A] or IS[B] analysis methods) used a similar classification scheme. Ranges of *in vitro* scores were assigned to one of four irritation categories. The major ranges of scores for each category are shown in **Table 2-5**.

Table 2-5 Summary of HET-CAM Score Ranges Used in Irritancy Classification

HET-CAM Score Range	Irritation Category
0-0.9	Nonirritant or Practically None
1-4.9	Weak or Slight Irritation
5-8.9 or 5-9.9	Moderate Irritation
9-21 or 10-21	Strong or Severe Irritation

From Luepke (1985) and Kalweit et al. (1987).

In addition to the IS ranges listed above, several other irritation classification schemes have been described. For example, van Erp and colleagues (1990) used the following classification scheme for IS values: Nonirritant ≤ 0.5 ; Slight Irritant = 0.5 to 3.4; Moderate Irritant = 3.5 to 4.9; Severe Irritant ≥ 5 . Gilleron et al. (1996, 1997) classified test substances as either nonirritant or irritant. In these studies, substances inducing HET-CAM IS values from 0 to 4.9 were defined as nonirritant while those inducing scores of 5.0 and above were defined as irritants.

One of the test method protocols correlated *in vitro* HET-CAM test method scores with a classification system used by a regulatory agency. The studies conducted by Gettings and colleagues (1991, 1994, 1996) used different criteria to classify test formulations according to the FHSA classification system (CPSC 1988). Using this system, in which substances are defined as irritants or nonirritants, a range of IS values were used. The range depended on the test method protocol and the data analysis method. **Table 2-6** provides the decision criteria used by the various test method protocol analysis methods to classify a test substance IS value as an irritant according to the FHSA classification system.

According to the published reports, the *in vitro* values used to classify a test substance as an irritant or nonirritant were determined *post hoc* and were deliberately chosen to minimize the number of false positives or false negatives (Gettings et al. 1994, 1996).

Table 2-6 Decision Criteria used by Gettings and Colleagues to Define Test Substances as Irritants According to the FHSA Classification System

Literature Source	IS Value for Irritants
Gettings et al. 1991 (HET-CAM I)	> 10
Gettings et al. 1991 (HET-CAM II)	> 300 ¹
Gettings et al. 1994 (HET-CAM)	≥ 5
Gettings et al. 1994 (Chorioallantoic Membrane Assay)	≥ 4.8
Gettings et al. 1996 (HET-CAM I)	≥ 5.1
Gettings et al. 1996 (HET-CAM III)	≥ 4.83

¹The analysis method used for this evaluation was different than previously described (see **Section 2.2.12**). The mathematical model of Bartnik et al. (1987) was used for this evaluation. However, information could not be located on this model to assess the range of scores that could be obtained.

2.2.13.2 Irritation Classification Based on Q-Score

Ranges of Q-Scores were assigned to one of three or four irritation categories. A summary of the ranges of scores used in the classifying the irritancy potential of a test substance is shown in **Table 2-7**.

Table 2-7 Summary of Q-Scores Used in Irritancy Classification

Q-Score (Range 1) ¹	Q-Score (Range 2) ²	Irritation Category
< 1.5	-	Nonirritant
-	≤ 0.8	Slight
1.5 ≤ Q < 2	0.8 < Q < 1.2	Moderate
-	1.2 ≤ Q < 2	Irritant
≥ 2	≥ 2	Severe

¹From Balls et al. (1995).

²From Brantom et al. (1997).

No information was provided on the correlation of Q-Scores to irritation categories defined by the GHS (UN 2003), EPA (1996), or EU (2001) ocular irritation classification systems.

2.2.13.3 Irritation Classification Based on Mean Coagulation Detection Time

A range of mean coagulation times when using a 10% solution (mtc10) values were assigned to one of two EU irritation categories. The studies conducted by Spielmann and colleagues (1996) used different criteria to classify test substances according to the EU classification system (EU 1992). Using this system, in which substances are defined as R41 or Remainder (R36 and nonirritants), a range of mtc10 values were used. The ranges provided in **Table 2-8** were based on different analyses and sets of data used in the evaluation. **Table 2-8** provides the decision criteria used to classify a test substance mtc10 value as an irritant according to the EU classification system.

Table 2-8 Summary of mtc10 Values Used in Irritancy Classification

mtc10 (Range 1)	mtc10 (Range 2)	mtc10 (Range 3)	Irritation Category
<174 seconds	< 139 seconds	< 50 seconds	R41
≥ 174 seconds	≥ 139 seconds	≥ 50 seconds	Remainder

From Spielmann et al. (1996).

2.2.13.4 Irritation Classification Based on IS and ITC

Several studies based irritation classifications on combinations of IS and ITC values. **Table 2-9** provides an example of a combination of scores defined in one study for various irritation categories (Spielmann et al. 1996).

Table 2-9 Summary of ITC and IS combinations Used in Irritancy Classification

ITC (% Concentration)	Irritation Score (10% Concentration)	Irritation Category (EU Category) ¹
> 10%	< 16	None/slight (Nonirritant)
> 10%	> 16	Moderate (Nonirritant)
< 10%	< 16	Moderate (Nonirritant)
< 10%	> 16	Irritant (R36)
< 2.5%	< 16	Irritant (R36)
1% < ITC ≤ 2.5%	≥ 16	Severe (R41)
≤ 1%	-	Severe (R41)

From Spielmann et al. (1996).

¹EU (1992, 2001).

The combination of IS and ITC values used to define various irritation categories were similar, but not identical, between test method protocols. For example, **Table 2-9** shows that when a test substance had an ITC greater than 10% and the IS was less than 16, the test substance was classified as none/slight (nonirritant). In contrast, another test method protocol classified the same combination as a moderate irritant (Spielmann 1995).

Two of the reviewed test method protocols utilized IS and ITC scores to assign a classification that corresponded to an *in vivo* irritation classification system used by a regulatory agency. The study conducted by Spielmann and colleagues (1996) classified test substances according the classification system used by the EU (1992). The correlation of IS and ITC values to this classification system is provided in **Table 2-9**. Spielmann and colleagues (1996) reported that the prediction model was developed empirically and was not supported by statistical methods.

The study conducted by Gettings et al. (1996) classified test formulations according to the classification system defined in the FHSA (CPSC 1988). For this analysis, the ratio of the IS value to the ITC value was determined. A final value equal to or greater than three was

defined as an irritant, as defined by the FHSA (CPSC 1988). According to the published reports, the *in vitro* values used to classify as an irritant or nonirritant were determined *post hoc* and were deliberately chosen to minimize the number of false positives and false negatives (Gettings et al. 1996).

2.2.13.5 Irritation Classification Based on S-Score

Ranges of S-Scores were assigned to one of three or four irritation categories. A summary of the classification system using this value is shown in **Table 2-10**.

Table 2-10 Summary of S-Scores Used in Irritancy Classification

S-Score (Range 1) ¹	S-Score (Range 2) ²	Irritation Category
< 6	-	Nonirritant
-	< 6	Slight
$6 \leq S < 15$	$6 \leq S \leq 12$	Moderate
-	$12 < S < 16$	Irritating
≥ 15	≥ 16	Severe

¹From Balls et al. (1995).

²From Brantom et al. (1997).

No information was provided on the correlation of S-Scores to irritation categories defined by the GHS (UN 2003), EPA (1996), or EU (2001) ocular irritation classification systems.

2.2.14 Information and Data that Will be Included in the Study Report and Availability of Standard Forms for Data Collection and Submission

There was no description provided in the studies reviewed about the information and data that was included in the study reports for this test method. There are no known standard forms for data collection and submission.

It would seem appropriate that the test report include the following information, if relevant to the conduct of the study:

Test and Control Substances

- Chemical name(s) such as the structural name used by the Chemical Abstracts Service (CAS), followed by other names, if known
- The CAS Registry Number (RN), if known
- Purity and composition of the substance or preparation (in percentage(s) by weight)
- Physicochemical properties such as physical state, volatility, pH, stability, chemical class, water solubility relevant to the conduct of the study
- Treatment of the test/control substances prior to testing, if applicable (e.g., warming, grinding)
- Stability, if known

Information Concerning the Sponsor and the Test Facility

- Name and address of the Sponsor
- Name and address of any test facilities involved
- Name and address of the Study Director

*Justification of the Test Method and Protocol Used**Test Method Integrity*

- The procedure used to ensure the integrity (i.e., accuracy and reliability) of the test method over time (e.g., periodic testing of proficiency substances, use of historical negative and positive control data)

Criteria for an Acceptable Test

- Acceptable concurrent negative control ranges based on historical data
- Acceptable concurrent positive control ranges based on historical data
- If applicable, acceptable concurrent benchmark control ranges based on historical data

Test Conditions

- Experimental starting and completion dates
- Details of test procedure used
- Test concentration(s) used
- Description of any modifications of the test procedure
- Reference to historical data of the model (e.g., negative and positive controls, proficiency substances, benchmark substances)
- Description of evaluation criteria used

Results

- Tabulation of data from individual test samples (e.g., irritancy scores for the test substance and the positive, negative, and benchmark controls, reported in tabular form, including data from replicate repeat experiments as appropriate, and means and the standard deviation for each experiment)

*Description of Other Effects Observed**Discussion of the Results**Conclusion**A Quality Assurance Statement for GLP-Compliant Studies*

- This statement indicates all inspections made during the study, and the dates any results were reported to the Study Director. This statement also serves to confirm that the final report reflects the raw data.

Additional reporting requirements for GLP-compliant studies are provided in the relevant guidelines (e.g., OECD 1998; EPA 2003a, 2003b; FDA 2003).

2.3 Basis for Selection of the Test Method System

The HET-CAM is proposed to provide information on the effects that may occur in the conjunctiva following exposure to a test substance. Published reviews note that chicken-embryo models have long been used as models by embryotoxicologists and virologists. (Parish 1985; Luepke and Kemper 1986). Extending the use of chicken embryos, the HET-CAM test method was proposed by Luepke (1985) and Luepke and Kemper (1986). It was assumed that acute effects induced by a test substance on the small blood vessels and proteins of this soft tissue membrane are similar to effects induced by the same test substance in the eye of a treated rabbit. The CAM has been proposed as a model for a living membrane (such as the conjunctiva) since it comprises a functional vasculature.

Additionally, evaluation of coagulation (i.e., protein denaturation) may reflect corneal damage that may be produced by the test substance. It has been inferred that only very mildly irritating substances are capable of inducing conjunctival effects in the absence of corneal effects (Prinsen M, personal communication). Therefore, this would not appear to limit the effectiveness of the HET-CAM with respect to predicting severe irritants.

2.4 Proprietary Components

There are no proprietary components used in the HET-CAM test method.

2.5 Basis for the Number of Replicates and Repeat Experiments

2.5.1 Sample Replicates

The number of replicate eggs used per test substance or test concentration varied between protocols. Several protocols did not describe or discuss the number of replicates used in the method. Of those test method protocols that did describe the number of replicate eggs used per test substance or test concentration, most of the methods described using three, four, or six eggs. Three test method protocols described using 10-15 or 20 replicate eggs per test substance (Brantner et al. 2002; Demrici et al. 2003, 2004). Comparison of results from studies conducted with three eggs to those conducted with six eggs previously indicated no significant difference (Spielmann H, personal communication).

2.5.2 Experimental Replicates

Few studies reviewed provided information on the number of replicate experiments conducted for each test substance. A sample of the reviewed studies indicates that one conducted a single experiment, five repeated the experiment two times, four repeated the experiment three times, and three repeated the experiment four times. Additionally, one investigator stated that a total of four eggs were used per test substance and that the experiment could either be conducted a single time with four eggs, or could be conducted two times with two eggs used in each experiment. There was no discussion provided in any of the sources on the optimal number of repeat experiments and no rationale was provided for selecting a specific number of repeats. However, based on sound scientific judgment, it

would seem reasonable to expect that equivocal responses or divergent results among test cornea would mandate repeating the experiment.

2.6 Compliance with Good Laboratory Practices

Compliance with GLP guidelines only can be evaluated by the information that was provided in the published reports. No attempt was made to review original records to assess the quality of the data presented. Based on the available information, the only reports that were identified to have followed GLP guidelines or used data obtained according to GLP guidelines were Gettings et al. (1991, 1994, 1996), Spielmann et al. (1993, 1996), Balls et al. (1995), Brantom et al. (1997), Hagino et al. (1999), and Steiling et al. (1999).

Conducting studies under GLP guidelines increases confidence in the quality and reliability of test data. Furthermore, if data using this test method is to be submitted to the EPA or another agency in response to Federal testing requirements, then compliance with appropriate GLP guidelines will be required.

2.7 Study Acceptance Criteria

There was no description provided in the test method protocols reviewed about the study acceptance criteria.

A test should be considered acceptable if the negative/solvent and positive controls each give an values that falls within the classification of nonirritating or irritating, respectively. A test also should be considered acceptable if the benchmark control (if used) produces an irritation response that is within acceptable limits of historical data.

[This Page Intentionally Left Blank]